İstatistiksel Analiz, Olasılık ve Rassal Değişkenler

Dr. Cahit Karakuş

2020

İçindekiler

Giriş

Bilgi yığının görünen yüzüne bakarak yapılan değerlendirmeler ve analizler organizasyonları öyle sıkıntılara sokar ki, kurtulmaya çalıştıkça batarlar. Güç faktörü olarak stratejik bir öneme sahip olan bilginin günümüzde etkinliği giderek artmaktadır. Egemenlik mücadelesinde bilgi toplamak stratejik öneme sahiptir. Bilginin gücünü bilmeyenler, karanlıkta el yordamıyla yön bulmaya çalışırlar. Öte yandan günümüzde toplanan bilginin boyutu o kadar hızlı büyümektedir ki, yığının içerisinden gerekli olan bilgiyi bulabilmek, işlemek, tasnif etmek ve zamanında erişebilmek, analiz ederek çıkarımlar elde etmek çok fazla önemsenir hâle gelmiştir. Hayati olan nokta herkesin gözü önünde bulunan bilgi yığınları içerisinden, kimsenin dikkatini çekmeyen, kimsenin akıl edemediği örtülü veya kapalı bilgiyi bulup anlamlandırabilmektir. Amaç bilgi yığınından oluşturulan bünyede değişim, sapma ve dönüşümün izlerini bulmaktır.

Bilgi yığınının oluşturduğu bünyenin kestirim yapmayı öğrenebilmesi için veriyi ölçen, sayısallaştıran, saklayan ve kıyaslayarak sınıflandıran yetenekler kazanması gerekmektedir. Laplace, olasılık teorisini şu şekilde açıklıyordu: Bir durumun olasılığını hesaplamak için kurulan denklemler, sonuçtan emin olmayı sağlamıyordu, sadece hata payı en az olan sonucu bulmaya yarıyordu; yani hata payını ortadan kaldırmaya değil, en aza indirmeye çalışıyordu, çünkü hatasız bir denklem kurmak mümkün değildi.

Kestirim yapmak aynı zamanda risk satın alınarak seçim yapmaktır. Kestirim yapma sadece ve sadece kazanma üzerine olursa, kayıpların yıkıcı etkisi beklenilenden çok daha fazla olacaktır. Sağlıklı kestirim yapabilmek için süreçlere ait işlevler sürekli ölçülmeli, bilgiler toplanmalı ve analiz edilmelidir. Toplanan bilgilerin doğru değerlendirilmesi, veri analizine dayalı grafiksel gösterimlere dayalı yorumlar yapan yeteneklerin geliştirilmesi ile mümkün olabilmektedir.

1. Temel Kavramlar

Araştırma: Bir problemin nedenlerinin saptanması, çözüm yollarının planlanması, uygulamaya konulması ve sonuçlarının değerlendirilmesine yönelik yapılan çalışmalardır**.** Araştırmanın konusu ve amacı belirgin olmalıdır. Zaman, personel ve maliyet dikkate alınarak, araştırmanın sınırı iyi belirlenmelidir. Araştırmada görev alan araştırıcılar ya da uzmanlar yeterli bilgi düzeyine sahip olmalıdır. Araştırmanın sonucunda elde edilen bilgiler doğru biçinde değerlendirilmelidir.

Analiz (Çözümleme, Tahlil):
Bir olayın bileşenlerinin hepsinin veya bir kısmının ve neler olduğunu ortaya koymaktır. Karmaşık bir konuyu veya maddeyi daha iyi anlamak için daha küçük parçalara bölme işlemidir.

Sentez:
Element veya başka maddeleri bir araya getirerek yapay olarak bileşik cisimler oluşturma, bireşim. Yalından karmaşık olana giden düşünme biçimidir.

Popülasyon (Kitle, Ana Kitle, Ana Kütle, Yığın) : Araştırma kapsamına giren, aynı özellikleri taşıyan birimlerin ya da bireylerin oluşturduğu topluluğa kitle denir. Kitlenin büyüklüğü araştırmanın özelliğine göre değişir. Nüfus sayımı için kitle Türkiye'dir. Esenyurt'daki üniversite öğrencilerinin giderleri için kitle Üniversitesi öğrencileridir.

Tamsayım: Bir araştırma kapsamında, kitledeki tüm birimlerine ulaşılarak istenen bilginin elde edilmesi işlemidir.

Gözlem (Denek): Örneklemde yer alan her birime gözlem ya da denek denir. Gözlem (ya da denek) sayısı aşağıdaki biçimde simgeleştirilmektedir.
Kitledeki Gözlem Sayısı : **N**
Örneklemdeki Gözlem Sayısı: **n**

Örneklem: Belirli yöntemler kullanılarak, aynı özellikleri taşıyan bir kısım verinin seçildiği topluluğa örneklem denir.

Örnekleme: Nitelik ve nicelik yönünden temsil edebilecek bir kümenin çekilmesi işlemidir.

Kesikli (Süreksiz) Değişken: Tanımlı olduğu aralıklarda ayrık değerler alan değişkenlerdir.

Sürekli Değişken: Tanımlı olduğu aralıkta tüm değerleri (sonsuz sayıda) alabilen değişkenlerdir.

Nicel (Kantitatif) Değişkenler: Ölçüm sonucu değerleri saptanan sayısal özelliklerini belirten değişkenlerdir. Sayılabilir veya ölçülebilir büyüklüklerdir.

Nitel (Kalitatif) Değişkenler: Karakteristik özelliklerini, durumlarını ve pozisyonlarını belirten değişkenlerdir. Sayılamayan, birimlendirilemeyen ve ölçülebilir olmayan büyüklüklerdir. Bir şeyin nasıl olduğunu belirten, onu başka şeylerden ayıran özelliktir (sıfat).

Sıklık (Frekans) Dağılımı: Verilerin gösterdiği dağılıma sıklık (frekans) dağılımı denir. Aynı peryoda sahip sinyalin bütün içerisinde kaç kez tekrar etmesidir.

Sınıf: Eşit ya da birbirine yakın değerli deneklerin oluşturduğu her bir gruba sınıf denir. Sınıf sayısı, k ile gösterilir.. Araştırmacı tarafından belirtilen sınıf sayısı, çok sayıda veri olduğunda aşağıda verilen Sturges'in formülü ile de bulunabilir. $k=1+3.3\log(n)$

2. Veri Analizi

Matematik ve bilgisayardaki gelişmeler, karşılaşılan problemlerin matematiksel olarak modellenip çözülerek gerçek hayata yansıtma olanağı vermiştir. Verilerin toplanması, işlenip düzenlenmesi, çözümlenmesi, tablo veya grafikler şeklinde yorumlanmasının temel amacı verilerin analiz edilerek sorgulamalara yanıt vermek ve geleceğe yönelik tahminde bulunmaktır.

İş süreçlerinde oluşan fırsatlar, hatalar, sapmalar, belirsizlikler ve değişkenler hakkında bilgi edinmek ve olasılık hesaplama teknikleri kullanılarak gelecek hakkında tahminde bulunmak için ölçme ve kıyaslama yapılır. Ölçerseniz yönetirsiniz. Gelecek hakkında tahminde bulunurken, rakamsallaştırılmış ölçümler manipule edilirerek anlamlandırılır. Problem çözmeye yönelik akıl geliştirmede rakamsal analiz kadar, geçmişsel tecrübelerden benzer davranışların raporlanması ve yorumlanması gerekmektedir.

Veri analizi, ham verilerden karar vermeye yönelik yararlı bilgilere dönüşüm sürecidir. Sorulara yanıt vermek için veriler toplanır, hipotezleri test edilir veya teorileri kabul ya da reddetmek için analiz edilir. Verilerden anlam çıkarma tekniğinde, bakış açısı değiştirilmelidir. Böylece problemin varlığının neden olduğu kısır döngüsel bakış açısına yönelten düşünce bloklarından kurtarılır. Yapısal veri analizi, problemi küçük parçalara ayırma işlemidir. Matematikte çoğu ispatlar oldukça fazla dolambaçlar içeren süreçler ortaya çıkarır. Bir teoremi ispatlamaya başlayan birisi özünden sapmış bir şekilde birçok yol dolaşarak çözüm ararken kaybolur. Geriye dönük veri analizinde (Retrograde analysis) geriye dönük çözüm metotları geliştirilir.

Toplanan ham verilerin çeşitli tekniklerle analiz edilerek çözümlenmesi ile bulgular(kanıt bilgiler) elde edilir. Bulgular, araştırmacının konu ile ilgili tartışmalarda kullanıldığı kanıtlardır. Problemin çözümü için bulguların yorumlanması ve çeşitli bakış açılarının değerlendirilmesi gerekir.

Değişkenler arasındaki ilişkileri belirlemek için, algoritmalar adı verilen çeşitli matematiksel formüller veya matematiksel modeller verilere uygulanabilir.

Neden veri anlizi?

- Bilgi toplayarak gözlem yapmak.
- Sorgulanmaya rakamlar ile yanıt vermek.
- Mevcut durum hakkında yorum yapmak.
- Neler olduğu hakkında kestirim yapmak ya da tahmin etmek
- Gelecek için öngörüde bulumak.

Veri Analizinde Bünye Oluşturmak:

Veri analizinde öğrenen zekanın temelini oluşturan kurumsal hafıza ve sorgulama kültürü yok ise beyin de yoktur. Kurumsal hafıza, geçmişi hatırlayan ve değerleriyle birlikte tecrübeye dönüştürendir. Ne kadar büyük ve gelişmiş olurlarsa olsunlar, bilgiyi doğru yönetemeyenler yüzeysel değerlendirme hatasına düştüklerinde kaybetmeye mahkûmdurlar. Bilgi yığınları kurumların arşividir. Arşivler, kurumların geçmişini ve geleceğini aynı anda aydınlatan değerlerdir. Arşivleme yetersizliği birikimleri yok eder, karışıklık çıkmasına neden olur. Bulunamayan ya da kaybolan bilgi ve belgeler yüzünden organizasyonlar çok büyük zararlara uğrarlar. Bu nedenle veri analiz yapacak öğrenen zekayı oluşturmak için, bilgi yığını içerisindeki değişimlerin izlerini algılama yeteneği geliştirilmelidir. Belirsizliklerin, hataların, eksik bilgi sayısının oldukça fazla olduğu değişimlerin izlerinde doğruluğu arttırmak için araştırma yapmaya yönelik akıl geliştirilmelidir. Yığın içerisinde aranan bilginin bulunabilmesi ve analiz edilebilmesi için ekip olmayı becerebilecek problem çözmeye odaklı ortak akıl geliştirilmelidir. Birlikte bilgi yığınları içerisinde analiz yapan ekip değişimleri ve sapmaları bulurken iş bölümü yapılmasını öğrenerek süreç yönetmeye ve problem çözmeye yönelik katılımcı bir akıl geliştirirler. Başarıya giden yolda ekip olma ve ekiplerin birbirlerini algılamaları hedefe yönelik katılımcı akıl ile mümkündür. Kalite gücünü fark eden akıl sadece tehditleri değil fırsatları yakalamada da farklı olmak gerektiğini hissederek aynı anda tek bir noktaya odaklanabilmelidir.

Problem çözmeye yönelik geliştirilen ortak akıl karar vermede, alternatifler arasından birini tercih ederken, problemlerin çözüm alternatifleri arasından seçim yapılmış olur. Karar verme denilen seçim yapmak, çoğunlukla problem çözmeyle ilişkilidir. Bu nedenle karar vermek risk almaktır. Karar verme sürecinde:

- Bilinçlenme şüreci (şuur) oluşur.
- İzleme ve değerlendirmeler yapılır.
- Akıl yürütülür.
- Belirginlik, belirsizlilik ikilemindeki muğlaklık denizinde var olmaya çabalanılır.
- Olasılık yüzdeleri ile faydalar ve zararlar kestirilir.
- Risk alınır.
- Sonunda duygusallık yaşanır (spişmanlık duyup duymama)

Karar verirken karşılaşılması olası değişiklikler, tuzaklar ve belirsizlikler hakkında yorumlar yapılırken, geçmişte elde edilen deneyimlere dayalı sorgulamalar yapılır. Öte yandan yönlendirmelerin yoğun baskısı altında karar verirken sosyal medya, patron, sevgili, anne, baba ya da yönetenlerin sizi yönlendirmeye çalıştığını görürsünüz. Onlara göre önerdikleri seçeneklerden biri kesinlikle sizin için daha parlak göründüğüdür. Eğer onların seçeneklerinden birine karar verirseniz sıralı oyunlar başlar ve süreçlerin bir yerlerinde verdiğiniz karar, sizi hiçbir zaman istememiş olduğunuz bir yere götürdüğünü ve tuzağa düşürdüğünü anlarsınız. Doğal olarak karar verme, belli bir olasılık stratejisine uygun olarak yapılmalıdır. Karar vermek, yönetmektir.

Karar vermede rekabet, avantaj, kar veya başarı elde etmek için yapılan mücadeledir. Rekabet durumunda; taraflar kendileri için en iyisini yapmaya çalışırken aynı zamanda başkalarından da daha iyi yapmaya çalışırlar. İşbirliği hem kendi amaç ve çıkarlarını korumak, hem de başkalarının çıkar ve amaçlarına dikkat ederek birlikte çalışmaktır. İşbirliği, rekabetin zıddı bir durumu ifade edip karşılıklı güveni artırıcı bir özelliğe sahiptir.

Karar vermede ikilem ise iki şıktan birini seçme mecburiyetinde kalmak zorunluğunda kararsız kalmak, ve istenmeyene karar vermektir. Alternatiflerden birisi daha cazipse seçim yapmak kolaylaşmaktadır.

2.1. Veri - Bilgi

Veri yönetenler, araştırmaya yönelik bilinç geliştirenlerdir. Makine, biyoloji ve beyin konusunda yaşanan bilimsel ve teknolojik gelişmeler insanı, çevremizi ve ötesinde kainatı çözmemizi sağlıyor. Bilim, algoritmalardan ibaret olan organizmaların şifresini çözme yeteneğine kavuştu. Biyokimyasal veriler elektronik sinyallere çevrilerek bilgisayarlarda analiz edilebilmesi sağlandı. ***Bilginin verdiği gücü kullanmak isteyenler ile o güçten korkanlar, insan bedenine hükmetme yeteneğiyle yaşamın geleceğine karar vermeye yöneldiler.*** Saldırganlıklar, deneylerin birer parcası mı? Verinin kimin elinde bulunduğu bu yüzden her zamankinden daha önemli.

Bugün insanların çoğu veri denince tek şeyin bilgisayarlar olduğunu sanıyorlar, bilgilerine saldırıdan endişe ediyorlar. Ancak beden çok daha büyük bir hedef. Asıl hedef ise beyindir. Yaşarken bıraktığımız dijital izlerimiz, birileri tarafından sınıflandırılmamıza yardımcı mı oluyor?

Veri, anlam kazanmamış, ilişkilendirilmemis, özümlenmemiş, işlenmemiş gerçekler ya da bilgi parçacıklarıdır. Herhangi bir içerikten yoksun formlardadırlar.

Enformasyon sözcüğü, İngilizce'deki 'information' sözcüğünün Türkçe'ye uyarlanmiş halidir. Enformasyon, veriye değer katılarak, verinin anlamlandırılmasıdır. Belli bir amaç için birbiriyle ilişkili verilerin biraraya getirilmesi, düzenlenmesi sonucu oluşur ve bir anlam taşır. Kurumsal olarak bakıldığında enformasyon, anlamı olan veritabanıdır.

Enformasyon veriden doğmaktadır ve enformasyon da bilgiye dönüşmektedir. Genel olarak bilgi, veri ve enformasyonun yorumlanmasıyla ortaya çıkar.

Verinin bilgiye dönüşme süreci:

- Toplama, Sınıflandırma, Düzenleme, Özetleme, Saklama, Yeniden elde etme
- İletme
- Analiz, Yorum

Veri tabanı, verilerin depolanmasını, değiştirilmesini, silinmesini, erişilmesini kolaylaştırmak için sistematik olarak dosyalar biçiminde düzenlenmiş veri topluluklarını ifade eder. Veritabanı yönetim sistemi, birbirleriyle ilişkili ve benzer verileri biraraya getiren, verilere erişimi ve verilerin yönetilmesini sağlayan sistem ve yazılımlardır.

Verilerin günümüzde hız, çeşitlilik, kapasite (hacim) açısından büyük artış göstermesi ve bu artışa teknolojinin de destek vererek, yeni çözümler üretmesi ile birlikte “Büyük Veri” kavramı ortaya çıkmıştır. Büyük veri genel olarak veri kümelerini anlatmak için kullanılan bir terimdir. Verinin devasa boyutları ile bundan fayda sağlamak için gereken analizlerin karmaşıklığının birleşmesi, yeni sınıf teknolojilerin ve bunları yönetecek araçların gelişmesine neden olmuştur.

2.2. Değişkenlerin Ölçümü

Matematikte değişken, kümenin herhangi bir elemanını temsil etmek için kullanılan bir semboldür. Sayılara ek olarak değişkenler, genellikle vektörleri, matrisleri ve fonksiyonları temsil etmek için kullanılır. Değişken kullanarak cebirsel hesaplamalar yapmak, tek bir hesaplamada bir dizi problemi çözmeyi sağlar. Tipik bir örnek, verilen her denklemin, denklem katsayılarının sayısal değerlerini, onları temsil eden değişkenlerin yerine basitçe değiştirerek, her ikinci dereceden denklemi çözmesine izin veren ikinci dereceden formüldür ($y=ax^2 + bx +c$). Denklemlerde bilinmeyenler genellikle x, y ve z; bilinenler ise a, b ve c harfleriyle gösterilir. Kesirli, ondalıklı, tamsayılı, ve karekter olarak tanımlanırlar.

Değişkenlerin alabileceği değerlerin nasıl ölçüldüğünü belirlemek, veri analizi için çok önemlidir. Nominal ölçekte veriler sadece niteliklerine göre isimlendirilip gruplandırılabilir. Cinsiyet, medeni durum, din, parça numarası, çalışma durumu, meslek vb. kalitatif veriler nominal ölçekle ölçülürler. Bu verilerin frekansları ve modları belirlenebilir. Nominal bir değişkende ölçüm düzeyleri arasında bir sıralama ya da uzaklık-yakınlık gibi belirli bir mesafe yoktur.

Ordinal ölçeğin amacı, bir konu hakkındaki düşünceleri belirli bir öncelik sırasına koymaya hizmet etmektir. Örneğin restaurant, otel, hastane gibi hizmet işletmeleri yiyecek kalitesi, servis, tesisler, konukseverlik gibi özellikler açısından “mükemmel”, “iyi”, “orta”, “kötü”, “çok kötü” olarak sınıflandırılabilirler. Bu ölçekle ölçülen verilerin frekansı, modu, medyanı, yüzdelik değeri ve korelasyonu hesaplanabilir. Aralık Ölçeğinde veriler için ölçekte tanımlanan birimlerin birbirinden eşit uzaklıkta bulunduğu bir ölçü birimi ve rastgele bir başlangıç noktası tanımlanabilir. Bu ölçekle ölçülen verilerin aritmetik ortalama, standart sapma, yatıklık ve basıklık ölçüleri hesaplanabilir.Oran Ölçeği, aralıklı ölçeğin özelliklerini taşımasının yanı sıra verilerden iki ölçümlemenin oranlanmasıyla anlamlı kıyaslamalar elde edilir. Uzunluk, parasal değerler, zaman, güç değerleri oran ölçeğiyle ölçülürler.

Ölçmede dikkat edilmesi gereken husus, ölçek güvenilirliği ve ölçek geçerliliğidir. Kullanılan ölçeklerin istenen bilgiyi toplamaya elverişli olup olmadığı büyük önem taşır.

2.3. Örnekleme

Bir araştırmada ana kitlenin tümünü gözlem altına alma ya da ana kitleyi tam olarak sayma yerine, ana kitleyi nitelik ve nicelik yönünden temsil eden bir örneğin alınması yoluna gidilir. Böylece örnekten elde edilen bilgilerin, belirli olasılık kademelerinde ana kitle için de geçerli olduğu kabul edilebilir. Bu anlamda örnekleme; ana kitleyi nitelik ve nicelik yönünden temsil edebilecek bir kümenin oluşturulması işlemidir.

Söz gelimi, analog sinyalin genliği, frekans, faz ve zamanla değişir. Verilerin geçerli olabilmesi, analog sinyallerden sağlıklı örnekleme yapılması ile mümkündür. O halde doğru örnek alabilmek için örnekleme sayısı ve örnekleme zaman aralığı doğru belirlenmelidir. Amaç, toplanan verilerdeki örnekleme aralığı ile gerçek sinyalin farklılığını minimize etmektir. Davranışların analog sinyal özellikleri, sınıfların sınıf sayısı, aralığı ve sınırları doğru belirlenmelidir.

$X(t) = A*\cos(2\pi ft+\phi)$

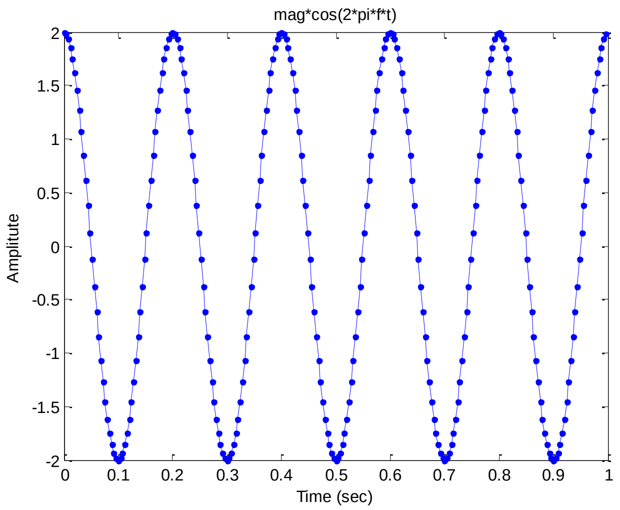
Burada, A: genliği, f: frekansı, ϕ: fazı temsil etmektedir. Zamanla (t) genliği değişen analog sinyalin genel gösterim denklemidir.

$t_i = 0{:}T_s{:}1-T_s$

$T_s=1/f_s$

$fs \geq 2f$ (Nyquist teoremi), Bir peryottaki örnek sayısı, Ps=20 ise fs=Ps * f alınır.

```
clear all
 close all
  mag  = 2;      % magnitude (arbitrary units)
  f = 5;      % frequency in Hz
  Ps=50;         % number of sampling on a periot
  samp = f*Ps;    % sampling rate in Hz
  t        = 0:1/samp:1-1/samp;  % time (1s of data)
  N        = length(t)
  x = mag*cos(2*pi*f*t);     % the signal equation
  figure
  plot(t,x,'.-');
  xlabel('Time (sec)');
  ylabel('Amplitute');
  title('mag*cos(2*pi*f*t)')
```



Örneklemenin diğer bir amacı ise tahmin yapmaktır. Örneğin, bir çuval dolusu pirinç içinden alınacak bir avuç pirincin incelenmesiyle tüm çuval dolusu pirinç hakkında çeşitli özellikleri itibariyle genelleme yapılabilir. Burada, çuvaldaki toplam pirinç miktarı ana kitle (popülasyon)'dir. İncelemek amacıyla aldığımız bir avuç dolusu pirinç ise örnektir.

Örnekleme kuramı, ana kitleyi oluşturan her davranışın örneğe girme şansının eşit olmasını öngörür. Örnekleme kuramına göre, örneklerden elde edilen bilgilerin matematik ve istatistik tekniklerle test edilip genelleme yapılabilmesi için örneklemenin uygun olarak yapılması gerekir. Ana kitle hakkında tahminde bulunurken yapılan tahminin geçerli olabilmesi için örnekleme tesadüfi (rastgele) olabillir. Tesadüfi örnekleme kavramı, örnekte yer alacak değerlerin belirlenmesinde hiçbir dış etkinin rolünün olmamasını ifade eder.

Elde edilecek bilgi, ana kitleyi oluşturan değişik özellikteki gruplara göre farklılık gösteriyorsa, ana kitleyi oluşturan sınıflara göre örnekleme yöntemi kullanılır. Kümelere göre örneklemede kitle önce kümelere ayrılır, sonra kümelerden örnekler seçilir. Bu tür örnekleme, sağlıklı bir ana kitle çerçevesinin elde bulunmaması ya da çok büyük ana kitleden çekim yapmanın çok zor ve maliyetinin yüksek olması durumunda uygulanır. Bu tür örnekleme tesadüfi örnekleme olmasına karşın, ana kitleyi ne ölçüde temsil ettiği tartışılabilir. Bu nedenle de çok dikkatli yapılması gerekir.

Sistematik örnekleme, seçim işlemlerinin kolay olması nedeniyle özellikle ana kitlenin büyük olduğu durumlarda kullanılan bir örnekleme yöntemidir. Çok sayıda birim içeren kayıt sistemlerinin incelenmesinde. *Örneğin,* hasta dosyaları, hasta ya da işçi kayıtları, kayıt defterleri, fişler , listeler gibi. Bu yöntemde başlangıç sayısı dağılımı büyük oranda etkiler.

Örnek büyüklüğünün belirlenmesi:

Örnek büyüklüğünün hesaplanmasında aşağıdaki formüller kullanılabilir.

$$n = \left.\frac{Z^2\ \sigma^2}{(\bar{x}-\mu)^2}\right\} \ Ana\ kitle\ varyansı\ biliniyorsa$$

$$n = \left.\frac{Z^2\ s^2}{(\bar{x}-\mu)^2}\right\} \ Ana\ kitle\ varyansı\ bilinmiyorsa$$

Z : belirlenen güven düzeyi için Z tablosundan bakılan değer,
σ^2: ana kitle varyansı,
$\bar{x}$: örnek ortalaması,
μ : ana kitle ortalaması,
$(\bar{x}-\mu)$: göze alınan örnekleme hatası.

Örnek : Bir iş kolunda çalışan işçilerin ortalama saat ücretleri gerçek ana kitle ortalama saat ücretinden en fazla 100 TL sapma gösterecek şekilde ve %90 güven aralığında tahmin edilmek isteniyor. Geçmiş kayıtlara dayanarak bu iş kolu için hesaplanan standart sapma 500 olduğuna göre örnek büyüklüğü ne olmalıdır.

$$n = \frac{z^2\sigma^2}{(\bar{x}-\mu)^2} = \frac{(1{,}64)^2(500)^2}{(100)^2} = \frac{(2{,}69)(250.000)}{10.000} = 67{,}25 \cong 68$$

Bu sonuca göre, gerçek ortalamanın $\mp$100 TL sınırları içinde ve % 90 güven aralığında istenen amacı gerçekleştirmek için örnek büyüklüğü 68 olmalıdır.

Örnek büyüklüğü de belirlendikten sonra örnekleme planı oluşturulur.

Örnekleme yöntemlerine dayalı olarak yapılan tahminlerde örnekleme hataları yapılır. Bunlar örnek sayısının yeterli olmamasından kaynaklanan tesadüfi olarak yapılan hatalar ve sistematik olarak adlandırılan hatalar ise; örnekleme sürecindeki hatalardan kaynaklanır ve sonradan giderilmeleri mümkün değildir. Çünkü,

- Örnekleme yöntemi doğru seçilmemiştir,
- Ana kitle yanlış tanımlanmıştır,
- Örnekleme çerçevesinin yanlış belirlenmiştir,
- Örnekler doğru seçilmemiştir,
- Örnek büyüklüğü doğru hesaplanmamıştır.

2.4. Verilerin Toplanması

Birincil ve ikincil veri kaynaklarının araştırılması ve verilerin toplanması için kullanılan bazı yöntemler mevcuttur. Bu yöntemler; literatür araştırması, gözlem, deney ve anket yöntemleridir.

Literatür Araştırması

Araştırma konusu ile ilgili benzer bilimsel araştırmalar ve raporlar, indeksler vb. kaynaklar bilimsel literatürde, ilgili kuruluşların sitelerinde detaylı olarak araştırılmalıdır. Bu şekilde bulunabilecek kaynaklardan veriler hazır olarak elde edilebileceği gibi, veri toplama araçlarının tasarımı için faydalı bilgiler de elde edilebilir.

Gözlem: Olayları belirlenen sırada sistemli ve amaçlı bir biçimde inceleyerek bilgi toplama yöntemidir.

Her tür bilginin toplanmasına uygun değildir. Mekanik ve elektronik araçların kullanılmaması halinde, olayın ayrıntılarını yakalamak veya hatırlamak gerrekir. Gözlem altında olduklarını anlayanlar, olağan davranışlarını değiştirebilirler, Bazı durumlarda maliyetleri yüksektir.

Deney: Bir hipotezin sınanması amacı ile koşulları deneyi yapan tarafından hazırlanan ve bağımsız değişkenin bağımlı değişken üzerindeki etkinin şiddetini ya da yönünü ortaya koymayı amaçlayan bir gözlem türüdür.

Veri toplama teknikleri:

- Bilgi kaynakları ve verilerin toplanacağı ortamlar
- Çevresel riskler
- Yeni donanım ve/veya hizmet girdisi
- Cihaz ve malzeme konum takibi
- Kablolama (Modernleştirme ve sadeleştirme)
- Yedekleme, Güvenlik, Enerji
- Çevresel Riskler

2.5. Verilerin Doğrulanması

Teknolojinin hızla gelişmesiyle artan ölçüm cihazlarına orantılı olarak veri sayısı ve türleri de artmaktadır. Aynı zamanda bir çok işlem elektronik ortamda kayıt edilmekte, bu kayıtlar saklanabilmekte ve anında erişilebilmektedir. Günümüzde veri tabanları büyük boyutlarından ve birçok farklı kaynaktan gelmelerinden dolayı gürültülü, eksik, tutarsız, çelişkili veriler ile doludur. Hatalı, eksik ve kayıp veriler ise fark edemezseniz, bilinmezlik denizinde sizleri yıkıma götürmek için sinsice tuzaklar hazırlamaktadır. O halde verilerin doğrulanmasında mükemmel doğruluk, kesinlik ve belirlilik mümkün değildir. Önyargılar genellikle “bilinmeyenlerdir”. Güven aralığı önemlidir. Sistematik hataların (önyargıların) yakalanması zordur çünkü genellikle bu hataların farkında olunmaz.

Verilerin doğrulanmasında sorgulanma önemsenmelidir. Sorgulamazsanız, duygusallıktan dolayı inanırsınız ya da inandırılırsınız. İnandırılma sadece kişisel değildir, günümüzde yazılımlar, sistemler de sizleri inandırma üzerine manipülasyon yapmaktadır.

Eksik Veri: Veri tabanında, çeşitli sebeplerle bazı verilerin eksik olması durumudur.
Hatalı veri: Ölçülen bir değerdeki hata ya da yanlış nitelik değerleri, hatalı veri toplama gereçlerinden, veri girişi ve veri iletimi problemlerinden, teknolojik kısıtlar ve tutarsızlıklardan dolayı veri tabanındaki bazı verilerin tanımı ile çelişmektedir.
Uyuşma sorunu: Veri tabanındaki birden fazla alandan gelen verilerin birbiri ile uyuşmaması durumudur.

Veriyi ayıklama:

- Yığın içerisinde yinelenenleri bulma ve kaldırma
- Hatalı olanları belirleme
- Gereksiz, anlamsız verileri belirleme

Hata:

Sistematik hata kaynakları: Eksik veri, kayıp veri, yanlılık, bilinmezlik, belirsizlik,

Hata: Önyargı veya sistematik hata, rastgele hatalar, hassasiyet, değişkenlik.

Sorununa doğru ve zamanında tanı konulması ve bunun önlem alınmasında aksama olması. Hata olabilir diye sorununa yönelik olmayan gereksiz çözümlerin engellenmesi için çok önemlidir. Hataların bir kısmının farkına varmıyoruz. Hatayı kabul etmeyen bir sistemden, hataları kabul eden ve oluşmasını engellemeye çalışan bir sisteme doğru yol almamız ile mümkün. Ancak en büyük vurguyu, ekip çalışmasının önemi almakta. Tüm işlevsel süreçlerin vereceği bilgilerin dikkatli izlenmesi, bazı durumlarda doğru tanıya ulaşmada çok önemli ipuçları verir. Bilişim teknolojilerinin tanısal süreçlere entegre edilmesi ve tanı hatalarını tanınması, azaltılması ve nedenlerinin öğrenilmesi için yöntemlerin geliştirilmesi, tanısal performansı arttıracak yöntemlerin ve kültürün geliştirilmesi, tanı hatalarının bildirilmesini teşvik edecek bir ortamın ve yükümlülük sisteminin hataların azaltılması hedefine yönelik olarak kurulması, ve tanı süreci ve tanı hataları ile ilgili daha çok araştırmanın yapılmasının teşvik edilmesi, raporda vurgulanan diğer tanı hatalarını azaltıcı öneriler. Tanı hatalarını görmezden gelmek veya kaçmak yerine bir daha oluşmaması için kayda geçerek gerekli önemleri almak, tanı sürecinde yer alan herkes ile etkili ve profesyonel bir iletişimi sağlamak ve hepsinden önemlisi tüm sistemi bu yönde yeniden oluşturmak. Görüldüğü gibi hiç de kolay bir süreç değil, ama durumun farkında olmak bu süreç için en önemli adım.
Veri yığının davranışından elde edilen sapmalar ve tepkilerden kestirimler yapılarak performans izlenir. Hatalar ve tıkanmalar bulunduğunda fonksiyon düzeltilir. Böylece sistem sürekli öğrenen yapıya dönüşür. Algılayıcıların ekip olarak doğru fonksiyonu belirlemeyi öğrenmesi için kestirilmiş değerler ile kullanım sonrası ortaya çıkan hataların düzeltilerek güncellenmesi gerekir. Tüm olayların temelinde değişkenlikler vardır ve hataların büyük bir bölümü değişkenlikten kaynaklanmaktadır. Değişkenliğin özelliği belirlenirken hata kaynakları doğru tespit edilmelidir. Belirsizlik yaratan değişkenleri izlemek için tekrar eden değişkenler ayrıştırılmalıdır. Toplanan verilerden değişken olanlar kıyaslanarak belirlenir. Tüm olayların temelinde değişkenlikler vardır. Hataların büyük bir bölümü değişkenlikten kaynaklanır. Değişkenlerin özelliği belirlenmeli, hata kaynakları tespit edilmelidir. Belirlenmiş tekrar eden değişkenler elimine edildikten sonra kararlı belirsiz değişkenlerin azaltılması önemsenmelidir.

Çok yoğun bilginin toplandığı bir ortamda en doğru yaklaşım, öncelikle hatalı olanların ayıklanmasını da içeren öğrenen algoritmaların geliştirilmesidir. Öğrenen algoritmalar tarafından sınıflandırılan bilgiler içerisinde aranan bilgiye hızlıca erişim sağlanmalıdır. Bilgilerin saklandığı bellek ortamlar, hem maliyet hem de kapasite büyümesinde sıkıntı gösterdiğinden bellek alanını verimli kullanacak sıkıştırma teknikleri geliştirilmelidir.

Belirsizlik:

- Bilimsel belirsizlik
- Sistematik Belirsizlik (Önyargı):
- Tahmin belirsizliği: Model belirsizliği, Parametre belirsizliği
- Parametre Belirsizliği: İstatistiksel (rastgele), Sistematik (önyargı)

İstatistiksel Belirsizlik: Rastgele! Mümkün olduğu durumlarda tekrarlanan örnekleme ve bunu
izleyen istatistiksek analiz ile belirlenebilir. Pek çok parametre için, aynı işletim koşulları altında tekrar tekrar örnekleme yapmak ya da veri toplamak mümkün olmayabilir. İstatistiksel belirsizliği genel olarak zaman içerisinde ortalamaya ulaştığı varsayılır. İstatistiksek belirsizlik veri içerisindeki "gürültüdür". İstatistiksel parametre belirsizliği, ölçüm ekipmanının nominal kesinliğine yakın olabilmektedir.

Sistematik Belirsizlik (Önyargı): Ölçülen değerler sürekli olarak 'gerçek' değerden daha yüksek (ya da daha düşüktür). Önyargılar veri örnekleme ve istatistiksel analiz yoluyla tespit edilemez. Önyargılar aşağıdakiler ile belirlenebilir:

- Veri kalitesi incelemeleri ve diğer Kalite Güvencesi / Kalite Kontrol önlemleri
- Verinin başka bağımsız veri setleri ile karşılaştırılması
- Önyargılar zaman içerisinde ortalamaya ulaşmaz, bu nedenle rastgele belirsizliklere kıyasla daha ciddi bir sorun teşkil ederler.
- Eğer veri kalitesi düşerse (veya yükselirse) ve geçmiş veriler revize edilemezse, önyargılar zaman içerisinde artabilir (ya da azalabilir).
- Zaman içerisinde önyargılarda meydana gelen değişiklikler, emisyon trendleri ve tahmin edilen emisyon indirimleri üzerindeki etkileri nedeniyle özellikle sıkıntılıdır.

Parametre Belirsizliklerinin Sebepleri: Ölçüm cihazlarındaki rastgele hatalar (paralaks hatası, sıcaklık değişikliği, vb.) Ölçüm cihazları sistematik önyargılar oluşturabilir. Kesin olmayan kalibrasyon, hatalı ölçüm ekipmanı, çevresel faktörler, operatör hatası, mükerrer sayım, verinin dışarıda bırakılması, vb. Parametreler temsili olmayan örneklere de dayalı olabilir. Yakıt teslimatı haftalık olarak yapılırken, yakıt örneklerinin aylık olarak alınması. Verinin süreç başlangıç ve kapanma koşullarını ya da düzensiz işletim koşullarını açıklamaması.

Veriler eksik olabilir, tekrarlayabilir veya hatalar içerebilir. Veri temizleme gereksinimi, verilerin elde edilmesi ve saklanması ile ilgili sorunlardan kaynaklanacaktır. Veri temizleme, bu hataları önleme ve düzeltme işlemidir. Veri temizlemenin görevleri arasında, kayıt eşleştirme, verilerin yanlışlığını belirleme, mevcut verilerin genel kalitesi, tekilleştirme ve sütun bölümlemesi bulunur. Bu tür veri problemleri, çeşitli analitik teknikler aracılığıyla da tespit edilebilir. Belirli eşik değerlerin üstünde veya altında olağandışı miktarlar incelenebilir. Aykırı değer tespiti için Nicel - Sayısal veri yöntemleri yanlış girilen muhtemel verileri kaldırmak için kullanılabilir. Örneğin bir bölgenin yıllık sıcaklık değişiminde Temmuz ayının 15 ile 20 si arası eksik ise elimde en az 5 yıılık değişim var ise bu eksik veriyi tamamlamak mümkündür. Aritmetik ortalama, stokastik, türevsel değişimlerden eksik veri tamanlanabilir.

Önyargı veya Sistematik hata: Doğruluk eksikliği. Birçok farklı ölçümün ortalaması güvenilir bir miktarda ve yönde gerçek değerden farklılık göstermektedir. Dahil edilen ilgili bütün süreçlerin yakalanmasındaki başarısızlıklardan veya kullanılabilir mevcut verilerin gerçek dünyadaki durum ve koşulları tam anlamıyla yansıtmamasından veya kullanılan araçlardan kaynaklı hatalardan ortaya çıkmaktadır.

- Rastgele hatalar: Ortalama değerin üzerinde veya altındaki rastgele varyasyon. Rastgele hata, kesinlikle ters orantılıdır. Genellikle rastgele hataların miktarı ortalama bir değer ile belirtilmektedir, fakat ortalama değerde önyargı olabilir veya olmayabilir. Bundan dolayı, rastgele hata sistematik hatayla karşılaştırılabilen farklı bir kavramdır.
- Ölçme hatası
- Örnekleme hatası

Yoğun bilgi yığını içerisinde belirsizliğin bölgesi tanımlanmalıdır. Belirsizlik, nedenini bilememek, sonuçları tahmin edememek, sistemi anlayamamak, herhangi bir fikir yürütememek, sorulara cevap verememek, verilen cevaplarla tatmin olamamaktır. Deprem ne zaman olacak? Yanardağ ne zaman faaliyete geçecek? Dünyanın sonu ne zaman gelecek?

2.6. Veri Madenciliği ve Sınıflandırma

Çok büyük boyutlardaki bilgi yığını içerisinden aranan verinin bulunup çıkarılması için sınıflandırma, veri madenciliği ve veri füzyonu yöntemleri kullanılmaktadır. Veri madenciliği yararlı bilginin büyük miktardaki veri yığını arasından bulunup çıkarılmasıdır.

Veri madenciliğinde aranan bilgi toplanan bilgi yığının göstermiş olduğu davranışın meydana getirdiği izlerin türünü göre hızlıca bulunması gerekir. Verilerini kayıt edildiği ortamlardaki ve dolaştığı veri kanallardaki trafik analizinden iz bulan yönetim sistemleri tasarlanmalıdır. Merkezi bilgi işlem birimi algılayıcıların geliştirdiği davranışlardan bilgilerin toplanmaya başladığı anı keşif ederek erken uyarı oluşturması gerekir. Böylece bilgiler toplanmaya başladığı anda alınacak pozisyonda belirlenmiş ve davranışın vereceği tepkilerin sınıflandırılması yapılmış olacaktır. Diğer taraftan alınacak pozisyonu belirlemek için erken uyarı mesajları da elde edilmiş olacaktır.

Merkezi işlem birimi, bilgi toplama ve transfer etme birimleri arasındaki iletişim ortamında en kestirme yolların tablosunu sürekli güncelleyerek veri tabanında tutması gerekmektedir. Kimin hangi zamanda kime en hızlı veri ileteceği güzergahı tüm algılayıcılar ve toplama istasyonları bilecektir.

Veri füzyonu ise algılayıcılardan gelen bilgileri kaynaştıran ve bünyeye birleştiren algoritmalardır. Hedeflerin davranışlarını bulmada, tanımlamada ve takip etmede gerekli olan bilgileri toplayan ve sentez yapmayı öğrenen veri füzyonu algoritmaları kullanılmaktadır. Bilgileri birleştirme işlevi, başta insanlar olmak üzere canlıların her zaman farkında olmadan yaptıkları sürekli bir işlemdir. Bir hareketin davranışının nedenini ve vereceği tepkileri kestirebilmek için toplanan verilerden yaşayan bir organizma oluşturulması gerekir. Öğrenen algoritmalar ile sürekli kendini geliştirerek yaşayan organizmanın tepkisel davranışını doğru kestirebilmek için diğer algılayıcılardan gelen bilgilerin organizmaya bütünleştirilmesi gerekmektedir. Örneğin tehditlere ait hedeflerin oluşturduğu bilinmeyen sayıdaki izler toplanan bilgilerin bütünleştirilmesi ile hedeflerin yerleri ve davranışları belirlenebilir. Dağınık noktalara yerleştirilmiş çok sayıdaki algılayıcılardan gelen bilgiler hem çok karmaşık hem de çok fazla çeşit içerdiklerinden dolayı, toplanan bilgiler analiz edilirken karmaşık algoritmalar ve paralel işlemciler kullanılır.

Toplanan bilgilerden öğrenen organizma oluşturabilmek için organize olabilen, mevcut bilgi kaynaklarını bir araya getirerek iyi işleyen düzeni kurabilmek gerekir. Bilgileri toplayan, sınıflandıran, bütünleştirerek organizmalar oluşturan organizasyonuz bunları yapabilecek ekip yeteneği geliştirilebilmelidir. Organizmanın ekip olabilmesi için denge oyununda geleceği kestirmesi ve öngörüde bulunması gerekir. Küçücük dağınık bilgi organizmaları organize olup organ gibi davranmaya ve organlardan da hisseden canlılar gibi davranmayı öğrenmeleri gerekmektedir.

2.7. Veri Analizinde Akıl Oyunları

Karar verenlerin diğer düşüncelerle etkileşimini ve özellikle rekabet halinde olduğu durumları modelleyen bir yaklaşım olması akıl oyunu kuramının en temel özelliğidir. Akıl oyunları davranışların değiştirilmek istenmediği denge noktasını bulmaya çalışır. Oyun kuramı fikri üzerine Nash dengesi geliştirilmiştir. Akıl oyununa ilişkin yöntemler eleştiriden uzak değildir. Oyunların çoğu karşılıklı etkileşim ve rekabet içerisindedir. Akıl oyunları kuramının en temel özelliği karar verenlerin düşüncelerini ve rekabet halindeki sosyal durumlarını modelleyen bir yaklaşım olmasıdır.

Birden fazla sığınağın bulunduğu bir savaş alanında bir askerin tepesinde daireler çizen uçağın içerisinde bomba bırakmak için fırsat kollayan bir pilot düşünün. Normalde asker çevredeki en sağlam görünüşlü sığınağı seçmesi ve orada saklanması gerekir. Aynı anda pilot da askerin en iyi sığınağı seçeceğini düşünerek orayı bombalamak isteyecektir. Bunu bilen asker o denli sağlam görünmeyen ikinci sığınağı seçmesi gerekmez mi? Eğer ikisi de çok akıllıysa, olasılıklara dayanan stratejiler izlerler. Örneğin asker çevredeki çeşitli sığınaklar arasında ona en fazla kurtulma şansı verecek özelliklere sahip olanları arar, bundan sonra nereye saklanacağını belirlemeye çabalar. Pilot da askeri vurma şansı en yüksek düzeyde olduğu sığınağı belirlemek için benzer biçimde olasılıklardan yararlanır. Bu saçma gelebilir ama ikisi de akılcı davranabiliyorsa yapacakları budur. Doğal olarak asker hareketlerini gizlemezse pilotun işi kolaylaşır, buna karşılık pilot da nereyi bombalamayı tasarladığını askere sezdirmemeye çalışmalıdır.

Oyun kuramı, insanların karar verirken etkileşim içinde oldukları ile giriştikleri çatışma ve işbirliğinin modellendiği bir yaklaşımdır. Oyun, gerçek bir işletme probleminin ya da durumunun özet biçiminde modellenmesiyle ortaya çıkmaktadır. Oyun teorisindeki ünlü 'Nash dengesi', 1928 doğumlu Nash tarafından geliştirilmiştir. Nash, 1948'de kendini Princeton'da matematik öğrencisi olarak buldu. O dönemde dünyanın en iyi matematikçileri, fizikçileri, mantıkçıları oradaydı. Nash doğuştan rekabetçi bir insandı. Oyun teorisi ile ilgilenmeye başlamasının nedeni matematik dalında henüz çözülmemiş çok sayıda problemin mevcut olmasıydı. Bu yeni matematik dalının kurucusu John Von Neumann'dı. Macaristan Yahudisi bir zengin ailenin çocuğu olan dahi matematikçi, 1928 yılında yayımladığı bir makalesiyle bu yeni matematik dalını kurmuştu.

2015 yılında bir kazada ölene kadar konferanslar ve ders veren John Forbes Nash Jr. -dahi matematikçi, rasyonel davranış teorisinin mucidi, düşünen makineyi öngören adamdır. Hem rekabetçi, hem işbirliğine dayalı davranışları modelleyebilmesi oyun teorisinin en önemli katkılarından birisidir.

Tutuklunun Açmazı (Mahkum Teoremi), Oyunlar Teorisi, esas olarak iki teorem üstüne kurulu. Bunlardan birincisini, yani minimum - maksimum teoremi adıyla bilinen teoremi, geçen yüzyılın bir başka önemli matematikçisi John Von Neuman geliştirdi. İkincisi ve çok daha önemlisini ise Nash geliştirdi. Buna da 'Nash Dengesi' denir. Nash dengesiyle ilgili teorem hemen dönemin en iyi beyinleri tarafından test edildi. Bu testlerden biri için geliştirilen 'oyun'lardan birinin adı 'Tutuklunun açmazıydı. Bu oyunu, Nash'in doktora hocası Al Tucker icat etmişti. Oyun şöyleydi: Aynı suçtan ötürü iki kişi tutuklanır ve ayrı ayrı odalarda sorgulanır. Her tutukluya üç seçenek verilir: İtiraf etmek, Ötekini suçlamak, Sessiz kalmak. Tutuklu açısından en iyi seçenek itiraf etmektir. Eğer öteki tutuklu da itiraf ederse, en azından çok ağır bir ceza almaktan kurtulacaktır, yok öteki sessiz kalırsa yegâne tanık olarak cezadan da kurtulabilecektir. Yani, itiraf 'baskın strateji'dir. Ama işe bakın ki, eğer birlikte olsalar, ya da işbirliği yapabilseler, her iki tutuklu da kendi iyilikleri için sessiz kalacaktır. Yani, işbirliksiz (non-cooperative) oyundaki baskın (dominant) strateji ile işbirlikli oyundaki baskın strateji birbirinden epey farklıydı. 'Tutuklunun açmazı' oyunu, Nash'in denge kavramıyla çelişiyordu. Çünkü Nash, her oyuncunun kendi en iyi stratejisini izleyeceğini, çünkü öteki oyuncuların da öyle yapacağını varsayar. Oysa oyun bunun illa ki böyle olmayacağını gösteriyordu. Bilindiği gibi sıfır ya da sabit toplamlı olmayan oyunlarda her oyuncunun ayrı ayrı kazancı vardır diğer bir ifadeyle oyuncuların kazanç ve kayıpları birbirine eşit değildir dolayısıyla toplamları da sıfır ya da sabit bir sayıya eşit değildir. Bu tür oyunlar sıfır toplamlı oyunlara dönüştürülemez.

Nash dengesinde amaç ön görüde bulunmaktır. Kısaca öngörmektir. Nash, herhangi bir stratejik etkileşimde, bir oyuncunun en iyi seçiminin ya da hamlesinin, öteki oyuncuların ne yapacaklarına dair inancına sıkı sıkıya bağlı olduğunu fark etti. Nash, her oyuncunun, öteki oyuncuların yapabileceği hamle seçeneklerine bakarak en uygun hamleyi seçtiği duruma bakmamızı önerdi.

İki oyuncu, birbirinden bağımsız olarak ve aynı anda, 180'le 300 lira arasında bir miktar seçsinler. Bu oyunda seçilen en düşük miktar her iki oyuncaya ödenecektir. İki miktar arasındaki fark büyük miktarı seçenden alınacak küçük miktarı seçene verilecektir. İki oyuncu aynı miktarı seçerse, ikisine de seçtikleri miktar ödenecek, ayıca transfer yapılmayacaktır. Örneğin seçilen büyük miktar 230, küçük miktar 200 olsun. Bu durumda küçük miktarı seçene 230, büyük miktarı seçene 170 ödenecektir. Sonuç olarak her ikisi de 180 i seçecektir. Bu oyunda fark değil de belirlenen miktar büyük miktardan seçenden alıp küçük miktara seçene verilirse. Belirlenen miktarın büyüklüğü, karar vermede önemli rol oynayacaktır. Belirlenen miktar küçüldükçe seçilecek miktar 300'e doğru olacaktır.

Etkileşimli karar vermede, ekip içerisindeki her bir kişinin ödemesi gereken bedel ve elde edilecek fayda kararlarına yansır. Karar vericiler bir ekibin oyunculardır. Seçimleri diğer kişilerin tercihlerine bağlı olarak farklılık gösterir. Bu durumda taraflar davranış biçimleri ile ilgili olarak birtakım kurallar üzerine anlaşmayı tercih edebilirler. Dolayısıyla her karar bir risk içerdiğinden etkileşimli karar ortamında ya da çatışma altında karar verme durumunda ekip üyelerinin riske karşı tutumları önem kazanmaktadır.

Karar verici olarak alternatifler arasından tercih yaparken, **birbirleri ile etkileşim içinde işbirliği modellerine dayalı** akıl oyunları kuramının çok iyi bilinmesi ve değişimlerin sürekli izlenmesi ve sorgulanması gerekir.

Oyunda denge kuramı faydaya dayalı beklentilerin karar üzerindeki etkisi göz önüne alınır. Etkileşim temelinde karar verilirken ortaya çıkan sorunların çözülmesinde oyun kuramının kuralları geçerli olmaktadır. Ödenmesi gereken bedel ve elde edilecek fayda kararlara yansır. Karar vericiler oyunculardır. Seçim diğer etkileyicilerin tercihlerine bağlı olarak farklılık gösterir. Bu durumda etkileyici taraflar birtakım kurallar üzerine anlaşmayı tercih edebilecekleri davranışlar sergilerler. Dolayısıyla her karar bir risk içerdiğinden karar verilen ortamında ya da çatışma altında karar verme durumu da riske karşı alınacak tutumları belirler.

3. İstistiksel Veri Analizi

İstatistik: Bilgilerin sistematik olarak toplanması ve işlenmesi sonucunda risklere ve fırsatlara yönelik tahminde bulunmayı ve yorum yapmayı sağlayan bilim dalıdır.
İstatistik, verilerin toplanması, işlenmesi, analiz ve yorumunda kullanılan metodlar bütünüdür.

Neden istatistiksel veri analizi?

- Gözlem yapmak, bilgi toplamak.
- Sorgulanmaya rakamlarla yanıt vermek.
- Yorum yapmak. Kestirim yapmak, Tahmin etmek
- Öngörüde bulunmak.

Sayılabilir veya ölçülebilir özellikleri (değişkenleri) içeren, aralarında bir çok benzerlikler olmakla beraber farklılıklar da bulunan nesnelere veya olaylara "istatistik birimi" denir. Eğer, sayılamayan veya ölçülemeyen nesneler veya olaylar söz konusu olduğunda bunlar istatistik birimi oluşturmazlar.

Araştırmaya ilişkin tanımlanan istatistik birimlerin tümünün oluşturduğu topluluğa *anakütle* denir. Örneğin; bir yılda fabrikada üretilen çamaşır makinelerinden bozuk olanlara ilişkin yapılan bir çalışmada, arızalı makinelerin her biri istatistik birim iken bu makinelerin tümünün oluşturduğu topluluğa anakütle denir.

İstatistiksel kalite kontrol; en az maliyetle, zamanında ve doğru veri üretmektir. Böylece, İstatistiksel analiz düzeltici ve önleyici faaliyetlerin başlatılabilmesi için verilere dayalı karar verme olanağı sağlar. İstatistiksel süreç kontrolünde Dr. E.Deming'in yorumu; Kalitesizliğin temelinde ise değişkenlik yatar. Kalite birdenbire sağlanamaz, sistem süreçlerinin kontrol altına alınması ancak istatistiksel süreç analizi ile mümkün olur.

İstatistiksel süreç analizi, problemlerin önceden belirlenmesine imkan sağlar, değişkenlikler azalır, kalite gelişir, hurda oranı azalır, etkin kapasite kullanımı artar, birim maliyet düşer, kontrol faaliyetleri azalır, kalitesizlik maliyetleri düşer, makine veya süreç yeterliliğinin izlenmesine imkan sağlar, düzeltici ve önleyici faaliyet ihtiyaçlarını belirler. Problem çözülünce sürecin istatistiksel olarak kontrol altında olduğunu görmek için analize devam edilmelidir.

İstatistiğin Kullanım Alanları

İstatistik günlük hayatımızın hemen her alanında kullanılmaktadır. Aşağıda istatistiğin kullanım alanlarından bazılarına yer verilmiştir.

- IMKB 'de hisse senetlerinin analizi
- Siyasi partilerin seçimlere ilişkin kamuoyu yoklamaları
- Kalite Kontrol
- Şirketin muhasebe kayıtlarının denetlenmesi
- Pazarlama araştırmaları
- Ekonomik göstergelerin takibi
- Tıbbi araştırmalar
- Mühendislik araştırmaları
- Bilimsel çalışmalar
- Uzay araştırmaları
- Demografik (Nüfus özellikleri) araştırmalar

İstatistik Türleri :

Tanımlayıcı (betimleyici) istatistik, sayısal verileri sınıflama ve özetlemede kullanılan yordamlardır. Verileri tablo, grafik veya sayısal olarak anlamlı bir biçimde özetler. Bazı veriler frekans dağılımı olarak düzenlenebilir. Verilerden ortalama değer ve bazı özel orta değerler hesaplanabilir. Örneğin, medyan bir grup sayısal veriyi ikiye bölen (%50 - %50) orta noktadaki değerdir.

Öngörüye dayalı (tahminleyici) istatistik, gözlem yapılarak (ölçülmüş) elde edilen verilerle, gelecekteki durumlar için sonuç çıkarır ve popülasyon (yığın davranışı) özellikle sapma hakkında öngörüde bulunur.

Açıklayıcı istatistikte kullanılan yöntemler:

- Frekans Tabloları
- Şekiller ve Grafikler
- Histogram ve Frekans Poliganları
- Sütun ve Pasta Grafikleri

3.1. İstatistiksel Yorumlama

Örnekten elde edilmiş istatistiksel analiz yöntemleri ile örnek özelliklerine dayanılarak populasyon parametreleri hakkında genellemeler yapmak gerekir. Bu işleme istatistiksel yorumlama denir. İstatistiksel yorumlama iki tip problemin genellenmesinden oluşur.

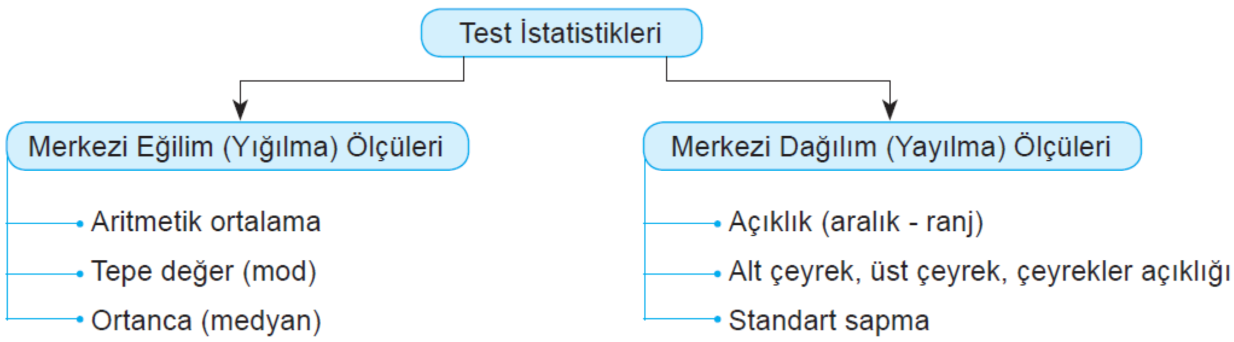
1. Tahmin
2. Hipotez testi

Hipotez testi yapılırken örnek istatistiğine karşılık gelen (değeri bilinmeye çalışılan) populasyon parametresine uygun olup olmadığının saptanmasına çalışılır.

Bir istatistik yardımı ile parametre tahmini yapılırken mutlaka belli bir seviyede belirsizlik olacaktır. Sınırlı fayda çalışılan örnek istatistiği ile populasyon parametresi arasında bir fark oluşur. Bu durumda tahmin yapılırken hata yapma riski ile karşı karşıya kalınır.

Bir hipotez kurulduğunda, bir tahmini kullanabilmek için bu tahmine ne derece güvenle bakıldığının bilinmesi gerekir. Diğer taraftan da hangi tür hatalar ile karşı karşıya kalındığının bilinmesi gerekir.

3.2. Merkezi Eğilim ve Dağılım Ölçümleri



Merkezi Eğilim Ölçüleri:

- Aritmetik Ortalama
- Ağırlıklı (Tartılı) Ortalama
- Geometrik Ortalama
- Harmonik Ortalama
- Medyan (Ortanca)
- Mod (Tepe Değeri)

Dağılım (Değişim) Ölçüleri:

- Değişim Genişliği
- Varyans
- Standart Sapma
- Varyasyon Katsayısı

Davranışların karakteristik eğilimi, örnek alınan değerlerin aritmetik, geometrik, harmonik ve ağırlıklı ortalaması hesaplanarak bulunur. Alınan örnekleme değerler arasında fark çok büyük ise aritmetik ortalama davranışın eğilimini doğru yansıtmaz. Bu gibi durumlarda medyan değerlendirilmesi yapılarak davranışın eğilimi belirlenebilir. Medyan değerlendirmesi, alınan örneklere ait veri değerleri büyükten küçüğe ya da küçükten büyüğe sıralandıktan sonra, tam ortadan veri dizisini 2 eşit frekansa ayıran değerdir. Bir veri setindeki bütün değerleri dikkate almayan (hassas olmayan) bir başka davranış eğilimi belirleme yöntemi ise Mod ölçümüdür. Mod ölçümü, bir veri setinde en sık olarak gözlenen veri değeridir.

Merkezi Eğilim Ölçüleri:

Veri yığının karakteristik eğilimi, yığından örnek alınan değerlerin aritmetik, geometrik, harmonik ve ağırlıklı ortalaması hesaplanarak bulunur. Örnek alama iki türlüdür. Birincisi analog sinyalden ikincisi ise veri yığından örnek alınır.

Aritmetik Ortalama:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + \ x_2 + \cdots + x_n}{n}$$

Alınan örnekleme değerlerinden bir ya da iki tanesi çok yüksek ya da düşük olursa aritmetik ortalama davranışın eğilimini yansıtmaz.

Örnek:
A=(3,4,7,6), aitmetik ortalama=5
B=(1,2,1,16), aitmetik ortalama=5

Yukarıdaki iki örnekten görüleceği üzere dağılımın yapısı hakkında sadece ağırlık ortalamanın verdiği bilginin yetersiz olduğu anlaşılmaktadır. Bu nedenle dağılımın değişkenlik ölçüleri: Ortalama mutlak sapma, varyans, standart sapma ve değişim katsayıları ile hesaplanmalıdır. Bu gibi durumlarda medyan değerlendirilmesi yapılarak davranışın eğilimi belirlenebilir. Medyan değerlendirmesi, alınan örneklere ait veri değerleri büyükten küçüğe ya da küçükten büyüğe sıralandıktan sonra, tam ortadan veri dizisini 2 eşit frekansa ayıran değerdir. Bir veri setindeki bütün değerleri dikkate almayan (hassas olmayan) bir başka davranış eğilimi belirleme yöntemi ise Mod ölçümüdür. Mod ölçümü, bir veri setinde en sık olarak gözlenen veri değeridir.

Yığın içerisindeki verilerin değişkenlik aralığı, ortalama sapma ve standart sapma gibi değişkenlik ölçüleri kullanılır. Davranışlardaki değişimin aralığı bir veri serisindeki en yüksek değer ile en düşük değer arasındaki farktan hesaplanır. Ortalama sapma, tüm veri değerlerinin aritmetik ortalamasından olan mutlak sapmalarının aritmetik ortalamasıdır. Bir yığın içerisindeki örnek değerlerin hangi mertebelerde çeşitlenerek değiştiğini gösteren ölçüt varyans olarak adlandırılır. Değişkenlik bulabilmek için davranışların nominal değerlerden sapmalarının iyi analiz edilmesi gerekir. Sınıflandırılmış verilerde, orta nokta her zaman davranışın ağırlıklı orta noktası olmayacağından, ham verilere göre gruplandırılmış değerlerde daha yüksek sapma değeri ölçülür. Belirli bir değişimin olma ihtimalinin ölçülmesi ve sapmaların çok iyi belirlenmesinde olasılık hesaplamaları ve istatistiksel yöntemler birlikte kullanılmaktadır.

Sistemin davranışını tanımlayan fonksiyon içindeki bir değişkene ait değişime karşılık fonksiyonun değerlerindeki değişimin oranı tablo ve grafiksel olarak gösterilir. Grafikte belirlenen bir noktaya yaklaşımın nasıl olacağının analiz edilmesi için o noktadaki teğetin eğimi bulunmalıdır. Ayrıca yorum yapmaya destek olması için formülü verilen bir fonksiyonun entegral ve türevinin grafiği de çizilmelidir. Değişkenleri bir araya toplayarak ölçmeyi az sayıda faktör ile açıklamayı amaçlayan faktör analizi değişken sayısını azaltır, ulaşılan sonuçları anlamlı kılar. Faktör analizi, ölçmenin nasıl gerçekleştiğini belirler. Değişkenler arasındaki ilişkileri ölçmek için regresyon analizi teknikleri kullanılmaktadır.

Örnek:
Yanda verilen örneklemi göz önünde bulundurun: 10, 8, 9, 83,10

Aritmetik Ortalamayı hesaplayın. $\mu = \Sigma X/n = 120/5=24$
X_1 yok sayılırsa, $\mu=27.5$
X_2 yok sayılırsa, $\mu=28$
X_3 yok sayılırsa, $\mu=27.75$
X_4 yok sayılırsa, $\mu=9.25$
X_5 yok sayılırsa, $\mu=27.5$

Geometrik ortalama:

Özellikle eşit zaman aralığı ile değişen oranların ortalamasının hesaplanmasında (örneğin, nüfus artışı, faiz gibi olaylarda ortalama artış hızını hesaplayabilmek için) geometrik ortalama kullanılır.

Geometrik ortalama:

$$GO = \sqrt[n]{x_1 x_2 x_3 \dots\ x_n}$$

Örnek:
Cep telefonu fiyatlarında ardışık 3 yıldaki fiyat artışı %10, %20 ve %5 olarak gerçekleşmiştir. Ortalama artış oranını hesaplayınız.

$$GO = \sqrt[3]{10 * 20 * 5} = 10$$

Harmonik Ortalama:

Verilerin terslerinin ortalamasının tersi harmonik ortalamayı verir.

$$HO = \frac{n}{\frac{1}{x_1}\frac{1}{x_2} \cdots \frac{1}{x_n}} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}}$$

Oran şeklinde türetilmiş verilerde, oran elde edilirken, harmonik ortalama kullanılır. Bunun tersi durumda ise aritmetik ortalama kullanılır.

Örneğin, *hız=yol/zaman* şeklinde ifade edilen olaylarda; zamanın değişken gidilen yolun sabit, olması halinde harmonik ortalama kullanılır.

Örnek:

Bir kamyon 1000km'lik bir yolu 10 saatte gitmiş, 12.5 saatte dönmüştür. Bu yolculukta kamyonun ortalama hızını hesaplayınız.

$HO = \frac{1}{\frac{1}{2}\left[\frac{1}{100}+\frac{1}{80}\right]}$ = 800/9=89 km/saat

Üç ortalama arasında: AO>GOHO ilişkisi vardır. Bütün değerlerin aynı olması halinde ilişki eşitlik halinde gerçekleşir.

Dağılım Genişliği:

Dağılım Genişliği (Aralık), Bir veri kümesindeki en büyük değer ile en küçük değer arasındaki farktır. DG= Maks(A) – Min(A)

A=(3,4,7,6), DG=7-3=5
B=(1,2,1,16), DG=16-1=15

Örnek:

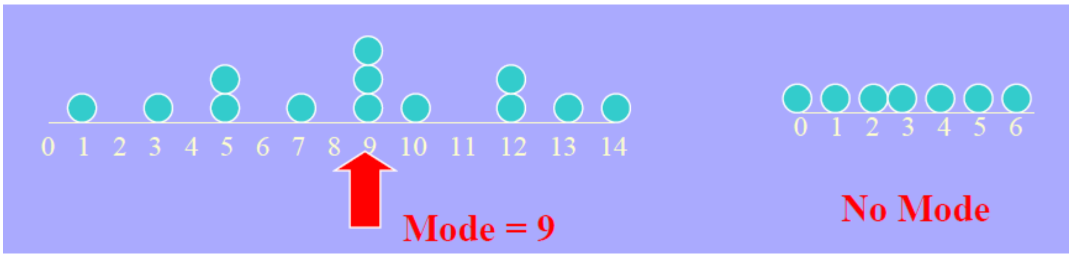
C=10, 8, 9, 83,10; DG=88-8:75

Büyükten küçüğe sıralama,
C=83, 10, 10, 9, 8; DG=83-8=75
C1=10, 10, 9, 8; DG=10-8=2
C2= 10, 9, 8; DG=10-8=2
C3=9, 8; DG=9-8=1

Dağılım genişliğinden maksimum olan değeri anomali bir değer olduğu görülür.

Mod (Tepe) değeri:

Mod dağılım kümesinde olasılığı en yüksek değerdir. Dizide en çok tekrarlanma sayısıdır. Dağılımda en yüksek olasılık değeri birden fazla nokta ile temsil ediliyorsa, mod bu değerlerin hepsine karşılık geldiğinden sonuç tek anlamlı olmaktan çıkar. Böylesi durumlarda dağılımın bimodal, trimodal ya da multimodal olduğundan söz edilir. Frekans, bir saniyedeki titreşim sayısı, sıklık sayısıdır.



Örnek:

"Bu cümlenin her bir kelimesindeki harfleri sayın ve modu verin."

Cümlenin içindeki ilk 10 harflerin sayılarını

B – 2	U – 2	C – 1	Ü – 1	M – 3	L – 3
E – 8	N – 5	İ – 7	H – 2	R – 4	K – 2
S – 2	D – 2	A – 2			

2 2 1 1 3 3 8 5 7 2 4 2 2 2 2

Verileri tararken, modun 2 olduğunu görüyoruz, çünkü ikişer defa tekrarlanan çok fazla harf olduğunu görüyoruz. Mod=2

Medyan – Ortanca:

Sınıflandırılmamış serilerde ortanca hesaplanırken, veriler öncelikle küçükten büyüğe doğru sıralanırlar. n çift ise n/2 değer ile n/2+1 değerlerin ortalaması, n tek ise (n+1)/2 değeri ortancadır.

Alt çeyrek, ortalamanın altındaki değerlerdir.

Üst çeyrek ortalamanın üstündeki değerlerdir.

Alt çeyrek ve üst çeyrek değerlerinden medyan bulunabilir:

Alt çeyrek değerlerinden maksimum olanı < Medyan < Üst çeyrek değerlerinden minumum olanı

Ortalama Mutlak Sapma (OMS):

OMS, Verilerin ortalamalarından sapmalarını gösteren bir dağılım ölçüsüdür. Değişkenlik düzeyinin anlaşılması için kullanılır.

$$OMS = \frac{\sum_{i=1}^{n} |x_i - \mu|}{n}$$

Kıyaslanırken aynı ortalamaya sahip olsalar bile, Ortalama Mutlak Sapmalarının farklı olduğu görülmektedir. OSM'si küçük olan veriler ortalamadan daha az sapar, yani değişkenliği daha az olması şeklindedir.

A=(3,4,7,6), OMS=(I3-5I + I4-5I + I7-5I + I6-5I)/4=6/4=1.5

B=(1,2,1,16), OMS=(I1-5I + I2-5I + I1-5I + I16-5I)/4=22/5=4.4

Örnek:

C=10, 8, 9, 83,10; OMS=(|10-24| + |8-24| + |9-24| +|83-24| +|10-24|)/5=23.6

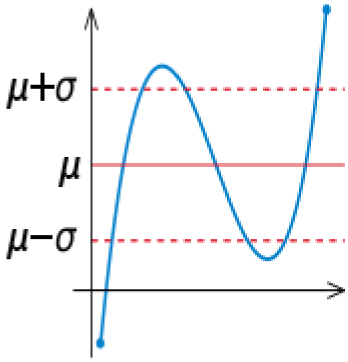
Anomali değer gözardı edilirse,

C1=10, 8, 9, 10; OMS=(|10-9.25| + |8-9.25| + |9-9.25| +|10-9.25|)/4=0.75

μ=9.25

Varyans - Standart Sapma:

Standart sapma, değerlerin aritmetik ortalamasından kaynaklanan kök ortalama karesi **(RMS)** sapmasıdır. Olasılık ve istatistikte, bir olasılık dağılımının standart sapması, rasgele değişken veya popülasyon veya değerlerin yayılmasının bir ölçüsüdür. Genellikle σ harfi ile belirtilir (küçük harf sigma). Standat sapma, varyansın karekökü olarak tanımlanır. Varyans, veriler ile aritmetik ortalama farklarının karlerinin toplamıdır. Ölçülen verilerin ortalamaya yayılmasını ölçer. Standart sapma, aritmetik ortalamadan olan sapmayı verir.



Veri değerleri aritmetik ortalamaya yakınsa, standart sapma küçüktür. Ayrıca, birçok veri noktası ortalamanın uzağındaysa, standart sapma büyüktür. Tüm veri değerleri eşitse, standart sapma sıfırdır.
Bir veri dağılımındaki değişimin önemli bir ölçüsü varyanstır. Varyansın karekökü alınarak standart sapma elde edilir.

Standart sapma dizideki herbir değerin aritmetik ortalamaya yakınlığını gösterir. Standart sapmanın küçük olması ortalamalarda sapmaların ve riskin az olduğunu, standart sapmanın büyük olması ortalamalarda sapmaların ve riskin çok olduğunu gösterir.

Eğer bir çalışmada ana kütlenin tümünden veri toplanırsa buna tamsayım denir. Örneklem bir ana kütleden seçilen belirli elemanların oluşturduğu veri grubudur.

Bir veri grubunun standart sapması:

Örnek için Varyans: $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{N-1}$

Örnek için Standart Sapma: $S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{N-1}}$

Bir veri grubunun standart sapması aşağıdaki gibi bulunur:

i. Veri grubunun aritmetik ortalaması bulunur.
ii. Her verinin aritmetik ortalama ile farkının kareleri toplanır.
iii. Bulunan toplam, veri sayısının 1 eksiğine bölünür ve karkökü alınır. Bu değer standart sapmadır.

Not: Örneklem için neden (n-1)'e bölünüyor ?
10 koltuk bulunan bir minübüse yolcular sırasıyla binsinler. İlk yolcunun 10 koltuktan birini seçme serbestliği var. İkinci yolcunun kalan 9 koltuktan birini seçme serbestliği var. Bu şekilde devam edilirse, 9. yolcunun kalan 2 koltuktan birini seçme serbestliği var. Ancak 10. yolcunun koltuk seçme serbestliği kalmamıştır. Demek ki 9 yolcunun seçme serbestliği varken 1 yolcunun yoktur.

Ana kütlenin standart sapması:

Ana Kütle için Varyans : $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{N}$

Ana Kütle için Standart Sapma: $S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{N}}$

A=(3,4,7,6),
Farkların karelerinin toplamı=$(3-5)^2 + (4-5)^2 + (7-5)^2 + (6-5)^2$) =(4+1+4+1)=10
σ=10/4=2.5

B=(1,2,1,16),
Farkların karelerinin toplamı=$(1-5)^2 + (2-5)^2 + (1-5)^2 + (16-5)^2$ =16+9+16+121=162
σ=162/4 = 40.5

Örnek:

Bir popülasyonun (Anakütle) standart sapmasının nasıl hesaplanacağını göstereceğiz.
Örneğimiz dört küçük çocuğun yaşını kullanacaktır: {5, 6, 8, 9}.
Aitmetik Ortalama, μ=28/4=7
Varyans, σ:

$$\sigma = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

$$\sigma = \frac{1}{4}\sum_{i=1}^{4}(x_i - 7)^2$$

$$\sigma = \frac{1}{4}[(5-7)^2 + (6-7)^2 + (8-7)^2 + (9-7)^2]$$

$$\sigma = \frac{1}{4}((-2)^2 + (-1)^2 + 1^2 + 2^2) = \frac{5}{2}$$

Örnek:

Yanda verilen örneklemi göz önünde bulundurun: 5,10,15,20

a. Aritmetik Ortalamayı hesaplayın.X = ΣX/n = 50/4 =12.5.

d. Varyansını ve Standart sapmasını hesaplayın.

X	X−X	(X−X)2
5	−7.5	56.25
10	-2.5	6.25
15	2.5	6.25
20	7.5	56.25
50	0	125

Varyans, s2 = Σ(X−X)2/(n−1)=125/3=41.667

Stadart Sapma, s = 6.455.

Örnek:

Aşağıdaki veri kümesini göz önünde bulundurun:

69, 40, 32, 74, 54, 47, 80, 25, 60, 69, 44, 67, 72, 35, 51, 82, 46, 200, 66, 27, 50, 70, 38, 85, 66, 81, 30, 68, 42, 78

Aritmetik Ortalamayı hesaplayın.

n = 30, Σx = 1848; bu nedenle Aritmetik Ortalama, **μ = 1848/30 = 61.60**.

Örnek:

Aşağıdaki veri kümesini göz önünde bulundurun:

6, 12, 9, 7, 8, 4, 3, 12, 11

a. Aritmetik Ortalamayı hesaplayın.
b. Medyan değerini bulun.
c. Mod değeri nedir.
d. Varyansını hesplayın.
e. Standart sapmasını hesaplayın.

a. X = ΣX/n = 72/9 = 8.

b. n=9 öğe sıralanır.
3 4 6 7 8 9 11 12 12

c. Medyan ortadki değerdir. 5. değer=8.
Mod değeri=12 (iki defa tekrarlanmıştır.)

d.

X	X−X	(X−X)^2
6	−2	4
12	4	16
9	1	1
7	−1	1
8	0	0
4	−4	16
3	−5	25
12	4	16
11	3	9
72	0	88

Varyans, $s^2 = \Sigma(X-X)^2/(n-1) = 88/8 = 11$.

Stadart Sapma, $s = 3.317$.

Değişim Katsayısını (DK):

Değişim katsayısı standart sapmanın karşılaştırmadaki dezavantajını ortadan kaldıran ve iki veya daha fazla serinin dağılımlarını karşılaştırarak hangi serideki birimlerin daha homojen olduklarını belirlemek amacıyla kullanılan bir ölçüdür.

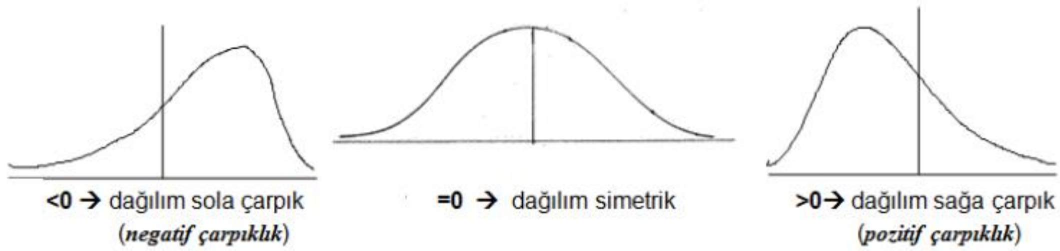
$$DK = \frac{s}{\mu}100$$

Verilerin ortalamaları ve standart sapmaları arasında büyük farklılıklar olduğunda, karşılaştırma için değişim katsayısının kullanılması daha uygun olacaktır.

Çarpıklık Katsayısı:

Dağılımın ortalamaya göre biçimine ilişkin bazı bilgileri çarpıklık ve basıklık katsayıları ile öğrenebiliriz. Çarpıklık Katsayısı aşağıdaki eşitlik ile hesaplanabilir,

$$ÇK = \frac{\sum_{i=1}^{n}(x_i - \mu)^3 / n}{S^3}$$



Merkezi Eğilim Ölçüleri Yardımıyla Serilerin Çarpıklığının Hesaplanması:

Serilerin mod ya da medyanının aritmetik ortalamadan farkının o serinin standart sapmasına bölünmesi ile serinin asimetrisi yani çarpıklığı hesaplanabilir. Bulucusunun adından dolayı **Pearson Asimetri Ölçüsü** denilen bu çarpıklık katsayıları (ÇK) aşağıdaki formüller yardımıyla hesaplanır.

$ÇK = (\mu - Mod)/\sigma$ ve $ÇK = \frac{3(\mu - Medyan)}{\sigma}$

ÇK=0 ise dağılım simetriktir.

ÇK<0 ise dağılım sola doğru çarpık ya da - yöne eğilimlidir.

ÇK>0 ise dağılım sağa doğru çarpık ya da + yöne eğilimlidir.

Basıklık Katsayısı:

$$BK = \frac{\sum_{i=1}^{n}(x_i - \mu)^4 / n}{S^4} - 3$$

BK=0 ise dağılımın yüksekliği standart normal dağılıma uygundur.
BK<0 ise dağılım standart normal dağılımdan daha basıktır.
BK>0 ise dağılım standart normal dağılımdan daha sivridir.

Örnek Büyüklüğünün Belirlenmesi:

Bazı durumlarda, yapılan tahminlerde belli bir doğruluk veya isabet derecesine ulaşmak hedef olarak alınabilir. Bu gibi durumlarda istenen hedefe ulaşabilmek için örnek büyüklüğünün hesaplanmasında aşağıdaki formüller kullanılabilir.

$$n = \frac{Z^2\,\sigma^2}{(\bar{x}-\mu)^2} \Big\} \; Ana\ kitle\ varyansı\ biliniyorsa$$

$$n = \frac{Z^2\,s^2}{(\bar{x}-\mu)^2} \Big\} \; Ana\ kitle\ varyansı\ bilinmiyorsa$$

Z : belirlenen güven düzeyi için Z tablosundan bakılan değer,
σ^2: ana kitle varyansı,
$\bar{x}$: örnek ortalaması,
μ : ana kitle ortalaması,
$(\bar{x}-\mu)$: göze alınan örnekleme hatası.

Örnek : Bir iş kolunda çalışan işçilerin ortalama saat ücretleri gerçek ana kitle ortalama saat ücretinden en fazla 100 TL sapma gösterecek şekilde ve %90 güven aralığında tahmin edilmek isteniyor. Geçmiş kayıtlara dayanarak bu iş kolu için hesaplanan standart sapma $\sigma = 500\,TL$ olduğuna göre örnek büyüklüğü ne olmalıdır.

$$n = \frac{z^2\sigma^2}{(\bar{x}-\mu)^2} = \frac{(1{,}64)^2(500)^2}{(100)^2} = \frac{(2{,}69)(250.000)}{10.000} = 67{,}25 \cong 68$$

Merkezi eğilim ölçülerinin kıyaslanması:

	A	B	C
Ortalama			
Maksimum			
Minumum			
Dağılım Genişliği			
OMS			
Varyans			
Standart Sapma			
Değişim Katsayısı			
Çarpıklık Katsayısı			
Basıklık Katsayısı			

- A ve B verileri aynı ortalama ve medyana sahip ise Ortalama mutlak sapma, varyans, standart sapma ve değişim katsayıları değerleri kıyaslanır.
- Ortalama mutlak sapma, varyans, standart sapma ve değişim katsayılarından daha homojen diyebilmek için hesaplanan katsayıların bir birine benzer olması gerekir.
- Çapıklık katsayıları A'ya ait verilerin, B, C'lere ait veriler ile kıyaslanarak yorum yapılır.
- Basıklık katsayıları standart normal dağılımdan daha basık olup olmadığı ya da birlerine göre basıklık durumları karşılaştırılır.

Örnek:

Bir fabrikada bulunan üç adet üretim bandında aylık üretilen ürün miktarları aylık olarak aşağıda verilmiştir.

	A	B	C
Ay	x_i	x_i	x_i
1	11	8	15
2	11	7	10
3	10	1	15
4	9	6	20
5	8	10	10
6	12	10	2
7	10	5	10
8	10	40	9
9	9	4	2
10	8	10	10
11	10	9	10
12	12	10	7

a)

a) Herbir üretim bandı için "Aritmetik Ortalamayı" hesaplayınız.

	A	B	C
Toplam	120	120	120
N	12	12	12
Aritmetik Ortalama	10	10	10

Dağılımın yapısı hakkında sadece ağırlık ortalamanın verdiği bilginin yetersiz olduğu anlaşılmaktadır. Bu nedenle dağılımın değişkenlik ölçüleri: Ortalama mutlak sapma, varyans, standart sapma ve değişim katsayıları ile hesaplanır.

b) Herbir üretim bandı için maksimum ve minumum üretim miktarını bulunuz ve "Dağılım Genişliğini" hesaplayınız..

	A	B	C
Maksimum	12	40	20
Minumum	8	1	2
Dağılım Genişliği	4	39	18

c) Herbir üretim bandı için "Medyan - Ortanca" değerini bulunuz.

	1	2	3	4	5	6	7	8	9	10	11	12
A	8	8	9	9	10	10	10	10	11	11	12	12
B	1	4	5	6	7	8	9	10	10	10	10	40
C	2	2	7	9	10	10	10	10	10	15	15	20

n=12 ve çift olduğundan 6'ıncı ve 7'inci değerlerin ortalaması alınır.

	A	B	C
Medyan – Ortanca	10	(8+9)/2=17/2	10

d) Herbir üretim bandı için "Ortalama Mutlak Sapmayı - OMS" hesaplayın.

	A	B	C
Ay	I Xi-Xo I	I Xi-Xo I	I Xi-Xo I
1	1	2	5
2	1	3	0
3	0	9	5
4	1	4	10
5	2	0	0
6	2	0	8
7	0	5	0
8	0	30	1
9	1	6	8
10	2	0	0
11	0	1	0
12	2	0	3
Toplam	12	60	40

	A	B	C
OMS	1	5	3.33

A, B ve C kıyaslanırken aynı ortalamaya sahip olsalar bile, Ortalama Mutlak Sapmalarının farklı olduğu görülmektedir. A'nın OSM'si B'nin ve C'nin OSM'sinden küçük olduğundan A'ya ait veriler ortalamadan daha az sapar, yani değişkenliği daha az olması şeklindedir.

e) Herbir üretim bandı için "Varyansları" ve "Standart Sapmaları" hesaplayınız.

	A	B	C
Ay	$(Xi-Xo)^2$	$(Xi-Xo)^2$	$(Xi-Xo)^2$
1	1.00	4.00	25.00
2	1.00	9.00	0.00
3	0.00	81.00	25.00
4	1.00	16.00	100.00
5	4.00	0.00	0.00
6	4.00	0.00	64.00
7	0.00	25.00	0.00
8	0.00	900.00	1.00
9	1.00	36.00	64.00
10	4.00	0.00	0.00
11	0.00	1.00	0.00
12	4.00	0.00	9.00
	20.00	1,072.00	288.00

Varyans - Standart Sapma:

	A	B	C
Varyans	20/11	1072/11	288/11
Standart Sapma	1.35	9.87	5.12

A'nın varyansı ya da standart sapması, B'nin ve C'nin varyansından daha küçük olduğundan A'nın değerleri daha homojendir.

f) Herbir üretim bandı için Değişim Katsayısını (DK) hesaplayınız.

	A	B	C
Değişim Katsayısı	13.48	98.72	51.17

Verilerin ortalamaları ve standart sapmaları arasında büyük farklılıklar olduğunda, karşılaştırma için değişim katsayısının kullanılması daha uygun olacaktır.

Örnek:

Aşağıda verilen tabloda hesaplanan değeleri yerine koyunuz. Kıyaslayınız.

	A	B	C
Ortalama	10.00	10.00	10.00
Maksimum	12.00	40.00	20.00
Minumum	8.00	1.00	2.00
Dağılım Genişliği	4.00	39.00	18.00
OMS	1.00	5.00	3.33
Varyans	1.82	97.45	26.18
Standart Sapma	1.35	9.87	5.12
Değişim Katsayısı	13.48	98.72	51.17
Çarpıklık Katsayısı	0.00	2.24	0.12
Basıklık Katsayısı	-1.29	4.18	-0.63

Merkezi eğilim ölçülerinin kıyaslanması:

- A ve B verileri aynı ortalama ve medyana sahip ise Ortalama mutlak sapma, varyans, standart sapma ve değişim katsayıları değerleri kıyaslanır.
- Ortalama mutlak sapma, varyans, standart sapma ve değişim katsayılarından daha homojen diyebilmek için hesaplanan katsayıların bir birine benzer olması gerekir.
- Çapıklık katsayıları A'ya ait verilerin, B, C'lere ait veriler ile kıyaslanarak yorum yapılır.
- Basıklık katsayıları standart normal dağılımdan daha basık olup olmadığı ya da birlerine göre basıklık durumları karşılaştırılır.

Örnek:

A, B, C üretim bandı için "Çarpıklık" ve "Basıklık" katsayıları aşağıda verilmiştir. Yorumlayınız**.**

	A	B	C
Çarpıklık Katsayısı	0.00	2.24	0.12
Basıklık Katsayısı	-1.29	4.18	-0.63

A simetriktir, B sağa çarpıktır. C ise A göre sağa B göre ise sola çarpıktır.

A basıktır. B sivridir. C ise A ya göre daha az basıktır.

Yüzdelik:

• **Percentiles**: Let *p* be any integer between 0 and 100. The *pth* percentile of data set is the data value at which *p* percent of the value in the data set are less than or equal to this value.

• **How to calculate percentiles**: Use the following steps for calculating percentiles for small data sets. For large data sets, we use computers to find percentiles.

• Step 1: Sort the data in ascending order (from smallest to largest)

• Step 2: Calculate i^{th}=$(p/100)n$, where *p* is the particular percentile you wish to calculate and *n* is the sample size.

• Step 3: If i is an integer, the pth percentile is the mean of the data values in positions i and i+1. If i is not an integer, then round up to the next integer and use the value in this position.

Example: Use the following set of stock prices (in dollars): 10, 7, 20, 12, 5, 15, 9, 18, 4, 12, 8, 14 Find the 10th percentile and the 50th percentile

Solutions:

• First sort the data in ascending order: **4, 5, 7, 8, 9, 10, 12, 12, 14, 15, 18, 20**

• There are 12 scores so, *n* = 12.

• To find the 10th percentile, we use the formula

$$i^{th} = \left(\frac{p}{100}\right)n = \left(\frac{10}{100}\right)12 = 1.2 \approx \text{Round Up}(1.2) = 2$$

• The 10th percentile is the number in the 2nd position.

- The 10th percentile is the number in the 2nd position.

Position	1st	**2nd**	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
Data	4	**5**	7	8	9	10	12	12	14	15	18	20

$$10\text{th Percentile} = P_{10} = 5$$

- To find the 50th percentile, we use the formula

$$i^{th} = \left(\frac{p}{100}\right)n = \left(\frac{50}{100}\right)12 = 6$$

- We need to find the 6th and 7th numbers in the sorted data set.
- Since the answer is an integer, we need to find the 6th and 7th number in the data set.

Position	1st	2nd	3rd	4th	5th	**6th**	**7th**	8th	9th	10th	11th	12th
Data	4	5	7	8	9	**10**	**12**	12	14	15	18	20

$$50\text{th Percentile} = P_{50} = \frac{10+12}{2} = 11$$

Box-and-whisker plot: Requires (**five-number summary**):

- Minimum entry
- First quartile= $Q_1 = P_{25}$
- Median= $Q_2 = P_{50}$
- Third quartile = $Q_3 = P_{25}$
- Maximum entry

3.3. İstatistik Uygulama

Başlıca merkez ve değişkenlik ölçülerini sınıflandırarak kısaca tanımlayınız ve aşağıdaki terimleri açıklayınız. İstatistik, parametre, ortalamanın standart hatası, örnek, populasyon, varyans.

Merkez Ölçüleri		Değişkenlik Ölçüleri
Analitik Olanlar	Analitik Olmayanlar	
Aritmetik Ortalama	Mod	Varyans
Harmonik Ortalama	Medyan	Varyasyon Katsayısı
Tartılı Ortalama		Standart Sapma
Kuadratik Ortalama		Değişim Aralığı(Range)

İstatistik: Sayısallaştırılmış bir örneği karakterize eden ölçümler.
Parametre: Veri yığınını karakterize eden ölçümler.
Örnek: Veri yığının sınıflandırılmış alt grubu
Veri yığını: Belli bir özelliğe ilişkin veri değerleri topluluğu
Varyans : Ortalamadan ayrılışların kareleri toplamının aritmetik ortalamasıdır.
Ortalamanın Standart Hatası : Örnek ortalamalarının örnekleme dağılışının standart sapmasına ortalamanın standart hatası denir.
Mod (Tepe) değeri: Bir dizide en çok tekrarlanan adeti mod (Tepe) değerini verir. Birden fazla sayının en çok tekrarlanma adeti aynı ise bu sayılar alınır.
Medyan (Ortanca): Dizinin terimleri büyükten küçüğe ya da küçükten büyüğe sıralandığında baştan ve sondan eşit uzaklıktaki sayıya medya (ortanca) denir. Dizinin tam ortasındaki sayı medyandır. Dizinin terim sayısı çift ise ortadaki iki terimin aritmetik ortalaması alınır.
Açıklık (Aralık)= En büyük değer – En küçük değer
Üst çevrek açıklığı= Ortanca değerin üst kısmını ifade eder.
Alt Çeyrek Açıklığı= Ortanca değerin alt kısmını ifade eder.
Ortanca=(alt çeyrek açıklığı + Üst çeyrek açıklığı)/2

Örnek:
10 öğrenci girdikleri sınavda aldıkları puan sırayla 10, 9, 2, 3, 4, 5, 7, 6, 5, 9 dır. Bu öğrencilerin aldıkların puanların aritmetik ortalaması nedir?

a) 5 b) 8 c) 7 d) 6 e) 4

$$AO = \frac{10+9+2+3+4+5+7+6+5+9}{10} = \frac{60}{10} = 6$$

Doğru cevap: d şıkkıdır.

Örnek:
10 öğrenci girdikleri sınavda aldıkları puan sırayla 10, 9, 2, 3, 4, 5, 7, 6, 5, 9 dır. Mod'unu (tepe değerini), Medya (Ortanca) değerini bulunuz.

Mod=5 ve 9 dur. (ikişer defa tekrarlanmışlar)

Açıklık (Aralık)= En büyük değer – En küçük değer=10-2=8

Sıralama: 2, 3, 4, 5, 5, 6, 7, 9, ,9, 10

Dizinin sayısı çift olduğundan, $Medyan = \frac{5+6}{2} = 5.5$
Üst çeyrek: 6,7,9,9,10
Alt çeyrek: 2,3,4,5,5

Örnek:
7,7,8,7,9,5,16,16,5,16,8 dizisinin Aritmetik ortalamasını, Mod (Tepe) değerini, Medya (Ortanca) değerini bulunuz.

AO=104/11
Mod: 7,16 (üçer adet)

Sıralama: 5,5,7,7,7, 8, 8,9,16,16,16
Medya (Ortanca)=8

Örnek:
İstanbul'da, 2020 yılının Nisan ayının 20 sinden itibaren günlük sıcaklıkların, 16, 11, 13, 11, 15, 16, 17, 22,22, 19, 19, 18, 19, 20 olacağı öngörülmüştür.

Aritmetik ortalamasını bulunuz. AO=238/14=17
Mod(Tepe)=19
Medyan=39/2

Örnek:

Aşağıdaki tabloda farklı iki ülkede 4 ay boyunca koronadan ölen hastaların sayısı verilmiştir. Hangi ülkede ölüm oranı yüksektir.

A: 12, 24,18, 22

B: 25, 23,12, 20

Standart sapma dizideki herbir değerin aritmetik ortalamaya yakınlığını gösterir. Standart sapmanın küçük olması ortalamalarda sapmaların ve riskin az olduğunu, standart sapmanın büyük olması ortalamalarda sapmaların ve riskin çok olduğunu gösterir.

A ülkesi için;

AO=76/4=19

Farkların kareleri toplamı= $(12-19)^2 + (24-19)^2 + (18-19)^2 + (22-19)^2 = 84$

Varyans=84/3=28

Standart sapma= 5.29

B ülkesi için;

AO=80/4=20

Farkların kareleri toplamı= $(25-20)^2 + (23-20)^2 + (12-20)^2 + (20-20)^2 = 98$

Varyans=98/3=32.66

Standart sapma= 5.71

B ülkesindeki standart sapma A ülkesindeki standart sapmasından daha yüksek olduğundan daha risklidir.

6) Bir sınıfta uygulanan bir testin ortalaması 65 ve standart sapması 10 olarak hesaplanmıştır. Bu testten 75 alan bir öğrencinin Z puanı kaçtır?

A) –1 B) 0 C) 1 D) 2 E) 3

6) ➣ z puanı

Aritmetik ortalaması sıfır (0), standart sapması bir (1) olan puanlara z puanı denir. Bu dağılımlara ise, standart normal dağılım denir.

$$z = \frac{X - \overline{X}}{S_x}$$

z = Z puanı

X = Bireyin puanı

$\overline{X}$ = Grubun puan ortalaması

S_X = Puan dağılımının standart sapması

Soruda verilenler;

$X = 75$

$\overline{X} = 65$

$S_X = 10$

Formülde bunları yerleştirirsek

$$z = \frac{X - \overline{X}}{S_X}$$

$$\Rightarrow \; z = \frac{75 - 65}{10}$$

$$\Rightarrow \; z = \frac{10}{10}$$

$\Rightarrow$ z = 1 olur.

Doğru Cevap: C şıkkı

7) Bir sınıfta uygulanan testin ortalaması 70 ve standart sapması 8 dir. Bu testten 80 puan alan Ahmet'in T puanı kaçtır?

A) 62 B) 62,5 C) 63 D) 63,5 E) 64

7) z puan dağılımının aritmetik ortalaması 50 ve standart sapması 10 olacak şekilde dönüştürülmesi ile elde edilen puanlara "T puanı" denir.

(Hesaplanan z puanı 0,3487 olabilir. Bu nedenle yorumlaması zor olabilir. Bunun için daha kolay yorumlanabilen bir standart puan olan T puanı kullanılır.)

$T = 10.z + 50$

$T = 10 \cdot \left(\frac{X - \overline{X}}{S_X}\right) + 50$ formülüyle hesaplanır.

Soruda verilenler;

$X = 80$

$\overline{X} = 70$

$S_X = 8$

İstenilen; T puan

$$z = \frac{X - \overline{X}}{S_X} = \frac{80 - 70}{8} = 1,25$$

$T = 10.z + 50$

$T = 10.1,25 + 50$

$T = 62,5$ olur.

Doğru Cevap: B şıkkı

8) Tablo: Öğrencinin dört farklı dersteki durumu

	Puan	Sınıfın Aritmetik Ortalaması	Sınıfın standart Sapması
Matematik	75	54	10
Türkçe	70	65	6
İngilizce	80	75	8
Fizik	65	55	15

Yukarıdaki verilen tabloya göre, öğrencinin en başarılı ve en başarısız olduğu dersler sırasıyla hangi şıkta doğru verilmiştir?

A) Matematik, Fizik

B) Türkçe, İngilizce

C) İngilizce, Türkçe

D) Matematik, İngilizce

E) Fizik, İngilizce

8) Öğrencinin farklı derslerdeki başarısının karşılaştırılarak en başarılı ve başarısız olduğu ders sorulduğu için öncelikle z puanlarını hesaplamalıyız.

Öğrencinin z puanının en büyük olduğu derste en başarılı, en düşük olduğu derste en başarısız olduğunu söyleyebiliriz.

Dersler	X	$\overline{X}$	S_X	
Matematik	75	54	10	$z_M = \frac{75-54}{10} = 2,1$
Türkçe	70	65	6	$z_T = \frac{70-65}{6} \cong 0,83$
İngilizce	80	75	8	$z_İ = \frac{80-75}{8} \cong 0,62$
Fizik	65	55	15	$z_F = \frac{65-55}{15} \cong 0,66$

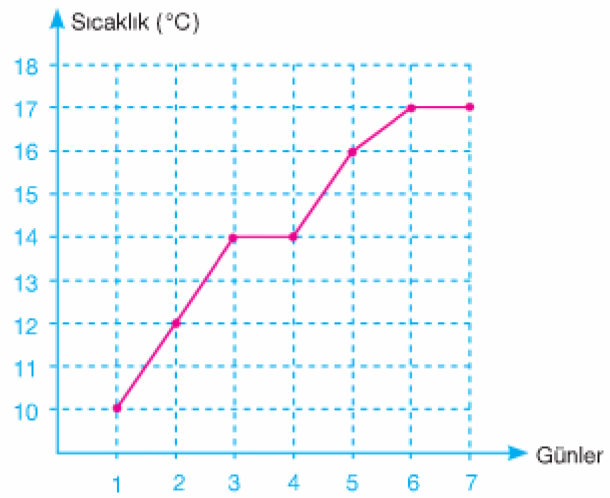
Yukarıdaki z puanlarına göre, bu öğrencinin

en başarılı olduğu ders Matematik ($z_M = 2,1$),

en başarısız olduğu ders İngilizce ($z_İ = 0,62$) dir.

Doğru Cevap: D şıkkı

9) Grafik : Nisan ayının ilk haftasına ait sıcaklık değerleri



Yukarıdaki grafik, bir ilin Nisan ayının ilk haftasına ait, sıcaklık değerlerini göstermektedir. Buna göre, aşağıdaki soruları cevaplayınız.

Bu haftadaki sıcaklık ortalaması kaç derecedir? (yaklaşık)

A) 14,28 B)15,36 C) 17,42 D) 18,54 E) 21,33

9) 1. gün ⟶ 10 °C
2. gün ⟶ 12 °C
3. gün ⟶ 14 °C
4. gün ⟶ 14 °C
5. gün ⟶ 16 °C
6. gün ⟶ 17 °C
7. gün ⟶ 17 °C

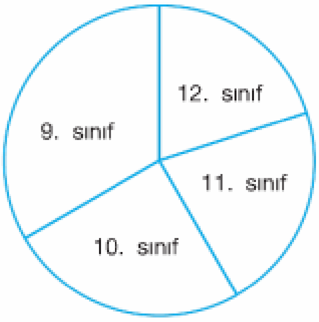
Buna göre,

$$\text{Ortalama} = \frac{10+12+14+14+16+17+17}{7}$$

$$= \frac{100}{7} \cong 14,28\,°C \text{ dir.}$$

Doğru Cevap: A şıkkı

10) Grafik : Bir okulun farklı sınıflarındaki öğrenci sayıları



Yandaki daire grafiğinde bir okulun farklı sınıflarındaki öğrenci sayıları verilmiştir. 12. sınıfta 240 öğrenci, 11. sınıfta 260 öğrenci, 10. sınıfta 300 öğrenci ve 9. sınıfta 400 öğrenci vardır.

Buna göre, aşağıdaki soruları cevaplayınız.

Dairesel grafikte 12. sınıf öğrencilerinin sayısına karşılık gelen merkez açısının ölçüsü kaç derecedir?

A) 30 B) 60 C) 72 D) 96 E) 108

10) Okuldaki toplam öğrenci sayısı,

$240 + 260 + 300 + 400 = 1200$ dür.

12. sınıftaki öğrencilerin sayısı 240 tır.

Buna göre,

1200 öğrenci	360°
240 öğrenci	x

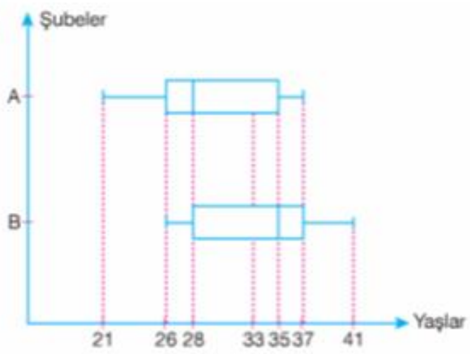
$$1200 . x = 240 . 360°$$

$$x = 72° \text{ olur.}$$

Doğru Cevap: C şıkkı

Örnek:

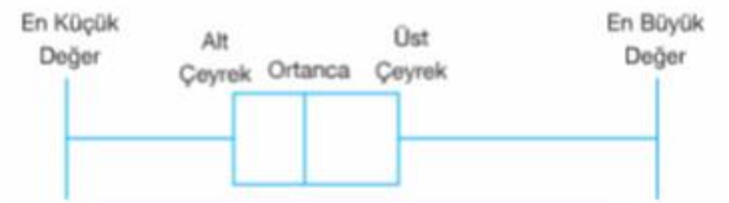
11)



	Medyan(A)	Açıklık(B)
A)	20	10
B)	22	11
C)	24	13
D)	26	14
E)	28	15

Yukarıdaki şekilde, bir şirketin A ve B şubelerinde çalışan kişilerin yaşlarına ait kutu grafiği verilmiştir. Buna göre,

A şubesine ait veri grubunun medyanı ve B şubesine ait veri grubunun açıklığı kaçtır?



	En Düşük Değer	Alt Çeyrek	Ortanca	Üst Çeyrek	En Büyük Değer
A	21	26	28	35	37
B	26	28	35	37	41

Üst çevrek açıklığı= Ortanca değerin üst kısmın ifade eder.
Alt Çeyrek Açıklığı= Ortanca değerin alt kısmın ifade eder.

Ortanca değeri üst çeyrek açıklığının altında, alt çeyrek açıklığını üstünde bir değerdir. Şıklarda 15 ile birlikte 28 uyumludur.

B, Açıklık (Aralık)= En büyük değer – En küçük değer=41-26=15

Frekans -Sıklık

Bir araştırma sonunda elde edilen sürekli veriler, düzenlenmemiş ham ya da sınıflandırılmamış verilerdir. Veriler düzenlenmemiş ham verilerdir.

Tablo1. Firmaların AR-GE faaliyetlerine ayırdıkları kaynak miktarları

Firma No	Miktar	Firma No	Miktar	Firma No	Miktar	Firma No	Miktar	Firma No	Miktar
1	13.5	**11**	8.0	**21**	8.2	**31**	9.6	**41**	7.1
2	8.4	**12**	7.9	**22**	8.0	**32**	7.2	**42**	13.2
3	10.5	**13**	6.8	**23**	7.7	**33**	8.8	**43**	7.7
4	9.0	**14**	9.5	**24**	7.4	**34**	11.3	**44**	5.9
5	9.2	**15**	8.1	**25**	6.5	**35**	8.5	**45**	5.2
6	9.7	**16**	13.5	**26**	9.5	**36**	9.4	**46**	5.6
7	6.6	**17**	9.9	**27**	8.2	**37**	10.5	**47**	11.7
8	10.6	**18**	6.9	**28**	6.9	**38**	6.9	**48**	6.0
9	10.1	**19**	7.5	**29**	7.2	**39**	6.5	**49**	7.8
10	7.1	**20**	11.1	**30**	8.2	**40**	7.5	**50**	6.5

Kaynak: (Balce ve Demir, 2007, s. 7)

Bu veriler miktarlarına göre küçükten büyüğe doğru sıralanırsa,

Tablo2. Sıralanmış Kaynak Miktarı Verileri

Firma No	Miktar	Firma No	Miktar	Firma No	Miktar	Firma No	Miktar	Firma No	Miktar
45	5.2	**28**	6.9	**43**	7.7	**35**	8.5	**9**	10.1
46	5.6	**38**	6.9	**49**	7.8	**33**	8.8	**3**	10.5
44	5.9	**10**	7.1	**12**	7.9	**4**	9.0	**37**	10.5
48	6.0	**41**	7.1	**11**	8.0	**5**	9.2	**8**	10.6
25	6.5	**29**	7.2	**22**	8.0	**36**	9.4	**20**	11.1
39	6.5	**32**	7.2	**15**	8.1	**14**	9.5	**34**	11.3
50	6.5	**24**	7.4	**21**	8.2	**26**	9.5	**47**	11.7
7	6.6	**19**	7.5	**27**	8.2	**31**	9.6	**42**	13.2
13	6.8	**40**	7.5	**30**	8.2	**6**	9.7	**1**	13.5
18	6.9	**23**	7.7	**2**	8.4	**17**	9.9	**16**	13.5

Kaynak: (Balce ve Demir, 2007, s. 7)

Dağılım Sınırları: Bir dağılımda (veri kümesinde) yer alan en küçük ve en büyük örnek değerleridir.
En büyük değer (Maksimum): 13.5 (dağılımın üst sınırı)
En küçük değer (Minimum) : 5.2 (dağılımın alt sınırı)

Dağılım Genişliği (DG): Dağılım sınırları arasındaki farktır.
DG=En büyük değer - En küçük değer =13.5 - 5.2 = 8.3

Sınıf: Eşit ya da birbirine yakın değerlerin oluşturduğu her bir gruba sınıf denir. Sınıf sayısı, k ile gösterilir. Sınıf sayısı, çok sayıda veri olduğunda aşağıda verilen Sturges'in formülü ile de bulunabilir.

k=1+3.3log(n)

k=1+3.3*log(50)=1+3.3*log(100/2)=1+6.6-3.3*0.3=6.61

$$x_{enb} = \textit{En büyük gözlem değeri}$$
$$x_{enk} = \textit{En küçük gözlem değeri}$$
$$k = \textit{Sınıf Sayısı}$$
$$h = \textit{Sınıf Aralığı Büyüklüğü}$$

$$k = 1 + (3.322) \cdot log(N)$$
$$h = (x_{enb} - x_{enk}) / k$$

Sınıf Aralığı: Ard Arda gelen iki sınıfın alt sınırları ya da üst sınırları arasındaki farktır. Sınıf aralığı, c ile gösterilir. Örneğimizde sınıf sayısı 8 olarak alınsın.
C= (Dağılım genişliği + a)/Sınıf Sayısı
a: veri kümesindeki verilerin ondalık kısmındaki hane sayısı ile ilgilidir. Örneğimizde tam kısımdan sonra 1 hane olduğu için a=0.1 alınır.
C=(8.3+0.1)/8=1.05

Çizelge 1. AR-GE faaliyetleri için ayrılan kaynak miktarı verileri için sıklık çizelgesi

Sınıf	Alt Sınır (AS)	Üst Sınır (ÜS)	Sınıf Orta Değeri (m_i)	Çeteleme	Sıklık (f_i)	Göreli Sıklık (p_i=f_i/n)
1	5,2	6,24	(5,20+6,24)/2=5,72	////	4	4/50=0,08
2	6,25	7,29	6,77	///// ///// //	12	0,24
3	7,3	8,34	7,82	///// ///// ///	13	0,26
4	8,35	9,39	8,87	/////	5	0,10
5	9,4	10,44	9,92	///// //	7	0,14
6	10,45	11,49	10,97	/////	5	0,10
7	11,5	12,54	12,02	/	1	0,02
8	12,55	13,59	13,07	///	3	0,06
				Toplam	**50**	**1,00**

Kaynak: (Balce ve Demir, 2007, s. 7)

Göreli Sıklık(Sıklık Yüzdesi): Her sınıfa düşen denek sayısının toplam denek sayısına göre yüzdesidir. Göreli Sıklıklar p_i ile gösterilir. Toplamları 1 olmalıdır.

$$p_i = \frac{f_i}{n} \;, \; i = 1,2,\ldots,k \qquad\qquad \sum_{i=1}^{n} p_i = 1$$

Örnek:

Sınıflar	Frekanslar (f_i)	Sınıf Orta Değerleri (m_i)	$f_i.m_i$
10-20	3	15	3.15=45
21-31	5	26	5.26=130
32-42	8	37	8.37=296
43-53	4	48	4.48=192
54-64	2	59	2.59=118
Toplam	**22**		**781**

Bu verilere göre aritmetik ortalama :

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i.m_i}{\sum_{i=1}^{n} f_i} = \frac{781}{22} = 35{,}5 \text{ bulunur.}$$

Örnek : 15, 22, 18, 19, 15, 18, 20, 15 serisi için tepe değerini bulmak için her değerin frekansı (sıklığı) bulunur:

Değer	Frekans
15	**3**
18	2
19	1
20	1
22	1

En fazla tekrarlanan değer 15'dir (3 kez) bu nedenle bu dağılımın tepe değeri (mod) 15'dir.

Örnek : 25, 12, 32, 18, 28, 25,42, 26 serisinin ortanca değerini bulunuz.
Öncelikle sayılar sıralanır:
12, 18, 25, 25, 26, 28, 32, 42
n=8 olduğu için n/2=8/2=4. Ve n/2+1=8/2+1=4+1= 5. Değerlerin ortalaması alınır:
4.değer=25, 5.değer=26 olduğundan
Ortanca=(25+26)/2=25,5 bulunur.

Örnek : 25, 12, 32, 18, 28, 25,42 serisinin ortanca değerini bulunuz.
Öncelikle sayılar sıralanır:
12, 18, 25, 25, 28, 32, 42
n=7 olduğu için (n+1)/2=(7+1)/2=8/2=4.değer ortancadır:
ortanca=25

Örnek :

Veriler (x_i)	**Frekanslar (f_i)**	**Birikimli Frekanslar**
10-20	3	3
21-31	5	8
32-42	8	16
43-53	4	20
54-64	2	22
Toplam	**22**	

n=22 çift sayı olduğu için n/2=22/2=11.eleman medyandır. 11.elemanı içeren sınıf medyan sınıfıdır. Bu durumda medyan sınıfı 32-42 sınıfıdır.

L=32, c=10, F=8, f_{med}=8 olduğuna göre:

$$Medyan = L + C.\frac{\left(\frac{n}{2}\right) - F}{f_{med}} = 32 + 10.\frac{\left(\frac{22}{2}\right) - 8}{8} = 32 + 10.\frac{11 - 8}{8} = 32 + \frac{30}{8}$$
$$= 32 + 3{,}75 = 35{,}75\ bulunur.$$

Ortalama, ortanca ve tepe değeri arasındaki bağıntı

Ortalama=Mod=Medyan ise sıklık dağılımı simetrikdir.

Ortalama<Medyan<Mod ise dağılım sola çarpıktır.

Ortalama>Medyan>Mod ise dağılım sağa çarpıktır.

Örnek.

X1 = [5,6,7,8,8,8,7,5,5,8,9,10]

a) Aritmetik ortalaması nedir?

b) Varyansı nedir?

c) Standart sapması nedir?

d) Varyasyon katsayısı nedir?

e) Modu nedir?

f) Medyanı nedir?

g) Ortalama, ortanca ve tepe değeri arasındaki bağıntı nedir?

Not = $\Sigma x=86$, $x^2=646$, n= 12

Cevap .

a) A.O. $= \frac{\Sigma x_i}{n} = \frac{86}{12} = 7.16$, n = Eleman say.

b) Varyans= $S^2 = \frac{1}{n-1}\left[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right] = \frac{1}{11}\left[646 - \frac{(86)^2}{12}\right] = 2.27$

c) Standart Sapma = S = $\sqrt{S^2} = \sqrt{2.7} = 1.64$

d) Varyasyon Katsayısı = VK = $\frac{S}{\overline{X}} * 100 = \frac{1.64}{7.16} * 100 = 22.9$

e) Mod en çok tekrarlanan değerdir.(Frekansı en çok olan değerdir).Mod = 8'dir

f) Medyan = ortancadır.n tekse ortada yer alan değer, n çift ise iki değerin ortalamasıdır.

Medyan = (8 + 7)/2 = 7.5

Soru .

Aşağıdaki veri seti için medyan, mod, ortalama, varyans, standart sapma, standart hata ve varyasyon katsayısını hesaplayınız.

X_i=[80, 70, 80, 78, 81, 75, 77, 74, 77, 68]

$\Sigma x_i = 760$; $\Sigma x_i^2 = 57.928$

Cevap .

a) Medyan = ortanca büyüklüğüne göre sıralanmış verilerde ortadaki değerlerdir.

68, 70, 74, 75, 77, 77, 78, 80, 80, 81.

Medyan = (77+77)/2=77

b) Mod= frekansı en büyük olan değerdir.Burada 77 ve 80'dir.2 kez tekrarlanmışlardır.

c) $\bar{x} = \frac{\Sigma x}{n} = \frac{760}{10} = 76'\text{dır.}$

d) $S^2 = \frac{1}{n-1}\left[\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right] = 18{,}66$

e) Standart Sapma = $\sqrt{S^2} = \sqrt{18.6} = 4.32$

f) Standart Hata populasyon varyansının n'de birinin kareköküne eşittir.

$S_{\overline{X}} = \frac{S}{\sqrt{n}} = \frac{4.32}{\sqrt{10}} = 1.36$

g) Varyasyon Katsayısı = VK = (S/X)*100=[(4.32/76)]*100=5.68

ÖRNEK: Aşağıdaki seriler **Elektronik** ve Bilgisayar bölümündeki öğrencilerin olasılık final notlarını vermektedir.

Elektronik (X_i)	Bilgisayar (Y_i)
30	**5**
35	**10**
40	**20**
45	**25**
50	**50**
55	**75**
60	**80**
65	**90**
70	**95**
$\sum \boldsymbol{X_i} = \mathbf{450}$	$\sum \boldsymbol{Y_i} = \mathbf{450}$

İki sınıfında aritmetik ortalaması ve medyanı hesaplandığında, sınıfların aynı aritmetik ortalamaya ve medyana sahip olduğu görülmektedir. Medyan bir grup sayısal veriyi ikiye bölen (%50 - %50) orta noktadaki değerdir. Sınıflandırılmamış serilerde medya - ortanca hesaplanırken, veriler öncelikle küçükten büyüğe doğru sıralanırlar. n çift ise n/2 değer ile n/2+1 değerlerin ortalaması, n tek ise (n+1)/2 değeri ortancadır.

$\overline{X_i} = \overline{Y_i} = \frac{450}{9} = 50$ ve Medyan = 50

Merkezi eğilim ölçülerine baktığımızda iki sınıf arasında bir fark gözükmüyor olsa da verilerin yayılımında farklılık olduğu dikkatlerden kaçmamaktadır. Elektronik bölümündeki öğrencilerin notları göreceli olarak Bilgisayar bölümündeki öğrencilerin notlarından birbirine daha yakındır. Diğer bir ifadeyle, Elektronik bölümündeki öğrencilerin notları Bilgisayar bölümündeki öğrencilerin notlarıyla karşılaştırıldığından ortalama değer etrafında daha yakın dağılmıştır. Bu nedenle serilerin diğer özelliklerini de ortaya koyacak merkezi eğilim ölçüleri dışında değişim ölçülerine ihtiyaç vardır.

Değişim Aralığı (Range):

Serideki en büyük gözlem değeri ile en küçük gözlem değeri arasındaki farka “Değişim Aralığı” denir. Formülü aşağıdaki gibi yazılır. R = Xmax – Xmin. Elektronik bölümü için R = 70 - 30 = 40 ve Bilgisayar bölümü için R = 95 - 5 = 90 olarak hesaplanır. Tüm gözlem değerlerinin birbirinden ya da ortalamadan farklarını ortaya koymaması açısından serinin dağılımı hakkında fazla bir bilgi içermez. Yukarıdaki örnekten de görüldüğü gibi özellikle serideki uç değerlerden aşırı etkilenmektedir.

3.4. Faktoriyel - Permütasyon - Kombinasyon

Faktoriyel

n! = 1 x 2 x ... x n-1 x n

0!=1

1!=1

n büyüdükçe, n!'li bulmak güçleşir. Bu durumda Stirling formülü kullanılarak n! yaklaşık olarak bulunur. $n! = \sqrt{2\pi n}\, n^n e^{-n}$ ile yaklaşık değeri hesaplanabilir.

Teorem: n tane birbirinden farklı nesne sıralandığında elde edilecek değişik düzen sayısı n!'dır.

İpuçları:

(n+1)!=(n+1) x n!

n!=n x (n-1)!

(n-1)!=(n-1) x (n-2)!

Dairesel Düzen: n tane farklı nesneyi bir daire çevresinde sıralarsak elde edilecek farklı düzenlerin sayısı (n-1)! bulunur.

Örnek: 7 kişi yuvarlak bir masa çevresinde kaç farklı düzende oturabilir?

(7-1)! =6! = 720

Örnek: 7 kişi bir gişe önünde kaç farklı düzende sıralanabilir? 7!

Fatoriyel sorusu:

$$\frac{11\,n!}{(n-2)!}\,\frac{(n-1)!}{(n+1)!}\ =?$$

$$\frac{11\,n!}{(n-2)!}\,\frac{(n-1)!}{(n+1)!} = \frac{11\,n(n-1)(n-2)!}{(n-2)!}\,\frac{(n-1)!}{(n+1)n(n-1)!} = \frac{11(n-1)}{(n+1)}$$

Permütasyon (Sıradüzen)

Nesnelerin değişik düzenlerde sıralanma sayısı nasıl bulunur? n tane nesneyi sıralama, belirli bir sırada düzenleme ilgi alanımıza giriyorsa, olası düzenlemelerin tümüne sıradüzen (permütasyon) adı verilir.

Örnek: isimlerinin baş harfleri a, b, c, d, e, f olan kişilerden ikili sıralama yapılacaktır. Hiçbir harf tekrarlanmadan ikili kaç farklı sıralama yapılır?

$$P(n,k) = P_k^n = \frac{n!}{(n-k)!}$$

$$P(6,2) = P_2^6 = \frac{6!}{(6-2)!} = 6 * 5 = 30$$

Not: Burada dikkat edilmesi gereken husus, ab ve ba ya da ac, ca … birbirlerine benzese de burada harflerin sıralamadaki yeri önemlidir.

Permütasyon sorusu:

A={1,2,3,4,5,6,7} sayıları ile etiketlenmiş ürünlerde 4 ve 6 nolu ürünler kusurludur. A ürün kümesinin üçlü permütasyonunlarının kaçında 4 ya da 6 nolu ürün bulunur.

$$P(n,r) = P_r^n = \frac{n!}{(n-r)!}$$

4 ve 6 nolu ürünleri ayırırsak kalan elemanlardan oluşturulacak permütasyonlar, P(5,3)=5!/2!=60 adettir.

Tamamında üçlü permütasyonlar oluşturursak,
P(7,3)=7!/4!=210

P(7,3)-P(5,3)=210-60=150 adet üçlü gruplamada 4 ya da 6 nolu ürünler bulunmaktadır.

Üçlü gruplamalarda 4 ya da 6 nolu ürünlerin bulunma olasılığı=150/210=5/7

Örnek: 1, 2 ve 8 rakamlarından kaç adet 2 rakamlı sayı oluşturulabilir ? (çekimler iadesiz)

$$P(3,2) = \frac{3!}{(3-2)!} = 6 \text{ adet}$$ sayılar: 12,18,28,21,81,82

Bu örneği çekimlerin iadeli yapıldığını kabul ederek hesaplarsak, her rakamın tekrarlanma olanağı doğduğu için permütasyon sayısı;

$_3P_2 = 3^2 = 9$ olacaktır. sayılar: 11,12,18,22,21,28,81,82,88

Örnek: 12 sandalyeye, 2 kişi kaç farklı yolla oturtulabilir?
Çözüm: 12 yerden 2'sini seçeceğiz. Ayrıca bu 2 kişinin kendi aralarında bir sıradüzeni (permütasyonu) vardır. Buna göre farklı düzenlerin sayısı

$$P(12,2) = \frac{12!}{(10)!}$$

P(12,2)=12*11=132

Bu durumda iki kişi iki farkıl şekil de de oturabilir.

Kombinasyon (Birleşim)

Permütasyonda sıralama sayısı önemli iken kombinasyonda sıralama sayısı değil seçim sayısı önemlidir.

$$C(n,k) = C_k^n = \frac{n!}{(n-k)!\,k!}$$

Permütasyonda dizilim (sıralama) önemlidir ancak kombinasyonda sadece seçme işlemi vardır dizilimin bir önemi yoktur.

Örneğin sayısal loto da 1 den 49 a kadar rakamlardan sadece 6 sı çıkar. Permutasyon olamaz.
3 – 21- 33 – 44 -45 – 49 ile 3 – 44 – 45- 49 -21 -33 aynıdır.
C(49,6)=49!/[(49-6)! 6!]

Örneğin; A, B ve C ile gösterilen üç nesneden ikişerli sıralanmasından elde edilecek farklı düzen sayılarını bulmak istersek sıralama sayısı önemli olduğundan permütasyon yardımıyla, 6 bulunur. Sıralama sayısı değil de seçim sayısı istersek AB, AC, BC gibi üç farklı seçim yapılabilir. Burada nesnelerin sırasını dikkate almadığımız için AB ile BA, AC ile CA, BC ile CB aynı seçimlerdir. Permütasyonda benzerlik vardır, fakat harflerin sıralamadaki yeri önemlidir. Eğer sıralamadaki yeri değil gruplama önemli ise kombinasyon kullanılır.

$$C(3,2) = C_2^3 = \frac{3!}{(3-2)!\,2!} = 3$$

(AB, AC, BC)

Kombinasyon (Gruplama) sorusu

4 adet Masa üstü bilgisayar (PC1, PC2, PC3, PC4 etiketli) ve 5 adet Dizüstü bilgisayar (Laptop: L1, L2, L3, L4, L5 etiketli) arasından

a) Oluşturulacak tüm bilgisayarlardan üçlü grup sayısı nedir?
b) İçinde Dizüstü bilgisayar bulunmadan oluşturulacak tüm 3’lü grup sayısı nedir?
c) İçinde disüstü bilgisayar bulunmayan üçlü grup olasılığı nedir?

4 adet Masaüstü bilgisayar (PC) ve 5 adet Dizüstü bilgisayar (Laptop) arasında

a) Oluşturulacak tüm bilgisayarlardan üçlü grup sayısı nedir?

$$C(9,3) = \frac{9!}{(9-3)!\,3!} = 84$$

b) İçinde Dizüstü bilgisayar bulunmadan oluşturulacak tüm 3’lü grup sayısı nedir?

$$C(4,3) = \frac{4!}{(4-3)!\,3!} = 4$$

c) İçinde disüstü bilgisayar bulunmayan üçlü grup olasılığı nedir?
C(4,3)/C(9,3)=4/84 =1/21

1) 15 sorudan oluşan bir sınavda 7 soruyu yanıtlamak zorunda olan bir öğrenci,

a) hiçbir koşul olmadan 7 soruyu kaç değişik şekilde seçer?

b) ilk 3 soru zorunlu ise 7 soruyu kaç değişik yolla seçer?

c) ilk 4 sorunun en az 3'nü yanıtlamak zorunda ise 7 soruyu kaç değişik yolla seçer?

Çözüm:

a) 15 sorudan 7 soruyu $\binom{15}{7}=\frac{15!}{7!.8!}$=6435 yolla seçer.

b) 15 sorudan ilk 3 soru zorunlu ise, seçim yapılabilecek soru sayısı (15-3)=12 olur. 7 soru yanıtlayacak ilk 3 sorudan sonra

7-3=4 soru daha yanıtlanması gerekir. Geriye kalan 12 sorudan 4 soruyu $\binom{12}{4}=\frac{12!}{4!8!}=495$ değişik yolla seçer.

Örnek:

9 farklı oyuncağı 4 çocuk arasında, 1.'ye 3, kalan çocuklara 2'şer oyuncak verecek şekilde kaç türlü dağıtabiliriz?

Çözüm: $\binom{9}{3,2,2,2}=\frac{9!}{3!2!2!2!}$=7560 farklı şekilde dağıtabiliriz.

ÖRNEK:

4 erkek ve 5 kız arasından en az biri kız olmak şartıyla 3 kişilik bir grup kaç farklı biçimde oluşturulabilir. ?

Oluşturulabilecek tüm 3'lü grupların sayısı, C(9,3)=9!/(3!.6!)=84

İçinde hiç kız bulunmadan oluşturabileceğimiz sadece yani erkeklerden oluşan üçlü grup sayısı, C(4,3)=4!/(3!.1!)=4

İçinde en az bir kız bulunan üçlü grup sayısı, 84-4=80 olarak bulunur.

Örnek: 9 soru arasından 6 soru kaç sekilde çekilebilir ?

$$C(9,6)=C_6^9=\frac{9!}{(9-6)!\,6!}=84$$

Örnek:

Bir sınıfta 15 erkek ve 10 kız var. Üç öğrenci rastgele seçilir. 1 kız ve 2 erkek çocuk seçilme olasılığını bulunuz.

Let S be the sample space and E be the event of selecting 1 girl and 2 boys.

Then, $n(S)$ = Number ways of selecting 3 students out of 25

$$= {}^{25}C_3$$

$$= \frac{(25 \times 24 \times 23)}{(3 \times 2 \times 1)}$$

$$= 2300.$$

$$n(E) = ({}^{10}C_1 \times {}^{15}C_2)$$

$$= \left[10 \times \frac{(15 \times 14)}{(2 \times 1)}\right]$$

$$= 1050.$$

$$\therefore\ P(E) = \frac{n(E)}{n(S)} = \frac{1050}{2300} = \frac{21}{46}.$$

ÖRNEK: A={1,2,3,4,5,7} kümesinin 3 elemanlı alt kümelerinin kaçında,

a) 1 bulunur ?
b) 1 ve 2 bulunur ?
c) 1 veya 2 bulunur ?

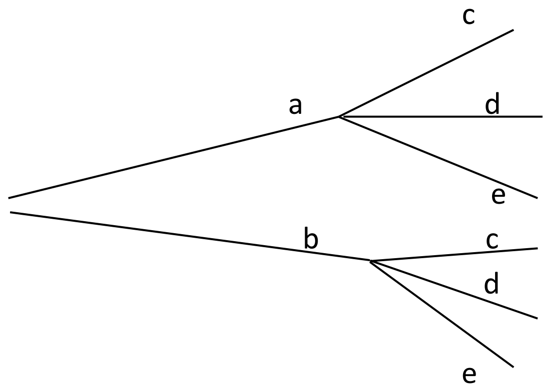
a) İstenen alt kümenin bir elemanı belirli olduğuna göre diğer iki eleman kalan sayılar arasından seçilir. Kümemizde toplam 6 eleman var. 1 tanesi belirli olduğundan 5 elemandan 3-1=2 eleman seçeriz.
C(5,2)=10 olarak bulunur.

b) Seçilmesi istenen 2 eleman belirli olduğuna göre, Diğer eleman kalan kümeden seçilir. 6 elemanımız vardı 6-2=4 elemandan 1 tanesi C(4,1)=4 farklı şekilde olur.

c) 3 elemanlı oluşturabileceğimiz tüm gruplardan 1 ve 2 elemanı dışında kalan 3 elemanlı alt kümelerin oluşturduğu 3 elemanlı alt kümeler çıkarılırsa istenilen şart sağlanır.
C(6,3)-C(4,3)=16 olur.

Permütasyon ve Kombinasyon Analizi:

Bileşik olayların olasılıklarının hesaplanmasını kolaylaştırmak için permütasyon(permutation) ve kombinasyon (combination) analizinden yararlanılır. Eğer bir olay $\mathbf{n_1}$ halde ortaya çıkabiliyorsa ve bu olaydan bağımsız diğer bir olay $\mathbf{n_2}$ halde ortaya çıkabiliyorsa her iki olay bir arada $\mathbf{n_1 * n_2}$ halde ortaya çıkabilecektir. Birinci olayın 2, ikinci olayın da 3 halde meydana çıkması durumunu ağaç diyagramı ile aşağıdaki gibi gösterebiliriz.



Birinci olayın şıkları a ve b, ikinci olayın şıkları c,d ve e ile gösterilmektedir. Her iki olay bir arada 6 defa ortaya çıkmaktadır.

a – c b – c a – d b – d a – e b - e

Örnek: Bir kutuda 4 kırmızı, 6 siyah top vardır. Kutudan rasgele 3 top çekildiğinde bunların;

- Üçünün de kırmızı olma olasılığı,
- Üçünün de siyah olma olasılığı,
- En az birinin kırmızı olma olasılığı nedir ?

Birinci, ikinci ve üçüncü çekişlerde kırmızı top çekme olaylarını E1, E2 ve E3 ile gösterelim. Olayların bir arada ortaya çıkma olasılığı;

P(E1 E2 E3) = P(E1) *P(E2\E1)* P(E3\E1E2) = (4/10)*(3/9)*(2/8) = (1/30)

Kombinasyonla hesaplarsak,

$$\frac{4 Kırmız\ toptan\ \text{üçünün}\ seçici}{10\ Toptan\ \text{üçünün}\ seçimi} = \frac{4C3}{10C3} = 1/30$$

En az birinin kırmızı olma olasılığı ise,1’den hepsinin siyah olma olasılığı çıkarılarak bulunur.

İki Terimli (Binom) Katsayılar

$(x+y)^2 = x^2 + 2xy + y^2$

$(x+y)^3 = x^3 + 3x2y + 3xy2 + y^3$

$(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4$

$$
\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1
\end{array}
$$

<u>Örnek:</u> $(x_1+x_2+x_3)^6$'nın açılımındaki $x_1^3x_2x_3^2$ teriminin katsayısı kaçtır?

$n=6$, $n_1=3$ $n_2=1$ $n_3=2$ olduğuna göre

$$\frac{6!}{3!\,1!\,!2!} = \frac{6.5.4.3!}{3!.1.2} = 60 \text{ bulunur.}$$

<u>Örnek:</u> $(x+y+z)^8$'in açılımındaki $x^2y^3z^3$ teriminin katsayısını bulunuz?

$n=8$ $n_1=2$ $n_2=3$ $n_3=3$

$$\frac{8!}{2!.3!.3!} = \frac{8.7.6.5.4.3!}{2.6.3!} = 560 \text{ bulunur.}$$

4. Olasılık

Olasılık, bir olayın gerçeklenebilme olasılığının oranının belirlenmesidir. Olasılık Kuramının doğuşu bir kumarbazın ihtirasıyla başlar. Chevalier de Méré adlı soylu bir Fransız, kumar oynayarak servetini büyütme ihtirasına kapılmıştır. Oynadığı oyunun kuralı şudur: Bir zarı dört atışta enaz bir kez 6 getiren kazanır. Ama Chevalier oyunun kuralını değiştirerek daha çok kazanmak istemektedir. Yeni kural şudur: Çift zarı 24 atışta bir tane düşeş (toplam 6+6=12) getiren kazanacaktır. Ama kısa sürede, bu kuralın daha az kazandırdığını gördü. Bunun nedenini arkadaşı Blaise Pascal (1623-1662)'a sordu. Pascal, o dönemin iyi matematikçilerinden biriydi. O ana kadar, matematik dünyası şans oyunlarının matematikle bir ilişkisi olduğunu bilmiyordu. Pascal, kendisine sorulan sorunun yanıtını, bir matematikçi gözüyle araştırdı. Sonunda basit ama kesin çözümü ortaya koydu. Eski kuralda Chevalier'in kazanma şansı %51.8 iken yeni kuralda %49.1 idi. Chevalier'in kaybetme nedeni buydu.

Pascal, bu basit problemi çözmekle yetinmedi. Sorunun gerisinde daha büyük bir matematik kuramın yattığını anlamıştı. Çağdaşı Pierre de Fermat ile mektuplaşarak fikir alışverişinde bulunmaya başladı. Sonunda, matematiğin önemli bir dalı olan Olasılık Kuramını yarattılar. Bu gün, olasılık kuramı, şans oyunlarına uygulanma özeliğini çoktan aşmış bilim, endüsri, ekonomi, spor, yönetim gibi çağdaş insanın yaşamını etkileyen her alana girmiştir. Örneğin bankacılık, sigortacılık, endüstride kalite kontrolü, genetik, gazların kinetik teorisi, istatistiksel mekanik, kuantum mekaniği gibi pek çok alan olasılık kuramı olmadan ayakta duramaz.

Olasılık Kuramını geliştiren önemli matematikçilerden bazıları şunlardır:
Blaise Pascal (1623-1662), Pierre de Fermat (1601-1665), Christiaan Huygens (1629-1695), Jakob Bernoulli (1645-1705), Abraham de Moivre (1667-1754), Daniel Bernoulli (1700-1782), Comte de Buffon (1707-1788), Pierre-Simon Laplace (1749-1827), Augustus De Morgan (1806-1871), Thomas Bayes (1702-1761), Andrei Andreyvich Markov (1856-1922), Richard von Mises (1883-1953).

Matematiğin bir dalı olarak olasılık kuramının doğuşu 17.yüzyılın ortalarına raslar. 20.yüzyılda olasılığın formalizasyonu yönünde, iki önemli adım atıldı. Birincisi, 1933 yılında *Andrey Nikolaevich Kolmogorov (1903–1987)* tarafından ortaya konulan aksiyomlardır. Bu aksiyomlar, olasılık kuramını, bir ölçüm uzayına taşıyor ve o zamana kadar olasılıkla ilgili yapılan bütün *discrete hesaplamaları* özel haller olarak içeren çok genel bir yapıya yükseltiyordu. İkincisi ise *Richard Threlkeld Cox (1898-1991)* tarafından ortaya konulan aksiyomlarıdır. Cox, olasılığı, daha basite indirgenemez bir ilkel (primitive) kavram olarak aldı ve onun sağladığı temel özelikleri ortaya koydu. 1960 lı yıllarda birbirlerinden habersiz olarak, Ray Solomonoff, Anrey Kolmogorov, Gregory J. Chaitin *algoritmik seçkisizlik* (randomness) kavramını ortaya attılar. Bu yeni kavram, olasılık kuramının dayandığı rasgelelik kavramına açıklık getirmekle kalmayıp, *informasyon* kuramına da yeni açılımlar getirdi. Gödel'in eksiklik teoremine benzer bir niteliğe sahip algoritmik seçkisizlik, günümüzde çok aktif bir konudur ve görünüşe göre olasılığı bulunduğu yerden alıp daha yükseklere taşıyacaktır.

Sonuçlarının kümesi belli olan, ancak hangi sonucun ortaya çıkacağı önceden söylenemeyen bir işleme "Rasgele Sonuçlu Deney" veya kısaca "Deney" denir.

Bir deneyin tüm olabilir sonuçlarının kümesine "**Örnek Uzay**" denir.

Örnek uzayın bir altkümesine "**Olay**" denir. Deneyler ile elde edilen sonuçların toplamına olay denir.

İki olay aynı anda meydana gelemiyorsa bu olaylara "**ayrık olaylar**" denir. Diğer bir ifadeyle kesişimleri boş küme olan olaylara ayrık olaylar denir.

Olasılık: Her olay, 0 ile 1 arasında değeri ile gerçeklenme ihtimali bulunan bir fonksiyondur. Bir olayın gerçeklenme olasılığının rakamsal değeri (P), olay sayısının (a), toplam olaylar sayısına (n) bölümüdür.

$$P = a / n$$

Örnek: Bir robada 100 adet bilye olsun. Bu bilyelerden 10 tanesinin beyaz olduğu bilindiğine göre çekilen bir bilyenin beyaz olma olasığı nedir.
P(B)=10/100=0.1

Çekilen bilyenin beyaz olmama olasılığı nedir? P(D)=90/100=0.9 (D: diğerleri)

Olasılık fonksiyonunun belirtildiği örnek uzaya, "olasılık uzayı" denir. Örnek uzaydaki her noktaya (basit rasgele olay) ve her alt kümeye (bileşik rasgele olay) olasılık uzayında bir nokta (bir olasılık) karşılık gelir.

Örnek:
1 ila 20 numaralı biletler karıştırılır ve daha sonra rastgele bir bilet çekilir. Çekilen biletin 3 veya 5'in katı olan bir sayıya sahip olma olasılığı nedir?
S = {1, 2, 3, 4, ..., 19, 20}.
E = 3 veya 5 in katı= {3, 6 , 9, 12, 15, 18, 5, 10, 20}.
P(E) = n(E)/n(S) = 9/20.

Olasılık Yasaları:

Olasılık yasalarının matematiksel ifadeleri, p olasılık fonksiyonu, A olay, A^c onun tümleyeni olmak üzere, aşağıdaki bağıntılarla verilir.

1. $0 \leq p(A) \leq 1$
2. $p(A^c) = 1 - p(A)$
3. $p(A \cap B) = p(A).p(B \mid A)$

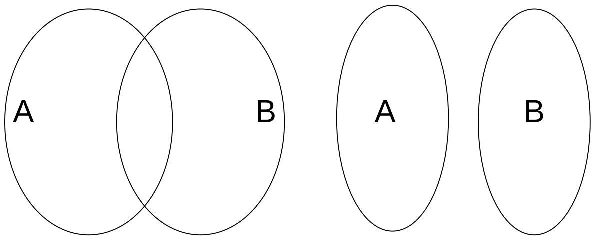
$$p(B|A) = \frac{p(A \cap B)}{p(A)}$$

A kümesindeki B olayının olma olasılığı, Kesişim olasılığının A olma olasılığına bölümüdür.

4) A ile B ayrık olmayan iki olay ise $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

5) A ile B ayrık iki olay ise $P(A \cup B) = P(A) + P(B)$

(Yani, A ile B ayrık iki olay ise $A \cap B = \emptyset$ *ve* $P(A \cap B) = 0$ dir.)



Örnek:

Bir torbada 9 mavi, 7 siyah, 4 sarı top vardır. Torbadan rastgele bir top çekersek, seçilen topun siyah olmama olasılığı nedir?

P(Siyah olma)=7/20

P(Siyah olmama)=20/20 – 7/20 =13/20

İadesiz olarak ilkinin mavi ikincisinin sarı olma olasılığı nedir?

İlkinin mavi olma olasılığı=9/20

İadesiz olarak ilkinin mavi ikincisinin sarı olma olasılığı=(9/20)*(4/19)=36/380=9/95

Örnek: Bir zarın bir kere atılışı deneyinde her olayın olasılığı 1 / 6 dır. Örnek uzay ise;

S = {1, 2,3,4,5,6 }

2 veya 5'in üste gelmesi olayını A ile gösterirsek;

A = { 2, 5 }   P(A) = 1/6 + 1/6 =2/6 = 1/3

A'nın ortaya çıkmama olasılığını(1,3,4 veya 6'nın üste gelmesi) A' ile gösterirsek;

P(A') = 1/6 + 1/6 + 1/6 + 1/6 = 4 / 6 = 2/ 3 veya

P(A') = 1- P(A) = 1 – 1/ 3 = 2 / 3

Örnekten anlaşılacağı gibi bir olayın olasılığı 0 ve 1 arasında değişmektedir. Olayın ortaya çıkması mümkün değilse olasılığı 0, ortaya çıkması muhakkak ise olasılığı 1 dir.

Örnek:

Bir A olayının olma olasılığı hangisi olamaz?

a) 0 b) ½ c) 1 d) 3/2

Örnek:

Bir zar atıldığında, üste gelen yüzünün tek sayı olma olasılığı nedir?

A={1,3,5}

P(A)=3/6=1/2

Örnek:

Bir kutuda 8 kırmızı, 7 mavi ve 6 yeşil top vardır. Bir top rastgele alınır. Ne kırmızı ne de yeşil olma olasılığı nedir?

N(S)=Toplam top sayısı=(8+7+6)=21

N(E)=Çekilen top ne kırmızı ne de yeşil=7

P(E)=N(E)/N(S)=7/21=173

Örnek:

Bir torbada 5 kırmızı, 10 siyah ve 15 beyaz ürünün renk kodları bulunmaktadır. Bu torbadan rastgele çekilecek bir ürünün

a) Herhangi bir renkte ürün gelme olasılığını hesaplayınız.
b) Kırmızı renkte ürün gelme olasılığını hesaplayınız.
c) Siyah renkte ürün gelme olasılığını hesaplayınız.
d) Beyaz renkte ürün gelme olasılığını hesaplayınız.
e) Mavi renkte ürün gelme olasılığını hesaplayınız.

Toplam ürün= 5 + 10 + 15 =30

a) P(ürün)=30/30=1
b) Toplam kırmızı renk kodlu ürün =5
   Torbadan rastgele çekilecek bir ürünün kırmızı gelme olasılığı=5/30=1/6
c) Toplam siyah renk kodlu ürün =10
   Torbadan rastgele çekilecek bir ürünün siyah gelme olasılığı=10/30=1/3
d) Toplam beyaz renk kodlu ürün =10
   Torbadan rastgele çekilecek bir ürünün beyaz gelme olasılığı=15/30=1/2
e) Toplam mavi renk kodlu ürün =0
   Torbadan rastgele çekilecek bir ürünün mavi gelme olasılığı=0/30=0

Dağılım:

Olaylar (ya da önermeler) üzerinde tanımlı ve bir olayın kaç kez olduğunu gösteren bir fonksiyondur. Olasılığın bir uygulaması olan istatistikte önem taşır.

Olasılığın incelenmesinde sonucu bilinemeyen deneyler gözönüne alınır. Bu deneylerin bir çoğu tekrar edilebilir deneylerdir ve rasgele deneyler(random experiment) olarak adlandırılır. Her deneyin sonucu gözlenebilir ve liste olarak yazılabilir. Üretilen makine parçaları arasından kusurlu olanları belirlemek, bir ana kitle içinden örnek seçmek tekrarlanabilir deneylerdir.

Örnek:

Madeni para ile iki defa atış yapalım. Bu iki atımda 4 durumdan biri ortaya çıkacaktır.

Y ile yazı , T ile tura gösterilmek üzere; ***YY YT TY TT.***

Deneylerin sonuçları **olay** olarak tanımlanmaktadır. Zar ile yapılan atışta olaylarımız veya sonuçlarımız 1,2,3,4,5 ve 6 nın üste gelmesi olacaktır. Bu sonuçların herbiri birer basit olaydır. Özetle **basit olay** tek bir deneyde tek bir sonuç veren olaylardır**. P(E)** ile basit olayın gerçekleşme olasılığı gösterilir. Madeni para ile tek atışta tura gelmesi basit olaya örnek olarak verilebilir. Bir deneyin bütün sonuçları o deneyin örnek uzayını meydana getirir. Örnek uzay **S** harfi ile gösterilir.

Bir deneyin olası sonuçlarından herbirini $\mathbf{p_i}$ ve örnek uzayı **P** ile gösterirsek bir zarın atılışında örnek uzay; **P= {p1 , p2, ...,p6 }** şeklindedir.

Örnek:

Bir lotarya oyununda üç basamaklı bir tam sayı rasgele seçilmektedir.Üç basamaklı sayıların herbiri olası bir sonuçdur. Örnek uzay: S = { 000,001,002,.........,998,999 }

Örnek:

Bir kutuda siyah ve sarı renkde toplar bulunmaktadır. Kutudan bir top çekip tekrar yerine koyarsak ve bu deneyi 100 defa tekrarlarsak çekilen topun sarı ya da siyah olması gerektiğinden deneyin 2 basit olayı olacaktır.

$E_1 =$siyah topun çekilmesi,

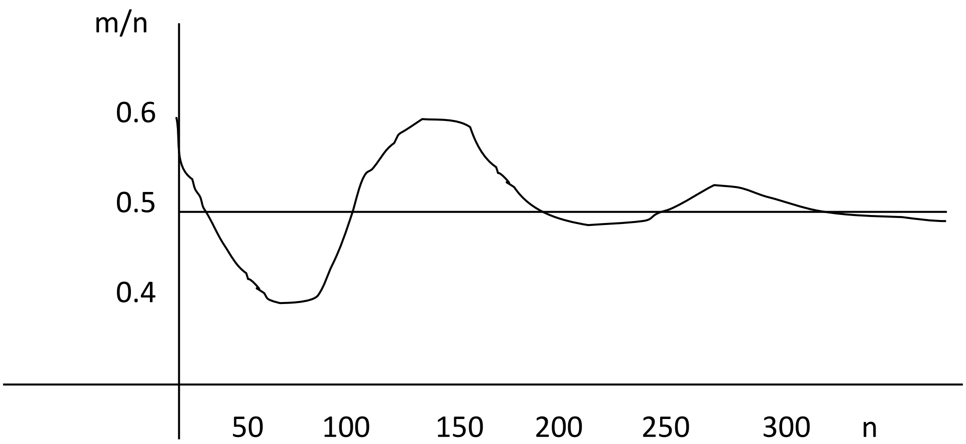
$E_2 =$sarı topun çekilmesi,

örnek uzay S = {E1 , E2 }

İki veya daha çok olayın birlikte veya birbiri ardına gerçekleşmesi ile bileşik olay meydana gelir. Bileşik olayın gerçekleşme olasılığı E1 , E2 iki olayı göstermek üzere **P(E1,E2)** şeklindedir. Para ile iki atış yapıldığında; ***YY YT TY TT s***onuçlarının çıkması bir bileşik olaydır. Aynı şekilde zarla yapılan 2 atışta 6 x 6=36 sonucu da bileşik olaydır. Basit ve bileşik olay tanımlarından başka olayı **bağımlı** ve **bağımsız** olay olarak da tanımlayabiliriz. Bir olayın ortaya çıkması diğer bir olayın veya olayların ortaya çıkmasına neden olmuyorsa bu olaylar **bağımsız** olaylardır. Zarın 2 defa atılmasında 1. atışın sonucu, 2. atışı etkilemeyeceği için bağımsız olay olarak adlandırılır. Bir olayın meydana gelmesi diğer bir olayın veya olayların meydana gelmesini etkiliyorsa bu tür olaylar **bağımlı** olaylardır. 52'lik iskambil destesinden iadesiz olarak arka arkaya 2 kağıt çekildiğinde, ikinci kartın sonucu birinci kartın sonucuna bağımlıdır.

Nisbi Frekansla Olasılık Belirleme:

Para ile atış yapıldığında deney sonuçları ***yazı*** ve ***tura*** basit olaylarından biri olacaktır. Deneyin 300 defa tekrarlandığını varsayarsak ilk atışlarda yazı olayının toplam atışa oranı % 50 'den uzak olacak, deney sayısı çoğaldıkça oran % 50'ye çok yakın seyredecektir. ***y*** ekseninde yazı gelme sayısının toplam atış sayısına oranı (m / n), ***x*** ekseninde de toplam atış sayısı(n) gösterildiğinde, aşağıdaki şekilde görüldüğü gibi atış sayısı arttıkça yazı sayısı (m) %50 cıvarında olacaktır.



Bir deney çok tekrar edilirse o olayın nisbi frekansı yaklaşık olarak o olayın olasılığına eşit olacaktır. Nisbi frekans bir olayın olasılığını vermemekte ancak gerçek olasılığın tahminini yapabilmemizi sağlamaktadır. Para atışında görüleceği gibi m ≤ n olacaktır. Yazı sayısı toplam atış sayısından küçük veya eşittir. Nisbi frekans;

$$m / n <= 1$$

atışlarda hiç yazı çıkmamışsa m=0 ve m / n =0 olacaktır. Bu durumda nisbi frekans;

$$0 \leq m / n \leq 1$$

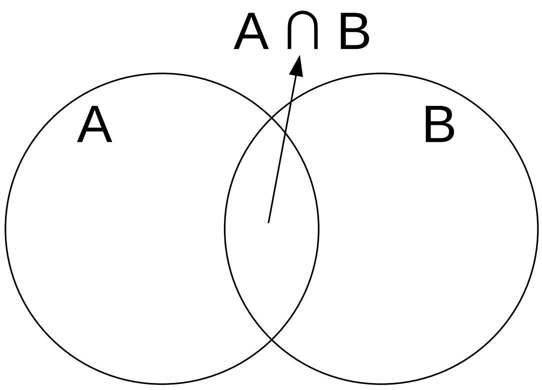
m / n, P(A)'nın bir tahmini olduğuna gore P(A) için de aynı şey söylenebilir;

$$0 \leq P(A) \leq 1$$

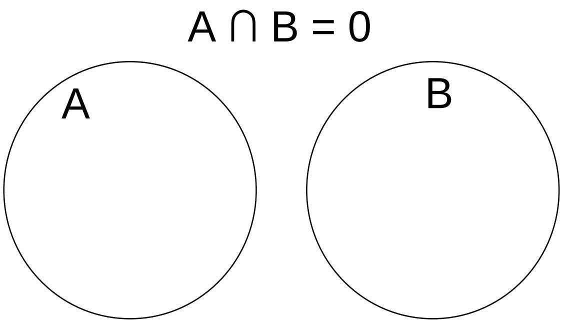
nisbi frekans yaklaşımına göre bir olayın olasılığının 1 olması o olayın ortaya çıkmasının kesin olduğu anlamına gelmemektedir. n'nin çok büyük olduğu durumlarda olay çok sayıda ortaya çıkmaktadır. Nisbi frekans yaklaşımı ancak tekrarlanabilir olaylara uygulanabilir. Tekrar edilemeyen olaylara uygulanması risklidir. Örneğin uzay deneyinin başarı olasılığının belirlenmesinde olay fazla tekrar edilemiyeceği için uygulamak zordur.

4.1. Olasılık ve Kümeler

Olasılıkla ilgili bu hesaplamalar kümelerle de ifade edilebilir. Bir deneyin sonuçlarını veya basit olaylarını bir örnek uzayı içinde noktalar olarak belirtebiliriz. Basit olaylar noktalarla gösterilebilirler ancak bunların toplamı olan bileşik olaylar bir küme veya noktalar grubu olarak gösterilirler. A ve B gibi birbirini engellemeyen iki bileşik olay ***Venn*** diyagramı ile aşağıdaki gibi çizilebilir:

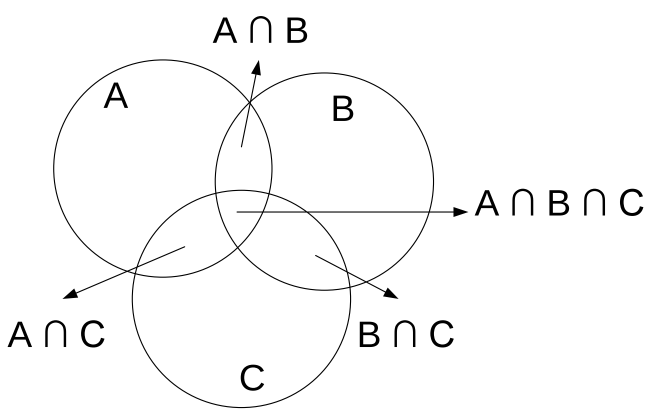


Birbirini engelleyen olaylarda ortak nokta yoktur yani P(AB) = 0 olmaktadır.



Küme notasyonu kullanılırken A + B olayı **A U B** (bileşim) olarak, AB olayı da **A∩ B** (kesişim) olarak gösterilir.

Venn Diyagramı:



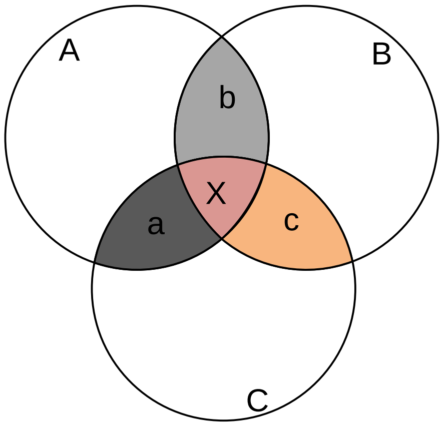
Örnek:

Okul kantininde 100 öğrenci üzerinden bir araştırma yapılmış ve aşağıdaki sonuçlar bulunmuştur. Çocuklardan 75'i dondurmadan, 50'si şekerden, 80'i tatlılardan, 60'ı hem tatlı hem de dondurmadan hoşlanmaktadır. 40 çocuk tatlı ve şekerden hoşlanırken, sadece 30'u üçünden de hoşlanmaktadır. Hem dondurma hem de şekerlerden hoşlanan kaç çocuk vardır?

A: dondurmadan hoşlanan çocukların kümesi, A1: Sadece dondurma

B: şekerden hoşlanan çocukların kümesi, B1: Sadece şeker

C: tatlılardan hoşlanan çocukların kümesi olsun, C1: Sadece tatlı



S(T)=100

S(A)=75

S(B)=50

S(C)=80

S(A ∩ C)=60

S(B ∩ C)=40

S(A ∩ B ∩ C)=30

X= S(A ∩ B ∩ C)=30

a= S(A ∩ C) – X = 60 – 30, a=30

c= S(B ∩ C) – X = 40 – 30, c=10

b= S(A ∩ B) – X,   b= S(A ∩ B) – 30

S(C)= C1+a+ X +c,   80=C1+70, C1=10

S(T)=A1+B1+C1+a+b+c+ X

100=A1+B1+10+30+b+10+30, A1+B1+b=20

S(B)=B1+b+c+ X

50=B1+b+10+30, B1+b=10

A1+B1+b=20, A1+10=20, A1=10

S(A)=A1+a+b+ X, 75=10+30+b+30, b=5

S(A ∩ B)=b + X =5+30=35

Örnek:

Bir konfeksiyon firması işçi alımında bir model oluşturmak istiyor. Bunun için şirket yöneticileri, gelecek beş yıl içinde yeni işçilerin %80'ninin kadın ve %30'unun da bekar olmasını istiyor. Her yeni beş işçiden birinin de evli erkek olmasını planlıyor. Yöneticilerin bekar bir kadını ise alması olasılığı nedir?

Burada 4 temel olay vardır:
A işçinin bekar olması
B işçinin evli olması
C işçinin kadın olması
D işçinin erkek olması

Bazı olasılıklar soruda verilmiş:
$P(C) = 0.80$, $P(D) = 0.20$,
$P(A) = 0,30$, $P(B) = 0,70$ olur. $P(B \cap D) = 1/5 = 0.20$
$P(A \cap C)=?$

$P(A \cap C)=P(C)-P(B \cap C)$ bulunabilir.
$P(B) =P(B \cap C) + P(B \cap D)$
$0.70 = P(B \cap C) + 0.20$, $P(B \cap C) = 0.50$
$P(A \cap C) = 0.80 – 0.50= 0.30$'u bekar kadın olur.

Örnek:

Bir öğrencinin cebinde kağıt para olarak birer adet 100, 50, 20,10, 5 TL bulunsun. Öğrenci cebindeki iki adet kağıt parayı düşürmüş olsun.

a) İkili paraların toplam örnek uzayı ne olur?
   (100 +50), (100 + 20), (100 + 10), (100+5):   4 durum
   (50 + 20), (50+10), (50 +5): 3 durum
   (20 + 10), (20+5): 2 durum
   (10 + 5): 1 durum
b) İki para düştüğünde 100TL'den büyük parayı düşürme olasılığı nedir?
   $P(>100)=4/10$

c) 50 TL'den küçük olma ihtimali nedir?
   $P(50<)=3/10$

4.2. Olasılık Kuralları

P Olasılık Fonksiyonunun Özellikleri:

Teorem: P(*A*) = 1-P(A)'dır.

Teorem: P(∅) =0'dır.

Teorem: Eğer $A \subset B$ ise $P(A) \leq P(B)$'dir.

Teorem: Herhangi bir A olayı için $P(A) \leq 1$'dir.

Teorem: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$'dir.

Teorem: A, B ve C olayları aynı S örneklem uzayında tanımlanmış üç olay olmak üzere bu üç olayın birleşim;

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap C) - P(B \cap C) - P(A \cap B) + P(A \cap B \cap C)$ olarak yazılır.

4.3. Toplama kuralı

Birbirlerini Engeleyen Olaylar:

A ve B birbirini engelleyen olaylar olduğunda, A olayının veya B olayının ortaya çıkması bu olayların basit olasılıklarının toplamına eşittir. (A+B) ile gösterilir. Birbirlerini engelleyen olayların olasılıkları toplam,

$$P(A+B) = P(A) + P(B)$$

P(A veya B) = P(A) + P(B)

Birbirlerini engelleyen olayların olasılıkları toplamı 1 dir.

Ayrık olaylar
(Birbirini Engelleyen olaylar)

- Bir mağazaya giren müşteri 0.2 olasılıkla lacivert, 0.3 olasılıkla siyah bir takım elbise alacaktır. Bu müşterinin mağazadan iki takımdan birisini alma olasılığı nedir?
- **$P(A \cup B) = P(A) + P(B) = 0.2 + 0.3 = 0.5$**

Örnek:

Kusursuz bir tavla zarı atıldığında 2 veya 3 gelmesi olasılığı nedir?

Çözüm: Bu olay birbirini engelleyen özellikte olup, herhangi bir anda sadece tek yüz ile karşılaşılacağından toplama kuralı kullanılmalıdır.

P(2 veya 3) =P(2)+ P(3)=1/6 + 1/6= 2/6=1/3

Örnek :

Bir paketleme fabrikası otomatik makineyle yarım kiloluk plastik torbalara bezelye, havuç, patates doldurmaktadır. Paketlerin çoğu yarım kilo gelmekte, ancak sebzelerin boyutları nedeniyle bazı paketler daha ağır yada hafif gelmektedir. Geçmişte yapılan kalite kontrol çalışmalarında Tablo 1'deki veriler elde edilmiştir.

Tablo 1 : Örnek 1 için veriler Ağırlık	Olay	Torba Sayısı	Olasılık
Eksik	E	100	0,025
Tam	T	3.600	0,900
Fazla	F	300	0,075
Toplam		4.000	1,000

Üretim sürecinden seçilen herhangi bir torbanın eksik yada fazla olma olasılığı nedir?

P(E)=0,025; P(T)=0,900; P(F)=0,075 olduğuna göre:

P(E yada F)=P(E)+P(F)=0,025+0,075=0,10 olacaktır.

Örnek:

12 adet Kırmızı, 4 adet Siyah, 4 Adet Yeşil ve 4 adet Mavi renkli GSM telefon kılıfları bir torbanın içerisine konulmuştur. Torbanın içerisinden 1 adet kılıf çekilmiştir. Çekilen kılıfın Kırmızı veya Yeşil olma olasılığını bulunuz.

Toplama:

- Ayrık olaylar:
  P(K veya Y)=P(K) + P(Y)
  Soru ayrık olay olasılık sorusudur.
  P(K)=12/24=1/2=0.5
  P(Y)=4/24=1/6
  P(K veya Y)=P(K) + P(Y)=12/24+4/24=16/24=2/3

Toplam Olasılık - Birbirlerini Engellemeyen Olaylar:

Eğer olaylar birbirini engellemiyorsa A olayının veya B olayının ortaya çıkması, ya A olayının ya B olayının ya da A ve B olaylarının her ikisinin birlikte gerçekleşmesi anlamındadır. A ve B birbirlerini engellemiyorsa;

P(A+B)=P(A) + P(B) – P(AB)

P(A veya B)=P(A) + P(B) – P(A ve B)

P(A veya B veya C)=P(A) + P(B) + P(C) – P(A ve B) – P(A ve C) – P(B ve C) + P(A ve B ve C)

$P(A \cup B \cup C)=P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

Örnek:

Bir deste iskambil kağıdından bir vale çekme olayı A ile, bir maça çekme olayı da B ile gösterilsin. Bu iki olay birbirini engellemediğinden A veya B nin olma olasılığı yani maça valesi çekme olasılığı vardır. Bu durumda A'nın veya B'nin olma olasılığı;

$$P(A+B)=(4/52) + (13/52)-(1/52)=4/13$$

Örnek:

Bir mağazadaki reyonda kutuların içerisinde aşağıda tabloda verilmiş ölçütlerde ve renklerde tişörtler konulmuştur. Bir adet tişörtün satıldığını biliyoruz.

	Mavi	Kırmızı	Beyaz	Siyah
Küçük	3	3	3	3
Orta	4	4	4	4
Büyük	3	3	3	3

a) Satılan tişörtün Mavi olma olasılığı nedir?
   10/40=1/4
b) Satılan tişörtün <u>Beyaz renkli veya Orta boy</u> olma olasılığı nedir?
   - **Birleşik olaylar:**
     P(B veya O)=P(B) + P(O) – P(B ve O)
     =10/40+16/40-4/40=22/40=11/20
c) Satılan tişörtün <u>Beyaz renkli ve Orta boy</u> olma olasılığı nedir?
   P(B ve O)=4/40

Örnek:

Küçük bir montaj atölyesinde 50 işçi çalışmaktadır. Her işçinin kendisine verilen işi zamanında ve son kontrole hazır biçimde bitirmesi beklenmektedir. Bazen işçiler performans ölçütlerine erişemeyerek ya geç kalır, ya da hatalı montaj gerçekleştirirler. Performans değerleme döneminin sonunda 50 işçiden 5'i geç bitirmiş, 6'sı hatalı montaj yapmış, 2'si hem gecikmiş hem de hatalı montaj yapmıştır. Bir işçinin geç yada hatalı montaj yapma olasılığı nedir?

P(G)=5/50=0,10; P(H)=6/50=0,12; P(G ve H)=2/50=0,04 olarak bulunur.

Bir işçinin geç yada hatalı montaj yapma olasılığı:

P(G yada H)= P(G) + P(H) - P(G ve H) = 0,10 + 0,12 – 0,04 = 0,18'dir.

Örnek:

Bir işin 2-4 günde bitirilmesi olasılığı 0.50, 4-6 günde bitirilmesi olasılığı 0.25, 2-6 günde bitirilmesi olasılığı 0.55 ise 4 günde bitirilmesi olasılığı nedir?

$A= \{(X=2) \cup (X=3) \cup (X=4)\}$ $\quad P(A)=0.50$

$B= \{(X=4) \cup (X=5) \cup (X=6)\}$ $\quad P(B)=0.25$

$A\cup B= \{(X=2) \cup (X=3) \cup (X=4) \cup (X=5) \cup (X=6)\}$ $\quad P(A\cup B)=0.55$

$A\cap B= \{(X=4)\}$ $\quad \underline{P(A\cap B)= ?}$

$\underline{P(X=4)} = P(A\cap B) = P(A) + P(B) – P(A\cup B) = 0.50 + 0.25 – 0.55 = \underline{0.20}$

4.4. Çarpma Kuralı

Çarpma kuralı - bağımsız olay:

Bağımsız olaylarda çarpma kuralı;

P(A ve B)= P(A) * P(B) şeklindedir.

Birbirinden bağımsız ve aynı zamanda meydana gelebilen olayların olasılığı, bu olayların ayrı ayrı ortaya çıkma olasılıkları çarpımına eşittir.

Örnek:

Kusursuz bir tavla zarı ve madeni para birlikte atıldığında, paranın yazı ve zarın 5 gelmesi olasılığı nedir?

Çözüm: Bu olaylar birlikte meydana gelebilen özellikte olup, birbirini engellemez. Bu nedenle çarpma kuralı kullanılmalıdır.

P(Yazı ve 5 gelme)= P(Yazı) x P(5 Gelme)=(1/2) x (1/6)=1/12

Örnek:

Bir piyangoda 8 boş, 2 ikramiyeli bilet vardır. Bu piyangodan 2 bilet alan bir kişinin ikramiye kazanma olasılığı nedir ? (iadesiz)

Birinci biletin kazanma olasılığı 2/10'dur. Birinci bilet ikramiye kazanırsa geriye 8 boş 1 ikramiyeli 9 bilet kalır. Ikinci biletin kazanma olasılığı 1/9'dur. Her iki biletin de ikramiye kazanma olasılığı:

P(B1ve B2)= (2/10)*(1/9) = 1/45

Örnek:

Bir kutuda 5 adet yeşil renkte, 3 adet de beyaz renkte top bulunmaktadır. Kutudan 2 top çekildiğinde her ikisinin yeşil olma olasılığı nedir ? (Toplar kutuya iade edilmiyor)

A: ilk çekilen topun yeşil olması olayı,

B: ikinci çekilen topun yeşil olması olayı gösterilsin.

A ve B bağlı olaylar olduğu için P(A ve B) hesaplanacaktır. 1. topun yeşil olma olasılığı :

P(A) = (5 / 8)

1. topun yeşil olması durumunda 2. topun yeşil olma olasılığı:

P(B) = (4 / 7)

P(A ve B) = P(A) * P(B) = (5/8)*(4/7) = 0.357

Örnek:

A kişisinin 15 yıl sonra hayatta kalma olasılığı %80, B kişisinin 15 yıl sonra hayatta kalma olasılığı %60 ise, her ikisinin 15 yıl sonra hayatta kalma olasılığı nedir?

P(A ve B)=0.80 * 0.60 = 0.48

Örnek:

Bir sınıfta 12 erkek, 4 kız öğrenci vardır. Üç kişi rasgele seçiliyor. Bunların üçünün de erkek olma olasılığı nedir?

P= (12/16)(11/15)(10/14)=11/28

Bunların ikisinin kız birinin erkek olma olasılığı nedir?

P= (4/16)(3/15)(12/14)=3/70

Örnek:

52 kartlık bir desteden arda arda 5 kart seçiliyor. Kartların hepsinin maça seçilme olasılığı nedir?

P= (13/52)*(12/51)*(11/50)*(10/49)(9/48)=33/66640

Örnek:

Bir kutuda 6 adet limon ve 4 adet portakal bulunmaktadır.

a) İadeli olarak sırayla torbadan çekilen iki narenciyenin ilkinin portakal ve diğerinin limon olma olasılığı nedir?

   İaedeli, Birbirinden bağımsız olaylar:

   P(P ve L)= P(P)*P(L), Portakal çekiliyor ve tekrar kutuya konuluyor.

   P(P ve L)=4/10 * 6/10=24/100=6/25

b) İadesiz olarak sırayla torbadan çekilen iki narenciyenin ilkinin portakal ve diğerinin limon olma olasılığı nedir?

   İaedesiz, Birbirine bağımlı olaylar:

   P(P ve L)= P(P)*P(L /P), Portakal çekildikten ve kutudan ayrıldıktan sonra geriye kalanlardan limon çekilme olasılığı.

   İlki portakal, ikincisi limon, P(P ve L)=4/10 * 6/9 =24/90=8/30=4/15

   İlki limon, ikincisi portakal, P(L ve P)=6/10 * 4 /9 =24/90=4/15

c) İadesiz olarak sırayla torbadan çekilen iki narenciyenin de limon olma olasılığı nedir?

   İaedesiz, Birbirine bağımlı olaylar: P(L ve L)= P(L)*P(L /L), Limon çekildikten ve kutudan ayrıldıktan sonra geriye kalanlardan limon çekilme olasılığı.

   P(L ve L)= 6/10 * 5/9=30/90=1/3

Örnek : 4 kırmızı ve 6 yeşil top bulunan bir kutudan iki defa ardı ardına iadeli çekiliş yapılacaktır. Çekilen ilk topun kırmızı, ikinci topun yeşil olması olasılığı nedir?

P(K)=4/10; P(Y)=6/10

Çekiliş iadeli olduğu için bu iki olay bağımsızdır. Bu nedenle bileşik olasılık:

$P(K\ ve\ Y) = P(K) \times P(Y) = \frac{4}{10} \times \frac{6}{10} = \frac{24}{100}$ olacaktır.

Örnek : 4 kırmızı ve 6 yeşil top bulunan bir kutudan iki defa ardı ardına iadesiz çekiliş yapılacaktır. Çekilen ilk topun kırmızı, ikinci topun yeşil olması olasılığı nedir?

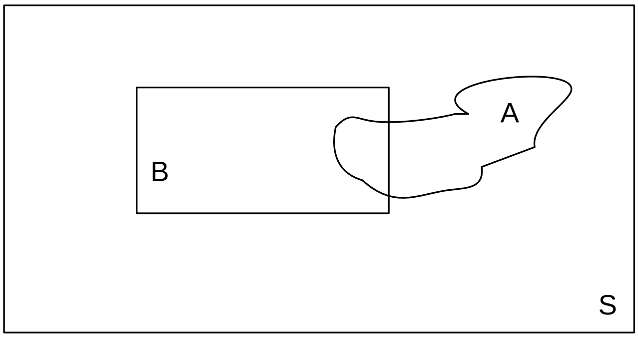
P(K)=4/10; P(Y/K)=6/9

Çekiliş iadeli olduğu için bu iki olay bağımsızdır. Bu nedenle bileşik olasılık:

$P(K\ ve\ Y) = P(K) \times P(Y/K) = \frac{4}{10} \times \frac{6}{9} = \frac{24}{90} = \frac{4}{15}$'dir.

4.5. Koşullu Olasılık

A ve B bağımlı olaylar ve A olayı olma olasılığı, P(A) olsun. Ardından A olayı olduktan sonra B olayının olma olasılığı, P(B \A) olsun. İki olayın sıralı olma olasılığı, yani ilkinde A olayı ikincisinde ise A olayı olduktan sonra B olayının birlikte sıralı olma koşullu olasılığı,
P(A∩B)=P(A) * P(B \A) ifadesi ile bulunur.



Örnek:
Bir torbada 3 mavi, 9 beyaz bilye bulunmaktadır. Üst üste yapılacak iki çekilişten birincisinde mavi, ikincisinde beyaz bilye gelme olasılığını hesaplayınız.
Çekiliş iadeli ise, bağımsız çarpma olayıdır. P(M ∩ B)= P(M) x P(B)=(3/12) x (9/12)=3/16
Çekiliş iadesiz ise, P(M ∩ B)=P(M) x P(B/M)= (3/12) x (9/11)=9/44

Örnek:
Bir torbada aynı büyüklükte 2 mavi, 4 kırmızı ve 5 siyah top vardır. Geri konulmamak şartıyla torbadan 3 top çekiliyor. Bu topların ilk ikisinin siyah üçüncüsünün mavi olma olasılığı nedir?
P(SSM)=(5/11) * (4/10) * (4/9)=8/99

Örnek:

Bir fabrikada üretilen parçalardan kusursuz 40 tanesi ve kusurlu 8 tanesi bir depoya konuyor. Çekilen yine yerine koyulmaksızın sırayla rastgele iki parça seçildiğinde her iki parçanın da kusurlu olması olasılığı nedir?

Çözüm: A ve B olayları aşağıdaki gibi tanımlansın.

A: İlk seçilen parça kusurludur. B: İkinci parça kusurludur.

O halde; $P(A) = \frac{8}{48} = \frac{1}{6}$ ve $P(B/A) = \frac{7}{47}$

öte yandan, $P(A \cap B) = P(A).P(B/A) = \frac{1}{6}.\frac{7}{47} = \frac{7}{282}$

Örnek:

Elimizde iki kavanoz var. Birinci kavanozda 6 mavi ve 3 sarı, ikinci kavaonozda ise 5 mavi ve 4 sarı top bulunmaktadır. Bu durumda rasgele çekilen bir kavanozdan sarı çekme olasılığı nedir?

P(C)= $P(A \cap C) + P(B \cap C)$

$P(A \cap C) = P(A) * P(C / A)$ =(1/2)*(3/9)=3/18

$P(B \cap C) = P(B) * P(C / B)$=(1/2)*(4/9)=4/18

P(C)=3/18 + 4/18=7/18

Örnek:

3 kız, 6 erkek öğrencinin olduğu bir gruptan iki öğrenci seçilecektir. Seçilen 1. öğrencinin kız, 2. öğrencinin erkek olması olasılığı nedir?

P(K)=3/9=1/3

P(KE)=(3/9) * (6/8)=18/72=1/4

Örnek:

Bir torbada 4 beyaz, 3 mavi, 2 sarı top vardır. Torbaya geri atılmamak koşulu ile arka arkaya 3 top çekiliyor. 1. topun beyaz, 2. topun mavi, 3. topun sarı renkli olma olasılığı nedir?

P(BMS)=(4/9)*(3/8)*(2/7)=1/21

Örnek:

Kusursuz bir tavla zarı atıldığında sonucun çift bir sayı olduğu biliniyor. Bu sayının 4 çıkma olasılığını hesaplayınız.

S={1,2,3,4,5,6} , A={2,4,6} , B={4}

P(B/A)= P(A ∩ B) /P(A) =(1/6) / (3/6)=1/3

Örnek:

Bir kutuda test edilen ürünlerin %30'si A-testinden, %48'i B-testinden kalmış. Ürünlerin, %24'ise hem A-testinden hem de B-testinden kalmış.

P(A)=0.3

P(B)=0.48

P(A ∩ B)=0.24

a) Ürünlerin A veya B testlerinde kalma olasılığı nedir?

P(M ᴜ T)=P(M)+P(T)-P(M∩T)==0.3+0.48-0.24=0.54

b) A testinde kalan bir ürünün B testinden kalma olasılığı

P(B \ A)= P(A ∩ B) / P(A)=0.24/0.3=0.8

c) B testinden kalan ürünün A testinden kalma olasılığı

$P(A \setminus B)= P(A \cap B) / P(B)=0.24/0.48=0.5$

Örnek:

Aynı anda iki hilesiz zar atılsın. Eğer gelen iki sayının toplamı 6 ise, bir zarda gelen sayının 2 olma olasılığı nedir? İki zar atıldıktan sonra toplamı 6 ise bir zarda 2 gelme olasılığı nedir?

Toplam 6 olan sayılar, A= { (1,5),(2,4), (3,3),(4,2), (5,1) }

Bir zarda 2 gelen sayılar, B={ (2,1), (2,2), (2,3),(2,4),(2,5),(2,6),(1,2),(3,2),(4,2),(5,2),(6,2) }

$A \cap B$ ={ (2,4), (4,2) }

$P(A)=5/36$

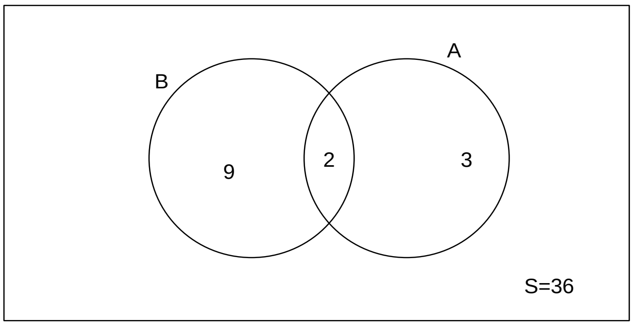
$P(B)=11/36$

$P(A \cap B)=2/36$

$P(A \setminus B)$, B geldikten sonra A gelme olasılığı = 2/11

$P(B \setminus A)$, A geldikten sonra B gelme olasılığı = 2/5

$P(B \setminus A)$= $A \cap B$ deki öğe sayısı / A'nın oluş sayısı



Örnek:

Bir sınıfta öğrencilerin %25'i Matematik, %20'si Türkçe, %10'ise hem Matematik hem Türkçe derslerinden kalmış.

$P(M)=0.25$

$P(T)=0.20$

$P(M \cap T)=0.10$

d) Öğrencilerin Matematik ya da Türkçe'den kalma olasılığı nedir?

$P(M \cup T)=P(M)+P(T)-P(M \cap T)=0.25+0.20-0.10=0.35$

e) Matematikten kalan bir öğrencinin Türkçe'den kalma olasılığı

$P(T \setminus M)= P(M \cap T) / P(M)=0.10/0.25=2/5$

f) Matematikten kalan bir öğrencinin Türkçe'den kalma olasılığı

$P(M \setminus T)= P(M \cap T) / P(T)=0.10/0.20=1/2$

Örnek:

Drone üreten 3 Makine için aşağıdaki hata ve üretim oranları veriliyor:

	Hatalı Drone oranı	İmalattaki oranı
A. Makine	%2	%35
B. Makine	%1	%25
C. Makine	%3	%40

a) Üretilen yaylardan rasgele birisi seçildiğinde bozuk olma olasılığını hesaplayınız.
P(H)=P(A)P(H|A)+P(B)P(H|B)+P(C)P(B|III)
P(H)=(35/100)*(2/100)+ (25/100)*(1/100)+(40/100)*(3/100)=215/10000=0.0215

b) Eğer bir drone hatalı çıktıysa bunun 3. makineden çıkma olasılığını hesaplayınız.
P(C|H)=P(C)P(B|III)/P(H)=120/215

Örnek: Okulumuz öğrencilerinden %45'i istatistik, %35'i bilgisayar derslerinden ve %25'i hem istatistik hem de bilgisayar derslerinden başarısızdır. Rasgele seçilen bir öğrencinin,

a) Bilgisayardan başarısız ise, istatistikten de başarısız olma olasılığını,
b) İstatistikten başarısız ise, bilgisayardan da başarısız olma olasılığını,
c) Bu iki dersten en az birinden başarısız olma olasılığını bulunuz.

İ, istatistik dersinden başarısız öğrencileri; ve B, bilgisayar dersinden başarısız öğrencileri göstersin.

$P(\dot{I})=0.45, \quad P(B)=0.35 \quad \text{ve } P(\dot{I}\cap B) = 0.25$

a) $P(\dot{I}|B) = \dfrac{p(\dot{I}\cap B)}{P(B)} = \dfrac{0.25}{0.35} = \dfrac{5}{7}$

b) $P(B|\dot{I}) = \dfrac{P(B\cap \dot{I})}{P(\dot{I})} = \dfrac{0.25}{0.45} = \dfrac{5}{9}$

c) $P(\dot{I}\cup B) = P(\dot{I}) + P(B) - P(\dot{I}\cap B) = 0.45 + 0.35 - 0.25 = 0.55$

Bulunur.

Örnek:

Elektrik - Elektronik mühendisliği bölümü 2. sınıf öğrencilerinin %25'i matematik dersinde, %15'i de hem matematik hem fizik dersinde üstün başarı göstermiştir. Bu sınıftan rastgele bir öğrenci seçildiğinde, seçilen öğrenci matematik dersinden üstün başarılı ise, fizik dersinden de üstün başarılı olma olasılığı nedir ? M: Matematik, F: Fizik.

P(M) = 0.25

P(M ve F) = 0.15

P(F \M)=P(M ve F) / P(M) =0.15 / 0.25=0.60

Örnek:

Suppose that A and B are events with probabilities: P(A)=1/3, P(B)=1/4, P(A ∩ B)=1/10. Find each of the following:

1. P(A | B) = P(A ∩ B)/P(B)=1/10/1/4=4/10
2. P(B | A) = P(A ∩ B)/P(A)=1/10/1/3=3/10
3. P(A' | B') = P(A' ∩ B')/P(B')=
   P((A U B)')/(1-P(B))=(1-P(A U B))/(1 – P(B))=
   (1 – (P(A)+P(B)-P(A ∩ B)))/(1-P(B))=
   (1 – (1/3+1/4-1/10))/(1-1/10)=(1-29/60)/9/10=
   31/60/9/10=31/54.

Örnek Yapılan bir çalışmada hastaların 0.20' si hem aspirin, hem de novaljin, 0.40' ı sadece aspirin ve 0.30' u da sadece novaljin kullanmaktadır. Rasgele seçilen bir hastanın aspirin kullandığı biliniyorsa, bu hastanın novaljin de kullanması olasılığı nedir?

A: Aspirin kullanma olayı
B: Novaljin kullanma olayı

$$P(A) = 0.40, \quad P(B) = 0.30, \quad P(A \cap B) = 0.20$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.20}{0.40+0.20}$$

Örnek:

Elektrik ampulü üreten bir fabrikanın üretiminin %20'si A tipi, %80'ide B tipi ampullerden oluşmaktadır. Hatalı üretim oranı A tipi ampullerde %36, B tipi ampullerde ise %18'dir. Rasgele seçilen bir ampulün hatalı olduğu bilindiğine göre bu ampulün A tipi olma olasılığı nedir?

Çözüm:

C : hatalı üretim oranını göstersin istenen olasılık P(A/C)=? dır

$$\mathbf{P(A / C) = \frac{P(A \cap C)}{P(C)}}$$

ancak bu formülün payındaki ifadenin değeri bilinmiyor.

P(C)=P(A) * P(C/A) + P(B) * P(C/B) = 0.20 * 0.36 + 0.80 * 0.18 = 0.216 ifadesi ve

$$\mathbf{P(C / A) = \frac{P(A \cap C)}{P(A)} \Rightarrow P(A \cap C) = P(A) * P(C / A)}$$ ifadesi yerine koyulursa

$$\mathbf{P(A / C) = \frac{P(A \cap C)}{P(C)} = \frac{P(A) * P(C / A)}{P(C)} = \frac{0.20 * 0.36}{0.216} = 0.333}$$ olarak bulunur

Örnek:

Elimizde bulunan iki kavanozdan birincisinde 3 mavi ve 4 sarı, ikincisinde 5 mavi 2 sarı top olsun. Bu durumda rasgele seçilen bir kavanozdan mavi top çekme olasılığı nedir?

A: Birinci kavanozu seçme durumu

B: İkinci kavanozu seçme durumu

C: Mavi top seçme durumu

$$P(C) = P(A)P(C \mid A) + P(B)P(C \mid B)$$

$$P(C) = P(A)P(C \mid A) + P(B)P(C \mid B) = \frac{1}{2}\frac{3}{7} + \frac{1}{2}\frac{5}{7} = \frac{4}{7}$$

4.6. Sonlu olasılıklar ve Ağaç diyagramı

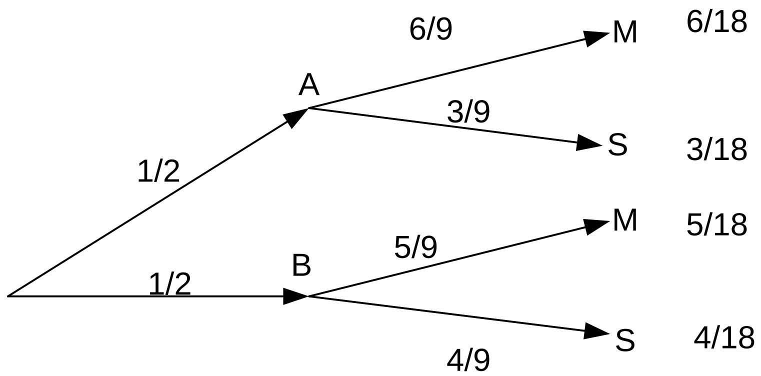
Her deneyin belirli olasılıklarla sonlu sayıda sonuçları olur. Sonlu sayıda olan olayların olasılığı ağaç diyagramı ile bulunabilir.

S örneklem uzayında A_1 olduktan sora A_2'nin olma olasılığı, A_2 olduktan sora A_3'nin olma olasılığı, ... , An-$_1$ olduktan sora A_n'nin olma olasılığı biliniyorsa,

$P(A_1 \cap A_2 \cap A_3 \cap .... \cap A_{n-1} \cap A_n) = P(A_1)\, P(A_2 \setminus A_1)\, P(A_3 \setminus A_2) .....\, P(A_n \setminus A_{n-1})$

Örnek:

Elimizde iki kavanoz var. Birinci kavanozda 6 mavi ve 3 sarı, ikinci kavaonozda ise 5 mavi ve 4 sarı top bulunmaktadır. Bu durumda rasgele çekilen bir kavanozdan sarı çekme olasılığı nedir?



Sarı top gelme olasılığı=3/18+4/18=7/18

Sarı topun birinci kavanozdan çekilme olasılığı= (3/18) /(7/18)=3/7

Sarı topun ikinci kavanozdan çekilme olasılığı= (4/18) /(7/18)=4/7

Örnek:

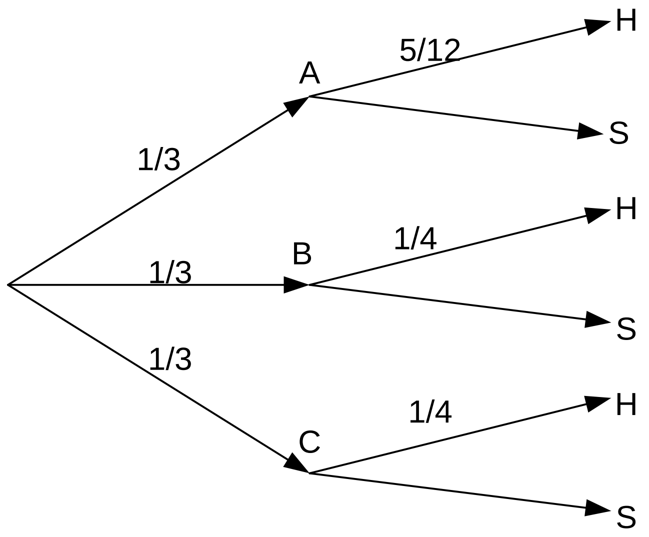
3 kutudaki çoraplar durumu aşağıda verilmiş olsun:

A kutusunda 5 hatalı toplam 12 çorap

B kutusunda 4 hatalı toplam 16 çorap

C kutusunda 6 hatalı toplam 24 çorap bulunmaktadır.

Rasgele bir kutu ve sonra rasgele bir çorap çekiliyor. Çekilen çorabın hatalı olma olasılığı nedir.



Ağaç herbir süreci tanımlar ve ağacın herbir dalı olasılık verir.

Çekilen çorabın hatalı olma olasılığı

P= 1/3 x 5/12 + 1/3 x 1/4 + 1/3 x 1/4 =5/36 + 1/12 + 1/12 =11/36

Hatalı çorabın A kutusunda çekilme olasılığı nedir?

P_A= (5 /36) / (11 /36)= 5/11

Hatalı çorabın B kutusunda çekilme olasılığı nedir?

P_B= (1 /12) / (11 /36)= 3/11

Hatalı çorabın C kutusunda çekilme olasılığı nedir?

P_C= (1 /12) / (11 /36)= 3/11

Örnek: Toplam olasılık formülü

Belli bir parça 3 makinede üretilmektedir. 1 nolu makinede hem 2 hem de 3 nolu makinelerdeki üretimin 2 katı kadar parça üretildiği bilinmektedir. Yine bilinir ki 1 ve 2 nolu makinelerdeki üretimin 0.02'si, 3 no lu makinedeki üretimin 0.04'ü kusurludur. Üretilen parçaların tümü bir depoya konuyor, sonra rastgele bir parça seçiliyor. Bu parçanın kusurlu bir parça olması olasılığı nedir?

Çözüm: Aşağıdaki olayları tanımlayalım.

A={Parça kusurludur}

B1={Parça 1 nolu makineden alındı}

B2={Parça 2 nolu makineden alındı}

B3={Parça 3 nolu makineden alındı}

P(A) olasılığını bulmak istiyoruz. O halde,

$P(A)=P(B1).P(A/B1)+P(B2).P(A/B2)+P(B3).P(A/B3)$

Burada $P(B1)=1/2$, $P(B2)=P(B3)=1/4$. Ayrıca $P(A/B1)=0.02$, $P(A/B2)=0.02$, $P(A/B3)=0.04$ olur. Bu değerleri formülde yerine koyarsak

$P(A)=0.025$ buluruz.

Örnek:

Bir fabrikada 2 adet makine bulunmaktadır. Herbir makinenin günlük üretim oranları A=%60, B=%40 olarak verillmiştir. Üretilenlerden kusurlu olanların yüzdesi ise A=%5, B=%2 olarak verilmiştir. Kusurlu parçalar birbirinden bağımsız olarak A, B makinelerinde üretilmektedir. Günün sonunda rasgele bir ürün seçiliyor ve bu ürünün kusurlu olma olasılığı nedir?

$P(A\cap H)=P(A)*P(H /A)$

$P(B\cap H)=P(B)*P(H /B)$

$P(T)= P(A\cap H)+P(B\cap H)$

Örnek:

Yarın havanın yağışlı olma olasılığı 0.4, güneşli olma olasılığı 0.6; Yarın yağmur yağarsa ertesi gün yağmur yağma olasılığı 0.3, güneşli olma olasılığı 0.7; Yarın güneşli olursa ertesi gün yağmur yağma olasılığı 0.2, güneşli olma olasılığı 0.8 dir. Bu durumda ertesi gün güneşli olma olasılığını dallanma yöntemi ile bulunuz.

4.7. Bayes Kuramı

Koşullu Olasılık:

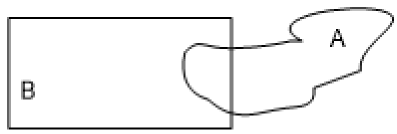
A ve B aynı örneklem uzayında tanımlanmış iki olay ve P(A)>0 olmak üzere A olayının gerçekleştiği varsayımı altında B olayının koşullu olasılığı,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

biçiminde tanımlanır. Bu tanımdan yararlanarak A ve B olaylarının birlikte gerçekleme olasılığını koşullu olasılık yardımı ile

$$P(A \cap B) = P(A)P(B|A)$$

biçiminde bulunur.



Örnek: Öğrencilerden %45'i istatistik %35'i bilgisayar derslerinden ve %25'i hem bilgisayar hem de istatistik derslerinden başarısız olmuştur. Rasgele seçlen bir öğrencinin

a) Bilgisayardan başarısız ise istatistikten başarısız olma olasılığını,
b) İstatistikten başarısız ise bilgisayardan başarısız olma olasılığını,
c) Bu iki dersten en az birinden başarısız olma olasılığını hesaplayınız.

İ: istatistik dersinden başarısız öğrencileri
B: Bilgisayar dersinden başarısız öğrencileri göstersin.
P(İ)=0.45, P(B)=0.35, P(A $\cap B) = 0.25$

a) $P(\text{İ}|B) = \frac{P(\text{İ}\cap B)}{P(B)} = \frac{0.25}{0.35} = \frac{5}{7}$

b) $P(B|\text{İ}) = \frac{P(\text{İ}\cap B)}{P(\text{İ})} = \frac{0.25}{0.45}\frac{5}{9}$

c) P(İ $\cup B) = P(i) + P(B) - P(\text{İ} \cap B) = 0.45 + 0.35 - 0.25 = 0.55$

Tam Bağımsız (İkiden çok olayların bağımsızlığı):

A, B ve C gibi üç olayın bağımsız olması için aşağıdaki eşitliklerin sağlanması gerekir.

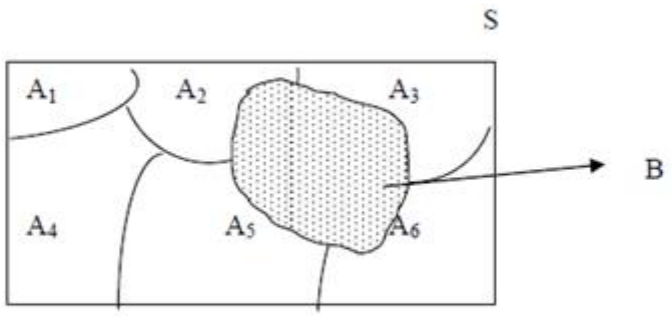
P(A∩B)=P(A)P(B)
P(A∩C)=P(A)P(C)
P(B∩C)=P(B)P(C)
P(A∩B∩C)= P(A)P(B)P(C)

N tane olayın bağımsız olabilmesi için

$P(A_1 \cap\ A_2 \ldots \cap A_k)$=P$(A_1)P(A_2)$ ... $P(A_k)$, k=2,3,4, ... , N

Toplam Olasılık Kuralı:

Bir B olayının olasılığı doğrudan hesaplanamadığı zaman toplam olasılık kuralından yararlanılır. Öğreneğin bir fabrikadaki 6 makine tarafından üretilen ürünlerden rasgele bir tanesi alındığında bu ürünün bozuk olma olasılığı araştırılsın. Burada B olayı, çekilen ürünün bozuk olması ise bu ürün A1, A2, ..., A6 makinlerinin birisinde üretilmiş olabilir.



Bir B olayı için,

$$P(B) = \sum_{i=1}^{n} P(B|Ai)P(Ai)$$

Eşitliğine toplam olasılık kuralı denir.

Burada, i=1,2, ... ,n

Ai $\cap Aj = 0, i \neq j \; için$

P(Ai)>0, i=1, ... , n

$$S = \bigcup_{i=1}^{n} A_i$$

dir.

Örnek:

Bir fabrikada üretilen malların %50'si 1.makineden, %30'u 2.makineden ve %20'si 3.makineden üretilmektedir. Bu makinlerin ürettikleri malların sırasıyla %3, %4, ve %5'in bozuk olduğu gözlenmiştir. Üretilen mallarından seçilen bir tanesinin bozuk olma olasılığı nedir?

Ai: Seçilen i.makinede üretilmiştir. (i=1,2,3)
B: Seçilen mal bozuktur.

P(A1)=0.50
P(A2)=0.30
P(A3)=0.20
P(B|A1)=0.03
P(B|A2)=0.04
P(B|A3)=0.05

Üretilen mallar bu üç makineden çıktığı içib B olayı A1, A2, A3 olaylarının birisiyle birlikte ortaya çıkar. Bu durumda ayrık üç olayın toplamı olarak yazılabilir.
$P(B)=P(B \cap A1) + P(B \cap A2) + P(B \cap A3)$
$P(B)= P(B|A1)P(A1) + P(B|A2)P(A2) + P(B|A3)P(A3)$
$P(B)= 0.03*0.50 + 0.04*0.30 + 0.05*0.20$
$P(B)=0.037$
Rasgele seçilen bir malın bozuk olması olasılığı 0.037'dır. Bir başka deyişle bu fabrikadan 1000 tane mal alınırsa, bu seçilen 1000 ürün içinde bouk olacakların beklenen değeri 37 olacaktır.

Bayes Teoremi:

Thomas Bayes tarafından geliştirilen, koşullu olasılıkların hesaplanmasında kullanılan bir teoremdir. Bir olayın ortaya çıkmasında birden fazla bağımsız nedenin etkili olması durumunda, bu nedenlerden herhangi birinin hangi bağımsız nedendenin meydana getirme olasılığını hesaplamada kolaylık sağlar.

P(Ai)>0 ve her i $\neq$ j için Ai ∩ Aj =0 olsun. S örneklem uzayında tanımlanmış herhangi bir B olayı için P(B)>0 olmak kaydıyla, B olayının gerçekleştiği varsayımı altında A olayının koşullu olasılığı,

$$P\left(A_j|B\right) = \frac{P\left(B|A_j\right)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}, \qquad 1 \leq j \leq n \; olur.$$

Örnek:

Bi şirket üç firmadan 100 makine satın almaktadır. 1.firmadan 60, 2.firmadan 30 ve 3.firmadan 10 oranında makine satın alınmaktadır. 1.firmadan gelen makinelerin %9'u, 2.firmadan gelen makinelerin 20'si ve 3.firmadan gelen makinlerin %6'sı arızalı çıkmıştır.

a) Şirkete sağlanan bir makinen arızalı çıkma olasılığı nedir?
b) Arızalı makinenen 2.firmadan gelmiş olma olasılığı nedir?

B: Bir makinenin arızalı çıkma olasılığı nedir?

Ai: Makinenin 1,2 ya da 3.firmadan gelme i=1,2,3 olayları olsun.

P(A1)=0.60, P(A2)=0.30, P(A3)=0.10

P(B|A1)=0.09, P(B|A1)=0.20, P(B|A1)=0.06

a) P(B) → makinenin arızalı olma olasılığı soruluyor.

P(B)=P(B|A1)P(A1) + P(B|A2)P(A2) + P(B|A3)P(A3)

P(B)=0.60*0.09 + 0.30*0.20 + 0.10 * 0.06

P(B)=0.12

Bu şirkete satın alınan makinelerin %12'si arızalı olacaktır.

b) Şirketin satın aldığı makine arızalı çıkmış ise bu makinenin 2.firmadan satın alma olasılığı Bayes teoriminden yararlanılarak bulunabilir.

$$P(A_2|B) = \frac{P(B|A_2)P(A_2)}{\sum_{i=1}^{3} P(B|A_i)P(A_i)} = \frac{0.3 * 0.20}{0.12} = 0.50$$

Şirket, makinelerin %30'unu 2.firmadan satın almasına karşın, arızalı makinelerin %50'si 2.firmadan gelmektedir.

Örnek:

A ülkesinde korona testi yapılan insanların %10 ununda sonuç pozitif çıkmaktadır ve sonucu pozitif çıkanların %6 sı ölmektedir. B ülkesinde korona testi yapılan insanların %12 sinde sonuç pozitif çıkmaktadır ve sonucu pozitif çıkanların %5 i ölmektedir. Ölen birisinin A ülkesinden olma olasılığı nedir?

$P(A/Ö)=P(A\cap Ö)/(P(A\cap Ö)+P(B\cap Ö))$

$P(A\cap Ö)=P(A)*P(Ö/A)$,

$P(B\cap Ö)=P(B)*P(Ö/B)$

$P(A)=0.10$

$P(Ö/A)=0.6$

$P(A\cap Ö)=0.06$

$P(B)=0.12$

$P(Ö/B)=0.5$

$P(B\cap Ö)=0.06$

$P(A/Ö)=(0.06)/(0.06+0.06)=0.06/0.12=6/12=1/2=0.5$

Örnek:

Bir fabrikada 2 adet makine bulunmaktadır. Herbir makinenin günlük üretim oranları A=%50, B=%50 olarak verillmiştir. Üretilenlerden kusurlu olanların yüzdesi ise A=%10, B=%5 olarak verilmiştir. Kusurlu parçalar birbirinden bağımsız olarak A, B makinelerinde üretilmektedir.

$$P(A\,/H) = \frac{P(A \cap H)}{P(A \cap H) + P(B \cap H)}$$

$$P(A \cap H) = P(A) * P(H\,/A)$$
$$P(B \cap H) = P(B) * P(H\,/B)$$

a) Günün sonunda rasgele bir ürün seçiliyor ve kusurlu olma olasılığı nedir?

Herbir üretim bandında üretim olasıkları hesaplayınız. P(A), P(B)
P(A)= 0.5
P(B)= 0.5

Herbir üretim bandı için A dan ve B den kaynaklanan hata olasılıklarını hesaplayınız. P(H/A), P(H/B)
P(H/A)=0.10
P(H/B)=0.05

Bu kusurlu motorun A, B makinesinden çıkma olasıklarını hesaplayınız. Kusurlu olma olasılığı P(T) ise
P(A∩H)=P(A)*P(H/A)=0.05
P(B∩H)=P(B)*P(H/B)=0.025
P(T)=0.075

P(A/H), toplam hatalı olanlardan A daki hatalı olanların olasılığı
P(A/H)= P(A∩H)/P(Toplam)=0.05/0.075=50/75=2/3

P(B/H), toplam hatalı olanlardan B daki hatalı olanların olasılığı
P(B/H)= P(B∩H)/P(Toplam)=0.025/0.075=25/75=1/3

Örnek:
Bir ülkede, vatandaşların %52'si kadın. Ayrıca kadınların %20'sinin ve erkeklerin %15'inin bir işi olmadığı da biliniyor. Öyleyse, bu ülkedeki aktif nüfustan rastgele seçilen bir kişinin işsiz P(U) olma olasılığı nedir?

Koşullu olasılık:
$P(U/K) = P(K \cap U) / P(K)$
$P(K \cap U)= P(U/K) \times P(K)$

Kişinin bir kadın olma olasılığı: P (K)
Kişinin bir erkek olma olasılığı: P (E)
Vatandaşların %52'sinin kadın olduğunu bildiğimiz için P (K) = 0,52
Vatandaşların %48'inin erkek olduğunu bildiğimiz için P (E) = 1-0.52=0,48

Koşullu olasılıkları gereği:

- Bir kadının işsiz olma olasılığı: **P (U | K) = 0,20**
- Bir erkeğin işsiz olma olasılığı: **P (U | E) = 0,15**

Toplam olasılık yasasını kullanarak şunları elde ederiz:
$P(U) = P(K)\,P(U \mid K) + P(E)\,P(U \mid E)$
$P(U) = 0{,}20 \times 0{,}52 + 0{,}15 \times 0{,}48$
$P(U) = 0{,}176$

Aktif nüfusun %52'sinin kadın olduğu biliniyor. Ayrıca, işsiz kadınların yüzdesinin %20 ve işsizlerin yüzdesinin %15 olduğu biliniyor.

Rastgele, işsiz bir insan seçme olasılığımızın 0,176 olduğu biliniyor. Öyleyse, Bayes teoremini uygularsak, elde edeceğimiz sonuç, işsiz insanların arasından rastgele seçilen bir kişinin kadın olma ihtimalini hesaplayın.

$P(K \mid U) = (P(K) * P(U \mid K) / P(U)) = (0{,}20 * 0{,}52) / 0{,}176 = 0{,}59$

Örnek:

Bir sistemde hata olduğu zaman sisteminin durma olasığı 0.90, hata olmadığı zaman sistemin çalışma olasılığı 0.80 ve herhangi bir anda hata olma olasılığı 0.01 dir.

a) Bir sistem durduğuna göre hata nedeniyle sistemin durma olaslığı nedir?
b) Hata olması ve sistemin çalışma olasılığı nedir?
a) A: Sistemin durma olasılığı
   B: Hata olması
   P(A\B)= hata olduğunda sistemin durma olasılığı=0.90
   P(A’\B’)= hata olmadığında sistemin çalışma olasılığı=0.80
   P(A\B’)=0.20
   P(B)=0.01
   P(B’)=0.99
   $P(B\backslash A) = \frac{P(A\backslash B)P(B)}{P(A\backslash B)P(B)+P(A\backslash B')P(B')} = \frac{(0.90)(0.01)}{(0.90)(0.01)+(0.2)(0.99)}$=0.04345
   Sistem durduğuna göre, hata nedeniyle olma olasılığı yaklaşık %4.3 dür.

b) Çarpım kuralı uygulanırsa
   P(A’\B)=1-P(A\B)=1-0.90=0.1=P(A’)
   P(A’∩B)=P(A’\B) P(B)=(0.1)(0.01)=0.0001

Örnek:

Bir araştırmaya göre her 100 yaşlıdan 1 tanesi, hastalanmakta ve yapılan test sonuçlarına göre, hastalıklı bir yaşlının testi %80 korona pozitif, sağlıklı bir yaşlının testi ise %10 korona pozitif sonuç vermektedir. Bu bilgilere göre test sonucu korona pozitif olan bir yaşlının gerçekten hasta olma olasılığı nedir?

A olayı yaşlının hastalanmasını, B olayı da yaşlığa yapılan testin korona pozitif çıkmasını göstersin.

P(A | B) = P(A)·P(B | A) / [P(A)·P(B | A) + P(A')·P(B | A')]

P(A | B) = (1 / 100)·0,8 / [(1 / 100)·0,8 + (99 / 100)·0,1] = 0.008/(0.008+0.099)=1/13.75

Örnek:

Büyük bir ofisde 3 adet fotokopi makinası kullanılmaktadır. Çekimlerin %60'ı 1.makinada, %30'u 2.makinada ve %10'u da 3.makinada yapılmaktadır. Makinaların fire oranları da 1.makine için %10, 2.makine için %20 ve 3. makine için %40 olmaktadır. Tüm kopyalar için fire oranı nedir ?

M1, M2, M3 makinaları, D hatalı kopyayı(fire) göstersin.

Makinaların kullanılma olasılıkları:

$P(M1) = 0.60$

$P(M2) = 0.30$

$P(M3) = 0.10$

Fire oranları:

$P(D\backslash M1) = 0.10$

$P(D\backslash M2) = 0.20$

$P(D\backslash M3) = 0.40$

M1 makinesinde çekildikten sonra fire olma olasılığı: $P(D\backslash M1)= P(M1 \cap D) / P(M1)$

M2 makinesinde çekildikten sonra fire olma olasılığı: $P(D\backslash M2)= P(M2 \cap D) / P(M2)$

M3 makinesinde çekildikten sonra fire olma olasılığı: $P(D\backslash M3)= P(M3 \cap D) / P(M3)$

Olayları A , B , C ile gösterirsek;

$A = M1 \cap D$, $B = M2 \cap D$, $C = M3 \cap D$ yazılabilir.

$P(A) = P(M1 \cap D) = P(M1)*P(D\backslash M1)= (0.60)*(0.10) = 0.06$

$P(B) = P(M2 \cap D) = P(M2)*P(D\backslash M2)= (0.30)*(0.20) = 0.06$,

$P(C) = P(M3 \cap D) = P(M3)*P(D\backslash M3)= (0.10)*(0.40) = 0.04$

Her makine için hesaplanan fire olasılıkların toplamı bize toplam fire olasılığını verecektir:

$P(D) = P(A) + P(B) + P(C) = 0.06 + 0.06 + 0.04 = 0.16$

Tüm kopyaların %16'ı fireli çıkmaktadır diyebiliriz.

Örnek:

Bir fabrikada 4 adet üretim bandı bulunmaktadır. Herbir üretim bandının günlük üretim oranları aşağıdaki tabloda verillmiştir.

	A	B	C	D
Üretim Oranı	%20	%40	%30	%10

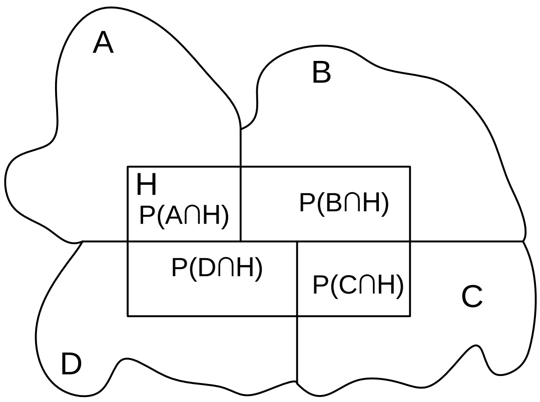
Üretilenlerden kusurlu olanların yüzdesi ise aşağıdaki tabloda verilmiştir.

	A	B	C	D
Kusurlu Yüzdeleri	%2	%1	%5	%4

Kusurlu parçalar birbirinden bağımsız olarak A, B, C, D üretim bandlarında üretilmektedir.

Küme tanımı yapıldığında; A, B, C, D ve K kümeleri mevcuttur. K kümesinde A, B, C, D üretim bandında üretilenlerden kusurlu ürünlerin oranları mevcuttur. Herbir üretim bandı için kusurlu ürün miktarı belirlidir ve ayrıktır (iadesiz).

Kusurlu ürün miktarı, ilgili üretim bandında üretilen ürün miktarı ile kusurlu yüzdesi çarpılarak bulunur. Bu nedenle herbir üretim band için hatalı olasılığı aşağıdaki çarpım ile hesaplanır.



$$P(A/H) = \frac{P(A \cap H)}{P(A \cap H) + P(B \cap H) + P(C \cap H) + P(D \cap H)}$$

$$P(A \cap H) = P(A) * P(H/A)$$
$$P(B \cap H) = P(B) * P(H/B)$$
$$P(C \cap H) = P(C) * P(H/C)$$
$$P(D \cap H) = P(D) * P(H/D)$$

Günün sonunda rasgele bir motor seçiliyor ve kusurlu olduğu görülüyor.

b) Herbir üretim bandında üretim olasıkları hesaplayınız. P(A), P(B), P(C), P(D)

P(A)= 0.2

P(B)= 0.4

P(C)= 0.3

P(D)= 0.1

c) Herbir üretim bandı için hata olasılıklarını hesaplayınız. P(H/A), P(H/B), P(H/C),

P(H/A)= 0.02

P(H/B)= 0.01

P(H/C)= 0.05

P(H/D)= 0.04

d) Bu kusurlu motorun A, B, C, D üretim bandında çıkma olasıklarını hesaplayınız. Yorumlayınız.

P(A∩H)=P(A)*P(H/A)=0.004

P(B∩H)=P(B)*P(H/B)=0.004

P(C∩H)=P(C)*P(H/C)=0.015

P(D∩H)=P(D)*P(H/D)=0.004

P(Toplam)=0.0270

P(A/H), toplam hatalı olanlardan A daki hatalı olanların olasılığı

P(A/H)= P(A∩H)/P(Toplam)=0.148

P(B/H), toplam hatalı olanlardan B daki hatalı olanların olasılığı

P(B/H)= P(B∩H)/P(Toplam)=0.148

P(C/H), toplam hatalı olanlardan C daki hatalı olanların olasılığı

P(C/H)= P(C∩H)/P(Toplam)=3.75

P(D/H), toplam hatalı olanlardan D daki hatalı olanların olasılığı

P(D/H)= P(D∩H)/P(Toplam)=0.148

e) Söz gelimi günün sonunda 1000 ürün üretilmiş olsun

- Herbir üretim bandı için yüzde oranlarını kullanarak üretim miktarlarını bulunuz.
- Herbir üretim bandı için verilen kusurlu oranlarını kullanarak her bir üretim bandı için kusurlu adetlerini bulunuz.

N_A=200, N_B=400, N_C=300, N_D=100 olur. Hatalı ürün miktarları ise;

$N_{AH}= N_A*2 /100=4$

$N_{BH}= N_B*1 /100=4$

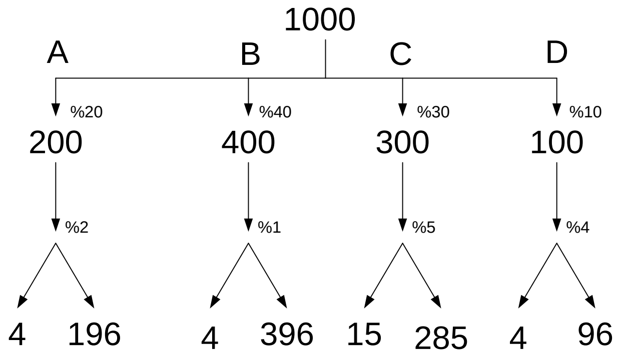
$N_{CH}= N_C*5 /100=15$

$N_{DH}= N_D*4 /100=4$

Rasgele seçilen bir motor kusurlu olduğu görülüyor. Kusurlu motorun A üretim bandında üretilme olasılığı nedir?

$P(H/A)= N_{AH} /(N_{AH} + N_{BH} + N_{CH} + N_{DH})=4/27=0.148$

f) Dallanma metotu ile çözüm



$P(H/A)= N_{AH} /(N_{AH} + N_{BH} + N_{CH} + N_{DH})=4/27=0.148$

$P(H/B)= N_{BH} /(N_{AH} + N_{BH} + N_{CH} + N_{DH})=4/27=0.148$

$P(H/C)= N_{CH} /(N_{AH} + N_{BH} + N_{CH} + N_{DH})=15/27=0.555$

$P(H/D)= N_{DH} /(N_{AH} + N_{BH} + N_{CH} + N_{DH})=4/27=0.148$

Example:

Bayes' theorem, fraction of women among students

From the proportion of students and women in the population and the fraction of students among women we compute the fraction of women among students:

$$
\begin{array}{lll}
P\{A\} & = 0.02 & \text{(fraction of students in the population)} \\
P\{B\} & = 0.5 & \text{(fraction of women in the population)} \\
P\{A \mid B\} & = 0.018 & \text{(fraction of students among women)} \\
P\{B \mid A\} & =? & \text{(fraction of women among students)}
\end{array}
$$

The dependence of the events A and B manifests itself in the difference of $P\{A\}$ and $P\{A \mid B\}$. Applying Bayes' theorem we obtain

$$
\begin{aligned}
P\{B \mid A\} &= \frac{P\{A \mid B\} P\{B\}}{P\{A\}} \\
&= \frac{0.018 \cdot 0.5}{0.02} = 0.45 \,.
\end{aligned}
$$

About 45% of the students are women.

Bayes' Theorem: *"Bayes' theorem relates the conditional and marginal probabilities of events A and B, where B has a non-zero probability"*

$$\text{posterior} \equiv P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}}$$

Example *"The department is formed by 60% men and 40% women. Men always wear trousers, women wear trousers or skirts in equal numbers".*

- A: I see a girl
- B: A person is wearing trousers
- The probability of meeting a girl with trousers is:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.5 \times 0.4}{0.5 \times 0.4 + 1 \times 0.6} = 0.25$$

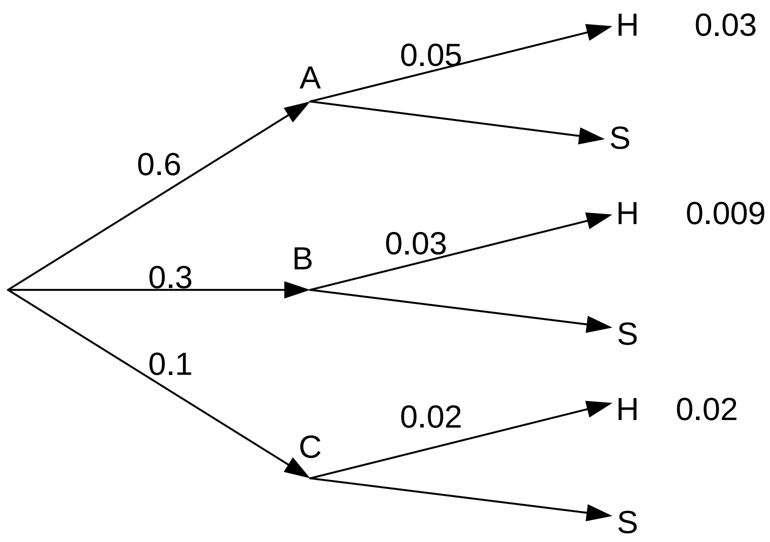
- Simple non-Bayesian probabilities would say: $0.4 \times 0.5 = 0.2$

Örnek:

Bir fabrikada ürünlerin % 60, %30, %10 A, B, C makinelerinde üretilmektedir. Bu makineler %5, %3, %2 oranında bozuk ürün vermektedir. Rasgele bir makiden seçilen ürünün bozuk olma olasılığı nedir.

P(x)=P(A) P(x\A)+ P(B) P(x\B)+ P(C) P(x\C)

P(x)=(0.6)(0.05)+(0.3)(0.03)+(0.1)(0.02)=0.059



Rasgele seçilen bir ürünün A makinesinde üretilme olasılığı nedir?

P(A \x)= P(A) P(x\A) / { P(A) P(x\A)+ P(B) P(x\B)+ P(C) P(x\C) }

P(A \x)= 0.030/0.059=30/59

Rasgele seçilen bir ürünün B makinesinde üretilme olasılığı nedir?

P(B \x)= P(B) P(x\B) / { P(A) P(x\A)+ P(B) P(x\B)+ P(C) P(x\C) }

P(B \x)= 0.009/0.059=9/59

Rasgele seçilen bir ürünün C makinesinde üretilme olasılığı nedir?

P(C \x)= P(C) P(x\C) / { P(A) P(x\A)+ P(B) P(x\B)+ P(C) P(x\C) }

P(C \x)= 0.020/0.059=20/59

Örnek:

A kutusunda üç beyaz ve iki siyah, C kutusunda dört beyaz ve dört siyah bilye bulunmaktadır. A kutusundan ardı ardına iki bilye alınarak, C kutusuna konmaktadır. Daha sonra, C kutusundan bir bilye alınmaktadır. C kutusundan alınan bu bilyenin beyaz olması olasılığını bulunuz.

Çözüm : *Olaylar:*

$$\begin{aligned} B_1 &= \{A \text{ kutusundan Beyaz top çekme}\} \\ S_1 &= \{A \text{ kutusundan Siyah top çekme}\} \\ B_2 &= \{C \text{ kutusundan Beyaz top çekme}\} \end{aligned}$$

olarak tanımlansın. Buna göre, A kutusundan ardı ardına alınan iki bilye için ilgili örneklem uzayı

$$S = \{S_1B_1, B_1S_1, B_1B_1, S_1S_1\}$$

olarak elde edilir. A kutusu için, tanımlanan örneklem uzayı noktalarının olasılıkları:

$$P(S_1B_1) = \frac{2}{5} * \frac{3}{4} = \frac{6}{20}, \qquad P(B_1S_1) = \frac{3}{5} * \frac{2}{4} = \frac{6}{20}$$
$$P(B_1B_1) = \frac{3}{5} * \frac{2}{4} = \frac{6}{20}, \qquad P(S_1S_1) = \frac{2}{5} * \frac{1}{4} = \frac{2}{20}$$

olur. Aranan olasılık toplam olasılık formülü yardımıyla,

$$\begin{aligned} P(B_2) &= P(S_1B_1)P(B_2/S_1B_1) + P(B_1S_1)P(B_2/B_1S_1) + P(B_1B_1)P(B_2/B_1B_1) + P(S_1S_1)P(B_2/S_1S_1) \\ &= \frac{6}{20} * \frac{5}{10} + \frac{6}{20} * \frac{5}{10} + \frac{6}{20} * \frac{6}{10} + \frac{2}{20} * \frac{4}{10} \;=\; \frac{13}{25} \end{aligned}$$

Örnek:

Bir okulda 1000 öğrencinin 300'ü kız, 700'ü erkektir. Bir spor gösterisi için, kızların 100'ünün, erkeklerin 300'ünün bileti vardır. Bir bilet bulunuyor. Bu biletin kız öğrenciye ait olması olasılığını bulunuz.

Çözüm : *Olaylar:*

$$\begin{aligned} K &= \{\text{Öğrencinin kız olması}\} \\ E &= \{\text{Öğrencinin erkek olması}\} \\ B &= \{\text{Öğrencinin bileti olması}\} \end{aligned}$$

olarak tanımlansın. Buna göre, soruda verilen olasılıklar:

$$P(K) = \frac{300}{1000} \qquad P(E) = \frac{700}{1000} \qquad P(B/K) = \frac{100}{300} \qquad P(B/E) = \frac{300}{700}$$

olur. Aranan olasılık Bayes teoremi yardımıyla,

$$P(K/B) = \frac{P(K)P(B/K)}{P(K)P(B/K) + P(E)P(B/E)} = \frac{\frac{300}{1000} * \frac{100}{300}}{\frac{300}{1000} * \frac{100}{300} + \frac{700}{1000} * \frac{300}{700}} = \frac{1}{4}$$

olarak elde edilir.

Örnek:

A kutusunda, 3 sağlam ve 4 bozuk transistör vardır. B kutusunda ise, 4 sağlam ve 3 bozuk transistör vardır. A kutusundan bir transistör alınıp, B kutusuna bırakılıyor. Daha sonra, B kutusundan bir transistör alınıyor ve sağlam olduğu görülüyor. A'dan B'ye geçirilen transistörün sağlam olması olasılığını bulunuz.

Çözüm : *Olaylar:*

$$\begin{aligned} A_1 &= \{A \text{ kutusundan sağlam transistör seçme}\} \\ A_2 &= \{A \text{ kutusundan bozuk transistör seçme}\} \\ B_1 &= \{B \text{ kutusundan sağlam transistör seçme}\} \end{aligned}$$

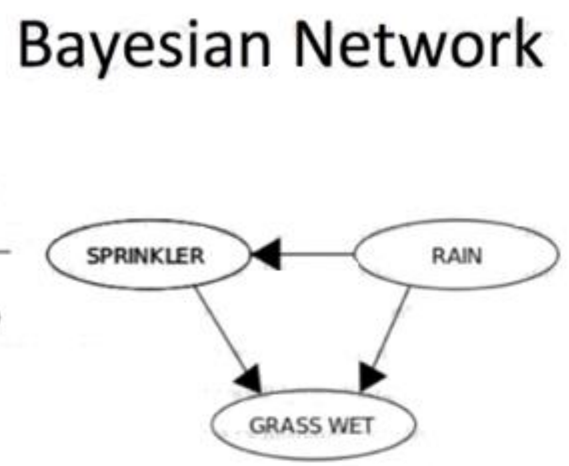
olarak tanımlansın. Buna göre, soruda verilen olasılıklar:

$$P(A_1) = \frac{3}{7} \qquad P(A_2) = \frac{4}{7} \qquad P(B_1/A_1) = \frac{5}{8} \qquad P(B_1/A_2) = \frac{4}{8}$$

olur. Bu durumda,

$$P(A_1/B_1) = \frac{P(A_1)P(B_1/A_1)}{P(A_1)P(B_1/A_1) + P(A_2)P(B_1/A_2)} = \frac{\frac{3}{7} * \frac{5}{8}}{\frac{3}{7} * \frac{5}{8} + \frac{4}{7} * \frac{4}{8}} = \frac{15}{31}$$

Örnek:



	SPRINKLER	
RAIN	T	F
F	0.4	0.6
T	0.01	0.99

RAIN	
T	F
0.2	0.8

		GRASS WET	
SPRINKLER	RAIN	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

5. Rassal Değişkenler ve Olasılık Yoğunluk Fonksiyonları

Hangi değeri alacağı önceden bilinmeyen ve belli olasılıklarla çeşitli değerler alabilen değişkenlere rastlantı (rastsal, rasgele) değişkenler adı verilir. Rassal İngilizce'de random kelimesinin karşılığıdır. Rastgele ile eş anlamlıdır.

Rassal denemelerin sonuçları sayılarla ifade edilebilir. Örneğin atılan bir zar, öğrencilerin notları, kişilerin gelirleri vb. Sayılarla ifade edilemeyen rassal deneme sonuçlarını bile sayılarla ifade edilebilir ve böylece olasılık fonksiyonları belirlenebilir. Örneğin bir malın kusurlu olup-olmaması, bir müsabakanın galibi-mağlubu, kişinin medeni hali ve eğitim durumu vb. Bir üretim sürecinde Arızalı mallar = 1 ve Arızasız mallar = 0 sayısal değerlerini verebiliriz.

Rassal değişken X ve rassal değişkenin alabileceği değerler ise x ile gösterilir. Stokastik değişken olarak da adlandırılan rassal değişkenler ikiye ayrılır: Ayrık (kesik) rassal değişkenler, Sürekli rassal değişkenler. Bir x rasgele değişkeninin alacağı değerlerinin sayısı sonlu veya sayılabilir ise X'e "**Kesikli Rasgele Değişken"** denir. Rassal denemenin sonuçları olan bir x rasgele değişkeni, belli bir aralıkta bütün değerleri alabiliyorsa yani sayı çizgisinde bir aralığı kaplıyorsa ise X'e "**Sürekli Rassal Değişkenler**" olarak adlandırılır. Bir X rasgele değişkeninin olanaklı değerleri bir aralıktan oluşur. ***Genellikle ölçümler sürekli rassal değişkenlerdir.***

Kesikli AnakütleDağılımları: **Kesikli düzgün dağılım:** Bernoulli dağılımı Binom dağılımı Poisson dağılımı Hipergeometrik dağılım Negatif binom dağılımı	**Sürekli Anakütle Dağılımları:** **Sürekli düzgün dağılım:** Normal dağılım Üstel dağılım Lognormaldağılım Gamma dağılımı Ki-kare dağılımı

Olasılık değerlerin hesabında, kesikli olasılık fonksiyonlarının kullanılması halinde toplam işareti (Σ) kullanılırken, olasılık yoğunluk fonksiyonun (sürekli) kullanılması halinde de integral işareti (∫) kullanılır.

Olasılık yoğunluk fonksiyonu:
bir rastlantı değişkenin alabileceği değerlerle, bu değerleri alabilmesi olasılıkları arasındaki ilişkiyi gösteren bir fonksiyondur.

$f(x) \geq 0$ ve $\int_{-\infty}^{+\infty} f(x)dx = 1$

koşulunu sağlayan f(x) fonksiyonuna x'in olasılık yoğunluk fonksiyonu denir.

X rassal değişkeninin R'deki değer kümesi A, sayılmaz küme ise x'e sürekli rassal değişken denir.

$$A = \{x | a \leq X \leq b\}$$

Rassal değişkenlerin bir fonksiyonu olan bu matematiksel modeller olasılık ya da olasılık yoğunluk fonksiyonları olarak adlandırılmaktadır.

X: sürekli rassal değişken
A: sürekli rassal değişkenin değer kümesi
X ∉ A için f(x) = 0 olmak üzere;
A_i = (a,b] ⊂ A için, her A_i ⊂ A için;
P{x∈ A_i } = $\int_a^b f(x)\, d_x$ özeliğini gerçekleyen f(x) ' e **x'in olasılık yoğunluk fonksiyonu** denir.

OLASILIK YOĞUNLUK FONKSİYONU ÖZELLİKLERİ:

1) Verilen fonksiyon matematiksel olarak X ∈ R için bir değer alabilir. Ancak tanım gereği X∉ A için f(x) = 0 olarak ele alınır.

2) ∀ x ∈ A için f(x) ≥ 0 olarak ele alınır.
3) P(x ∈ A)=$\int_a^b f(x)dx$= 1 olmalıdır.
4) ∀ (a,b] ⊂ A için ∀ (a< x ≤ b) =$\int_a^b f(x)\, d_x$ olmak üzere, A üzerinde f(x) süreklidir.

Beklenen değer (ağırlık Ortalama):

X bir rasgele değişken olmak üzere

a) Sürekli rasgele deşikeni için

μ = $E(X) = \int_{-\infty}^{+\infty} x f(x) dx$ değerine $\int_{-\infty}^{+\infty} |x| f(x) dx < \infty$ olduğunda X'in beklenen değeri denir. Kararlı olduğunun göstergesidir.

b) Kesikli rasgele değişken için,

μ = $E(\mathrm{X}) = \sum_{i=1}^{n} x_i f(x_i)$ değerine $\sum_{i=1}^{n} |x_i|\ f(x_i) < \infty$ olduğunda X'in beklenen değeri denir. Kararlı olduğunun göstergesidir.

Moment:

$E(X-\mu)^k$ değerine X'in a noktasına göre k. Momenti denir.

Varyans:

$E(X-\mu)^2$ değerine X'in varyansı denir, Var(X) ile gösterilir. Var(X)≥ 0 olmak üzere,

$$Var(\mathrm{X}) = E(x - E(\mathrm{X}))^2$$

dir. Beklenen değer, dağılım merkezi hakkında, varyans merkez etrafında yayılım hakkında bilgi verir.

X bir rassal (rasgele) değişken a ve b birer rasgele değişken olmak üzere beklenen değer ve varyans için aşağıdaki özellikler yazılabilir:

$$E(aX \pm b) = aE(X) \pm b$$

$$Var(aX \pm b) = a^2 Var(X)$$

Ağırlıklı ortalaması ya da beklenen değeri:

Kesikli rasgele değişken için,

$$\mu = E(\mathrm{X}) = \sum_{i=1}^{n} x_i f(x_i)$$

$$E(X^2) = \sum_{i=1}^{n} x_i^2 f(x_i)$$

Sürekli rasgele deşikeni için,

$\mu = E(X) = \int_{-\infty}^{+\infty} xf(x)dx$

$$E(x^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx$$

Not: E(X)>0 ise beklenti uygun, E(X)=0 ise beklenti ortada, E(X)<0 ise beklenti uygun değil.

$E(aX)=a\,E(X)$

$E(X+a)= E(X)+a$

$E(X+Y)= E(X) + E(Y)$

$Var(X) = E(X^2) - E^2(X) = E(X^2) - \mu^2$

$\sigma(X) = \sqrt{\mathrm{Var(X)}}$

Örnek $Y = 3X - 5,\ E(X) = 4,\ Var(X) = 2$ olmak üzere Y rasgele değişkeninin beklenen değer ve varyansını bulunuz.

$$E(Y) = E(3X - 5) = 3E(X) - 5 = 3.4 - 5 = 12 - 5 = 7$$

$$Var(Y) = Var(3X - 5) = 1$$

Örnek $Y = X^2 + 3X$ ve $E(X) = 10, Var(X) = 6$ olmak üzere $E(Y) = ?$

$$E(Y) = E(X^2 + 3X)$$

$$Var(X) = E(X^2) - (E(X))^2$$

$$6 = E(X^2) - 100$$

$$E(X^2) = 106$$

$$E(X^2 + 3X) = E(X^2) + 3E(X) = 106 + 3 \times 10 = 136$$

Örnek:

f(x) fonksiyonun standart sapmasını hesaplayınız.

$$f(x) = \begin{cases} 2(1-x) & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$

f(*x*) fonksiyonun olasık fonksiyonu olup olmadığını belirleyiniz. Not: f(x) fonksiyonun olasılık fonksiyonu olabilmesi için integralinin 1'e eşit olması gerekir.

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$\int_{-\infty}^{\infty} f(x)dx = 2\int_0^1 (1-x)dx = 2\left(x - \frac{x^2}{2}\right)\Big|_0^1 = 1$$

$f(x)$ fonksiyonun aritmetik ortalamasını hesap ediniz. Not: Rasgele fonksiyonların aritmetik ortalaması aşağıdaki ifade ile bulunur,

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

$$\begin{aligned}
\mathbb{E}(X) &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_{0}^{1} x \Big[2(1-x) \Big] dx \\
&= 2 \int_{0}^{1} \Big(x - x^2 \Big) dx \\
&= 2 \Big(\frac{x^2}{2} - \frac{x^3}{3} \Big) \Big|_{0}^{1} \\
&= 1/3
\end{aligned}$$

f(x) fonksiyonun varyansını ve standart sapmasını aşağıdaki ifadeleri kullanarak hesaplayınız

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} \big(x - \mathbb{E}(X) \big)^2 f(x) dx$$

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

$$\begin{aligned}
\mathrm{Var}(X) &= \int_{-\infty}^{\infty} \big(x - \mathbb{E}(X)\big)^2 f(x)dx \\
&= \int_0^1 \big(x - 1/3\big)^2 \cdot 2(1-x)dx \\
&= 2\int_0^1 (x^2 - \frac{2}{3}x + \frac{1}{9})(1-x)dx \\
&= 2\int_0^1 \big(-x^3 + \frac{5}{3}x^2 - \frac{7}{9}x + \frac{1}{9}\big)dx \\
&= 2\big(-\frac{1}{4}x^4 + \frac{5}{9}x^3 - \frac{7}{18}x^2 + \frac{1}{9}x\big)\Big|_0^1 \\
&= 2\big(-\frac{1}{4} + \frac{5}{9} - \frac{7}{18} + \frac{1}{9}\big) \\
&= \frac{1}{18}
\end{aligned}$$

Therefore, the standard deviation of *X* is

$$\begin{aligned}
\sigma &= \sqrt{\mathrm{Var}(X)} \\
&= \frac{1}{3\sqrt{2}}
\end{aligned}$$

f(x) fonksiyonun varyansını ve standart sapmasını aşağıdaki ifadeleri kullanarak hesaplayınız

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$$

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\
&= \int_{-\infty}^{\infty} x^2 f(x)dx - \left(\frac{1}{3}\right)^2 \\
&= 2\int_0^1 x^2(1-x)dx - \frac{1}{9} \\
&= 2\int_0^1 (x^2 - x^3)dx - \frac{1}{9} \\
&= 2\big(\frac{1}{3}x^3 - \frac{1}{4}x^4\big)\big|_0^1 - \frac{1}{9} \\
&= 2\big(\frac{1}{3} - \frac{1}{4}\big) - \frac{1}{9} \\
&= \frac{1}{18}
\end{aligned}$$

5.1. Kesikli Rassal Değişkenlerin Olasılık Fonksiyonu

Eğer örnek uzay belirli sayıda elemandan ya da sayılabilir sayıda elemanın sonsuz serisinden oluşuyorsa (mesela tam sayılar), buna kesikli örnek uzay denir. Kesikli örnek uzaydaki elemanları bir sayıyla eşleştiren fonksiyonlara kesikli rastgele değiken denir. Kesikli rastgele değişkenler için olasılık fonksiyonu tanımlıdır:

$$f(x)=P(X=x)$$

$$f(x) = P(\boldsymbol{X} \leq \boldsymbol{x}) = \sum_{\boldsymbol{x_i} \leq \boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x_i})$$

Kesikli bir rassal değişken X ve bu rassal değişkenin alabileceği değerler ise xi ile gösterilsin. Bu durumda kesikli bir X rassal değişkeninin belli bir x değerinin alma olasılığı x'in bir fonksiyonu olarak aşağıdaki şekilde ifade edilebilir.

Px (x) = P(X=x)

Yukarıdaki ifade kesikli X rassal değişkeninin olasılık dağılımını vermektedir. Belli olayla ilgili olasılık dağılımı bir kez hesaplandığında olasılık fonksiyonu da çıkarılmış olur.

Kesikli rassal değişkenin olasılık fonksiyonunda, x: kesikli rassal değişken, P(x): x rassal değişkenin olasılık fonksiyonu ise

Koşul-1: x_i örnek uzayında bir olaya karşı geldiğine göre olasılık değerleri 0 ile 1 arasındadır.

$0 \leq P(x_i) \leq 1 \ \forall X_i \in A$ dır.

Koşul-2: x rassal değişken X(S)=A dönüşümünü gerçekleştirip P(S)=1 olduğundan ayrık olayların toplam özelliğine göre tüm olayların olasılıkları toplamı 1'e eşittir.

$p(x) = \left\{ \left(\frac{a}{b}\right)^{x} \left(\frac{c}{d}\right)^{1-x} \right.$ bir olasılık fonksiyonu olabilmesi için

- $0 \leq a/b \leq 1$, $0 \leq c/d \leq 1$
- $a/b+c/d=1$ olmak zorundadır.

Örnek:

$p(x) = \left\{\left(\frac{3}{5}\right)^{x}\left(\frac{2}{5}\right)^{1-x}\right.$, x=0,1 değerleri almakta ve kesikli rassal değişkendir. Bu durumda *p(x) olasılık fonksiyonu mudur.*

Koşul-1:

P(X=0)=2/5, 0≤ P(X=0)≤ 1

P(X=1)=3/5, 0≤ P(X=1)≤ 1

Koşul-2:

2/5 + 3/5 =1

1. ve 2. koşullar sağlandığından dolayı p(x) bir olasılık fonksiyonudur.

Örnek:

$p(x) = \left(\frac{a}{c}\right)^{x}\left(\frac{b}{c}\right)^{1-x}$, x=0,1 değerleri almakta ve kesikli rassal değişkendir. Bu durumda *p(x) olasılık fonksiyonu ise a+b neye eşit olur?*

Koşul-1:

P(X=0)=b/c, 0≤ P(X=0)≤ 1

P(X=1)=a/c, 0≤ P(X=1)≤ 1

Koşul-2:

a/c + b/c =1 ise a+b=c olur.

Örnek:

Örneğin bir zar atıldığında gelen sonuç rassal değişken (X) iken, gelebilecek sonuçlar x_i =1, 2, 3, 4, 5 ve 6'dır ve her birinin gelme olasılığı 1/6'dır. Bu durumda bu olayın olasılık fonksiyonu aşağıdaki gibi olacaktır.

P_x (x) = P(X=x) = 1/6 > 0

$\sum_{x=1}^{n} P(X = x) = 1/6$ + 1/6 +1/6 + 1/6 +1/6 +1/6 = 1

Kesikli rassal değişkenlerde rassal örnekleme:

- N tane nesne arasından n tanelik bir örneklem seçilmesinin istendiğini düşünelim.
- n nesneli olanaklı her örneklemin seçilme şansını eşit kılan seçim sürecine rassal örnekleme (random sampling) denir.
- Amaç: Örneklem bilgisine dayanarak anakütleye ilişkin çıkarsamalar yapmak
- Bu çıkarsamalar anakütleden çekilen örneklem bilgisinin belli bir fonksiyonu olan bir istatistiğe dayanır.
- Bu istatistiğin örnekleme dağılımı, bu anakütleden çekilebilecek aynı büyüklükteki bütün örneklemlerde sözkonusu istatistiğin alabileceği değerlerin olasılık dağılımıdır.

$X(S)=(x_1, x_2, \ldots, x_n)$ sonlu raslantı değişkenin X(s) olasılık uzayında, x_i'nin olasılığı $f(x_i)$ biçiminde yazılır ve $P(X= x_i)$ olarak tanımlanan fonksiyona X'in dağılım ya da olasılık fonksiyonu denir. Burada f dağılımı, $f(x_i) \geq 0$ ve $\sum_{i=1}^{n} f(x_i) = 1$ koşullarını sağlar.

x_1	x_2	...	x_n
$f(x_1)$	$f(x_2)$	...	$f(x_n)$

Örnek:
Yığın: {6,9,12,15,18} olsun. f(x)=1/5 farzedelim. Olasılık fonksiyonu aşağıdaki şeklilde yazılabilir.

x	6	9	12	15	18
f(x) = P(X=x)	1/5	1/5	1/5	1/5	1/5

Yığının ortalamasını, varyansını ve medyanını bulalım.

$$E(\mathrm{X}) = \mu = \sum_{i=1}^{n} x_i f(x_i)$$

$$E(X^2) = \sum_{i=1}^{n} x_i^2 f(x_i)$$

E(x)=μ=6*1/5 + 9*1/5 + 12* 1/5 + 15* 1/5 + 18*1/5=60/5=12
$E(x^2)$=36/5 + 81/5 + 144/5 + 225/5 + 324/5 =162

Var(X) = σ^2 = $E(X^2)$- μ^2 =162 – 144 =18

Med(X)=12 (Ortadaki değer)

5 nesneli bu anakütleden n=3 nesneli örneklemler çekmek istendildiğinde, olanıklı tüm örneklemlerin toplam sayısı,

$$C_3^5 = \frac{5!}{3!\,(5-3)!} = 10$$

Örneklem ortalaması, $\bar{X}$ ve medyan ve örneklem varyansının s^2 örneklem dağılını bulalım. Örneğimizdeki populasyonda sadece 5 nesne bulunduğundan olanaklı tüm örneklemleri (10 tane) listeleyip, herbiri için örneklem istatistiklerini hesaplayabiliriz:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$m = medyan(x)$$

Olanaklı tüm örneklemler için örneklem istatistikleri:

Örneklem no	Örneklem değerleri	$\overline{x}$	m	s^2
1	6, 9, 12	9	9	9
2	6, 9, 15	10	9	21
3	6, 9, 18	11	9	39
4	6, 12, 15	11	12	21
5	6, 12, 18	12	12	36
6	6, 15, 18	13	15	39
7	9, 12, 15	12	12	9
8	9, 12, 18	13	12	21
9	9, 15, 18	14	15	21
10	12, 15, 18	15	15	9

Örnek:

Bir çift zar atılıyor. 1 ve 6 arasındaki sayılardan oluşan 36 sıralı çifti içeren sonlu eşit olasıklı S örneklem uzayı elde edilir. S={(1,1), (1,2), ... , (6,6)}, X(S)={1,2,3,4,5,6}. X, S'deki her (a,b) sayılarının en büyüklerine karşılık gelsin. X'in ağırlıklı ortalamasını bulunuz.

X'in f dağılımı,

f(1) = P(X =1) = P{(1,1)} = 1/36

f(2) = P(X =2) = P{(2,1), (2,2), (1,2)} = 3/36

f(3) = P(X =3) = P{(3,1), (3,2), (3,3), (2,3), (1,3)} = 5/36

f(4) = P(X =4) = 7/36

f(5) = P(X =5) = 9/36

f(6) = P(X =6) = 11/36

x_i	x_1	x_2	x_3	x_4	x_5	x_6
$f(x_i)$	1/36	3/36	5/36	7/36	9/36	11/36

$$E(\mathrm{X}) = \mu = \sum_{i=1}^{n} x_i f(x_i) = 1\,.\frac{1}{36} + 2\,.\frac{3}{36} + 3.\frac{5}{36} + 4\,.\frac{7}{36} + 5\,.\frac{9}{36} + 6\,.\frac{11}{36} = 4.47$$

Örnek:

4'ü kız, 12'si erkek olan bir sınıftan 4 nesneli örneklem rasgele çekiliyor. İçerisinde kız öğrencilerin olduğu *E* beklenen değeri bulunuz.

S, örneklem uzayı 4'lü toplam grup sayısı: $C_4^{16} = \frac{16!}{(16-4)!4!} = \frac{13*14*15*16}{24} = 1820$

0 kız öğrenci, 4 erkek öğrenci grup sayısı: $C_4^{12} = \frac{12!}{(12-4)!4!} = \frac{9*10*11*12}{24} = 495$

1 kız öğrenci, 3 erkek öğrenci grup sayısı: $C_1^{4} * C_3^{12} = \frac{4!}{(4-1)!1!} * \frac{12!}{(12-3)!3!} = 4 * \frac{10*11*12}{6} = 880$

2 kız öğrenci, 2 erkek öğrenci grup sayısı: $C_2^{4} * C_2^{12} = \frac{4!}{(4-2)!2!} * \frac{12!}{(12-2)!2!} = 6 * \frac{11*12}{2} = 396$

3 kız öğrenci, 1 erkek öğrenci grup sayısı: $C_3^{4} * C_1^{12} = \frac{4!}{(4-3)!3!} * \frac{12!}{(12-1)!1!} = 4 * \frac{12}{21} = 48$

4 kız öğrenci, 0 erkek öğrenci grup sayısı: $C_4^{4} * C_0^{12} = \frac{4!}{(4-4)!4!} * \frac{12!}{(12-0)!0!} = 1 * \frac{1}{1} = 1$

Toplam grup sayısı=495+880+396+48+1=1820

E=0*495/1820 + 1*880/1820 + 2* 396/880 + 3*48/1820 + 4* 1/1820=1

Örnek:

X rasgele değişkeninin olasılık fonksiyonu,

$f(x) = \begin{cases} cx & , \quad x = 1,2,3,4 \\ 0 & , \quad \text{diğer} \end{cases}$ **a)** olmak üzere, c değerini hesaplayınız.

a) f fonksiyonunun olasılık fonksiyonu olması için, $\sum f(x) = 1$

şartını sağlaması gerekir. Buna göre, $\sum_{x=1}^{4} cx = 1$ olmalıdır. Yani,

$c.1 + c.2 + c.3 + c.4 = 1$ $\quad 10.c = 1$ $\quad c = \dfrac{1}{10}$ olmalıdır. Buna göre,

$$f(x) = \begin{cases} \dfrac{1}{10}x & , \quad x = 1,2,3,4 \\ 0 & , \quad \text{diğer} \end{cases}$$

X	1	2	3	4
$P(X = x)$	0.1	0.2	0.3	0.4

biçiminde yazılabilir.

$$E(X) = \sum_{x=1}^{4} xf(x)$$
$$= 1 \times 0.1 + 2 \times 0.2 + 3 \times 0.3 + 4 \times 0.4$$
$$= 0.1 + 0.4 + 0.9 + 1.6 \quad = 3$$

c) $Var(X) = E(X^2) - \left(E(X)\right)^2$

$$E(X^2) = \sum_{x=1}^{4} x^2 f(x)$$
$$= 1 \times 0.1 + 2^2 \times 0.2 + 3^2 \times 0.3 + 4^2 \times 0.4$$
$$= 0.1 + 0.8 + 2.7 + 6.4 = 10$$

$$Var(X) = E(X^2) - \left(E(X)\right)^2 = 10 - 3^2 \quad = 1$$

d) $P(X = 1) = 0.1$

e) $P(2 < X \leq 4) = P(X = 3) + P(X = 4)$

$$= 0.3 + 0.4 = 0.7$$

f) $P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3)$

$$= 0.1 + 0.2 + 0.2 = 0.6$$

Kesikli Rassal Değişkenlerin Kümülatif Olasılık Fonksiyonu:

Kesikli bir rassal değişken olan X'in herhangi bir x_0 değerini aşmama olasılığı Kümülatif Olasılık Fonksiyonu ile aşağıdaki gibi gösterilebilir.
$F(x_0) = P(X \leq x_0) = \sum_{x \leq x_0}^{n} P(X = x)$

Kesikli bir rassal değişkenin kümülatif olasılık fonksiyonunun iki özelliği vardır.

- Her x_0 değeri için $0 \leq F(x_0) \leq 1$arasındadır.
- $x_0 < x_1$ ise , $F(x_0) \leq F(x_1)$ olur.

ÖRNEK: Atılan bir zarın 3'ten küçük gelme olasılıklarının fonksiyonunu yazınız.
Atılan bir zarın 3'ten küçük gelmesi 1 veya 2 gelmesi durumudur.
$F(x_0) = P(X \leq x_0) = \sum_{x=1}^{2} P(X = 1,\ X = 2)$ = 1/6 +1/6= 2/6

ÖRNEK: Atılan bir zarın 1'den büyük ve 6'dan küçük gelme olasılıklarının fonksiyonunu yazınız.
Atılan bir zarın1'den büyük ve 6'dan küçük gelmesi 2, 3, 4 veya 5 gelmesi durumudur.
$F(x_0) = P(X \leq x_0) = \sum_{x=2}^{5} P(X = 2,\ X = 3,\ X = 4,\ X = 5\ )$ = 4/6

Örnek: bir kutuda aşağıda verilen iskambil kağıtları bulunmaktadır:

Kupa: 2, 3, 4
Maça: 1, 2
Sinek: 3, 4
Karo: 1, 2, 3

a) Kutudan rasgele bir kağıt çekersek x rasgele değişkeni çekilen siyah kağıt sayısını göstermek üzere olasılık dağılımı ne olacaktır ?
Kutudaki toplam 10 kağıdın 4'ü siyah olduğu için çekilen kağıdın siyah olma olasılığı:
f(1) = P(x = 1) = 4 / 10

b) Kırmızı kağıt çekme olasılığı:
f(0) = P(x = 0) = 6 / 10

c) Kutudan rasgele bir kağıt çekersek ve **x** ile kağıdın üzerindeki sayıyı gösterirsek, **x** 1,2,3,4 değerlerini alabilme olasılıkları:
f(1) = P(x = 1) = 2 / 10 (2 as)
f(2) = P(x = 2) = 3 / 10 (3 adet ikili)
f(3) = P(x = 3) = 3 / 10 (3 adet üçlü)
f(4) = P(x = 4) = 2 / 10 (2 adet dörtlü)

d) iadeli olarak iki kağıt çektiğimizde x siyah kağıt sayısını göstermek üzere olasılık dağılımı nasıl olacaktır ? **K (** kırmızı) , **S** (siyah)

İki kağıt çekildiğinde sonuçlar:
KK , KS , SK , SS olacaktır. Olaylar bağımsız olduğundan olasılıklar K ve S olaylarının olasılıkları çarpımı ile hesaplanacaktır.

P(x = 0) = P(KK) = 6/10 * 6/10 = 9/25
P(x = 1) = P(KS) = 6/10 * 4/10 = 6/25
P(x = 1) = P(SK) = 4/10 * 6/10 = 6/25
P(x = 2) = P(SS) = 4/10 * 4/10 = 4/25

Örnek:
2 zarla atış yapıldığında x rastsal değişkeni zarların üste gelen yüzlerindeki sayıların toplamını gösterirse, şıklar 2 ve 12 arasında olacaktır. Olasılıklar şu şekildedir:

x(şıklar)	2	3	4	5	6	7	8	9	10	11	12
P(x)	1/36	2/36	3/36	4 /36	5/ 36	6/36	5/36	4/36	3/36	2/36	1/36

5.1.1. Bernoulli Dağılımı

Bir değer aralığında sadece tam sayılarla ifade edilebilen bazı değerleri alabilen ayrık değişkenler, ***Binom***, ***Hipergeometrik***, ***Poisson*** dağılımı göstermektedirler.

Birçok deneyde iki farklı sonuç ortaya çıkmaktadır. İki sonucu olan olayların olasılığının hesaplanmasında kullanılır. Bir sınav sonucu başarılı ve başarısız olmak üzere iki durumda tanımlanabilir ya da kalite kontrolu için alınan bir ürün sağlam veya kusurlu olarak iki şekilde ortaya çıkabilir. **İki sonucu olan deneyler bir defa denenirse "*Bernoulli Dağılımı*" olarak adlandırılır.** Bernoulli süreci, her deneyde birbirini engelleyen iki sonuçtan birinin gerçekleştiği bir süreçtir. Madeni para atışı deneyinde her deneyde yazı(Y) ve tura(T) mümkün sonuçlarından sadece biri gerçekleşir ve her atışta P(Y) ile P(T) olasılıkları deneyler birbirinden bağımsız olduğu için aynıdır (1 / 2).

Bernoulli deneylerinde ortaya çıkan iki sonuçtan biri *başarı* diğeri ise *başarısızlık* olarak adlandırılır. p başarı olasılığını, q'da başarısızlık olasılığını göstermektedir, 1- p = q.

Örneğin;

- Kusurlu ve kusursuz parçaların bulunduğu kutudan bir parçanın çekilmesi
- Öğrencinin bir derste başarılı veya başarısız olması

Öğrencinin dersinden geçme olasılığı p ise, dersinden kalma olasılığı (1-p) olacaktır. X raslantı değişkeni, x alabileceği tüm değerler ise başarı için 1, başarısızlık için 0 değerini alsın. X'in olasılık fonksiyonu;

$P(X=1)=p,$

$P(X=0)=1-p=q$

$\mathbf{P(X=x)=p^{x}(1-p)^{1-x}}$; x=0,1 ise bu dağılıma Bernoulli dağılımı denir.

Bernoulli Dağılımının Aritmetik Ortalaması ve Varyansı:

Bernoulli rassal değişkeninin aritmetik ortalaması: $\mu_x = E(x) = p$

Bernoulli rassal değişkeninin varyansı: $\sigma_x^2 = E[(X - \mu_x)^2] = p(1 - p) = pq$

Bir olayın olma olasılığı p olmak üzere, olmama olasılığı 1 − p olur. Bir olayın olma olasılığının olmama olasılığına oranına karşıtlığı denir.

$$karşıtlık = \frac{p}{q} = \frac{p}{1-p}$$

Örnek:

Ardışık üç Bernoulli denemesinde p(üçü de başarılı)=0.075 ise üçünün de başarısızlık olma olasılığını bulunuz. q = 1-p = 1-0.075 =0.925

ÖRNEK:

Aşağıdaki ifadenin Bernoulli olasılık dağılım fonksiyonu olması için q ne olmaldır:

$P (X = x) = p^x (q)^{1-x}$

q=1-p olur.

İfade verilen x hangi değerleri alır. (0 ve 1)

Örnek:

200 kişilik bir kitlede 120 kişi sigara içmiyor ve 80 kişi içiyor, yani %60'ı sigara içmiyor olsun. Bu kitleden rasgele bir kişi seçildiğinde seçilen kişinin sigara içmiyor olması olayının olasılığı p=0.6 olup, X rasgele değişkeni sigara içmeyen için 1, içen için 0 değerini alsın. X bir Bernoulli rasgele değişkenidir.

Sigara içmemenin içmeye karşıtlığı= $\frac{p}{1-p}$ = 0.6/0.4 = 3/2 Sigara içmeyen 3 kişiye karşılık 2 kişi sigara içmektedir.

Örnek:

Bir öğrenci, Fizik dersinden geçme olasılığının %70 olduğuna inanmaktadır. Olasılık dağılımının fonksiyonunu yazınız? Ortalamasını ve varyansını bulunuz?

Öğrenci dersten geçerse X rassal değişkeni x = 1 ve kalırsa x = 0 değerini alırsa, X rassal değişkeninin olasılık dağılımı şöyle yazılabilir:

P (x=1) = 0.7 ve P (x=0) = 0.3

Olasılık dağılım fonksiyonu:

$P (X = x) = p^x (1 – p)^{1-x} = 0.7^x (1 – 0.7)^{1-x} = 0.7^x *0.3^{1-x}$ olarak bulunur.

x'in alacağı 1 ve 0 değerlerine göre olasılık dağılımı aşağıdaki gibi bulunur.

$P (x = 1) = p^1 (1 – p)^{1-1} = 0.7^1 (1 – 0.7)^0 = 0.7^1 *0.3^0 = 0.7$

$P (x = 0) = p^0 (1 – p)^{1-0} = 0.7^0 (1 – 0.7)^1 = 0.7^0 *0.3^1= 0.3$

Aritmetik ortalaması: $\mu_x = E (X) = p = 0.7$

Varyansı: $\sigma_x^2 = E [(X - \mu_x)^2] = p (1 – p) = 0.7*0.3 =0.21$ olarak bulunur.

Örnek:

6 topta birini kırmızı bir top çizmek olarak tanımlanırsa, başarı olasılığı (P) 1/6 veya 0.17 olacaktır. Mavi olasılığı (mavi top çizimi) 5/6 veya 0,83 olacaktır. Herhangi bir Bernoulli çalışması için başarısızlık olasılığı her zaman 1 - P'dir.

Olasılık dağılım fonksiyonu:

$P (X = x) = p^x (1 – p)^{1-x} = 0.17^x (1 – 0.17)^{1-x} = 0.7^x *0.83^{1-x}$ olarak bulunur.

x'in alacağı 0 ve 1 değerlerine göre olasılık dağılımı aşağıdaki gibi bulunur.

$P (x = 0) = p^0 (1 – p)^{1-0} = 0.17^0 (1 – 0.17)^1 = 0.17^0 *0.83^1= 0.83$

$P (x = 1) = p^1 (1 – p)^{1-1} = 0.17^1 (1 – 0.17)^0 = 0.7^1 *0.83^0 = 0.17$

Aritmetik ortalaması: $\mu_x = E (X) = p = 0.17$

Varyansı: $\sigma_x^2 = E [(X - \mu_x)^2] = p (1 – p) = 0.17*0.83 =0.14$ olarak bulunur.

5.1.2. Binom Dağılımı

İki sonuçlu rassal deneme aynı koşullar altında n kere tekrarlanırsa “Binom Dağılımı” adı verilen dağılım elde edilir. Bernoulli dağılımının özel bir şeklidir.

Binom dağılım, Kesikli dağılımların en yaygın kullanılanıdır.
Atılan bir paranın yazı veya tura gelmesi, Montajdaki parçanın toleransa uygunluğu ve uygunsuzluğu, öğrencinin bir dersten başarılı veya başarısız olması gibi iki sonuçlu olayların olasılığının hesaplanmasında kullanılır.

Binom dağılımının sağlaması gereken şartlar:

- Deneme belirli sayıda (n) tekrarlanır.
- Her deneyin başarılı ve başarısız olmak üzere iki sonucu vardır.
- Deneyler birbirinden bağımsızdır.
- Başarı olasılığı (p) ve başarısızlık olasılığı q=1-p dir.
- n deneyde elde edilen başarılı sonuçlar x değişkenine atanır.

Aşağıda verilen deneylerde tanımlanan X, binom rasgele değişkenidir.

- Bir para 10 kez atılsın. X, Tura gelme sayısı
- İçinde 8 siyah ve 4 beyaz top bulunan bir kavanozda tekrar yerine koyarak 3 top çekilsin. X, Çekilen siyah topların sayısı.
- İçinde 3 kusurlu ve 7 kusursuz parça bulunan bir kutudan tekrar yerine koyarak 4 parça seçilsin. X, Seçilen kusurlu parçaların sayısı.

Deney **n** defa tekrarlanırsa toplam başarılı durum sayısı **x** ile gösterilen bir rasgele değişkendir. Bu değişken ***binom değişkeni*** adını alır. **x** değişkenini binom değişkeni olarak kabul edebilmemiz için tekrarlanan deneylerin birbirlerinin aynı olmaları, olasılıkların deneyden deneye değişmemesi, seçimlerin iadeli yapılması gerekir.

Binom dağılımındaki p ve q olasılıkları birbirine eşitse dağılımın şekli simetrik olacaktır. Para atışı deneyinde p ve q olasılıkları aynı olduğu için dağılım simetrik olacaktır. p’nin q’ya eşit olmadığı durumlarda ise, dağılımın şekli asimetriktir.

Öncelikle, n tane deneme olduğu için n tane iki olasılıklı sonuç olacaktır. İncelenen olay başarı ya da başarısızlık ise n tane denemede x tane başarılı ve (n-x) tane başarısız sonuç olacaktır ve denemeler birbirinden bağımsız olduklarından sonuçların herhangi bir dizilimin olasılığı, tekil sonuçların olasılıklarının çarpımına eşittir ve aşağıdaki gibidir.

$P(x, n-x) = p.p.p.........p.(1-p).(1-p).......(1-p) = p^x (1-p)^{n-x}$

Her bir denemenin başarılı olma olasılığı p ve başarısız olma olasılığı (1-p) dir. n tane rassal denemede x tane başarının (n-x) tane başarısızlıkla birlikte çok farklı dizilişlerde gerçekleşebilir yukarıda sadece bir tanesi dikkate alındı.

Diziliş sırası önemli değilse n rassal denemede x tane başarı içeren dizilişlerin sayısı:
$C_n^x = \frac{n!}{x!(n-x)!}$ (n rassal denemenin x taneli kombinasyonu) olarak bulunur.

n tane rassal denemenin x tanesinin başarılı olma Binom olasılık fonksiyonu:

$$\mathbf{P(X; n,p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}}$$

$$\mathbf{P(X; n,p) = C_n^x p^x (1-p)^{n-x}}$$

şeklinde yazılır ve bu formül yardımıyla hesaplanır. x = 0, 1, 2, 3, ……n
n: deneyin tekrarlanma sayısı
X: İstenen sonuç sayısı
p: İstenen başarılı sonucun olasılığı
q: Başarısızlık olasılığı

Binom Dağılımının Aritmetik Ortalaması, Varyansı ve Momentleri

- Aritmetik ortalaması: $\mu_x = E(X) = np$
- Varyansı: $\sigma^2 = E[(X - \mu_x)^2] = npq = np(1-p)$
- Momentleri:
  $\mu_1 = 0$
  $\mu_2 = \sigma_x^2$
  $\mu_3 = n.p.(1-p)[(1-p) - p]$
  $\mu_4 = n.p.(1-p)[(1-6 p(1-p)+3n.p(1-p)]$
- Asimetri (Çarpıklık) ve Basıklık Katsayıları:
  ÇK = $\frac{(1-p)-p}{\sqrt{np(1-p)}}$ ve BK = $3 + \frac{1-6p(1-p)}{np(1-p)}$

Örnek:
Bir para 10 defa atılsın. 4 defa yazı gelme olasılığını hesaplayınız
Binom dağılımın uygun olduğu rastgele olaylarda başarılı ve başarısız olarak iki durumun olduğu olaylarla ilgilenildiğinden:
başarılı: yazı gelmesi (p=0.5)
başarısız: yazı gelmemesi (q=0.5)
olarak tanımlama yapılabilir. n=10; X=4 olduğundan istenilen olasılık:

$$P(X; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$P(4;10,0.5)= \frac{10!}{4!(10-4)!} 0.5^4 (1-0.5)^6 = 0.205$$

Örnek:
Bir zarın 20 kez atılması durumunda tam 12 kez altı gelme olasılığını hesaplayınız.

başarılı: 6 gelmesi (p=1/6)
başarısız: 6 gelmemesi (q=5/6) olarak tanımlama yapılabilir.
n=20; X=12 olduğundan istenilen olasılık:

$$P(X; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$P(12;20,1/6)= \frac{20!}{12!(20-12)!} (1/6)^{12} (5/6)^8 = 0.0000135$$

Örnek:
Bir kutuda bulunan 10 tabletten 5 tanesi cep telefonudur. Bu kutudan yerine koyarak 3 tablet çekildiğinde 2 tanesinin cep telefonu olma olasılığı nedir? Ağırlık ortalama ve varyans değerlerini hesaplayınız. Yorumlayınız.

X: Çekilen tabletin aspirin olması

$$X \sim b(n = 3, p = \frac{1}{2})$$

$$P(X = 2) = \binom{3}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{3-2} = \frac{3!}{2!\,(3-2)!}\left(\frac{1}{2}\right)^3 = 0.375$$

Örnek 4 çocuklu bir ailede kız çocukların sayısı X rasgele değişkeni olsun.

$$X \sim b(n=4, p=\frac{1}{2})$$

olmak üzere,

$$f(x) = \binom{4}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{4-x} = \frac{1}{16}\binom{4}{x} \quad , \quad x = 0,1,2,3,4$$

dır. Olasılık tablosu,

x	0	1	2	3	4
$f(x) = \frac{1}{16}\binom{4}{x}$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

ve

$$E(X) = np = 2 \quad , \quad Var(X) = npq = 1$$

dır. 4'er çocuklu 160 ailenin kız çocuk sayısı bakımından dağılışı ne olur?

kız çocukların sayısı	aile sayısı (teorik sıklık , frekans)
0	10=160×P(X=0)
1	40=160×P(X=1)
2	60=160×P(X=2)
3	40=160×P(X=3)
4	10=160×P(X=4)
	160

ÖRNEK:

Bir madeni para 4 kere atılmaktadır. 0, 1, 2, 3 ve 4 tane yazı gelme olasılıklarını sırayla hesaplayınız.

Bu bir Binom dağılımıdır ve olasılık fonksiyonu
P(x; 4,0.5) = C_n^x p^x $(1\text{-}p)^{n-x}$ olarak yazılır.

0 tane yazı gelme olasılığı: P(0; 4,0.5) = $\frac{n!}{x!(n-x)!}$ p^x $(1\text{-}p)^{n-x}$ = $\frac{4!}{0!(4-0)!}$ $0.5^0\,0.5^4$ = 0.0625

1 tane yazı gelme olasılığı: P(1; 4,0.5) = $\frac{n!}{x!(n-x)!}$ p^x $(1\text{-}p)^{n-x}$ = $\frac{4!}{1!(4-1)!}$ $0.5^1\,0.5^3$ = 0.25

2 tane yazı gelme olasılığı: P(2; 4,0.5) = $\frac{n!}{x!(n-x)!}$ p^x $(1\text{-}p)^{n-x}$ = $\frac{4!}{2!(4-2)!}$ $0.5^2\,0.5^2$ = 0.375

3 tane yazı gelme olasılığı: P(3; 4,0.5) = $\frac{n!}{x!(n-x)!}$ p^x $(1\text{-}p)^{n-x}$ = $\frac{4!}{3!(4-3)!}$ $0.5^3\,0.5^1$ = 0.25

4 tane yazı gelme olasılığı: P(4; 4,0.5) = $\frac{n!}{x!(n-x)!}$ p^x $(1\text{-}p)^{n-x}$ = $\frac{4!}{4!(4-4)!}$ $0.5^4\,0.5^0$ = 0.0625

Örneğimizdeki gibi p = (1-p) ise simetrik binom dağılımı, p ≠ (1-p) ise asimetrik binom dağılımı söz konusudur.

Yukarıdaki örnekte aritmetik ortalamayı, varyansı, momentleri, çarpıklık ve basıklık katsayılarını hesaplayınız.

Aritmetik ortalaması: μ_x = E (X) = n.p = 4*0.5 = 2

Varyansı: σ_x^2 = E [(X - μ_x)^2] = 4*0.5*0.5 =1

Momentleri:

$\mu_1 = 0$

$\mu_2 = \sigma_x^2 = 1$

μ_3 = n.p.(1-p)[(1-p) – p] = 4*0.5*0.5 [0.5 – 0.5] = 0

μ_4 = n.p(1-p) [(1-6 p(1-p) + 3n.p(1-p)] =

4*0.5*0.5 [1- 6*0.5*0.5 + 3*4*0.5*0.5] = 2.5

Asimetri (Çarpıklık) ve Basıklık Katsayıları:

$$ÇK = \frac{(1-p)-p}{\sqrt{np(1-p)}} = \frac{(1-0.5)-0.5}{\sqrt{4*0.5*0.5}} = \frac{0}{1} = 0$$

$$BK = 3 + \frac{1-6p\,(1-p)}{np(1-p)} = 3 + \frac{1-6*0.5*0.5}{4*0.5*0.5} = 3 + \frac{-0.5}{1} = 2.5$$

Çarpıklık katsayısı bu olasılık dağılımının simetrik olduğunu, basıklık katsayısı ise normale göre basık olduğunu işaret etmektedir.

Örnek:

6 defa atılan bir madeni paranın tam olarak 2 defa tura gelmesi olasılığı nedir ?

Simetrik binom dağılımı, p ve q birbirine eşit.

n = 6 , x =2, p=0.50, q= 0.50

P(x=2) = $_6C_2\,(0.50)^2 * (0.50)^{6-2}$ = (6! / (2!(6 – 2)!))* $(0.50)^2 * (0.50)^4$ =0.23

Binom Dağılım tablosunu kullanarak hesaplarsak;

P(x;n,p)= P(2;6, 0.50) = (0.3438 – 0.1094) = 0.2344

Aynı 6 atışta en az 4 defa tura gelmesi olasılığı ise tablodan şu şekilde hesaplanabilir;

P(x ≥ 4) $\longrightarrow$ P(3; 6 , 0.50) = 0.6562

P(x ≥ 4) = 1 – 0.6562 = 0.3438

Veya

P(x=0)+P(x=1)+P(x=2) = 0.3438

Örnek:

Bir ansiklopediden rasgele seçilen 5 sayfadan en fazla 4 tanesinde baskı hatası olma olasılığı nedir ?

Asimetrik binom dağılımı. Baskı hatası oranı: p= (1500 / 5000) =0.30 q=0.70

$P(x \leq 4) = 1 – P(x=5)$

$P(x=5) = (5! / (5!*0!))*(0.30)^5 *(0.70)^0 = 0.00243$

$P(x \leq 4) = 1 – 0.00243 = 0.99757$

Örnek:

Bir kutuda bulunan ampullerden 3'ünün kusurlu olduğu bilinmektedir. Bu kutudan rasgele ve iadeli olarak 3 ampul çekildiğinde;

- İkisinin kusurlu olma olasılığı nedir ?
- Bu deney sonunda beklenen ortalama kusurlu sayısı ve standart sapma nedir ?

İki tanesinin kusurlu olma olasılığı;

p= 3 / 12 = 0.25

q = 0.75

$P(x=2) = (3! /(2!*1!))*(0.25)^2* (0.75)^1 = 0.1406$

Tablodan;

P(2;3, 0.25) =0.9844 bu kümülatif olduğu için P(x=1) bu değerden çıkarılır.

P(1;3, 0.25) =0.8438

(0.9844 – 0.8438) = 0.1406

ortalama kusurlu sayısı : E(x)= n*p = 3*(0.25) =0.75

standart sapma : σ = 0.75

Örnek:

Bir makinenin ürettiği parçaların %5'sinin kusurlu olduğu bilinmektedir. Bu makinenin ürettiği 6 ürün incelenmiştir.

$$P(x,6,0.05) = \begin{cases} \binom{6}{x}(0.05)^x(0.95)^{6-x} & , \quad x = 0,1,2,\dots,6 \\ 0, & diğer\ x'leriçin \end{cases}$$

a) Hiçbir ürünün kusursuz olma ihtimali

$$P(x=0) = \binom{6}{0}(0.05)^0(0.95)^{6-0}$$

$$\binom{6}{0} = \frac{6!}{0!\,(6-0)!} = 1$$

$$P(x=0) = 1 * 1 * (0.95)^6 = 0.735$$

b) Bir ürünün kusurlu çıkma ihtimali

$$P(x=1) = \binom{6}{1}(0.05)^1(0.095)^{6-1}$$

$$\binom{6}{1} = \frac{6!}{1!\,(6-1)!} = 6$$

$$P(x=1) = 6(0.05)^1(0.095)^5 = 0.22$$

c) En az iki ürünün kusurlu çıkma olasılığı

$$P(x \geq 2) = 1 - [P(x=0) + P(x=1)] = 1 - 0.735 - 0.22 = 0.045$$

Örnek:

Üretilen bir ürünün testlerden başarılı sonuçlanma olasılığı %80 dir. Üretilen 10 ürüne ilişkin yoğunluk fonksiyonları hesaplanacaktır.

a) Problem hangi yoğunluk fonksiyonunu işaret etmektedir.

Binom dağılımının olasılık fonksiyonu,

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0,1,\ldots,n$$

b) n, p ve q değerlerini belirleyiniz.

$$n = 10, p = 0.80$$

c) 6 ürünün başarılı olma olasılığı nedir?

$$P(X = 6) = \binom{10}{6} (0.80)^6 (0.20)^{10-6} = \frac{10!}{6!(10-6)!} (0.80)^6 (0.20)^4 = 0.088$$

d) En az 9 ürünün testlerden başarılı olma olasılığı nedir?

$$P(X \geq 9) = P(X = 9) + P(X = 10)$$

e) Ağırlık ortalama ve varyans değerlerini hesaplayınız. Yorumlayınız.

5.1.3. Poisson Dağılımı

Poisson olasılık dağılım fonksiyounu, aritmetik ortalamanın bir ifadesidir. Poisson ("puason" okunur) dağılımı, ender olaylar dağılımı olarak da bilinir. İstatistikte ve olasılık kuramında bir ayrık olasılık dağılımı olup belli bir sabit zaman birim aralığında meydana gelme sayısının olasılığını ifade eder. Bu zaman aralığında ortalama olay meydana gelme sayısının bilindiği ve herhangi bir olayla onu hemen takip eden olay arasındaki zaman farkının, önceki zaman farklarından bağımsız oluştuğu kabul edilir.

Poisson dağılımı Poisson süreci ile birlikte ortaya çıkar. Poisson süreci aralıklı karakterde olan (yani 0, 1, 2, 3 .. kere meydana çıkan) bazı olgularin bir birim zaman, alan, mekân veya hacimde sabit bir olasılıkla oluşması şeklini alır. Bu çeşit olaylara ve Poisson dağılımının uygulanmasına örnekler şunlardır:

- Prusya süvari birliklerinde her bir yıl at ve katır tepmeleri ile ölen asker sayısı: Bu klasik örnek 1868'de Ladislaus Josephovich Bortkiewicz tarafından bir kitapta yayınlanmış örnek olarak yıllarca askeri ve sivil yüksek okul öğrencilerine verilmiştir.
- Bir saat aralığında belli bir Internet sitesine gelen bağlantılar sayısı,
- Yarım saat içinde bir nakliyat deposuna yükleme-boşatılma için gelen kamyon sayısı,
- Belli bir trafik kavşağından 1 dakika içinde geçen otomobil sayısı,
- Belli bir zaman aralığında bir büyük binada yanıp çalışması duran florasan lambalarının sayısı,
- Bir hava alanına her saat inen uçak sayısı,
- Belirli bir trafik noktasında meydana gelen aylık trafik kazası sayısı,
- Bir üretim malındaki kusur sayısı,
- 1 cm^3 kandaki anormal hücre sayısı
- 10 m^2 kumaştaki kusur sayısı,

Verilmiş bir zaman aralığında bir alanda ya da hacimde başarıların sayısı X rasgele değişkeni olsun. Aşağıdaki koşulları sağlayan X e Poisson rasgele değişkeni denir.

- Deney, verilmiş bir zaman, alan ya da hacimde bir olayın (başarıların) elde ediliş sayılarının sayılmasıyla oluşur.
- İki ayrık birim zamanda alanda ya da hacimde elde edilecek başarıların sayısı birbirinden bağımsızdır.
- Bir birim zaman aralığı, alan ya da hacimdeki başarı olasılığı tüm birimler için aynıdır.
- Çok küçük bir zaman aralığı, alan ya da hacimde iki ya da daha çok başarının olması hemen hemen imkansızdır. Yani, bu durumda birden çok başarı olması olasılığı sıfıra yaklaşır.
- Bir birim zaman aralığı, alan ya da hacimde bir sonucun ortalama elde ediliş sayısı λ dır.

Poisson dağılımının genel odaklandığı rassal değişken bir sayılabilen olaydır; bu olay belli bir sabit uzunlukta olan (genellikle zaman) aralıkta ayrık olarak ortaya çıkar ve bu aralıkta gözlenen olayların sayısı Poisson dağılım için rassal değişkendir. Bu sabit aralıkta ortaya çıkan olaylar sayısının beklenen değeri (ortaya çıkmanın ortalama sayısı) λ olarak sabittir ve bu ortalama değer aralık uzunluğuna orantılıdır. Eğer her 4 dakikalık zaman aralığı içinde ortalama 5 olay meydana geliyorsa, sabit 8 dakikalık aralıkta ortalama 10 (=8x5/4) olay ortaya çıkar. Birbirini takip eden Poisson tipi olaylar arasındaki aralık karşılıklı ilişkili olarak bir üstel dağılım olur. X rassal değişken olmak üzere, alabileceği tüm değerler negatif olmayan bir tam sayı olan *x* ile gösterilirse (*x* = 0, 1, 2, 3...n) olayın ortaya çıkma olasılığı şöyle ifade edilir:

$$P(X = x, \lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

burada

- *e*, doğal logaritmanın tabanı (*e* = 2.71828...);
- *n*, olasılığı fonksiyon ile verilmekte olan olayın ortaya çıkma sayısı;
- *x*!, *x* için faktöriyel
- λ aritmetik ortalamadır, verilen sabit aralıkta ortaya çıkma sayısının beklenen değeri; bir pozitif gerçel sayıdır. Bu *x*'nin fonksiyonu Poisson dağılım için olasılık kütle fonksiyonu olur.

Poisson Dağılımının Aritmetik Ortalaması, Varyansı ve Momentleri:

- Aritmetik ortalaması: μ_x = E (X) = n.p = ʎ
- Varyansı: σ_x^2 = E [(X - μ_x)^2] = n.p. (1 – p) = ʎ
- Momentleri:

  $\mu_1 = 0$

  $\mu_2 = \sigma_x^2$ = ʎ

  μ_3 = ʎ

  μ_4 = ʎ + 3ʎ^2

- Asimetri (Çarpıklık) ve Basıklık Katsayıları:

  $$ÇK = \frac{\mu_3}{\sigma^3} = \frac{1}{\sqrt{ʎ}} \quad \text{ve} \quad BK = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{ʎ}$$

n≥ 100 ve n.p ≤10 koşulları sağlandığında Binom dağılımı, Poisson dağılımına yaklaşır ve Binom yerine Poisson dağılımı kullanılabilir. Poisson dağılımında ortalama olasılık bulmada Binom dağılımı kullanılır.

$$\lambda=\mu=n.p$$

Poisson dağılımı deney sayısnın çok fazla olduğu, meydana gelme olasılığının çok düşük olduğu problemlerde oldukça doğru öngörü sonuçları vermektedir. (p ≤ 0.01 ve λ=n.p ≤ 5)

Örnek:

10000 öğrencisi olan Esenyurt Üniversitesinde bir yıl içinde girilen hatalı notların aritmetik ortalaması 0.5'tür.

1. Aritmatrik ortalama, Λ = 0.5
2. n = 10000
3. p = $\frac{\Lambda}{n} = \frac{0.5}{10000}$ = 0.00005
4. Rassal denemenin beklenen sonucu hatalı-hatasız olmak üzere iki sonuçludur ve n kez tekrarlanabilir.

Bu şartlar altında bu olayın Poisson fonksiyonunu aşağıdaki gibi ifade edebilir.

$$P(x) = \frac{e^{-\Lambda}\,\Lambda^{x}}{x!} \quad , \quad x = 0, 1, 2, \ldots\ldots n$$

$$P(x) = \frac{e^{-0.5}\,(0.5)^{x}}{x!} \quad , x = 0, 1, 2, \ldots\ldots n$$

Bu Poisson dağılımının,

- Aritmetik ortalaması: $\mu_x = E(X) = n.p = \Lambda$
  $\mu_x = 0.5$

- Varyansı: $\sigma_x^2 = E[(X - \mu_x)^2] = n.p.(1 - p) = \Lambda$
  $\sigma_x^2 = 0.5$

- Momentleri:
  $\mu_1 = 0$
  $\mu_2 = \sigma_x^2 = \Lambda = 0.5$
  $\mu_3 = \Lambda = 0.5$
  $\mu_4 = \Lambda + 3\Lambda^2 = 1 + 3\,(0.5)^2 = 1.75$

- Asimetri (Çarpıklık) ve Basıklık Katsayıları:

$$ÇK = \frac{\mu_3}{\sigma^3} = \frac{1}{\sqrt{\Lambda}} = \frac{1}{\sqrt{0.5}} = 1.414$$

$$BK = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{\Lambda} = 3 + \frac{1}{0.5} = 5$$

Bir yılda gerçekleşebilecek hata sayısının (x) olasılıkları hesaplayabiliriz.

$$P(x=0) = \frac{e^{-0.5}\,(0.5)^{x}}{x!}$$

- Bir yıl boyunca hiç hata olmama olasılığı:

  P(x=0) = $\frac{e^{-0.5}\,(0.5)^0}{0!}$ = 0.6065

- Bir yıl boyunca bir adet hata olmama olasılığı:

  P(x=1) = $\frac{e^{-0.5}\,(0.5)^1}{1!}$ = 0.3033

- Bir yıl boyunca iki adet hata olmama olasılığı:

  P(x=2) = $\frac{0.6\,(0.5)^2}{2!}$ = 0.0758

- Bir yıl boyunca üç adet hata olmama olasılığı:

  P(x=3) = $\frac{0.6\,(0.5)^3}{3!}$ = 0.0126

Örnek:

Bir banka şubesine her yarım saat içinde ortalama 15 müşteri gelmektedir. Müşterilerin gelmelerinin Poisson sürecine uyduğu varsayılırsa, 10 dakika içinde bankaya en az 1 müşteri gelmesi olasılığını hesaplayınız.

30 dakikada 15 müşteri geliyorsa 10 dakikada ortalama 15/3 = 5 müşteri beklenecektir.

λ = 5 alınır,

P(en az 1) = 1 – P(x=0)

$P(x = 0) = \frac{e^{-5}\,5^0}{0!}$ =00674

P(en az 1) = 1 – 0.00674 = 0.99326

Matlab:

```
clear all
close all
lambda = 5;
x = 0:10;
P = poisspdf(x,lambda)
```

P = 0.0067 0.0337 0.0842 0.1404 0.1755 0.1755 0.1462 0.1044 0.0653 0.0363 0.0181

Örnek:

Bir elektrik santralında ayda ortalama 4 arıza meydana gelmektedir. Bu santralde bir ay içinde hiç arıza görülmemesi olasılığı nedir ?

P(x) = $\frac{e^{-\lambda}\lambda^x}{x!}$, x = 0, 1, 2,……n

λ=4

P(x=0)= 0.01832

Matlab:

```
clear all
close all
lambda = 4;
x = 0
P = poisspdf(x,lambda)
```

P = 0.0183

Örnek:

Bir üretim hattında parçaların ortalama %5'i kusurlu çıkmaktadır. Rasgele seçilen 22 parça içinde 2 tanesinin kusurlu olma olasılığı nedir ?

Binom dağılımı ile hesaplarsak:

P(x; n,p) = $\frac{n!}{x!(n-x)!}$ $p^x (1-p)^{n-x}$ = C_n^x $p^x (1-p)^{n-x}$

p=0.05,

n=22,

q=1-p=0.95,

x=2,

P(k=2)= 0.2070

Poisson dağılımı ile hesaplarsak;

Λ = np, beklenen sonucun ortalama gerçekleşme sayısını (Aritmetik ortalama),

np= λ = 22 * 0.05=1.1

P(x) = $\frac{e^{-\lambda}\lambda^x}{x!}$, x = 0, 1, 2,……n

P(x=2)=0.201

Burada, e = 2.71828 doğal logaritmanın tabanıdır.

Örnek:

Korona virüsüne yakalananların ortalama %2'si ölmektedir. Korona virüs testi pozitif çıkanlardan rasgele seçilen 10 kişi içinden 2 tanesinin ölme olasılığı nedir ?

Binom dağılımı ile hesaplarsak:

P(x; n,p) = $\frac{n!}{x!(n-x)!}$ p^x $(1\text{-}p)^{n\text{-}x}$ = C_n^x p^x $(1\text{-}p)^{n\text{-}x}$

p=0.02,

n=10,

q=1-p=0.98,

x=2,

P(X=2) = C_n^x p^x $(1\text{-}p)^{n\text{-}x}$ $= \frac{10!}{2!(10-2)!}$ 0.02^2 $(0.98)^{10\text{-}2}$ =0.015

Matlab:

```
clear all
close all
defects = 0:10;
P = binopdf(defects,10,.02)
```

P = 0.8171 0.1667 <u>0.0153</u> 0.0008 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

Poisson dağılımı ile hesaplarsak;

ʎ = np, beklenen sonucun ortalama gerçekleşme sayısını (Aritmetik ortalama),

λ = np= 10 * 0.02=0.2

P(x) = $\frac{e^{-ʎ}\,ʎ^x}{x!} = \frac{e^{-0.2}\;0.2^x}{x!}$, x = 0, 1, 2,……n

P(x=2)= $\frac{e^{-0.2}\;0.2^2}{2!} =$ 0.016

Burada, e = 2.71828 doğal logaritmanın tabanıdır.

Matlab:

```
clear all
close all
lambda = 0.2;
x = 0:10
P = poisspdf(x,lambda)
```

P = 0.8187 0.1637 0.0164 0.0011 0.0001 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

Örnek:

Bir fabrikanın ürettiği pillerin %1'i kusurlu çıkarsa, 200 birimlik üretimde 3 tane kusurlu pil çıkma olasılığı nedir ?

n=200, p=0.01, np= λ =200*0.01= 2

P(x) = $\frac{e^{-\lambda}\lambda^{x}}{x!}$, x = 0, 1, 2,……n

P(x=3)=0.18045

Örnek:

Bir sigorta şirketi 1000 kişiyi trafik kazalarına karşı sigortalamıştır. Bu kazalardaki ölüm oranı %1 olduğuna göre, sigorta şirketinin 5 kişiye para ödeme olasılığı nedir?

p=0.01, λ =1000*0.01=10

P(x=5)=0.0375

P(x=5) = $\frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-10}10^{5}}{5!}$ =0.0375

Örnek:

Bir sınıftaki öğrencilerin %2'sinin boyları 190 cm'nin üzerindedir. Rasgele seçilen 100 öğrenciden; 6'sının, en az 2'sinin boylarının 190 cm'den fazla olması olasılıklarını bulunuz.

P(x) = $\frac{e^{-\lambda}\lambda^{x}}{x!}$, x = 0, 1, 2,……n

λ =2, (100 öğrenciden 2 si)

P(x=6)=0.1203

P(x=0)=0.13534

P(x=1)=0.27068

P(x ≥2) = 1- P(x =0) – P(x =1) = 1 – 0.13584 - 0.27068 = 0.59398

Örnek:

Bir şehirdeki 30 yaşın üzerindeki nüfusun %5'nin üniversite mezunu olduğu bilinmektedir. 30 yaş üzerindekilerden rasgele seçilecek 100 kişi arasında;

a) 5 üniversite mezunu bulunması,
b) Hiç üniversite mezunu bulunmaması olasılığını hesaplayınız.

λ = n*p=100*0.05=5 ,

a) P(X=5) = $\frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-5}\,5^{5}}{5!}$ =0.175

b) P(X=0) = $\frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-5}\,5^{0}}{0!}$ =0.0067

Örnek:

Bir makinenin kusurlu ürün üretim oranı % 0.01 dir. Her saat başında üretim hattından alınan 100 parçanın incelenmesi sonucu 2'den fazla bozuk ürün üretldiğinde üretim durdurulacaktır. Üretimin durdurulma olasılığı nedir?

$$P(x,\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

P(x>2) = 1-P(x<=2) = 1 –[P(x=0) + P(x=1) + P(x=2)]

$$P(x>2) = 1 - [\frac{e^{-1}1^{0}}{0!} + \frac{e^{-1}1^{1}}{1!} + \frac{e^{-1}1^{2}}{2!}$$

P(x>2) =0.0803

Örnek:

Bir fabrikada çok nadir olan arızadan dolayı, bir hafta içinde arızalanan makine sayısı 4'dür. Belirli bir hafta için bu arızadan

a) Hiç bir makinenin arızalanmama olasılığı nedir?

$P(X = 0)$

b) En az iki makinenin arızalanma olasılığı nedir?

$$P(X \geq 2) = 1 - P(X < 2) = 1 - \left(P(X = 0) + P(X = 1)\right)$$

c) 3 makinenin arızalanma olasılığı nedir?

$P(X = 3)$

X, bir haftada arızalanan makine sayısı

$$f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad x = 0,1,2,\ldots, \lambda = 4$$

a) $P(X = 0) = \frac{e^{-4}4^{0}}{0!} = 0.0183$

b) $P(X \geq 2) = 1 - P(X < 2) = 1 - \left(P(X = 0) + P(X = 1)\right) = 1 - \left(\frac{e^{-4}4^{0}}{0!} + \frac{e^{-4}4^{1}}{1!}\right) =$

$= 1 - (0.0183 + 0.0733) = 1 - 0.0916 = 0.9084$

c) $P(X = 3) = \frac{e^{-4}4^{3}}{3!} = 0.195$

Örnek: Bir mağazaya Cumartesi günleri 5 dakikada ortalama olarak 4 müşteri gelmektedir. Bir Cumartesi günü bu mağazaya,

a) 5 dakika içinde 1 müşteri gelmesi olasılığını,

b)Yarım saate 2'den fazla müşteri gelmesi olasılığını,

a) $\lambda = 4$ P (x = 1) = ? $\quad P(X=1)=\dfrac{e^{-4}4^{1}}{1!}=4e^{-4}$

b) 5 dk'da 4 müşteri gelirse, 30 dk'da 24 müşteri gelir.

$\lambda = 24$ P (x > 2) = ?

P(x > 2) = 1 – [P(x=0)+P(x=1)+P(x=2)]

$$1-\left(\frac{e^{-24}24^{0}}{0!}+\frac{e^{-24}24^{1}}{1!}+\frac{e^{-24}24^{2}}{2!}\right)=1-313e^{-24}$$

Örnek:

Tehlikeli bir kavşağın güvenliği araştırılıyor. Geçmiş polis kayıtları bu kavşakta ayda ortalama beş kaza olduğunu göstermektedir. Kazaların sayısı bir Poisson dağılımına göre dağıtılır ve 0, 1, 2, 3 veya 4 kazadan herhangi bir ayda olma olasılığını hesaplayın.

Çözüm: Poisson formülünü kullanarak, kaza olmama olasılığını hesaplayabiliriz:

P(0) = $\frac{\lambda^n e^{-\lambda}}{x!}$ = $\frac{5^0 e^{-5}}{0!}$ = $\frac{(1)(0.0067)}{1}$ = 0.00674

Bir kaza olma olasılığı:

P(1) = $\frac{\lambda^n e^{-\lambda}}{x!}$ = $\frac{5^1 e^{-5}}{1!}$ = $\frac{(5)(0.0067)}{1}$ = 0.03370

İki kaza olma olasılığı:

P(2) = $\frac{\lambda^n e^{-\lambda}}{x!}$ = $\frac{5^2 e^{-5}}{2!}$ = $\frac{(25)(0.0067)}{2\times1}$ = 0.08425

Üç kaza olma olasılığı:

P(3) = $\frac{\lambda^n e^{-\lambda}}{x!}$ = $\frac{5^3 e^{-5}}{3!}$ = $\frac{(125)(0.0067)}{3\times2\times1}$ = 0.14042

Dört kaza olma olasılığı:

P(4) = $\frac{\lambda^n e^{-\lambda}}{x!}$ = $\frac{5^4 e^{-5}}{4!}$ = $\frac{(625)(0.0067)}{4\times3\times2\times1}$ = 0.17552

Herhangi bir ayda 0, 1 veya 2 kaza olasılığını bilmek istiyorsak, bu olasılıkları şu şekilde ekleyebiliriz:

P(0,1, or 2) = P(0) + P(1) + P(2) = 0.00674 + 0.03370 + 0.08425 = 0.12469

P(3 veya daha az kaza) = P(0, 1, 2, or 3) = P(0) + P(1) + P(2) + P(3)

= 0.00674 + 0.03370 + 0.08425 + 0.14042 = 0.26511

Üçten fazla olasılığını hesaplamak istiyorsak, 1- 0.26511=0.73489).

Önemli nokta: Poisson dağılımı, n, 20'den büyük ya da eşit olduğunda ve p, 0.05'e küçük ya da eşit olduğunda binom dağılımının iyi bir yaklaşımdır.

5.1.4. Hipergeometrik dağılım

Olasılık kuramında ve istatistikte, hipergeometrik dağılım sonlu bir ana kütle içinden tekrar geri koymadan birbiri arkasına *n* tane nesnenin çekilmesi işlemi için *başarı* sayısının dağılımını bir ayrık olasılık dağılımı şekilde betimler. Hipergeometrik Dağılım Varsayımları, n deneme benzer koşullarda tekrarlanabilir. Her denemenin iki mümkün sonucu vardır. Sonlu yığından iadesiz örnekleme yapılır. Örnekleme iadesiz olduğundan başarı olasılığı (p) deneyden deneye değişir.

Hipergeometrik Dağılımın Olasılık Fonksiyonu:

n : örnek hacmi
N : anakütle eleman sayısı
B : yığındaki başarı sayısı
x : örnekteki başarı sayısı
S = { x= 0,1, 2, 3, ...,n }

$$P(X = x) = \frac{C_x^B \, C_{n-x}^{N-B}}{C_n^N}$$

- Geri koyulmadan alınabilmesi mümkün örnek sayısı C_n^N'dir.
- Başarılı nesne sayısının *x* olması için C_x^B sayıda ihtimal bulunur;
- Geride kalanların başarısız olması için de C_{n-x}^{N-B} ihtimal mevcuttur.

Kombinasyon:

$$C(N, n) = C_n^N = \frac{N!}{(N-n)!\, n!}$$

Hipergeometrik Dağılımının Aritmetik Ortalaması, Varyansı

- Aritmetik ortalaması: μ_x = E (X) = n.p
- Varyansı: σ_x^2 = E [(X - μ_x)^2] = ($\frac{N-n}{N-1}$) n.p. (1 – p)

Burada p = $\frac{B}{N}$ olur.

Örnek:

10 kişilik bir sınıfta 6 öğrenci dersten başarılı olurken 4 öğrenci başarısız olmuştur. İadesiz olarak 5 öğrenci seçilmiş. Seçilen 5 öğrenciden, başarılı ve başarısız öğrencileri sınıflandıran bir olasılık değişkenleri şu şekilde gösterilebilir:

Anakütle eleman sayısı, N=10
Yığındaki başarı sayısı, B=6
Örnek hacmi, n=5
Örnekteki başarı sayısı, x
S = { x = 0,1, 2, 3, …..,n }

$$P(X = x) = \frac{C_x^B C_{n-x}^{N-B}}{C_n^N}$$

- Geri koyulmadan alınabilmesi mümkün örnek sayısı C_n^N'dir.
  $$C_n^N = \frac{10!}{5!\,(10-5)!} = 252$$
- Başarılı nesne sayısının *x* olması için C_x^B sayıda ihtimal bulunur;
- Geride kalanların başarısız olması için de C_{n-x}^{N-B} ihtimal mevcuttur.

a) Seçilen 5 öğrencinin de başarılı olma olasılığını bulunuz.

$$P\,(X = 5) = \frac{C_5^6\, C_0^4}{C_5^{10}} = \frac{\frac{6!}{5!\,(6-5)!} * \frac{(4)!}{(0)!\,(4)!}}{\frac{10!}{5!\,(10-5)!}} = \frac{6 * 1}{252} = 0.0238$$

b) 3 öğrencinin başarılı olma olasılığını bulunuz.

$$P\,(X = 5) = \frac{C_3^6\, C_{5-3}^{10-6}}{C_5^{10}} = \frac{\frac{6!}{3!\,(6-3)!} * \frac{(4)!}{(2)!\,(4-2)!}}{\frac{10!}{5!\,(10-5)!}} = \frac{20 * 6}{252} = 0.476$$

c) En fazla 2 öğrencinin geçmiş olma olasılığını bulunuz.

$$P\,(x\ \leq 2) = P\,(X = 0) + P\,(X = 1) + P\,(X = 2) = \frac{C_0^6\, C_{5-0}^{10-6}}{C_5^{10}} + \frac{C_1^6\, C_{5-1}^{10-6}}{C_5^{10}} + \frac{C_2^6\, C_{5-2}^{10-6}}{C_5^{10}}$$

$$P\,(x\ \leq 2) = \ 0.262$$

d) En az 3 öğrencinin geçmiş olma olasılığını bulunuz.

$$P\,(x\ \geq 3) = 1 - [P\,(X = 0) + P\,(X = 1) + P\,(X = 2)] = 1 - 0.262 = 0.738$$

e) Sırasıyla seçilen 1, 2, 3, 4 ve 5’inci öğrencinin başarısız olma olasılıklarını bulunuz.

Bu örnekte bulunması istenen seçilen öğrencinin başarısız öğrenci olma olasılığıdır. Başarısız öğrenci sayısı 4 olduğundan B=4 olacaktır.

$$P\ (x = 0) = \frac{C_0^4\ C_{5-0}^{10-4}}{C_5^{10}} = \frac{1 * 6}{252} = 0.0238$$

$$P\ (x = 1) = \frac{C_1^4\ C_{5-1}^{10-4}}{C_5^{10}} = \frac{4 * 15}{252} = 0.238$$

$$P\ (x = 2) = \frac{C_2^4\ C_{5-2}^{10-4}}{C_5^{10}} = \frac{6 * 20}{252} = 0.476$$

$$P\ (x = 3) = \frac{C_3^4\ C_{5-3}^{10-4}}{C_5^{10}} = \frac{1 * 6}{252} = 0.238$$

$$P\ (x = 4) = \frac{4\ C_{5-4}^{10-4}}{C_5^{10}} = \frac{1 * 6}{252} = 0.0238$$

Burada dikkat edilirse öğrencilerin başarısızlığıyla ilgili bütün olasılıklar hesaplanmıştır. Bu olasılıkların toplamı 1'e eşittir.

$$P\ (x = 0) + P\ (x = 1) + P\ (x = 2) + P\ (x = 3) + P\ (x = 4) = 1$$

f) Bu dağılımın başarılı ve başarısız öğrenci sayısına göre Aritmetik ortalaması ve standart sapmasını bulunuz.

<u>Başarılı öğrenci sayısına göre:</u>

N = 10, B = 6 ve n = 5. Burada p = $\frac{B}{N}$ = $\frac{6}{10}$ = 0.6 olacaktır.

- Aritmetik ortalaması: μ_x = E (X) = n.p = 5*0.6 = 3
- Varyansı: σ_x^2 = E [(X - μ_x)^2] = ($\frac{N-n}{N-1}$) n.p. (1 – p) = ($\frac{10-5}{10-1}$) 3 (1 – 0.6) = 0.666

<u>Başarısız öğrenci sayısına göre:</u>

N = 10, B = 4 ve n = 5. Burada p = $\frac{B}{N}$ = $\frac{4}{10}$ = 0.4 olacaktır.

- Aritmetik ortalaması: μ_x = E (X) = n.p = 5*0.4 = 2
- Varyansı: σ_x^2 = E [(X - μ_x)^2] = ($\frac{N-n}{N-1}$) n.p. (1 – p) = ($\frac{10-5}{10-1}$) 2 (1 – 0.4) = 0.666

Örnek:

Hastaneye gelen 100 kişiden 20 tanesinin koronavirüsü testi pozitif çıkmaktadır. İadesiz rasgele seçilen 12 kişiden 5 kişinin koronavirüs testinin pozitif çıkma olasılığı nedir?

N=100, B=20, n=12, x=0,1,2,….12

$$P\ (x = 5) = \frac{C_x^B C_{n-x}^{N-B}}{C_n^N} = \frac{C_5^{20}\ C_7^{80}}{C_{12}^{100}} = 0.047$$

Hipergeometrik Dağılımının Aritmetik Ortalaması ve Varyansını hesaplayınız.

Burada p = $\frac{B}{N}$ olur, p=B/N=20/100=0.2

- Aritmetik ortalaması: μ_x = E (X) = n.p=12*0.2=2.4
- Varyansı: σ_x^2 = E [(X - μ_x)^2] = ($\frac{N-n}{N-1}$) n.p. (1 – p) =(88/99)*12*0.2*0.8=1.71

Örnek:

Bir makineden üretilen 100 ürün içinde 60 tanesi testten geçmiştir. İadesiz seçilen 8 üründen 5 tanesinin testten geçme olasılığı nedir?

N=100, B=60, n=8, x=5

$$P\ (X = x) = \frac{C_x^B C_{n-x}^{N-B}}{C_n^N} = \frac{C_x^{60}\ C_{8-x}^{100-60}}{C_8^{100}}$$ x=0,1,2,3, ..., 8

$$P\ (X = 5) = \frac{C_5^{60}\ C_3^{40}}{C_8^{100}}$$ = 0.29

5.2. Sürekli Rassal Değişkenlerin Olasılık Fonksiyonu

X sürekli rassal değişken ise; X ∈ {-∞,∞} olmak üzere, f(x) fonksiyonu bu rassal değişkenin olasılık yoğunluk fonksiyonu olarak nitelendirilmektedir.

$$F(b) = \int_{-\infty}^{b} f(x)dx$$

$$P(X \in B) = \int_{B} f(x)dx$$

$$P(a < x \leq b) = \int_{a}^{b} f(x)dx$$

$$P(b) = P(b < x \leq b) = \int_{b}^{b} f(x)dx = 0$$

$\frac{dF(b)}{db} = f(b)$ Toplam dağılım fonksiyonun türevi olasılık yoğunluk fonksiyonunu verir.

Örnek:

X rassal değişkeninin olasılık yoğunluk fonksiyonu

$$f(x) = \begin{cases} 2x \ ; & 0 \leq x \leq 1 \\ 0 \ ; & dd \end{cases}$$

a) Bu bir olasılık yoğunluk fonksiyonu mudur?

b) $P(\frac{1}{2} < X \leq 1) = ?$

c) $P(X \leq \frac{3}{4}) = ?$

a) $A = \{x \mid 0 \leq x \leq 1\}$ $\forall x \in A$ için $f(x) \geq 0$ olur.

$\int_0^1 2x\,dx = x^2 = 1$ olduğundan f(x) bir olasılık yoğunluk fonksiyonudur.

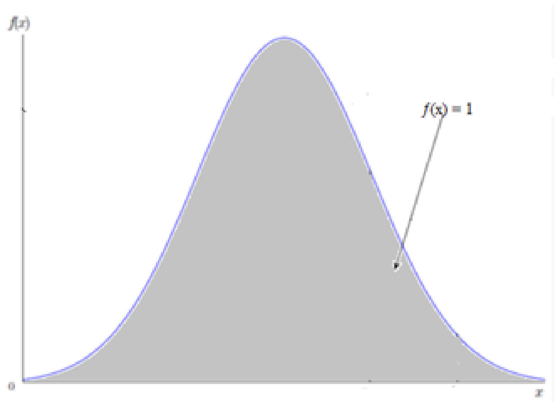
b) $P(\frac{1}{2} < X \leq 1) = \int_{\frac{1}{2}}^{1} 2x\,dx = x^2 \,|_{1/2}^{1} = 1 - \frac{1}{4} = \frac{3}{4}$

c) $P(X \leq \frac{3}{4}) = \int_{-\infty}^{\frac{3}{4}} f(x)\,dx = \int_{-\infty}^{0} \cancel{f(x)}\,dx + \int_{0}^{\frac{3}{4}} f(x)\,dx = \frac{9}{16}$

X sürekli rassal değişkeninin belli bir aralıkta olma olasılığı “Olasılık yoğunluk fonksiyonu” ile hesaplanabilir. Sürekli rassal bir değişkenin “Olasılık yoğunluk fonksiyonu” aşağıdaki özelliklere sahiptir:

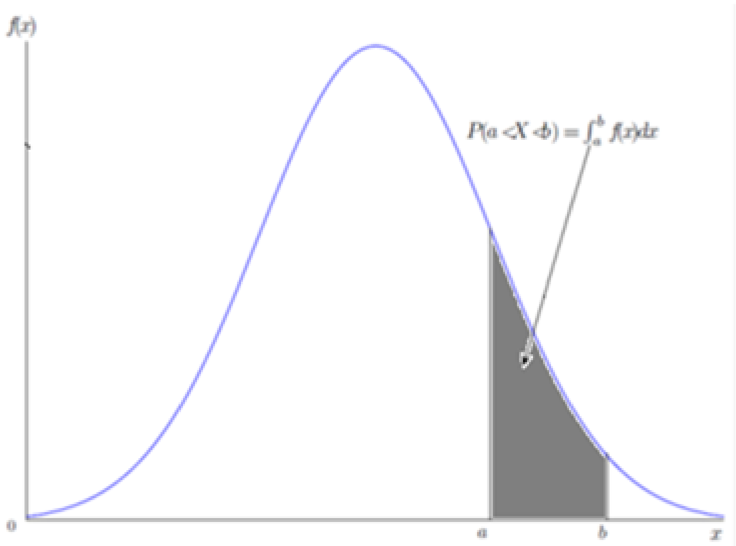
- Olasılık yoğunluk fonksiyonu ƒ(x) ile gösterilir. x’in bütün değerleri için ƒ(x) ≥ 0 (eğri yatay ekseni kesmez).
- Olasılık yoğunluk fonksiyonu eğrisi altında kalan alan 1’e eşittir.

$$\int_{-\infty}^{\infty} f\,(x)dx = 1$$



X rassal değişkeninin a ve b değerleri arasında bir değer olma olasılığı, Olasılık Yoğunluk Fonksiyonu altında bu iki değer arasında kalan alandır. (a<b)

P (a < X < b) = $\int_a^b f\,(x)dx$



ÖRNEK:

$$f(x) = \begin{cases} cx^2, & 0 \leq x \leq 3 \\ 0, & diğer\ x\ değerleri\ için \end{cases}$$

Yukarıdaki fonksiyonun "olasılık yoğunluk fonksiyonu" olabilmesi için c hangi değeri almalıdır?

Olasılık yoğunluk fonksiyonunda integral 0 ile sonsuz arasındadır.

$$f(x) = \int_{-\infty}^{x} f(x)dx = \int_{0}^{3} cx^2 dx = 1$$

$$= \mathrm{c}\left(\frac{\mathrm{x}^3}{3}\right)|_0^3 = \mathrm{c}\left(\frac{3^3}{3}\right) - \mathrm{c}\left(\frac{0^3}{3}\right) = \frac{27}{3}\ c = \frac{27}{3}\ c = 1'den$$

$$c = \frac{3}{27}\ \text{bulunur.}$$

Örnek:

$$f(x) = \begin{cases} \frac{3}{27}x^2, & 0 \leq x \leq 3 \\ 0, & diğer\ x\ değerleri\ için \end{cases}$$

F (x) kümülatif olasılık dağılım fonksiyonunu bulunuz?

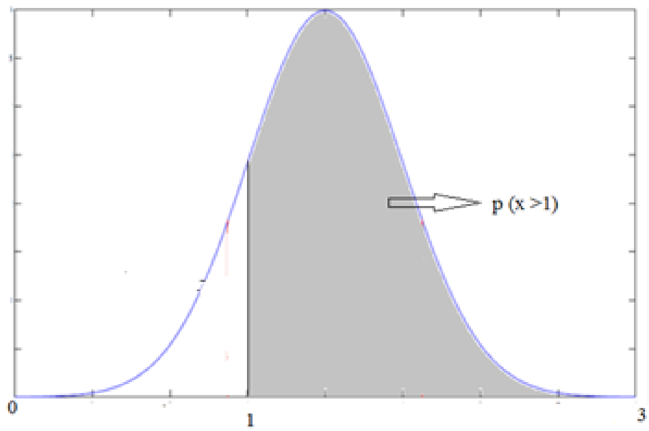
$$F(x) = \int_{-\infty}^{x} f(x)dx = \int_{0}^{x} \frac{3}{27}x^2 dx = \ = \frac{3}{27}\left(\frac{\mathrm{x}^3}{3}\right)|_0^{\mathrm{x}} = \frac{3}{27}\left(\frac{\mathrm{x}^3}{3}\right) = \frac{x^3}{27}$$

F (2) 'yi hesaplayınız?

$$F(2) = p\ (x\ \leq 2) = \frac{3}{27}\left(\frac{2^3}{3}\right) = \frac{8}{27} = 0.3$$

X rassal değişkeninin 2'den küçük olma olasılığı yüzde 30'dur.

P (x ≥ 1) ‘in olasılık yoğunluk fonksiyonunu hesaplayınız?

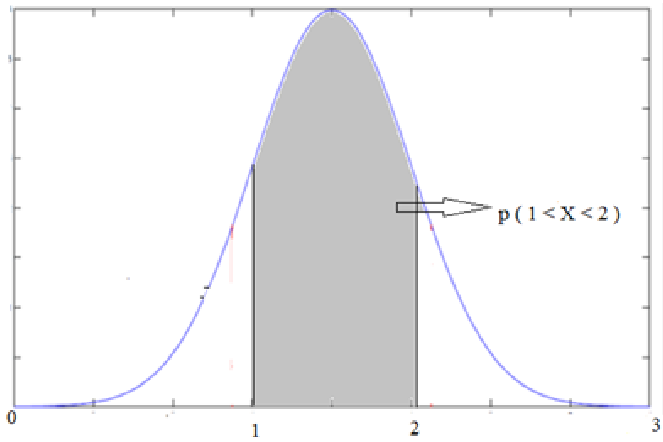


$$p\ (x\ \geq 1) = \int_1^3 \frac{3}{27} x^2 dx\ = \frac{3}{27}\left(\frac{x^3}{3}\right)|_1^3 = \frac{3}{27}\left(\frac{3^3}{3}\right) - \frac{3}{27}\left(\frac{1^3}{3}\right) =$$

$$= \frac{27}{27} - \frac{1}{27} = \frac{26}{27}\ = 0.96$$

X rassal değişkeninin 1’den büyük olma olasılığı yüzde 96’dır.

P (1 ≤ X ≤ 2) ‘in olasılık yoğunluk fonksiyonunu hesaplayınız?



$$p\ (1\ \leq x\ \leq 2) = \int_1^2 \frac{3}{27} x^2 dx\ = \frac{3}{27}\left(\frac{x^3}{3}\right)|_1^2$$

$$= \frac{3}{27}\left(\frac{2^3}{3}\right) - \frac{3}{27}\left(\frac{1^3}{3}\right) = \frac{8}{27} - \frac{1}{27}\ = \ \frac{7}{27} = 0.26$$

X rassal değişkeninin 1ile 2 arasında olma olasılığı yüzde 26’dır.

Sorular:

- f(t)=at+2 fonksiyonu t, 0 ile 1 arasında olasılık yoğunluk fonksiyonu olabilmesi için a değeri ne olmaldır. -2
- f(t)=at^2+2 fonksiyonu t, 0 ile 1 arasında olasılık yoğunluk fonksiyonu olabilmesi için a değeri ne olmaldır.-3
- f(t)=at^3+2 fonksiyonu t, 0 ile 1 arasında olasılık yoğunluk fonksiyonu olabilmesi için a değeri ne olmaldır.-4
- f(t)=at^4+2 fonksiyonu t, 0 ile 1 arasında olasılık yoğunluk fonksiyonu olabilmesi için a değeri ne olmaldır.-5
- f(t)=-2t+2 fonksiyonu t, 0 ile 1 arasında olasılık yoğunluk fonksiyonudur. Beklenen değeri ya da aritmetik ortalaması nedir?1/3
- f(t)=-3t^2+2 fonksiyonu t, 0 ile 1 arasında olasılık yoğunluk fonksiyonudur. Beklenen değeri ya da aritmetik ortalaması nedir?1/4
- f(t)=-4t^3+2 fonksiyonu t, 0 ile 1 arasında olasılık yoğunluk fonksiyonudur. Beklenen değeri ya da aritmetik ortalaması nedir?1/5
- f(t)=-5t^4+2 fonksiyonu t, 0 ile 1 arasında olasılık yoğunluk fonksiyonudur. Beklenen değeri ya da aritmetik ortalaması nedir?1/6
- f(t)=ae^(-2t) fonksiyonu t, 0 ile ꝏ arasında olasılık yoğunluk fonksiyonu olabilmesi için a değeri ne olmaldır. -2
- f(t)=ae^(-4t) fonksiyonu t, 0 ile ꝏ arasında olasılık yoğunluk fonksiyonu olabilmesi için a değeri ne olmaldır.-4
- f(t)=ae^(-20t) fonksiyonu t, 0 ile ꝏ arasında olasılık yoğunluk fonksiyonu olabilmesi için a değeri ne olmaldır.-20

Örnek:

Let $X \sim \text{uniform}(0,1)$. Find $E(X)$.

X has range $[0,1]$ and density $f(x) = 1$. Therefore,

$$E(X) = \int_0^1 x\,dx = \left.\frac{x^2}{2}\right|_0^1 = \boxed{\frac{1}{2}}.$$

Not surprisingly the mean is at the midpoint of the range.

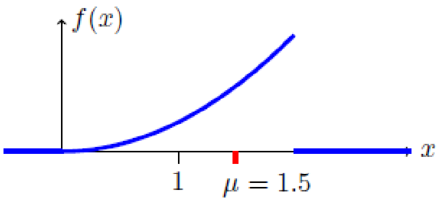
Örnek:

Let X have range $[0,2]$ and density $\frac{3}{8}x^2$. Find $E(X)$.

$$E(X) = \int_0^2 x f(x)\,dx = \int_0^2 \frac{3}{8}x^3\,dx = \left.\frac{3x^4}{32}\right|_0^2 = \boxed{\frac{3}{2}}.$$

Does it make sense that this X has mean is in the right half of its range?

Yes. Since the probability density increases as x increases over the range, the average value of x should be in the right half of the range.



Örnek:

Probability Density Function, pdf: f(x)

Let $X \sim \exp(\lambda)$. Find $E(X)$.

The range of X is $[0,\infty)$ and its pdf is $f(x) = \lambda e^{-\lambda x}$. Therefore

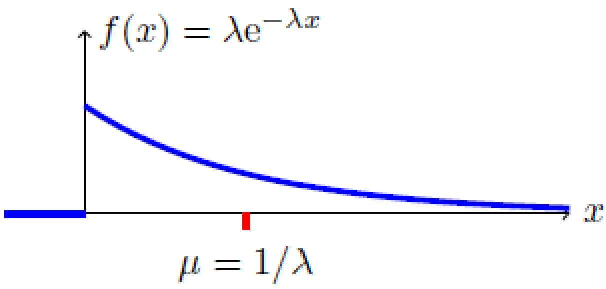
$$E(X) = \int_0^\infty x f(x)\,dx = \int_0^\infty \lambda x e^{-\lambda x}\,dx$$

(using integration by parts with $u = x$, $v' = \lambda e^{-\lambda x} \Rightarrow u' = 1$, $v = -e^{-\lambda x}$)

$$= -x e^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x}\,dx$$

$$= 0 - \left.\frac{e^{-\lambda x}}{\lambda}\right|_0^\infty = \frac{1}{\lambda}.$$

We used the fact that $xe^{-\lambda x}$ and $e^{-\lambda x}$ go to 0 as $x \to \infty$.



Mean of an exponential random variable

Örnek:

Let $X \sim \exp(\lambda)$. Find $E(X^2)$.

$$E(X^2) = \int_0^\infty x^2 f(x)\, dx = \int_0^\infty \lambda x^2 e^{-\lambda x}\, dx$$

(using integration by parts with $u = x^2$, $v' = \lambda e^{-\lambda x} \Rightarrow u' = 2x$, $v = -e^{-\lambda x}$)

$$= -x^2 e^{-\lambda x}\Big|_0^\infty + \int_0^\infty 2x e^{-\lambda x}\, dx$$

(the first term is 0, for the second term use integration by parts: $u = 2x$, $v' = e^{-\lambda x} \Rightarrow u' = 2$, $v = -\frac{e^{-\lambda x}}{\lambda}$)

$$= -2x\frac{e^{-\lambda x}}{\lambda}\Big|_0^\infty + \int_0^\infty \frac{e^{-\lambda x}}{\lambda}\, dx$$
$$= 0 - 2\frac{e^{-\lambda x}}{\lambda^2}\Big|_0^\infty = \frac{2}{\lambda^2}.$$

Örnek:

Let $X \sim \exp(\lambda)$. Find $\mathrm{Var}(X)$ and σ_X.

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x}\, dx = \frac{1}{\lambda} \quad \text{and} \quad E(X^2) = \int_0^\infty x^2\lambda e^{-\lambda x}\, dx = \frac{2}{\lambda^2}.$$

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \quad \text{and} \quad \sigma_X = \frac{1}{\lambda}.$$

We could have skipped Property 3 and computed this directly from $\mathrm{Var}(X) = \int_0^\infty (x - 1/\lambda)^2 \lambda e^{-\lambda x}\, dx$.

Örnek:

Soru 2.6.4: y herhangi bir kişinin ölüm yaşını göstermek üzere, yaşam süresini tanımlayan olasılık yoğunluk fonksiyonu

$f(y) = 3.10^{-9}.y^2(100-y)^2; 0 \le y \le 100$

olarak verilmiş olsun. Bir kişi en azından 70 yaşında ise bu kişinin 80 ile 85 yaşları arasında ölmesi olasılığı nedir?

$P(80 \le Y \le 85 \mid Y \ge 70) = ?$

$$P(80 \le Y \le 85 \mid Y \ge 70) = \frac{P(80 \le Y \le 85)}{P(Y \ge 70)}$$

$$\frac{\int_{80}^{85} 3.10^{-9} y^2(100-y)^2 dy}{\int_{70}^{100} 3.10^{-9} y^2(100-y)^2 dy} = \frac{0.031}{0.163} = \frac{31}{163} \text{ bulunur.}$$

Örnek X rasgele değişkeninin olasılık yoğunluk fonksiyonu,

$f(x) = \begin{cases} cx & , \quad 0 \le x \le 5 \\ 0 & , \quad dy \end{cases}$ olmak üzere,

a) $\int_0^5 cxdx = 1 \quad c\int_0^5 xdx = 1 \quad c\frac{x^2}{2}|_0^5 = 1 \quad c = \frac{2}{25}$

b) $f(x) = \begin{cases} \frac{2}{25}x & , \quad 0 \le x \le 5 \\ 0 & , \quad dy \end{cases}$

$E(X) = \int_0^5 x\frac{2}{25}xdx = \int_0^5 \frac{2}{25}x^2dx = \frac{2}{25}\int_0^5 x^2dx = \frac{2}{25}\frac{x^3}{3}|_0^5 = \frac{10}{3}$

c) $Var(X) = E(X^2) - (E(X))^2$

$E(X^2) = \int_0^5 x^2\frac{2}{25}xdx = \int_0^5 \frac{2}{25}x^3dx = \frac{2}{25}\frac{x^4}{4}|_0^5 = \frac{25}{2}$

$Var(X) = \frac{25}{2} - \left(\frac{10}{3}\right)^2 = \frac{25}{18}$

d) $P(1 \le X \le 3) = \int_1^3 \frac{2}{25}xdx = \frac{2}{25}\frac{x^2}{2}|_1^3 = \frac{2}{25}\left(\frac{9}{2} - \frac{1}{2}\right) = \frac{8}{25}$

e) $P(2 \le X \le 4) = \int_2^4 \frac{2}{25}xdx = \frac{2}{25}\frac{x^2}{2}|_2^4 = \frac{2}{25}\left(\frac{16}{2} - \frac{4}{2}\right) = \frac{12}{25}$

f) $P(X \le 3) = \int_0^3 \frac{2}{25}xdx = \frac{2}{25}\frac{x^2}{2}|_0^3 = \frac{2}{25}\left(\frac{9}{2} - 0\right) = \frac{9}{25}$

5.2.1. Düzgün Dağılım Fonksiyonu

X sürekli rastgele değişken belirli bir aralıktaki her değerinin meydana gelme olasılığı eşit ise bu rastgele değişkenin dağılım düzgün (Ünifrom) dağılımdır.

- Düzgün dağılım fonksiyonu:

$$f(x) = \frac{1}{b-a} \qquad a \leq x \leq b$$

- Düzgün dağılımın beklenen değeri:

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{x^2}{2(b-a)} \Big|_\alpha^\beta = \frac{b^2 - a^2}{2(b-a)}$$

$$E(X) = \frac{(b-a)(b+a)}{2(b-a)} \qquad E(X) = \frac{b+a}{2} \text{ olur.}$$

- Varyans için önce $E(X^2)$ hesaplanır.

$$E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{x^3}{3(b-a)} \Big|_\alpha^\beta = \frac{b^3 - a^3}{3(b-a)}$$

$$E(X^2) = \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} \qquad E(X^2) = \frac{b^2 + ab + a^2}{3}$$

- Düzgün dağılımın varyansı:

$$Var(X) = E(X^2) - [E(X)]^2 \text{ idi}$$

$$Var(X) = \frac{b^2 + ab + a^2}{3} - [\frac{b+a}{2}]^2$$

$$Var(X) = \frac{(b-a)^2}{12}$$

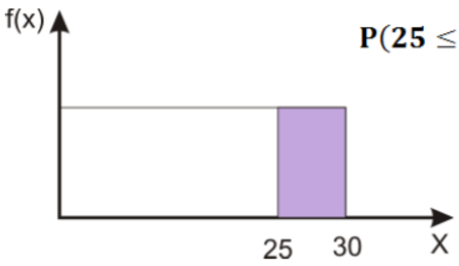
p = unifcdf (x, a, b), karşılık gelen alt uç noktayı (minimum), a ve üst uç noktayı (maksimum) kullanarak x cinsinden her bir değerdeki düzgün cdf'yi döndürür, b. x, a ve b, hepsi aynı boyuta sahip vektörler, matrisler veya çok boyutlu diziler olabilir. Skaler bir girdi, diğer girdilerle aynı boyutlara sahip sabit bir matrise genişletilir.

Örnek:

Süper marketteki kasaya 30 dakikalık periyotta bir müşteri gelmiştir. Bu müşterinin son 5 dakikada gelmiş olma ihtimalini hesaplayınız.

Olasılık yoğunluk fonksiyonu:

$$\mathbf{f(x)} = \begin{cases} \dfrac{\mathbf{1}}{\mathbf{b-a}} & \mathbf{a \leq x \leq b} \\ \mathbf{0} & \text{diğer} \end{cases} \quad \Rightarrow \quad \mathbf{f(x)} = \begin{cases} \dfrac{\mathbf{1}}{\mathbf{30-0}} & \mathbf{0 \leq x \leq 30} \\ \mathbf{0} & \text{diğer} \end{cases}$$



$$\mathbf{P(25 \leq X \leq 30)} = \int_{25}^{30} \mathbf{f(x)dx} = \int_{25}^{30} \frac{\mathbf{1}}{\mathbf{30}} \mathbf{dx} = \frac{\mathbf{30-25}}{\mathbf{30}} = \mathbf{1/6}$$

Bu örnek MATLAB komutu yardımıyla kolaylıkla hesaplanabilir: prob=unifcdf(5,0,30)
Continuous uniform cumulative distribution function: unicdf

Sorular:

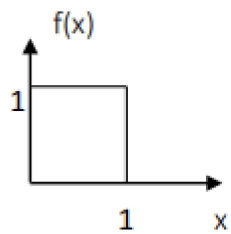
- Bir fabrikada 8 saatlik çalışma periyotunda bir ürün üretilmiştir. Bu ürünün olasılık fonksiyonu nedir? 1/8
- Bir fabrikada 10 saatlik çalışma periyotunda iki ürün üretilmiştir. Bu ürünün olasılık fonksiyonu nedir?1/5
- Bir fabrikada 10 saatlik çalışma periyotunda 4 ürün üretilmiştir. Bu ürünün olasılık fonksiyonu nedir?2/5
- Bir fabrikada 8 saatlik çalışma periyotunda bir ürün üretilmiştir. Bu ürünün son 2 saatte üretilme olasılığını hesaplayınız.1/4
- Bir fabrikada 10 saatlik çalışma periyotunda bir ürün üretilmiştir. Bu ürünün son 2 saatte üretilme olasılığını hesaplayınız.1/5
- Bir fabrikada 4 saatlik çalışma periyotunda bir ürün üretilmiştir. Bu ürünün son 2 saatte üretilme olasılığını hesaplayınız. 1/2

Örnek:

0 ile 1 arasında düzgün dağılıma sahip x rassal değişkeni için;

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & d.d \end{cases}$$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{1} f(x)dx = \int_{0}^{1} 1\, dx = 1$$



5.2.2. Üstel Dağılım Fonksiyonu

- Üstel dağılım fonksiyonu:

$$f(x)=\frac{1}{\mu}e^{-\frac{x}{\mu}} \quad x>0$$

- Beklenen değer:

$$E(X)=\int_0^\infty x\frac{1}{\mu}e^{-\frac{x}{\mu}}dx=\frac{1}{\mu}\int_0^\infty xe^{-\frac{x}{\mu}}dx \qquad u=x \quad du=dx \qquad dv=e^{-\frac{x}{\mu}}dx \quad v=-\mu e^{-\frac{x}{\mu}}$$

$$\int udv=uv-\int vdu \quad \text{kismi integrasyon islemi ile}$$

$$E(X)=\frac{1}{\mu}\left[-x\mu e^{-\frac{x}{\mu}}\Big|_0^\infty+\int_0^\infty \mu e^{-\frac{x}{\mu}}dx\right]=xe^{-\frac{x}{\mu}}\Big|_0^\infty-\mu e^{-\frac{x}{\mu}}\Big|_0^\infty$$

$$E(X)=\mu \quad \text{elde edilir .}$$

$$Var(X)=\mu^2 \text{ olur.}$$

Örnek:

Bir işletmenin üretilmiş olduğu elektronik cihazların arızasız çalışma sürelerinin (saat cinsinden) üstel dağılıma uyduğu görülmüştür ve ortalama arızasız çalışma süresinin 24 saat olduğu hesaplanmıştır. Buna göre

- a)Rastgele seçilen bir cihazın en az 12 saat arızasız çalışma olasılığını hesaplayınız
- b) En fazla 36 saat arızasız çalışması olasılığını bulunuz ?
- c) Seçilen cihazın 30 saatten fazla çalışma olasılığı %80 olabilmesi için bu cihazların ortalama arızasız çalışma süresi ne olmalıdır?

$$f(x)=\frac{1}{24}e^{-\frac{x}{24}} \quad x>0$$

a)

$$P(X\geq 12)=\int_{12}^\infty \frac{1}{24}e^{-\frac{x}{24}}dx=-e^{-\frac{x}{24}}\Big|_{12}^\infty=e^{-\frac{12}{24}}=e^{-0,5}=0,6065$$

b)

$$P(0<x<36)=\int_0^{36}\frac{1}{24}e^{-\frac{x}{24}}dx=-e^{-\frac{x}{24}}\Big|_0^{36}=1-e^{-1,5}=1-0,2231=0,7769$$

c)

$$\int_{30}^\infty \frac{1}{\lambda}e^{-(\frac{x}{\lambda})}dx=\frac{1}{\lambda}(-.\lambda.)e^{-(\frac{x}{\lambda})}\Big|_{30}^\infty=0,8\Rightarrow -e^{-(\frac{x}{\lambda})}\Big|_{30}^\infty=0,8$$

$$-e^{-\frac{\infty}{\lambda}}+e^{-\frac{30}{\lambda}}=0,8 \qquad e^{-\frac{30}{\lambda}}=0,8\Rightarrow -\frac{30}{\lambda}=\ln 0,8\Rightarrow \lambda=134 \text{ saat}$$

5.2.3. Normal Dağılım (Gauss)

Bir rasgele deneyin rasgele değişkenlerinin alabileceği tüm olası sonuçların toplamı 1'dir ve sonuçlara nasıl dağıldığını gösteren dağılıma **olasılık dağılımı** dedir.

Olasılık dağılımınd bir X tesadüfi değişkeninin alabileceği değerler,
x1 , x2 , x3 ,…. , xn olursa, bu değerlerinin olasılıkları da, P(x1), P(x2), P(x3),…. , P(xn)
olarak tanılanır.

X tesadüfi değişkeninin alabileceği bütün mümkün değerleri olasılıkları ile birlikte aşağıdaki gibi gösterebiliriz:

$X=x_i$	x1	x2	x3	……	xn
$P(X=x_i)$	P(x1)	P(x2)	P(x3)	……	P(xn)

X tesadüfi değişkeni bir anda x_i değerlerinden yalnızca birini alabildiğini varsayalım. Bu durumda, X tesadüfi değişkeninin x_i değerini alma olasılığı,
P(X=xi)= P(xi), şeklinde gösterilecektir.

Kesişim kümesi bulunmayan olaylarda, mümkün tüm sonuçların olasılıkları toplamı "1" dir:

$$\sum_{i=1}^{n} P(x_i) = 1$$

Bir rasgele deneyin tüm mümkün sonuçlarını olasılıklarına göre dağıtırsak bir **olasılık dağılımı** elde ederiz.

Olasılık dağılımları, ait oldukları tesadüfi değişkenin kesikli ya da sürekli olması durumuna göre, kesikli ve sürekli olasılık dağılımları şeklinde iki ana gruba ayrılır. Tesadüfi değişkenin kesikli olması halinde, kesikli olasılık dağılımı söz konusu olmaktadır ve binom, hipergeometrik ve poisson dağılımları uygulamada en çok karşılaştığımız kesikli olasılık dağılımlarıdır. Tesadüfi değişkenin sürekli olması durumunda ise, sürekli olasılık dağılımı ortaya çıkmaktadır.

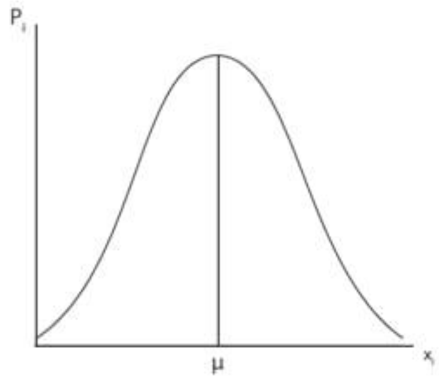
Normal Dağılım:

Gözlem değerlerinin dağılımında, ortalama değer etrafında toplanma ve ortalamadan uzaklaştıkça gözlem sayısının azaldığı gözlenir. Örneğin, zeka testi geniş bir kitleye uyguladığında, ortalama zekâ testi skoru etrafında gözlem değerlerinin toplandığını ve ortalama değerden uzaklaştıkça sayının yavaş yavaş azaldığını, en düşük seviyede ve en yüksek seviyede zekâ skoruna sahip olanların sayısının ise çok çok azaldığı gözlemlenir. Bu durum, olağan, beklenen ya da normal bir durum olarak değerlendirilir.

Normal dağılım sürekli bir olasılık dağılımıdır. Bilindiği gibi normal sözcüğü, olağan, sıradan ve yaygın olarak görülen vb. anlamlarda kullanılan bir sözcüktür. Normal dağılıma "normal" denmesinin de kökeninde yaygın görülen bir dağılım olması yatar.

Tüm olasılık dağılımlarında, dağılımı karakterize etmek yani nitelemek üzere iki önemli ölçü kullanılır. Bu ölçüler, dağılımın beklenen değeri, diğer deyişle dağılımın ortalaması ve dağılımın standart sapması veya varyansıdır. Normal dağılım da, ortalaması; μ ve standart sapması; σ ya da varyansı σ^2 ile ifade edilen sürekli bir dağılımdır. Normal dağılımın şekli simetriktir ve şeklinin çana benzemesi sebebiyle çan eğrisi olarak da bilinir.

Normal dağıılmda sürekli bir tesadüfi değişken söz konusu olduğundan ve süreklilik durumu, değişkenin iki değer arasında sonsuz değer alabilme özelliği olarak tanımlandığından, tesadüfi değişkenin alabileceği değerler birbirine çok yaklaşmaktadır. Bu durum, dağılımın şeklini bir eğri formuna dönüştürür. Zira, tesadüfi değişkenin alabileceği değerler birbirine o kadar yakındır ki, tesadüfi değişkenin alacağı değerleri gösteren X eksenindeki noktalar birbirine çok yaklaşmıştır ve normal dağılımın şekli de bu sebeple eğri biçimindedir. Aşağıda verilen şekilden de görüleceği gibi normal dağılım, dağılımın ortalama değeri (μ) üzerinde maksimum olasılığa sahip, simetrik ve dağılımın sağ ve sol kuyruklarının X eksenine asimptot olduğu yani, dağılımın kuyruklarının sonsuzda X eksenini kestiğinin varsayıldığı, bir dağılımdır.



Şekil: Normal Dağılım

Normal dağılımda olasılık hesabı, iki değer arasındaki alan hesabıyla yapılır ve eğrinin altındaki toplam alan, toplam olasılık olan ve tam sistem değerini ifade eden 1'e eşittir.

Dağılımın orta noktası olan ortalama yani μ değeri, eğrinin altında kalan alanı iki eşit parçaya ayırmakta olup, ortalamanın sağında ve solunda kalan alanlar 0,5 olasılığı ifade eder. Yani, tesadüfen seçilecek bir birimin ortalamanın altında değer alması olasılığı 0,5'dir, benzer şekilde ortalamanın üstünde değer alma olasılığı da yine 0,5 olmaktadır.

Normal dağılımın olasılık yoğunluk fonksiyonu:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(X-\mu)^2}{\sigma^2}}$$

formülüyle ifade edilmektedir. Ayrıca, yukarıdaki normal dağılım kısaca, $X \sim N\,(\mu,\sigma)$ şeklinde yazılarak, X tesadüfi değişkenin μ ortalaması ve σ standart sapması ise normal dağılmaktadır. Yukarıdaki normal olasılık yoğunluk fonksiyonunun $-\infty$ *ile* $+\infty$ arasında tanımlı bir fonksiyon olduğunu görüyoruz.

Normal dağılımın formülünde yer alan ifadeler:

- X, sürekli bir tesadüfi değişkeni,
- μ, tesadüfi değişkenlerin ortalamasını,
- σ, tesadüfi değişkenlerin standart sapmasını,
- π, matematikte sabit pi sayısını, 3.14159265…
- e, doğal logaritma tabanını, 2,71828…

ifade etmektedir.

Formülden de görüleceği gibi, normal dağılım fonksiyonunu niteleyen iki ölçü bulunmaktadır; ortalama ve standart sapma. Dolayısıyla, formülde yer alan diğer tüm değerlerin sabit olması sebebiyle, fonksiyonun belirleyicisi ya da niteleyicisi ortalama ve standart sapma iadeleri olmaktadır. Bu sebeple de ortalaması ve standart sapması farklı sonsuz sayıda normal dağılımdan söz edilebilmektedir.

Normal dağılım sürekli bir dağılım olduğundan olasılıklar, alan hesabı yapılarak bulunur. Normal dağılım fonksiyonunun altında kalan alanın toplam olasılık değeri olan "1" e eşit olduğunu ifade etmiştik. Normal dağılım fonksiyonunun $-\infty \ ile \ +\infty$ aralığında tanımlı olması ve toplam alanın da bir olması sebebiyle, dağılım fonksiyonunun söz konusu tanım aralığında integrali alındığında,

$$\int_{-\infty}^{+\infty} f(X) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}((X_i-\mu)/\sigma)^2} = 1$$

eşitliği gerçekleşir.

Öte yandan, olasılık hesabının, alan hesabı ile yapılması sebebiyle a < b olmak koşuluyla a ve b gibi iki değer arasındaki olasılık, yani a ile b arasındaki alan hesabının da,

$$\int_{a}^{b} f(X) = \int_{a}^{b} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}((X_i-\mu)/\sigma)^2}$$

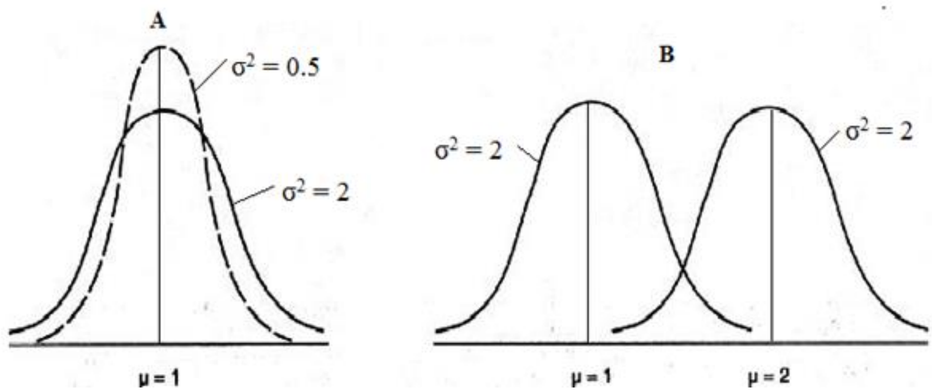
biçimindeki integralin hesaplanmasıyla yapılması gerekecektir.

Normal Dağılımın Özellikleri:

X rassal değişkeninin ortalaması μ ve varyansı σ^2 olan normal dağılıma uyduğu düşünülürse aşağıdaki özelliklere sahip olacaktır.

1. X rassal değişkeninin ortalaması E (X) = μ
2. X rassal değişkeninin varyansı Var (X) = E [(X- μ)^2] = σ^2
3. Skewness = 0 (Basıklık katsayısı)
4. Kurtosis = 0   (Çarpıklık Katsayısı)

X rassal değişkeninin dağılımın ortalaması ve varyansı bu değişkenin olasılık fonksiyonunun şeklini belirler. Dağılımın iki parametresi olan ortalama ve varyansı değiştikçe normal dağılım grafiğinin şeklide aşağıdaki örneklerde görüldüğü üzere değişecektir.

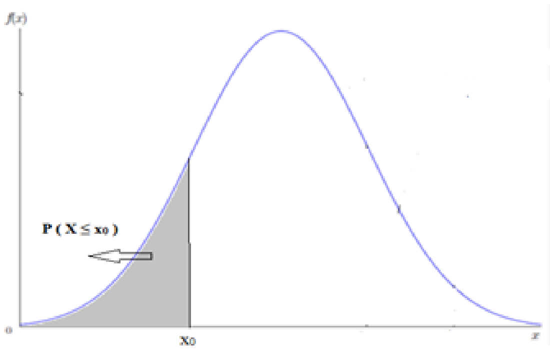


Şekil A'da normal dağılıma sahip ortalamaları aynı varyansları farklı iki tane olasılık yoğunluk fonksiyonu verilmektedir. Şekilden de görüldüğü üzere varyans azaldıkça olasılık yoğunluk fonksiyonu sivrileşmektedir (Basıklığı azalmaktadır).

Şekil B'de normal dağılıma sahip varyansları aynı ortalamaları farklı iki tane olasılık yoğunluk fonksiyonu verilmektedir. Şekilden de görüldüğü üzere ortalama artıkça yoğunluk fonksiyonu yana kayarken biçiminde herhangi bir değişme söz konusu değildir.

Normal Rassal Değişkenlerin Kümülatif (Birikimli) Olasılık Fonksiyonu:

Ortalaması μ ve varyansı σ^2 olan normal dağılıma sahip bir X rassal değişkeninin Kümülatif olasılık fonksiyonu F **(x) = P (X ≤ x_0)** olarak gösterilir. X rassal değişkeninin belli bir x_0 değerinde küçük olma olasılığını gösterir.

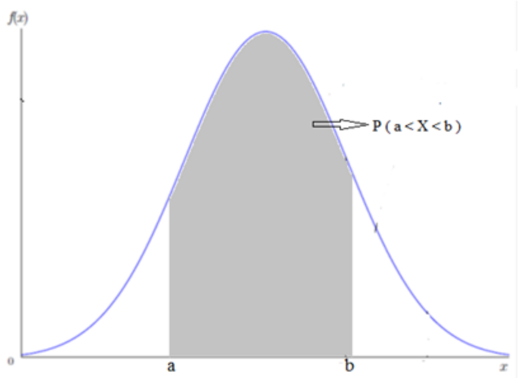


Yukarıdaki şekilde gri taralı alan, normal dağılıma sahip X rassal değişkeninin herhangi bir x_0 değerinden küçük olma olasılığını vermektedir. Bu alan aşağıdaki genel formül yardımıyla bulunabilir.

$$F(x) = P\,(X \leq \; x_0) = \int_{-\infty}^{x_0} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad dx$$

Normal Rassal Değişkenlerin Aralıklı Olasılık Fonksiyonu:

Ortalaması µ ve varyansı σ2 olan normal dağılıma sahip bir X rassal değişkeninin aralıklı olasılık fonksiyonu F (a) – F (b) = P (a < X < b) olarak gösterilir. X rassal değişkeninin a ve b gibi iki değer arasında olma olasılığını gösterir (a < b koşuluyla).



Yukarıdaki şekilde gri taralı alan, normal dağılıma sahip X rassal değişkeninin a ve b gibi iki değer arasında olma olasılığını vermektedir. Bu alan aşağıdaki genel formül yardımıyla bulunabilir.

$$f(x) = P\ (a < X < b) = \int_{a}^{b} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad dx$$

5.2.4. Standart normal dağılım

Ortalaması ve standart sapması farklı olan her normal dağılım için olasılık hesabı, integral çözümlemesi ile yapılacak ve yukarıdaki denklemin integrali de ortalaması ve standart sapması farklı her normal dağılım için tekrar tekrar hesaplanacaktır. Bu oldukça zahmetli integral hesaplarını yaparak olasılıkları hesaplamak yerine, z dönüşümü ile herhangi bir normal dağılımı standart normal dağılıma dönüştürmek ve standart normal dağılım için hazırlanmış olasılık değerleri tablosunu kullanarak ilgili alanı yani olasılığı kolaylıkla hesaplamak mümkündür.

$X \sim N\ (\mu,\sigma)$ gösterimi gereği μ ortalama ve σ standart sapma ise normal dağılan X tesadüfi değişkenine ilişkin her X_i değerine karşılık gelecek bir z_i değeri hesaplanmakta, başka bir deyişle X tesadüfi değişkeni z tesadüfi değişkenine dönüştürülmektedir. Standart z değişkeni de $z \sim N\ (0,1)$ sıfır ortalama ve bir standart sapma ile normal dağılmaktadır.

Normal dağılım fonksiyonundaki,

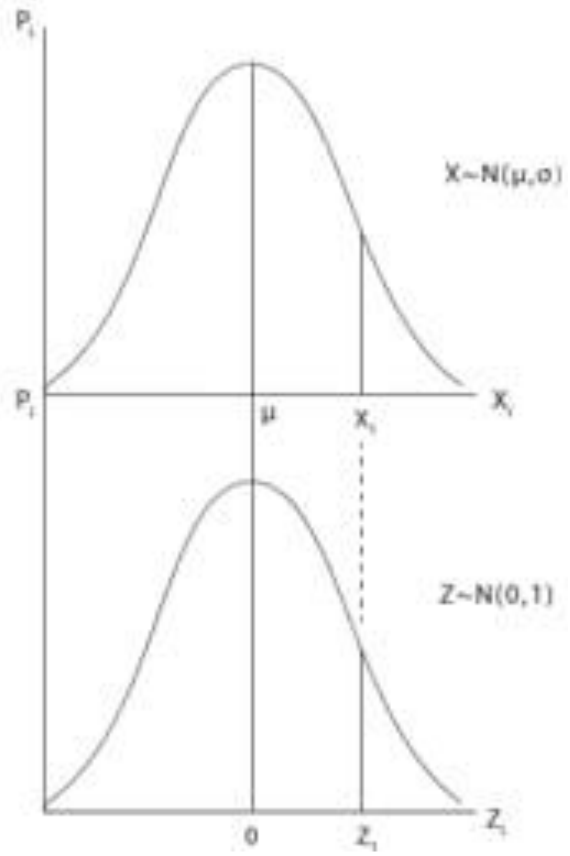
$$((X_i - \mu)/\sigma)^2$$

kısmı z dönüşümün yani, standart değer dönüşümün formülüdür. Z dönüşümü,

$$z = \frac{X_i - \mu}{\sigma}$$

formülüyle yapılmaktadır ve bir değerin kendi ortalamasından kaç standart sapma uzağa düştüğünü gösteren bir ölçüdür.

Formüllerde yer alan $\overline{X}$ ve μ değeri dağılımın ya da tesadüfi değişkenlerin ortalamasını ifade etmektedir.

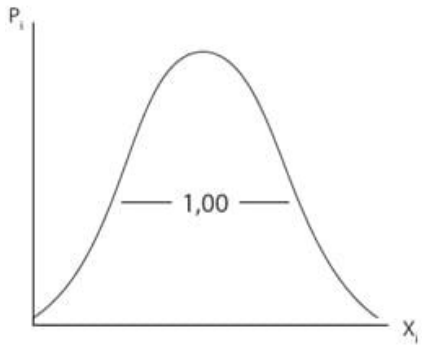


Şekil: Normal Dağılım ve Standart Normal Dağılım

X rassal değişkenin normal dağılıma sahip olduğundan olasılıklar Z (Normal dağılım tablosu) tablosu yardımıyla daha kolay hesaplanabilir. Z = $\frac{x-\mu}{\sigma}$ olarak ifade edilebilir.

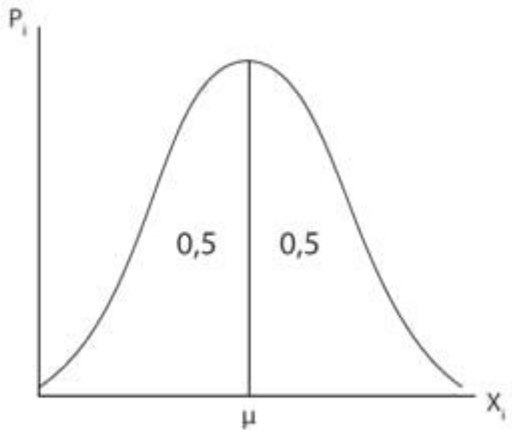
Standart Normal Dağılım Tablosu Kullanılarak Olasılıkların Belirlenmesi:
Normal dağılıma sahip bir tesadüfi değişkenin belirli değerleri alabilmesine yönelik olasılıkların, standart normal dağılım dönüşümü, normal olasılık değerleri tablosu kullanılarak elde edilir.

Olasılığın temel kuralı:
Normal eğrinin altında kalan toplam alan, daima “1”dir.



Şekil: Normal Eğri Altındaki Alan

Normal dağılım söz konusu olduğunda aritmetik ortalama, eğri altındaki alanı iki eşit kısma ayırır. Yani, ortalamanın sağında 0,5 ve solunda 0,5 olasılıkla gerçekleşecek iki alan oluşur.



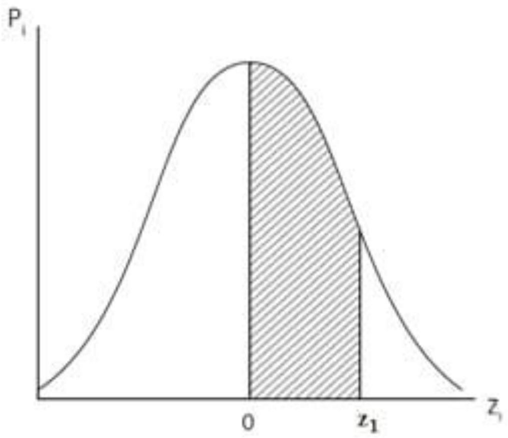
Şekil: Normal Eğri Altındaki Alan ve Ortalama

Standart normal dağılım tablosu ortalama ile ilgili z değeri arasındaki alanı yani olasılığı vermektedir. Dolayısıyla, bunun dışında olasılık hesabı yapmak gerektiğinde yukarıda bahsedilen eğri altındaki toplam alanın “1” olduğu ve ortalamanın eğri altındaki alanı 0,5 olasılıklı iki eşit kısma ayırdığı unutulmamalıdır. **Bu nedenle tablodan hesaplanan değerlerden 0.5 değeri çıkarılarak bulunur.**

Temelinde sürekli bir değişkenin bulunması sebebiyle, normal dağılımda tesadüfi değişkenin alabileceği değerler birbirine o derece yaklaşmıştır ki tek bir değer için olasılık sıfırdır. Dolayısıyla, $P(X<40)$ ya da $p(X\leq40)$ arasında bir fark oluşmamaktadır.

Standart normal eğri alanları tablosu sadece pozitif z değerleri için hazırlanmıştır. Normal dağılımın simetrik olması ve aritmetik ortalamanın eğri alanını iki eşit kısma ayırması sebebiyle, z değerinin pozitif olması üstünde bir değer söz konusu olduğunu, z değerinin negatif olması ise ortalamanın altında bir değer söz konusu olduğunu gösterir. Bir z değerinin pozitif ya da negatif olması sadece ortalamanın altında ya da üstünde bir değer olduğunu ifade eder, + ya da – bölgede olsa da ortalamaya uzaklık aynı olacağı için olasılık değerleri de aynı olacaktır.

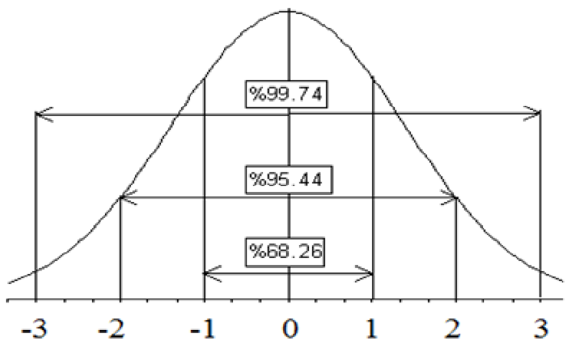
Standart normal dağılım tablosunda yer alan olasılıkların hangi bölgeye ilişkin olduğunu gösteren şekil aşağıda verilmektedir.

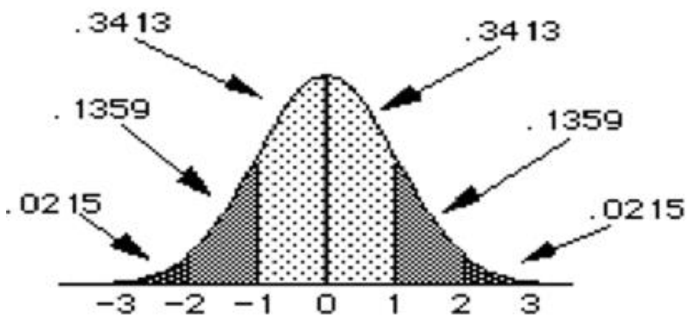


Şekil: Standart Normal Dağılım Tablosunda Okunan Olasılıkların

Standart normal dağılım Özellikleri:

1) Dağılım ortalamaya göre simetriktir. %50'si sağda, %50'si soldadır.
2) Normal dağılım eğrisinin altında kalan alan 1'e eşittir.
3) Aritmetik ortalama, ortanca, tepe değer (mod) birbirine eşittir ve maksimum yüksekliğin bulunduğu yerdedir.
4) Normal dağılımı gösteren değişkenlerin aldıkları değerlerin;
   - Gözlemlerin %68 ' i ortalama ile $\pm$ 1 standart sapma aralığına,
   - Gözlemlerin %95' i ortalama ile $\pm$ 2 standart sapma aralığına, ve
   - Gözlemlerin %99 ' i ortalama ile $\pm$ 3 standart sapma aralığına düşer.



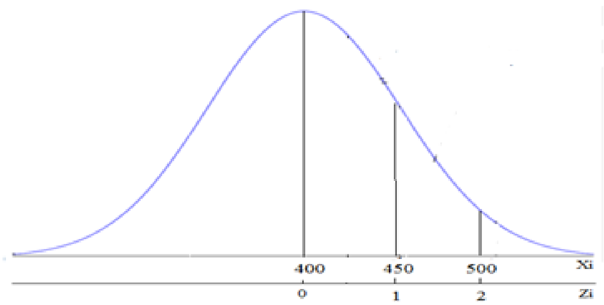


Normal Dağılımın Standart Normal Dağılıma Çevrilmesi:
Normal dağılımlar, standart normal dağılımlara kolaylıkla çevrilebilir. Bu çevirme işleminden sonra olasılıklar, standart normal dağılım (Z) tablosu yardımıyla bulunur.

Normal dağılıma (gauss) sahip olan bir fonksiyonu aritmetik ortalaması, μ ve standart sapması, σ biliniyorsa olasılık fonksiyonu standart normal dağılıma çevrilir. Seçilen değer, X, ortalamadan $Z=(X-\mu)/\sigma$, kadar standart sapmanın üstündedir ya da altındadır diye yorumlanır.

Örnek:
Üniversite öğrencilerinin gelirleri normal dağılıma sahiptir. Ortalamasının, μ=400 ve standart sapmanın, σ= 50 olduğu bilindiğine göre tesadüfü seçilen bir öğrencinin geliri X= 500 TL olsun. Bu normal dağılımı standart normal dağılıma çeviriniz.

$$Z = \frac{X - \mu}{\sigma} = \frac{500-400}{50} = 2$$



Bu durumda seçilen öğrencinin geliri, ortalamadan 2 standart sapma daha yüksektir.

Z tablosu:

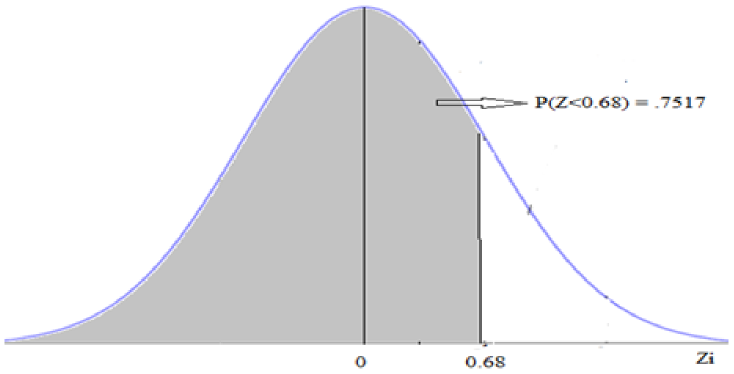
- Artı eksi 3.49 arasında değişiyor.
- Bu, teorik evrenin %99.98'ine karşılık geliyor.
- Z tablosu 1/10'luk aralarla standart sapmayı gösteriyor
- Araştırmacılar Z tablosundaki birkaç değerle ilgilenir. Çünkü çoğu hipotez testlerinde %95 ve %99'luk alanlarla ilgileniyor.

Aşağıdaki tabloda çeşitli Z değerleri için normal eğrilerin alanlarını hesaplanmıştır. Bu tablolar yardımıyla doğrudan olasılıklar hesaplanabilir.

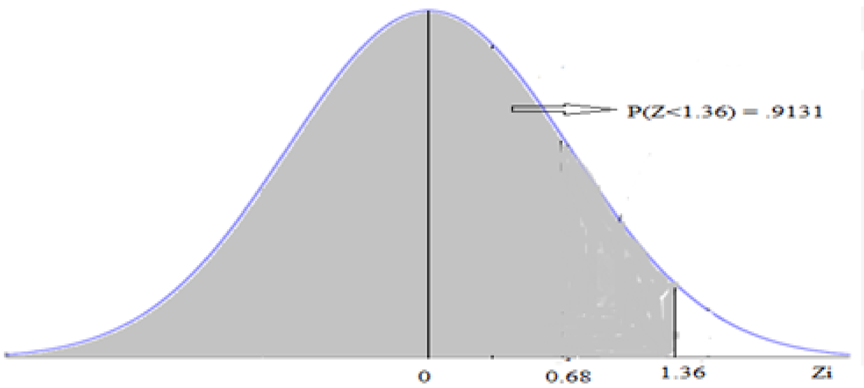
Tablo. Standart normal dağılım tablosu

	A	B	C	D	E	F	G	H	I	J	K
1	**z**	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
2	**0,00**	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
3	**0,10**	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
4	**0,20**	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
5	**0,30**	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
6	**0,40**	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
7	**0,50**	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
8	**0,60**	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
9	**0,70**	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
10	**0,80**	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
11	**0,90**	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
12	**1,00**	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
13	**1,10**	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
14	**1,20**	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
15	**1,30**	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
16	**1,40**	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
17	**1,50**	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
18	**1,60**	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
19	**1,70**	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
20	**1,80**	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
21	**1,90**	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
22	**2,00**	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
23	**2,10**	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
24	**2,20**	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
25	**2,30**	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
26	**2,40**	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
27	**2,50**	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
28	**2,60**	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
29	**2,70**	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
30	**2,80**	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
31	**2,90**	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
32	**3,00**	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Örnek: Standart normal dağılıma dönüştürülen bir normal dağılımın P(Z<0.68) olduğunu varsayalım. 0.68 için .6 mavi satır ve .08 kırmız sütunun kesiştiği nokta (.6 + .08 = .68) aranır. Böylece P(Z<0.68) = .7517 bulunur.



Örnek: Standart normal dağılıma dönüştürülen bir normal dağılımın P(Z<1.36) olduğunu varsayalım. 1.3 ve .06 noktalarının (1.3 + .06 = 1.24) kesiştiği yerde aranır. Böylece P(Z<1.36) = .9131 bulunur.

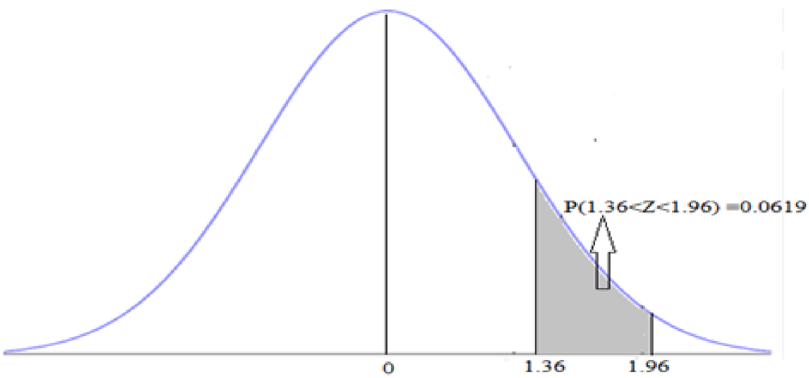


Örnek: Standart normal dağılıma dönüştürülen bir normal dağılımın P(Z>2.18) olduğunu varsayalım. Bu olasılığı tabloda arayabilmenin tek koşulu P(Z<Z_0) şeklinde yazılabilmesidir. P(Z>2.18) = 1 - P(Z<2.18) = 1- 0.9854 = 0.0146 şeklinde bulunur.

Örnek: Standart normal dağılıma dönüştürülen bir normal dağılımın P(Z< -1.36) olduğunu varsayalım. Bu olasılığı tabloda arayabilmenin tek koşulu aranan değerin pozitif olmasıdır. P(Z<-1.36) = 1 - P(Z<1.36) = 1- 0.9131 = 0.0869 şeklinde bulunur.

Örnek: Standart normal dağılıma dönüştürülen bir normal dağılımın P(Z>-2.96) olduğunu varsayalım. Normal dağılımın simetri özelliğinden bu dağılım P(Z<2.96) şeklinde yazılabilir.

Örnek: Standart normal dağılıma dönüştürülen bir normal dağılımın P(1.36<Z<1.96) olduğunu varsayalım. Bu aralıklardaki olasılık P(Z<1.96) - P(Z<1.36) = 0.9750- 0.9131 =0.0619 şeklinde bulunur.



- Standart normal dağılıma dönüştürülen bir normal dağılımın P(-1.26<Z<1.26) olduğunu varsayalım. Bu aralıklardaki olasılık P(Z<1.26) - [1 - P(Z<1.26)] = 2P(Z<1.28) - 1 şeklinde bulunur.

Örnek:

Bir imalathanede üretilen millerin çaplarının ortalaması 3.0005 inç ve standart sapmalarının ise 0.001 inç olan normal dağılıma uyduğu tespit edilmiştir. Üretilen miller eğer 3.00 +/- 0.002inç aralığının dışında iseler bu miller hatalı üretim kabul edilmektedir. Buna göre toplam üretimdeki hatalı ürün miktarını bulunuz.

İstenilen olasılık ifadesi: $\boldsymbol{P}(\textbf{hatalı ürün}) = \mathbf{1 - P(2.998 \leq X \leq 3.002)}$

Bu olasılık değerini hesaplamak için X sürekli normal değişkeni standart normal hale dönüştürülür:

$$\mathbf{Z = \frac{X - \mu}{\sigma} = \rightarrow Z_1 = \frac{2.998 - 3.0005}{0.001} = -2.5}$$

$$\mathbf{Z_2 = \frac{3.002 - 3.0005}{0.001} = 1.5}$$

$$\mathbf{P(\textbf{hatalı ürün}) = 1 - P(-2.5 \leq Z \leq 1.5) = 1 - 0.927 = 0.073}$$

Örnek:

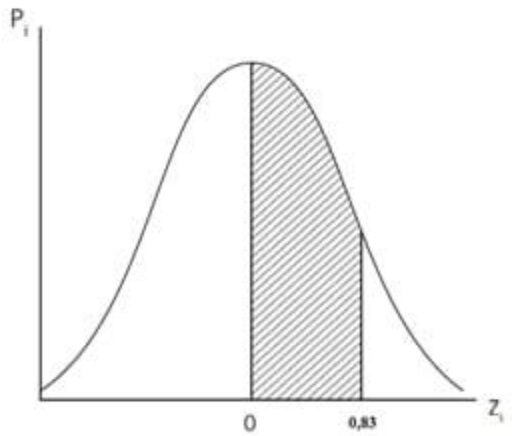
Bir kurum personel alımı için ilana çıkmış ve iş için başvuran çok sayıda adaya mesleki alanlarında hazırlanan bir test uygulamıştır. Adayların uygulanan testte aldıkları puanların ortalaması 60 ve standart sapması 12 hesaplanmıştır. Adayların aldıkları puanların normal dağıldığı bilindiğine göre, aşağıdaki şıklarda istenen olasılıkları hesaplayınız.

1) Tesadüfen seçilecek bir adayın 60-70 arasında puan almış olma olasılığı nedir?

Bunun için öncelikle 70 değerine karşılık gelen z standart değerinin hesaplanması yeterli olacaktı. Çünkü 60 değeri ortalama değer olduğundan Z_{60}=0 çıkacaktır.

$$z_{70} = \frac{X_i - \mu}{\sigma} = \frac{70 - 60}{12} = +0{,}83$$

Şekilden de görüldüğü gibi, istenen olasılık herhangi bir adayın 60-70 aralığında, yani standart z değeri olarak da 0-0,83 aralığında puan almış olma olasılığının hesaplanmasıdır.



Şekil: Ortalama ile ortalamadan büyük bir değer arasındaki alan için olasılık hesabı

Standart normal dağılım tablosunda +0,83 z değerine karşılık gelen olasılık değerini belirlenmesi için, tablonun 0,0 ile başlayıp aşağıya doğru devam eden ilk sütununda 0,8 değerinin yer aldığı satırı buluyoruz. z değerimiz 0,83 olduğu için 0,83 - 0,80 = 0,03 değerini de 0,00 ile başlayan ilk satırda belirliyor ve sonra bu iki sütun ve satırın kesişiminde yer alan değeri buluyor ve okuyoruz. Kesişim noktasında 0,7967-0.5=0.2967 değeri yer almaktadır. Dolayısıyla, tesadüfen seçilecek bir öğrencinin 60-70 arasında puan almış olma olasılığı % 29,67'dir. Başka bir deyişle sınava giren adayların % 29,67'sinin 60-70 arasında puan alması beklenir.

2) Tesadüfen seçilecek bir adayın 45-60 arasında puan almış olması olasılığı nedir?

Yine öncelikle 45'in z standart değerini hesaplamamız gerekecektir. Dağılımın ortalamasının 60 ve standart sapmasının da 12 olduğundan 45'e karşılık gelen z değeri,

$$z_{45} = \frac{X_i - \mu}{\sigma} = \frac{45 - 60}{12} = -1{,}25$$

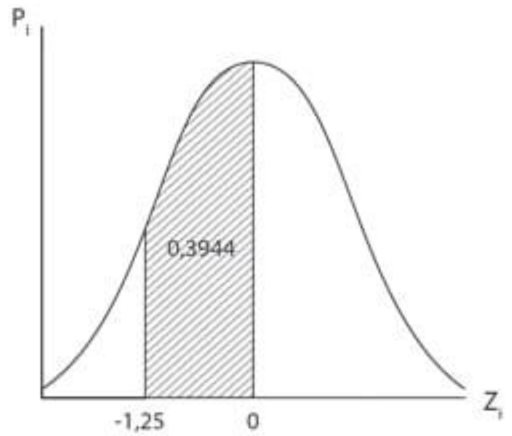
hesaplanmış olur.

z değerinin negatif çıkmasının, söz konusu puanın ortalamanın altında bir puan olduğunu göstermek dışında bir anlamı yoktur. Zira, normal dağılım simetrik bir dağılım olduğundan ve ortalama bu dağılımı tam ortadan iki eşit kısma ayırdığından, - 1,25 değeri ile + 1,25 değeri ortalamaya eşit uzaklıktadır.

Standart normal dağılım tablosu sadece pozitif z değerleri için düzenlenmiştir. Ancak, dağılımın simetrik olması sebebiyle negatif z değerleri için de aynı olasılıklar geçerlidir. Bu sebeple, z değeri negatif olduğunda da, işaret ihmal edilerek, z değerine karşılık gelen alan standart normal dağılım tablosundan aynı şekilde bulunur. z değerinin negatif olması sebebiyle, tablodan okunan alanın, ortalama ile ortalamanın altında yer alan söz konusu z değeri arasındaki alan olduğu uygulamacı tarafından bilinir.

Şimdi, sınava giren adaylar arasından tesadüfen seçilecek bir adayın, 45 ile 60 arasında değer alma olasılığını yani P(45 ≤ X ≤60)'ı belirleyelim (standart değer olarak da P(-1,25≤ z≤+1,25)). Bunun için, standart normal dağılım tablosunun 0,0 ile başlayan ilk sütununda 1,2 değerinin bulunduğu satırı belirliyoruz. Daha sonra 1,25 - 1,2 = 0,05 değerini de 0,0 ile başlayan ilk satırda arıyoruz. Daha sonra, ilk sütunda yer alan 1,2 satırı ile ilk satırda yer alan 0,05 sütununun kesiştiği hücredeki değeri okuyoruz. Bu değer, 0,3944 olarak tespit ediliyor.

Dolayısıyla, sınava giren adaylar arasından tesadüfen seçilecek birinin 45-60 arasında puan almış olma olasılığı % 39,44'dür ya da diğer deyişle, sınava giren adayların % 39,44'ü 45-60 arasında puan almıştır.



Şekil: Ortalama ile ortalamadan küçük bir değer arasındaki olasılığın belirlenmesi

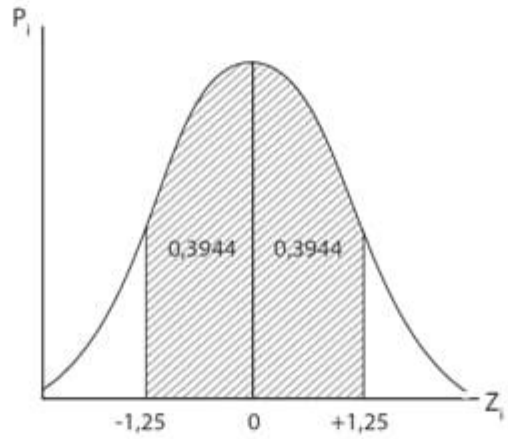
3) Tesadüfen seçilecek bir adayın 45-75 arasında puan almış olma olasılığı nedir?

Bu örnekte, 45 değeri ortalama değer olan 60'ın altında, 75 değeri ise üzerinde yer almaktadır. Her zaman olduğu gibi öncelikle 45 ve 75 puanlarına karşılık gelen z değerlerini hesaplamamız gerekecektir.

$$z_{45} = \frac{X_i - \mu}{\sigma} = \frac{45 - 60}{12} = -\,1{,}25$$

$$z_{75} = \frac{X_i - \mu}{\sigma} = \frac{75 - 60}{12} = +\,1{,}25$$

Görüldüğü gibi, 45 ve 75 puanlarına karşılık gelen z değerleri sırasıyla – 1,25 ve + 1,25 hesaplanmıştır. Şimdi bu değerleri normal dağılım üzerinde göstererek, hesaplamamız gereken alanı tarayalım.



Şekil : P(-1,25 ≤ z ≤ 1,25)

Standart normal olasılık değerleri tablosundan + 1,25 için olasılık değeri, ilk sütunda 1,2; ilk satırda ise (1,25 - 1,2=0,05) 0,05 değerini bularak ilgili satır ve sütunların kesişim noktasındaki olasılık 0,3944 olarak tespit ediliyor. Bu değerin 0 ile +1,25 arasındaki alanın olasılığını vermekte olduğuna dikkat ediniz.

Öte yandan, -1,25 için de tablodan olasılık değerinin 0,3944 olduğu anlaşılıyor bu değer de ortalamanın solunda kalan alan –1,25 ile 0 arasındaki alanın olasılığını veriyor.

Şekilden de görüleceği gibi, ortalamanın sağında kalan +1,25 değeri ile ortalama arasındaki alan 0,3944; ortalamanın solunda kalan yani ortalama ile –1,25 değeri arasındaki alanda 0,3944 olduğundan –1,25 ≤ 0 ≤ +1,25 arasındaki toplam taralı alan iki alanın toplamı alınarak bulunacaktır. Dolayısıyla – 1,25 ≤ 0 ≤ +1,25 arasındaki toplam alan, 0,3944 + 0,3944 =0,7888 % 78,88 hesaplanmaktadır.

Dolayısıyla, adaylar arasından tesadüfen seçilecek bir adayın 45 ile 75 arasında puan alması olasılığı % 78,88'dir. Başka bir deyişle sınava giren adayların % 78,88'i 45-75 aralığında puan almıştır.

4) Tesadüfen seçilecek bir adayın 80'den fazla puan almış olma olasılığı nedir?

80 puana karşılık gelen z değeri,

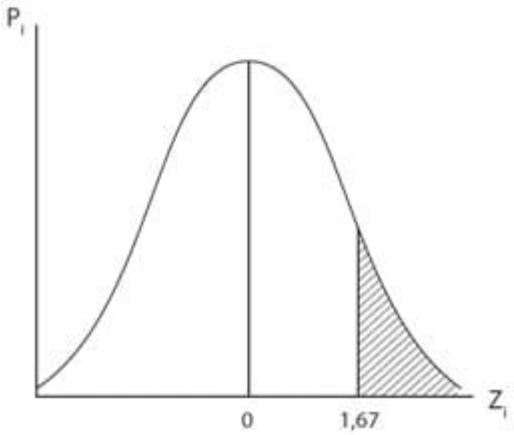
$$z_{80} = \frac{X_i - \mu}{\sigma} = \frac{80 - 60}{12} = +\,1{,}67$$

olarak hesaplanıyor. Adayın 80 puandan fazla alma olasılığı istendiğine göre, P(X > 80) olasılığının hesaplanması gerekiyor. Dolayısıyla, z > 1,67 bölgesinin olasılığını bilmek gerekiyor. z=1,67 değerine karşılık gelen olasılığı standart normal dağılım tablosundan, 0,4525 olarak belirliyoruz (İlk sütunda yer alan 1,6 değeri ile ilk satırda yer alan (1,67-1,6=0,07) 0,07 değerlerinin satır ve sütunlarını kesiştirdiğimizde 0,4525 değerine ulaşıyoruz).

0,4525 değeri ortalama ile 80 arasındaki alana ilişkin olasılık değeridir. Standart değer olarak da (0 - 1,67) arasındaki alanı ifade eder. Oysa bizden istenen 80'den fazla puan almış olma olasılığıdır ki bu $z > 1,67$ bölgesidir.

Ortalamanın sağında kalan alanın 0,5 olduğunu hatırlayalım. Bu durumda, $z > 1,67$ bölgesinin alanını bulmak için iki alanın farkını alarak istenen olasılığı, 0,5 – 0,4525=0,0475 hesaplanır.

Dolayısıyla, tesadüfen seçilecek bir adayın 80'den fazla puan almış olma olasılığı, % 4,75 olmaktadır. Başka bir deyişle, bu sınava giren adayların % 4,75'i 80 puanın üzerinde puan almıştır.



Şekil: P(z > 1,67)

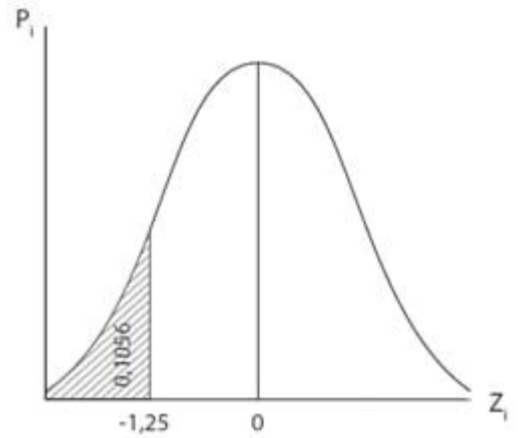
5) Tesadüfen seçilecek bir adayın 45'den az puan alma olasılığı nedir?

Standart değer dönüşümü ile, 45'e karşılık gelen z değerini,

$$z_{45} = \frac{X_i - \mu}{\sigma} = \frac{45 - 60}{12} = -1,25$$

hesaplıyoruz. Bu değerin standart normal dağılım olasılıkları tablosundaki olasılık değeri, ilk sütunda yer alan 1,2 değeri ile ilk satırda yer alan 0,05 değerlerinin satır ve sütunlarının kesiştirilmesi sonucunda 0,3944 olarak belirleniyor. Bu değer, ortalama (60) ile 45 arasındaki alanın olasılığıdır. Dolayısıyla, 45'in sol tarafındaki alanı bulmak için, ortalamanın solunda kalan tüm alanın 0,5 olduğu hatırlanarak ve iki alanın farkının alınması gerekir.

Ortalamanın solunda kalan tüm alandan yani 0,5'den , ortalama ile -1,25 değeri arasındaki alan çıkarılarak olasılık, 0,5 – 0,3944 = 0,1056 olarak hesaplanır. Bu durumda, tesadüfen seçilecek bir adayın 45'den az puan alma olasılığı % 10,56'dır, Başka bir deyişle sınava giren adayların % 10,56'sı 45'den az puan almıştır.



Şekil: P(-1,25 < z)

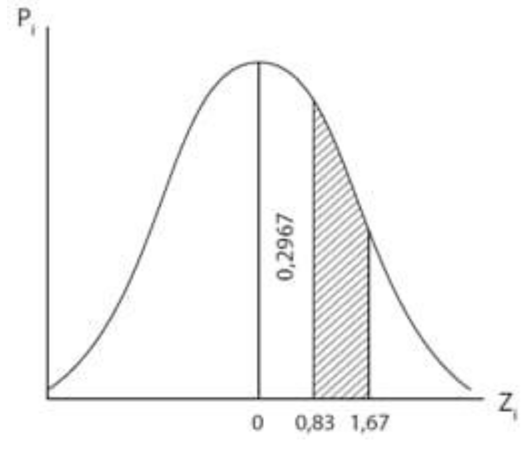
6) Tesadüfen seçilecek bir adayın70-80 arasında puan alması olasılığı nedir?

Öncelikle 70 ve 80 puanlarının standart değer karşılıklarını bulalım:

$$z_{70} = \frac{X_i - \mu}{\sigma} = \frac{70 - 60}{12} = +\,0{,}83$$

$$z_{80} = \frac{X_i - \mu}{\sigma} = \frac{80 - 60}{12} = +\,1{,}67$$

Her iki z değeri de pozitiftir ve ortalamanın sağında kalan bölgede yer almaktadır. Şekil üzerinde, z değerlerini göstererek, bizden istenen alanı belirleyelim. Şekilden 70 ve 80’e karşılık gelen z değerleri ve istenen taralı alan takip edilebilir.



Şekil: P(0,83 ≤ z ≤ 1,67)

Bu örnekte olasılığı istenen alan, 70 ≤ X ≤ 80 arasında kalan alandır ve bu alan z değeri olarak da 0,83 ≤ z ≤ 1,67 aralığına karşılık gelmektedir.

0 ≤ z ≤ 1,67 alanı standart normal dağılım tablosundan 0,4525;

0 ≤ z ≤ 0,83 alanı ise yine standart normal dağılım tablosundan 0,2967 olarak belirleniyor.
Bu durumda şekilden de izlenebileceği gibi 0,83 ≤ z ≤ 1,67 arasındaki alan, büyük alandan küçük alanın çıkartılması suretiyle, 0,4525 – 0,2967 = 0,1558 hesaplanı.

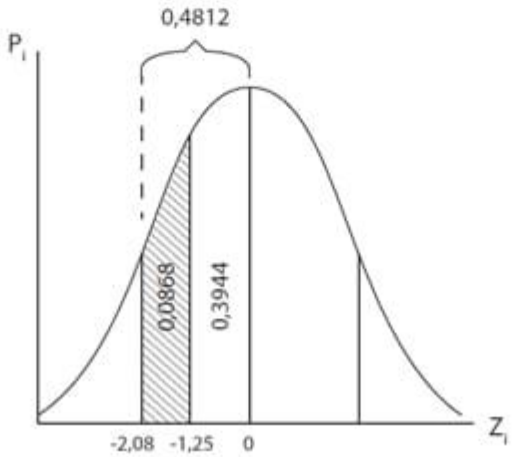
Dolayısıyla, iş başvurusu yapan adaylar arasından seçilecek herhangi birinin 70 ile 80 arasında puan alma olasılığı % 15,58 olarak hesaplanmakta, başka bir deyişle adayların % 15,58'i 70-80 arasında puan almış olmaktadır.

7) Tesadüfen çekilecek bir adayın 35-45 arasında puan almış olması olasılığı nedir?
Öncelikle 35 ve 45 puanlarına karşılık gelen z değerlerini hesaplayalım:

$$z_{35} = \frac{X_i - \mu}{\sigma} = \frac{35 - 60}{12} = -\,2{,}08$$

$$z_{45} = \frac{X_i - \mu}{\sigma} = \frac{45 - 60}{12} = -\,1{,}25$$

35 ve 45 puanlarına karşılık gelen her iki z değeri de negatif işaretlidir zira, her iki değer de ortalamanın altındadır.
Şekil'den de görüleceği gibi, 35-45 aralığında puan almış alma olasılığını hesaplayabilmek için iki z değerine karşılık gelen alanları bulmak ve sonra farkını almak gerekecektir.



Şekil: P(-2,08 ≤ z ≤ -1,25)

45'e karşılık gelen z değeri olan 1,25 değerinin standart normal dağılım tablosundaki karşılığı 0,3944'dür ve bu olasılık herhangi bir adayın 45-60 arasında puan alma olasılığıdır. 35'e karşılık gelen z değeri ise 2,08 olup bu değerin standart normal dağılım tablosundaki olasılık değeri de 0,4812'dir ve bu olasılık 35-60 arasında puan alma olasılığını vermektedir.

O halde, Şekilden de görüleceği gibi, 35-45 arasında puan alanların olasılığını belirlemek için 35-60 puan alanından 45-60 puan alanını çıkarmak gerekecektir. Bu durumda, tesadüfen seçilecek bir adayın 35-45 arasında puan almış olma olasılığı, 0,4812 – 0,3944 = 0,0868 biçiminde hesaplanmaktadır.

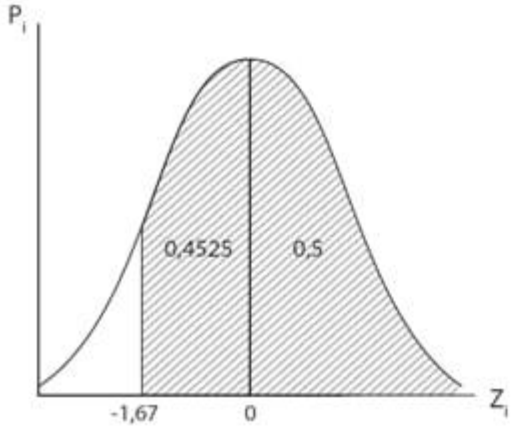
Bu değer aynı zamanda, test sınavına giren personel adaylarının % 8,68'inin 35-45 puan aralığında puan aldığının da göstergesidir.

8) Tesadüfen seçilecek bir adayın 40'dan fazla puan almış olma olasılığı nedir?

Önce, 40 puana karşılık gelen standart değeri hesaplayalım:

$$z_{40} = \frac{X_i - \mu}{\sigma} = \frac{40 - 60}{12} = -1{,}67$$

Z=1,67 değerine karşılık gelen standart normal dağılım tablosundaki olasılık değeri 0,4525'dir ve bu değer, ilgili z değeri ile ortalamayı ifade eden sıfır değeri arasındaki alanı; yani puan ölçeğinde 40-60 arasındaki alanı vermektedir. Bu durum Şekilden de anlaşılmaktadır.



Şekil: P(z > -1,67)

Standart normal dağılım olasılıkları tablosunda, z=1,67 değerinin olasılığı, tablonun ilk sütununda 1,6; ilk satırında da 0,07 değerlerinin bulunması ve bu iki değerin bulunduğu satır ve sütunların birbiriyle kesiştiği noktada yer alan değeri tespit etmek suretiyle bulunur. Nitekim, 1,6 ile 0,07 değerlerinin kesiştiği noktada 0,4525 değeri bulunmaktadır.

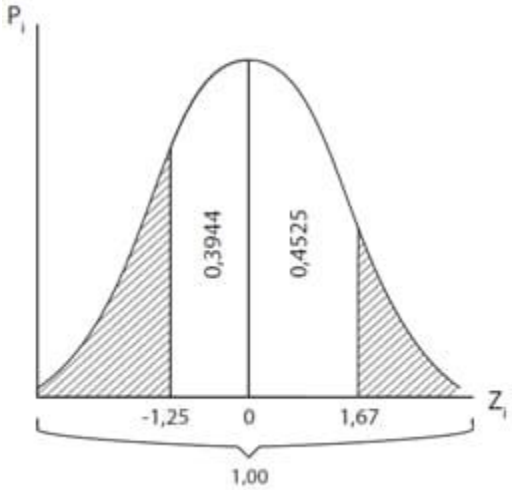
0,4525 olasılığı, ilgili z değeri ile ortalama arasındaki alanı verir. Yani, 40-60 arasında kalan alan 0,4525 olmaktadır. Oysa bizden istenen 40 puanın üstünde puan alanların olasılığıdır. Bu nedenle, ortalamanın eğri altındaki toplam alanı ikiye böldüğü ve her bir alanın 0,5 olduğu kuralını hatırlayarak 40 puanın üstünde alma olasılığını, 0,5 + 0,4525 = 0,9525 hesaplanır. Dolayısıyla, tesadüfen seçilecek bir adayın 40 puanın üstünde puan almış olma olasılığı % 95,25'dir veya başka bir deyişle, sınava girenlerin % 95,25'i 40 puandan fazla puan almıştır.

9) Tesadüfen seçilecek bir adayın 45 puandan az ya da 80 puandan fazla puan almış olma olasılığı nedir?

45 ve 60 puanlara karşılık gelen z değerlerini bularak, şeklimizi çizelim ve istenen olasılık bölgesini tarayalım.

$$z_{45} = \frac{X_i - \mu}{\sigma} = \frac{45 - 60}{12} = -1{,}25$$

$$z_{80} = \frac{X_i - \mu}{\sigma} = \frac{80 - 60}{12} = +1{,}67$$



Şekil: P(-1,25 < z ; z > 1,67)

Şekilden de görüleceği gibi, bizden istenen tesadüfen seçilecek bir adayın 45 puandan az ya da 80 puandan fazla almış olma olasılığıdır. Dolayısıyla, taralı bölgelerden de anlaşılacağı gibi eğrinin -1,25 değerinin solunda ve +1,67 değerinin sağında kalan alanlarının toplamı istenmektedir.

-1,25 standart değeri için standart normal olasılık değerleri tablosundan olasılığı 0,3944 ve +1,67 standart değeri için de olasılığı 0,4525 olarak belirliyoruz. Bu olasılık değerleri ilgili z değeri ile ortalamayı ifade eden sıfır değeri arasındaki alanları göstermektedir.

Bizden istenen kuyruklarda yer alan taralı bölgeler olduğuna göre, bu aşamada çözüm iki değişik yaklaşımla bulunabilir. Şöyle ki, Önce z değerlerine karşılık gelen olasılıklar toplanarak, bulunan değer toplam olasılık olan “1”den çıkarılır. Bu noktada eğrinin altında kalan toplam alanın “1” olduğunu hatırlayalım.

Bu çözüm yöntemi benimsenirse istenen olasılık,

P(X < 45 veya X> 80) = 1 - (0,3944 + 0,4525)

P(X < 45 veya X> 80) = 1 - 0,8469

P(X < 45 veya X> 80) = 0,1531

olarak hesaplanır.

İkinci bir yaklaşımda ise istenen olasılık, z değerlerine göre standart normal dağılım tablosundan elde edilen olasılıkları, ortalamanın sağında ve solunda kalan alanların 0,5 olasılığa sahip olduğunu dikkate alıp ayrı ayrı 0,5'den çıkararak kuyruk bölgelerinin alanını ayrı ayrı bulmak ve daha sonra çıkan sonuçları toplamak suretiyle de hesaplanabilir.

Şimdi önce, tesadüfen seçilecek bir adayın 40 puandan az almış olma olasılığını hesaplayalım. Ortalamanın solunda kalan alan 0,5 olduğuna göre, istenen kuyruk bölgesinin olasılığı, 0,5 – 0,3944 = 0,1056 olarak hesaplanır.

Tesadüfen seçilecek bir adayın 80 puandan yüksek puan almış olma olasılığı ise, benzer mantıkla, 0,5 – 0, 4525 = 0,0475 olarak hesaplanmaktadır.

Son aşamada da, tesadüfen seçilecek bir adayın 45'den az ya da 80'den çok puan almış olma olasılığını bulmak için yukarıda hesapladığımız iki kuyruk bölgesi olasılığını toplamamız gerekecektir. Dolayısıyla, 0,1056 + 0, 0475 = 0,1531 sonucuna ulaşılmaktadır. Bu sonuca göre, sınava giren adaylar arasından tesadüfen seçilecek herhangi birinin 45 puandan düşük ya da 80 puandan yüksek puan almış olma olasılığı % 15,31'dir. Başka bir deyişle, sınava giren adayların % 15,31'i 45 puandan az ya da 80 puandan yüksek puan almıştır.

Yukarıdaki örnekte, normal dağılımla ilgili olarak karşılaşılabilecek tüm olasılık hesabı versiyonlarına yönelik çözümlere yer verilmiştir.

Örnek:

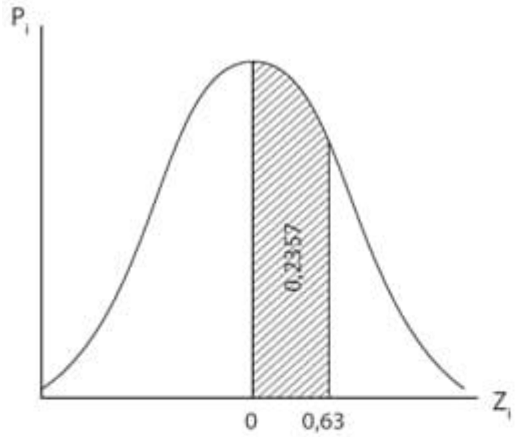
Bir margarin fabrikasında üretilmekte olan margarinlerin ağırlıklarının ortalaması 250 gr (μ= 250 gr) ve standart sapması 8 gr (σ =8 gr) olarak normal dağıldığı bilinmektedir. Günlük üretim içinden tesadüfen seçilecek bir margarin paketinin,

1. 250-255 gr arasında bulunma olasılığını hesaplayınız.
2. 255 gr'dan fazla olması olasılığını hesaplayınız.
3. 255 gr'dan az olması olasılığını hesaplayınız.

1) Tesadüfen seçilecek bir margarinin 250-255 gr arasında bulunma olasılığını hesaplayabilmek için öncelikle, 255 değerine karşılık gelen standart z değerini hesaplamalıyız.

$$z_{255} = \frac{X_i - \mu}{\sigma} = \frac{255 - 250}{8} = +0,625 \cong 0,63$$

Standart normal eğri alanları tablosundan, z = 0,63 değerinin alanını (ilk sütunda 0,6 ve ilk satırda 0,03 değerlerini bulup, satır ve sütunlarının kesişim noktasına bakarak) 0,2357 olarak belirliyoruz. Bu alan ortalama ile ilgili değer arasındaki alanı verdiğinden, tesadüfen seçilecek bir margarinin 250-255 gr arasında bulunma olasılığını % 23,57 olarak hesaplamış olunur.

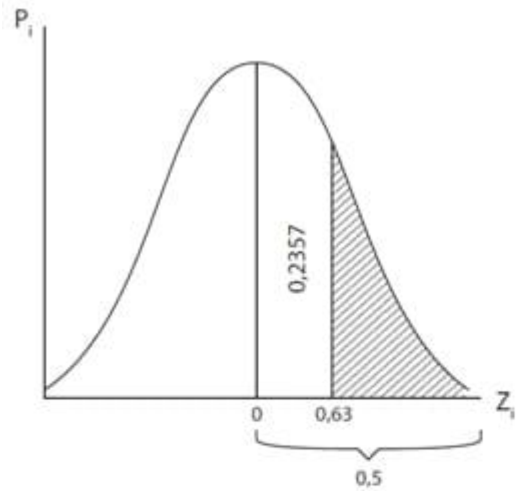


Şekil: P (0 ≤ z ≤ 0,63)

2) Tesadüfen seçilecek bir margarinin 255 gr'dan fazla olma olasılığını bulmak için de, ortalamanın sağında kalan bütün alan 0,5 olduğundan istenen olasılığı,

0,5 – 0, 2357 = 0,2643

hesaplıyoruz. Tesadüfen seçilecek bir margarinin 255 gr'dan fazla olması olasılığı % 26,43'tür. Başka bir deyişle, bu fabrikada üretilmekte olan margarinlerin % 26,43'ü 255 gramın üzerinde ağırlığa sahiptir.

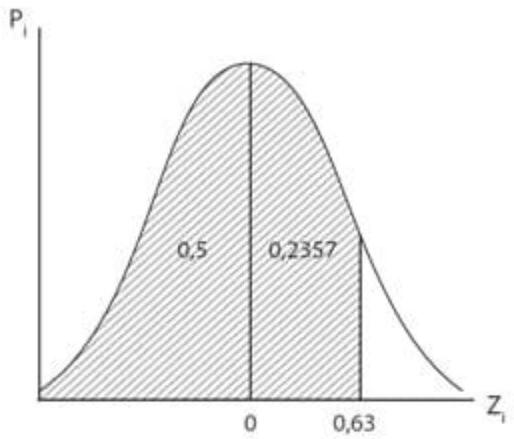


Şekil: P (z > 0,63)

3) Tesadüfen seçilecek bir margarinin 255 gramdan az olması olasılığını hesaplamak için de ortalamanın solunda kalan alanın 0,5 olduğunu hatırlayarak, bu değeri 250-255 gr aralığının olasılığı olan 0,2357 değeri ile toplamamız gerekecektir:

0,5 + 0,2357 = 0,7357

Dolayısıyla, tesadüfen seçilecek bir margarinin 255 gramdan az ağırlığa sahip olması olasılığı % 73,57 olmaktadır. Başka bir deyişle bu fabrikada üretilen margarinlerin % 73,57’si 255 gramdan az ağırlığa sahiptir.



Şekil: P (z < 0,63)

Normal Dağılım Olasılıklarını Kullanarak Beklenen Değer Hesabı

Normal dağılıma sahip tesadüfi değişkenlerin belirli bir aralıkta değer alma olasılıklarını hesaplamayı öğrendik. Bunun yanında, kitlenin içinde kaç birimin ilgili aralıkta değer alacağını da belirleyebiliriz. Başka bir deyişle, hesapladığımız olasılık değeri üzerinden, ilgili aralıkta yer alacak birimlerin frekansının ne olacağını bulabiliriz.

Olasılık bölümü içinde öğrendiğimiz beklenen değer kavramını hatırlayalım. Beklenen değer, gözlem sayısı ile olasılığın çarpımıdır:

E(X) = n.p

Dolayısıyla, normal dağılım çözümlemelerinde de hesaplanan olasılık toplam frekans ile çarpılarak ilgili olasılık alanına düşen gözlem sayısı ya da birim sayısı kolaylıkla hesaplanabilir

Şimdi yukarıda çözümünü yaptığımız margarin fabrikası örneğinde bir günde toplam 1000 adet margarin üretimi yapıldığını varsayalım.

Tesadüfen seçilecek bir margarinin 250-255 gram arasında olması olasılığını 0,2357 olarak hesaplamıştık. Bu durumda, günlük üretim içinde 250-255 gram aralığında ağırlığa sahip margarin sayısı da,

E(X) = n.p

E(X) = 1000 . 0,2357 = 235,7≅ 236 olarak hesaplanacaktır.

Benzer şekilde, kaç adet margarinin 255 gramdan fazla olacağını da belirleyebiliriz. Bunun için, tesadüfen seçilecek bir margarinin 255 gramdan fazla olması olasılığı olan 0,2643 değerini toplam frekans ile çarpmamız yeterli olacaktır.

E(X) = n.p

E(X) = 1000 . 0,2643 = 264,3≅ 264

Dolayısıyla bir gün içinde üretilen 1000 adet margarinin 264 tanesinin 255 gramdan daha fazla ağırlığa sahip olmasını bekleyebiliriz.

Örnek:

Bir konserve fabrikasında üretilmekte olan konservelerin ağırlığı normal dağılıyor. Ortalama ağırlık 500 gr ve standart sapma 10 gr olduğuna göre tesadüfen seçilecek bir konserve kutusunun,

1. 525 gramdan fazla olması olasılığı nedir?
2. Günde 5000 adet konserve üretilmesi durumunda kaç konservenin 525 gramdan fazla olması beklenir?

1) Tesadüfen seçilecek bir konserve kutusunun 525 gramdan fazla olması olasılığını hesaplayabilmek için öncelikle 525 değerine karşılık gelen z değerini hesaplıyoruz.

$$z_{525} = \frac{X_i - \mu}{\sigma} = \frac{525 - 500}{10} = +2{,}5$$

Standart normal eğri alanları tablosunda z=2,5 standart değerine karşılık gelen alan, tablonun ilk sütununda bulunan 2,5 değeri ile ilk satırdaki 0,00 değerlerinin kesişim noktasında yer alan 0,4938 olasılığını vermektedir.

Bu durumda, tesadüfen seçilecek bir kutunun 525 gramdan fazla olması olasılığı,
0,5 – 0, 4938 = 0,0062 olarak hesaplanmaktadır.

Tesadüfen seçilecek bir konservenin 525 gramdan fazla olması olasılığı % 0,62 olmaktadır.
b-Günde 5000 konserve üretilmesi halinde,
E(X) = n.p = 5000 . 0,0062
E(X) = 31 adet konservenin 525 gramdan fazla olması beklenebilir.

5.2.5. Güven Aralığı

Ana kütle Ortalamasının Güven aralığı: Ana kütlenin tüm birimlerini incelemek olanaksızdır. Bu gibi durumda Ana kütlenin ortalaması hesaplanamaz. Ancak örneklemler yardımıyla ana kütlenin ortalaması ya da ana kütlenin ortalamasının içinde bulunduğu sınırlar, seçilen yanılma olasılığında tahmin edilir.

Kitle ortalamasının içinde bulunduğu sınırlara “Ana kütlenin Ortalamasının Güven Aralığı ya da Güven Sınırları” adı verilir.
Normal dağılımdan alınan n birimli bir örneklemin ortalamasının ana kütlenin ortalamasına yani μ’ye eşit olması beklenemez. Bu durumda μ için güven aralığı tahmin edilebilir.

Ana kütlenin varyansı σ^2’nin bilindiğinde Ana kütlenin ortalamasının Aralık Tahmini:
Örneklem ortalaması $\overline{X}$’nin dağılımı $\overline{X}$ ~ N(μ ; σ^2 / n) ‘dir. Buna göre aralık tahminde kullanılacak olan istatistiğin dağılımı standart normal dağılımdır ve aşağıdaki gibi hesaplanır;

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

Normal dağılımın simetri özelliğinden güven aralığının boyu (1-α), ortalamanın iki yanına eşit dağıtıldığında, başka bir ifadeyle α’nın α/2 ve α/2 şeklinde iki tarafa bölündüğü durumda minimum olacaktır. Dolayısıyla Güven Aralığı;

$$P(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = (1 - \alpha)$$

Ana kütlenin varyansı σ^2’nin bilinmediğinde ana kütlenin ortalamasının Aralık Tahmini:

Ana kütlenin varyansı σ^2 bilinmediğinde Ana kütlenin ortalamasının aralık tahmininde

$$Z = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$

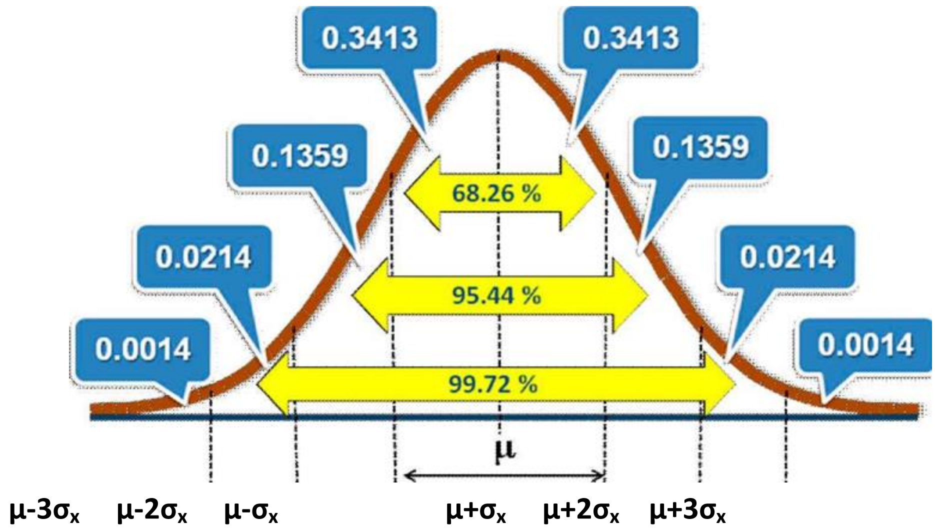
istatistiği Z dağılımına sahiptir.

Örnek Ortalamalarının Dağılımı

Toplam birim sayısı N, varyansı σ^2, aritmetik ortalaması μ ve dağılımı normal olan bir ana kütleyi $N(\mu, \sigma^2)$ şeklinde gösterelim. Örneklerin basit tesadüfî örnekleme yöntemine göre alındıklarını varsayarak, örnekler için örnekleme dağılımlarını görelim: Bir yığından n büyüklüğünde çekilmesi mümkün tüm örneklerin çekilsin ve her biri için aritmetik ortalamanın hesaplansın; C_N^n sayısı kadar olan bu örnek ortalamalarının dağılımı, teorik bir dağılımdır. Söz konusu dağılım, örnek ortalamalarından oluştuğu için buna "örnek ortalamalarının örnekleme dağılımı" adı verilir.

Örnek ortalamaları ana kütle ortalaması etrafında normal bir dağılım gösterirler. Standart hata ise örnek ortalamaları normal dağılımının standart sapmasından başka bir şey değildir. Bu normal dağılımın ortalaması ile standart sapması arasındaki ilişkiden hareketle herhangi bir örnek ortalamasının belirli olasılık kademelerine göre bulunabileceği sınırlar tahmin edilebilir.

Şekil 1 Normal dağılan örnek ortalamalarının çeşitli standart hata sınırları



Şekil 1 de görüleceği gibi, örnek ortalamalarının % 99.7 si ana kütle ortalamasından ± 3 $\boldsymbol{\sigma_x}$ standart hata sınırları arasında bulunur. Bunun gibi diğer olasılık kademeleri için de benzer açıklamalar yapılabilir. Buna göre, ana kütle ortalaması μ ve örnek ortalamaları standart hatası $\boldsymbol{\sigma_x}$ şeklinde gösterilecek olursa örnek ortalamaları şu şekilde hesaplanır:

% 68.3' ü	**$\mu - \sigma_x$**	**ile**	**$\mu + \sigma_x$**	Sınırları arasında
% 95' i	**$\mu - 1.96\ \sigma_x$**	**ile**	**$\mu + 1.96\ \sigma_x$**	Sınırları arasında
% 95.5' i	**$\mu - 2\ \sigma_x$**	**ile**	**$\mu + 2\ \sigma_x$**	Sınırları arasında
%99' u	**$\mu - 2.58\ \sigma_x$**	**ile**	**$\mu. + 2.58\ \sigma_x$**	Sınırları arasında
% 99.7' si	**$\mu - 3\ \sigma_x$**	**ile**	**$\mu + 3\ \sigma_x$**	Sınırları arasında

Dağılımı normal ve varyansı belli olan yığından çekilen örnek ortalamaları ana kütle ortalaması etrafında normal dağılım gösterir. Bu dağılımın ortalaması (örnek ortalamalarının ortalaması) ana kütle ortalamasına eşittir ve $\overline{X}$ = µ şeklinde gösterilir.
Örnek ortalamaları dağılımının standart sapması, standart hata olarak bilinir ve aşağıdaki formülle hesaplanır:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

Formüldeki $\sqrt{\frac{N-n}{N-1}}$ ifadesi, düzeltme faktörü olup, $\frac{n}{N} < 0.05$ ise, sonucu etkilemeyeceği için, standart hata hesabında ihmal edilebilir. Böyle durumlarda standart hata aşağıdaki gibi belirlenmektedir:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

σ bilinmiyorsa $\boldsymbol{\sigma_x = \frac{s}{\sqrt{n}}}$ olarak yazılır.

Ana kütle ortalaması bilinmiyorsa örneklemden yola çıkarak belli bir güven aralığında tahmin edilebilir.

Örnek.

Ürün ağırlıkları ile ilgili değişkenliğin σ = 5 gr. olduğu bilinmektedir. Örnekleme oranı %1 olacak şekilde alınan 100 birimlik (n) örneğin ortalaması 100 gr. ($\overline{X}$ bulunduğuna göre, ana kütledeki ürünlerin ortalama ağırlığını belirli olasılık kademelerine (güven aralığına) göre tahmin edelim:. Buna göre örneklem standart hata;

$$\boldsymbol{\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{100}} = 0.5\ gr}$$

Örnek ortalamaları ana kütle ortalaması etrafında normal dağıldığına göre; normal dağılımın özelliklerinden yararlanarak, ana kütle birimlerin ortalama ağırlığı,

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

şeklinde tahmin edilebilir. Formüldeki "Z" ifadesi belirli olasılık kademelerindeki standart normal dağılımın kritik değerleridir. Çeşitli olasılık kademeleri için yığın ortalamasının tahmini ise aşağıda gösterildiği gibi hesaplanır:

% 68 olasılıkla

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$100 - (1)(0.5) \le \mu \le 100 + (1)(0.5)$$

$$99.5gr. \le \mu \le 100.5gr.$$

% 95 olasılıkla

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$100 - (1.96)(0.5) \le \mu \le 100 + (1.96)(0.5)$$

$$99.02gr. \le \mu \le 100.98gr.$$

% 99 olasılıkla

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$100 - (2.58)(0.5) \le \mu \le 100 + (2.58)(0.5)$$

$$98.71gr. \le \mu \le 101.29gr.$$

Örnek:

Bir üniversitedeki ortalama bir erkek öğrencinin ağırlığı 80 kilogram. Burada belirli bir güven aralığı içerisinde üniversitedeki erkek öğrencilerin ağırlığının ne kadar isabetli tahmin edebileceği test edilecek.

Bu, hipotezi test etmek için veri toplama kullanılacak. Rastgele 1.000 erkek öğrenci seçildiğini varsayalım.

$$\underbrace{\bar{x}}_{\text{mean}} = \frac{\sum x \text{ — data values}}{n \text{ — sample size}}$$

$$\underbrace{\sigma}_{\text{standard deviation}} = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

Örneklem ortalamasının ve örneklem standart sapmasının hesaplanması:

Seçilen kitle parametresini tahmin etmek için kullanılmak istenen bir örnek istatistik (ör., örneklem ortalama, örneklem standart sapma) seçilir. Bir kitle parametresi, belirli bir kitle karakteristiğini temsil eden bir değerdir.

Örneklem ortalama ve örneklem standart sapması bulunabilir:

Verinin örneklem ortalamasını hesaplamak için seçilen 1.000 erkeğin ağırlıklarını toplanır ve sonucu erkeklerin sayısı olan 1.000'e bölünür. Böylece 80 kilogram olan ortalama ağırlığı elde edilmelidir.

Örneklem standart sapmasını hesaplamak için verinin ortalamasının bulunması gerekir. Sonra, verinin varyansını veya ortalamadan farklarının karesinin ortalamasının bulunması gerekir. Bu sayı bulunduktan sonra onun karekökünü alınır. Buradaki standart sapmanın 13,6 kilo olduğunu varsayılır.

İstenen güven düzeyini seçilir.

En yaygın güven düzeyleri yüzde 90, yüzde 95 ve yüzde 99'dur. Bu, ayrıca problem sırasında da verilebilir. %95'i seçildiğini varsayalım.

Hata payı hesaplanır.

Hata payı aşağıdaki formül kullanılarak bulunabilir:

$\mathbf{Z_{a/2} * \sigma/\surd(n).}$

$Z_{a/2}$ = güven katsayısı. Burada a = güven düzeyi, σ = standard sapma ve n = örneklem büyüklüğüdür.

Kritik değeri veya $Z_{a/2}$'yi bulmak için:

Güven düzeyi %95'dir. Yüzdelik ondalığa çevirilirse 0,95 olur ve bu 2'ye bölünürse 0,475 elde edilir. Ardından, z tablosuna bakarak 0,475 değerine karşılık gelen değer bulunur. En yakın değerin 1,9 satırı ve 0,06 sütunu kesişimindeki 1,96 olduğu görülecektir.

Standart hatayı bulmak için standart sapma 13,6 alınır ve bu örneklem boyutu olan 1.000'in kareköküne böl. Sonuç 13,6/31,6 veya 0,43 kg olur. 1,96 ile 0,43 (kritik değer bölü standart hata) çarpılarak 0,84 hata payı elde edilir.

Güven aralığını belirtilir.

Güven aralığını bulmak için,

$\mathbf{\bar{x} \pm Z_{a/2} * \sigma/\surd(n).}$

Burada $\bar{x}$ ortalamayı temsil eder.

Güven aralığını belirtmek için ortalama alınır (80) ve yanına ± ve hata payı yazılır.

Cevap: 80 ± 0.84.

Alt sınır: 80 - 0.84 = 79.16

Üst sınır: 80 + 0.84= 80.84'tür.

Örnek 2.

64 kişi üzerinde yapılan bir araştırmada Türkiye'de insanların ortalama günde 8 dakika televizyon seyrettikleri ve ana kütlenin standart sapmasının 4 dakika olduğu bilindiğine göre % 95 güven aralığında ana kütle ortalamasını tahmin ediniz.

İlk olarak örneklem standart sapması bulunur.

$$\boldsymbol{\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{64}} = 0.5\ dakika}$$

% 95 olasılıkla $$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$4 - (1.96)(0.5) \leq \mu \leq 4 + (1.96)(0.5)$$

$$3.02 dakika \leq \mu \leq 4.98 dakika$$

Yukarıdaki sonucu % 95 güven düzeyinde Türkiye'de insanlar günde 3.02 ile 4.98 dakika arasında televizyon seyretmektedir şeklinde yorumlayabiliriz.

Örnek 3.

Sakarya Üniversitesinde 49 öğrenci üzerinde yapılan çalışmada IQ ortalaması 110 ve standart sapması 14 bulunmuştur. % 99 güven aralığında SAÜ öğrencilerin IQ ortalaması hangi aralıkta olacaktır.

% 99 olasılıkla $$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$110 - (2.58)(2) \leq \mu \leq 110 + (2.58)(2)$$

$$104.76 \leq \mu \leq 115.16$$

Yukarıdaki sonucu % 99 güven düzeyinde SAÜ öğrencilerinin IQ 104.76 ile 115.16 arasındadır şeklinde yorumlayabiliriz.

5.2.6. Hipotez Testi

Hipotez, bir durum hakkında ileri sürülen bir varsayımdır. İstatistik hipotez, ana kütlenin durumu hakkında ileri sürülen bir varsayımdır. Burada bu hipotez daima yapılacak bir istatistik test sonucuna göre kabul veya reddedilecek şekilde formüle edilir. İstatistik test ise örnek istatistiklerini kullanarak bir hipotezin geçerli olup olmadığını ortaya koyma işlemidir.

Hipotez testinde amaç, karar verme sürecinin belirlenmesidir. Bunun için önce hipotezlerin ortaya konması gerekir. İstatistikte H_0'a sıfır hipotezi, H_1'e de alternatif veya araştırma hipotezi denir.

Yığındaki değerlerin farklı ve örneklem seçiminin tesadüfi sebepleri ileri gelecek kadar küçükse sıfır hipotezini kabul etmek, tesadüfi sebeplerden ileri gelmeyecek kadar büyükse reddetmek gerekir.

Önce, hipotezler kurulur. Hipotezler, sıfır hipotezi ve araştırma (alternatif) hipotezi olmak üzere iki tanedir. Sıfır hipotezi, yığın parametresinin bilinen veya belirlenmiş değerini gösterir. Alternatif hipotez ise, araştırmayı yönlendiren yani kanıtlanmak istenen asıl hipotezdir.

Hipotezler, biri red edildiğinde diğeri kabul edilecek şekilde düzenlenir. Sıfır hipotezinde, ana kütle parametresinin belirli bir değere eşit olduğu ifade edilir. Alternatifinde ise kanıtlanacak duruma göre ana kütle parametresinin belirli bir değerden büyük, küçük ya da farklı olduğu ileri sürülür.

Örneğin,
"Ürünlerin ortalama ağırlığı 100 gr. dan farklıdır" şeklindeki araştırma hipotezi sıfır hipotezi ile birlikte şu şekilde kurulabilir:
$H_0 : \mu = 100$ gr.
$H_1 : \mu \neq 100$ gr.

"Ürünlerin ortalama ağırlığı 100 gr. dan hafiftir" şeklindeki bir hipotez şu şekilde kurulabilir:
$H_0 : \mu = 100$ gr.
$H_1 : \mu < 100$ gr.

"Ürünlerin ortalama ağırlığı 100 gr. dan ağırdır" şeklindeki hipotez ise şu şekilde gösterilir:
$H_0 : \mu = 100$ gr.
$H_1 : \mu > 100$ gr.

Yukarıda verilen bilgiler kullanılarak bir karar vermek, hipotez testlerinin ikinci aşamasıdır. Karar verme aşamasında sıfır hipotezi ya kabul edilir ya da karşı hipotezin lehine sıfır hipotezi red edilir. Hipotez testinde sıfır hipotezinin reddi ya da kabulü örneklem bilgisine dayanır. Kitle için kabul edilen değer ile örneklemden elde edilen değer arasında ne kadar çok fark var ise, sıfır hipotezinin reddi de o kadar kuvvetlidir. Ancak, örneklem kitleden rasgele çekilmiş olduğu için örneklemden bulunacak değerin kitle değeri ile aynı olması beklenemez. Güven aralıkları konusunda görüldüğü gibi aynı kitleden çekilen örneklemlerin kitle bilgisini içermeme olasılığı α kadar olacaktır. Bu durumda sıfır hipotezi için karar sürecinde aşağıda verilen tabloda belirtilen sonuçlarla karşılaşılır.

Sıfır Hipotezi

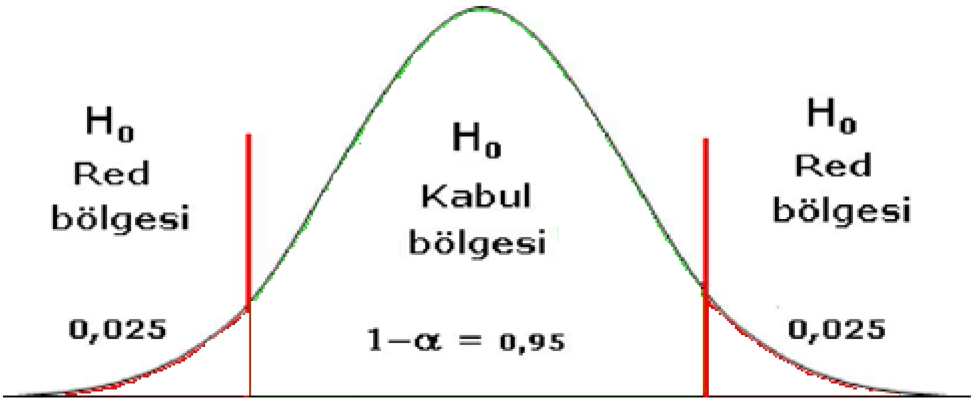
Karar	**Doğru**	**Yanlış**
Red	Hatalı Karar I. Tür Hata=α	Doğru Karar
Kabul	Doğru Karar	Hatalı Karar II. Tür Hata=β

Tablodan görüldüğü gibi, sıfır hipotezi Doğru ve biz bu hipotezi red etmiş isek, yanlış bir karar vermişizdir. Bu hatalı karar I Tür hata adını alır. I Tür hata yapma olasılığı α olacaktır. α'ya hipotezin **anlamlılık düzeyi** denir. Karar ya kabul ya da red edileceğine göre, kararı kabul etme olasılığı da (1-α) olur. Bu tabloya göre, yapılacak ikinci hata II Tür hata olarak adlandırılır. Belli bir karar kuralı için, sıfır hipotezi yanlış iken kabul etme olasılığı β ile gösterilir. Yanlış bir hipotezi red etme olasılığı da (1-β) dır. (1-β)' ya **testin gücü** denir.

Güven Sınırlarının Belirlenmesi

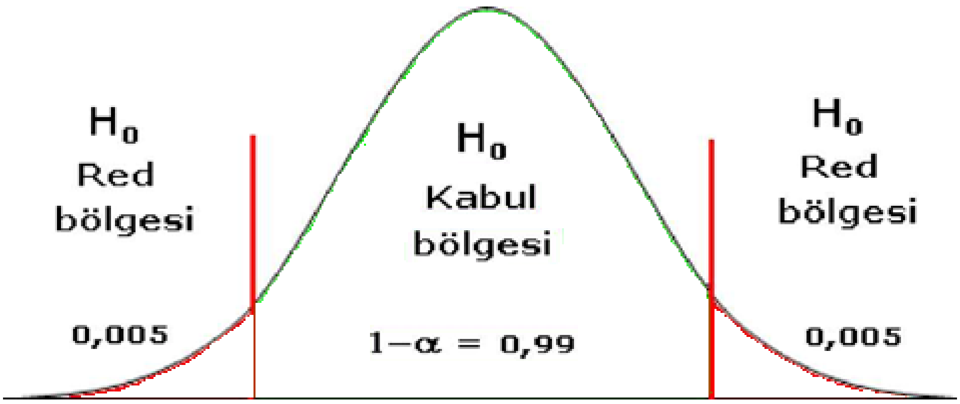
Anlam düzeyi, önem seviyesi şeklinde de ifade edilebilir. Örnek ortalamaları yığın ortalaması etrafında normal dağılış gösterdiğinden, örnek ortalamalarının %95' i $Z = \pm 1.96$ sınırları arasında, %99' u ise $Z = \pm 2.58$ sınırları arasında kalıyor demektir. Güven sınırları dışında kalan alanlar ise anlam düzeyi olarak bilinir ve bunlar sırasıyla % 10, % 5 ve % 1 dir.

Z_k' nin kritik değerleri önem düzeyine göre aşağıda verilmiştir.			
Anlam Düzeyi	**Sol Kuyruk Testi** **$H_1 < 100$**	**Sağ Kuyruk Testi** **$H_1 > 100$**	**Çift Yönlü Test** **$H_1 = 100$**
0.10	-1.28	+1.28	±1.65
0.05	-1.65	+1.65	±1.96
0.01	-2.33	+2.33	±2.58



-1.96 **0** **+1.96**

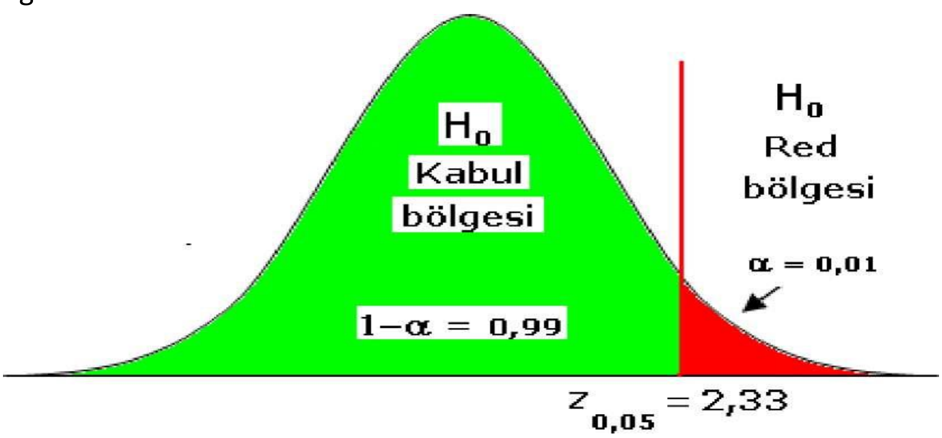
Şekil 1 Çift taraflı test, % 5 anlam düzeyine göre kabul ve red bölgeleri



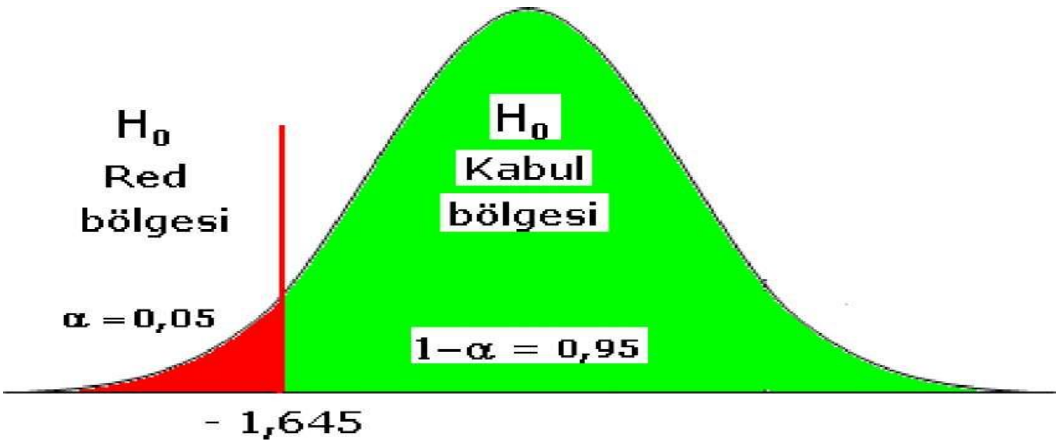
-2.58 **0** **+2.58**

Şekil 2 Çift taraflı test, % 1 anlam düzeyine göre kabul ve red bölgeleri

Güven sınırları, normal eğrinin her iki ucunda ve bu sınırların dışında kalan olasılıkların toplamı olduğundan, her bir uçtaki alan güven sınırının yarısı (%5 için %2.5, %1 için %0.5) kadardır. Tek taraflı testlerde güven sınırını belirleyen alan, testin yönüne bağlı ve normal eğrinin sadece bir ucundadır.



Şekil 3 Tek taraflı test, %1 anlam düzeyine göre red ve kabul bölgeleri



Şekil 4 Tek taraflı test, %5 anlam düzeyine göre red ve kabul bölgeleri

Uygulamada genellikle %5 veya %1 güven sınırlarından biri seçilir. Deney ya da uygulamanın özelliğine bağlı olarak %10, %0.5 veya %0.1 gibi farklı sınırlar da kullanılabilir.

1.1.1. Hipotez Testinin Sınanması

Bir hipotez kurulduktan sonra 2 aşamada test edilir.

1. **Aşama:** Örneklemden yola çıkılarak Z_h degeri hesaplanır.

$$\mathbf{Z_h} = \frac{\mathbf{X} - \mu}{\boldsymbol{\sigma}_{\bar{\mathbf{x}}}}$$

2. **Aşama:** Hesaplanan Z_h değeri Z_k değeri ile karşılaştırılır.

$Z_h > Z_k$ ise H_0 hipotezi reddedilir ve alternatif hipotez kabul edilir.

Hipotez testinde kullanacağımız yukarıda değindiğimiz bilgileri özetlersek;

Sıfır Hipotezi (H_0): Tersine yeterli kanıt bulununcaya kadar doğru kabul edilen fikirdir.

Alternatif Hipotez (H_1): Sıfır hipotezi karşısında test edilen, sıfır hipotezi red edildiğinde kabul edilen hipotez

Tek Yanlı Karşıt Hipotez: Ana kütlede ilgilendiğimiz parametre için sıfır hipotezince, belirlenen bir değerden küçük ya da büyük olanaklı bütün değerleri içeren karşıt hipotez

Çift Yanlı Karşıt Hipotez: Ana Kütlede ilgilendiğimiz parametre için sıfır hipotezince belirlenen değer dışında olanaklı bütün değerleri içeren karşıt hipotez

Hipotez Testi Kararı: Araştırmacıyı örneklem kanıtına dayanarak, sıfır hipotezini kabul ya da reddetmeye götürecek şekilde geliştirilmiş karar kuralı

I Tür Hata: Doğru hipotezin red edilmesi

II Tür Hata: Yanlış hipotezin kabul edilmesi

Anlamlılık Düzeyi: Doğru olan sıfır hipotezini reddetme olasılığı, α.

Testin Gücü: Yanlış bir sıfır hipotezinin reddedilme olasılığı, 1-α.

Örnek:

Bir alçı dolum makinesi μ_o=20 kg ortalama ağırlıklı alçı dolumu yaparken arıza yapar. Tamirci getirip tamir ettirir. Acaba yine μ_o=20kglık dolum yapabilecek midir? Deneme yapıp görmek gerekir.

40 torba basit örneklem yöntemine göre seçilip bu 40 alçı torbası ağırlıkları şöyle ölçülmüştür: X_1 = 19,8 kg, X_1 = 20,5 kg, X_2= 21,2 kg, X_3= 18,9 kg, ... , X_{40}= 20,8 kg

Örneklem ististikleri şöyle hesaplanmıştır:
n = 40 torba
Örneklem ortalaması : $\bar{x}$ = 21,4 kg
Örneklem standard sapması: σ = 3,2 kg
σ_x = 3.2/√40 =0.506kg
$\bar{x} \pm \sigma_x$ = 21,4±0,506 kg

Buradan sonra hipotez tesleri sürecine geçilir.

Hipotezler:
Ho: Elimizdeki örneklem anakütle ortalaması "Mo = 20 kg" olan bir anakütleden çekilmiş bir rassal örneklem olup, örneklem ortalaması X- değeri anakütle ortalamasına eşit olarak kabul edilebilir. Aradaki 1,4 kg lık fark ise tesadüfe bağlanabilecek, önemli olmayan, anlam taşımayan çok küçük bir farktır. Dolayısıyla X- = Mo yazabiliriz. Yani elimizdeki örneklemin ait olduğu anakütle ortalamasını M ile gösteririz.

H1: Bu örneklem "Mo = 20 kg" olan bir anakütleden çekilmiş bir rassal örneklem olamaz. Aradaki 1,4 kg lık fark tesadüfe bağlı değil, ayarlamanın yapılmamış olması nedeni ile gerçekleşmiştir. Bu kadarlık farkın tesadüfen ortaya çıkmış olması olasılığı çok küçüktür. Dolayısıyla dolum ayarı iyi olmadığı için istenenden daha hafif ya da daha ağır dolumlarla karşılaşmamız olasıdır. Bu örneklemin çekilmiş olduğu anakütle 20 kg olamaz. Örneklemimiz kendine ait başka bir anakütleden çekilmiş olmalıdır.

İstatistiksel anlamlılık düzeyinin belirlenmesi (Risk düzeyi, Yanılgı Payı, Hata payı)
α nın saptanması.
Hatasız bir test yapamayacağımız için her testte bir miktar yanılma riskimiz vardır. Bunu 0,05 ; 0,01 ; 0,005 ; 0,0001;... gibi bir düzey olarak benimseyebiliriz. Yanılma payımız küçüldükçe, teste olan güven düzeyimiz yükselir. O nedenle istatistikçiler olabildiğince az yanılma ile test yapmak isterler. Yine de α =0,05 ve α=0,01 düzeyleri en çok kullanılanlardır.
α=0,05 olsun. Testin güven düzeyi = 1 - α = 0,95 olur.

Örnekleme dağılımının belirlenmesi
Elimizdeki veriler tartma yoluyla elde edilmiş sürekli, nitelik, nicel bir değişkene aittir. Bu tip veriler genelde normal dağılım gösterirler. Yani örneklemimiz "normal dağılım" lı bir anakütleden çekilmiştir. Anakütle sonsuz büyüklüktedir. Seçim iadesiz seçimdir ve tamamen rassal bir süreçle yapılmıştır. Yani torbaların ağırlıkları birbirini etkilememiştir. n>30 olduğu için büyük bir örneklem ile çalışıyoruz. Aynı anakütleden n=40 birimli pek çok sayıda örneklem çekmiş olsak, bunların X- ortalama dağılımı bir normal dağılım olur. Bu ortalamaların ortalaması anakütle ortalamasını verir. "kg" biriminden kurtulmak için X- ortalama değerlerini standardize edersek, verilerimiz z değerlerine dönüşür ve dağılımımız bir standart normal dağılım olan *z dağılımı* na dönüşür.

Ret alanının belirlenmesi
Kritik değerin saptanması
Ret alanı demek; normal dağılım eğrisi altında seçtiğimiz güven alanı (H_o'ın kabul alanı) dışında kalan H_o'ın reddedilmesini sağlayan küçük alanlardır. Ret alanı çift yönlü olabilir. (eksi taraf, artı taraf) veya tek taraflı olabilir. (Yani ya sol tarafta ya da sağ tarafta) Bunun anlaşılması için H_1 hipotezine bakarız.

Test istatistiği
Elimizdeki örnekleme ait Z_h değeri örneklemin bir istatistiğidir. Bu istatistik yardımıyla hipotez testini sonuçlandıracağız. O nedenle, Z_h değerine Test İstatistiği adını veriyoruz.

$$Z_h = \frac{\bar{x} + \mu_0}{\sigma / \sqrt{n}}$$

Z_h= (21,4-20)/0,51 = 2,74

Karşılaştırma, sonuç ve yorum
Bir hipotez testinde; z_h **<** z_α ise; H_o **kabul edilir**. Bu elimizdeki X-in, M ye yakın kabul edilebilecek bir konumda (H_o'ın kabul alanında) bulunduğunu gösterir.
Eğer z_h **>** z_α ise; H_o **reddedilir**. Elimizdeki örneklemin, M_o ortalamalı bir anakütleden çekilmiş rassal bir örneklem olmayacağı çünkü böyle bir şeyin gerçekleşmesi olasılığının *çok küçük* (p<0,05 veya p<0,01) olduğu sonucuna ulaşılır.

Sonuç
z_h = 2,74 **>** $z_{0,05}$ = 1,96 --> *Ho RET*
Bu duruma göre: elimizdeki örneklemin ortalaması, ilgilendiğim anakütlenin ortalamasından çok uzağa düşen bir büyüklüktedir. O nedenle iki ortalama arasındaki farkı z değerine dönüştürdüğümde, bulduğum z_h = 2,74 değeri de $z_{0,05}$ = 1,96 nın ötesine düşmüştür. Yani %5'lik ret alanına düşmüştür. Bu durumda X- = M_o biçiminde ifade ettiğim ve oradan

$M=M_o$ düzeyine yükselttiğim H_o hipotezini kabul edemem. Demek ki, bu makine hatalı dolum yapmakta, ortalaması 20 kg olan dolumlar gerçekleştirememektedir. Aynı deneyi n=40 olan 100 örneklem ile tekrarlarsam, bunun 95inde gene aynı sonuçla karşılaşmayı beklerim. Belki yalnızca 5inde makinenin ayarı iyiymiş gibi hatalı bir sonuca ulaşabilirim.
Dolayısıyla; verdiğim kararın doğru olması olasılığı %95 iken hatalı olması olasılığı en fazla %5 tir.

Test sonucundaki değerlendirmeler ve yorum

1) $z_h<z_\alpha$ olduğunda, H_o hipotezini *kabul ediyoruz* ve;

- Bu iki örneklemin çekilmiş olduğu anakütle ortalamalarının birbirlerine *eşit* olduklarını,
- Bu iki anakütlenin *aynı anakütleden* çekilmiş birer rassal örneklem olduğunu,
- İki örneklem ortalaması arasında gözlediğimiz farkın bir olasılık eseri olarak ortaya çıkmış, istatistik bakımından anlamlı olmayan, önemli olmayan küçük bir fark olduğunu düşünürüz.

2) $z_h>z_\alpha$ olduğunda, H_o hipotezini *reddediyoruz* ve;

- Ho hipotezine ait olan düşüncemizin tersini kabul ediyoruz, yani H_1'i kabul ediyoruz.
- Bu büyüklükteki z_h değerinin *olasılığa bağlı olarak* ortaya çıkmış olması olasılığı (ihtimali) çok düşüktür. Bu olasılık (p değeri) seçtiğimiz α değerinden de *küçüktür*. Bu kadar küçük bir olasılıkla ortaya çıkan bu z değerini artık rastgeleliğe değil anakütlenin gerçekten farklı olmasına bağlarız.

Hipotez Testinin Sınanması Örnekleri

Örnek 1: Bir fırının ürettiği ekmek ortalama ağırlığı 500 gram olduğu iddia edilmektedir. Fırını denetleyen belediye yetkilileri 100 adet örneğin ortalama ağırlığını 490 gram ve standart hatasını 30 gram bulmuşlardır. %1 anlam düzeyinde (%99 güven aralığında) ekmeğin ortalama ağırlığı 500 gram kabul edilebilir mi test ediniz.

Hipotez testini kuralım.

H_0 : µ = 500 gr.

H_1 : µ ≠ 500 gr.

1. **Aşama:** Örneklemden yola çıkılarak Z_h degeri hesaplanır.

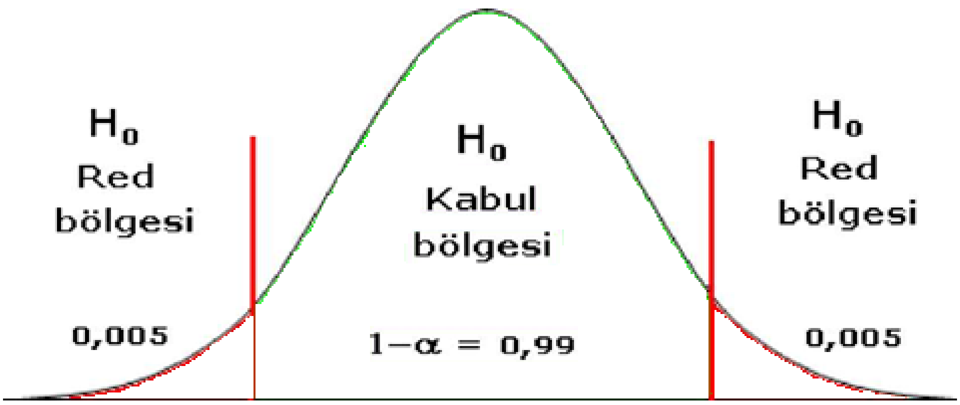
$$\mathbf{Z_h} = \frac{\mathbf{X} - \mu}{\sigma_{\bar{x}}} = \frac{\mathbf{490 - 500}}{\frac{\mathbf{30}}{\mathbf{10}}} = \mathbf{-3.3}$$

2. **Aşama:** Hesaplanan Z_h değeri Z_k değeri ile karşılaştırılır.

Z_h (-3.3) > Z_k (-2.58) oldugundan H_0 hipotezi reddedilir ve alternatif hipotez kabul edilir.

Yorum: Fırının ürettiği ekmeklerin ortalama ağırlığı 500 gramdan farklıdır.

Not: Z değerlerini mutlak değerler olarak karşılaştırıyoruz.



-3.3 -2.58 0 +2.58

Örnek 2: Ağrı kesici bir ilacın ortalama 60 dakikadan daha az bir sürede etkisini göstereceği olduğu iddia ediliyor. Rasgele seçilen hastalardan 64'üne ilgili ilaç veriliyor ve ortalama etki süresi 63 ve standart hatası 12 bulunuyor. α=0.05 anlamlılık düzeyinde (%95 güven aralığı) iddianın doğruluğunu test ediniz.

Hipotez testini kuralım:
H_0 : $\mu \leq 60$ dakika
H_1 : $\mu > 60$ dakika

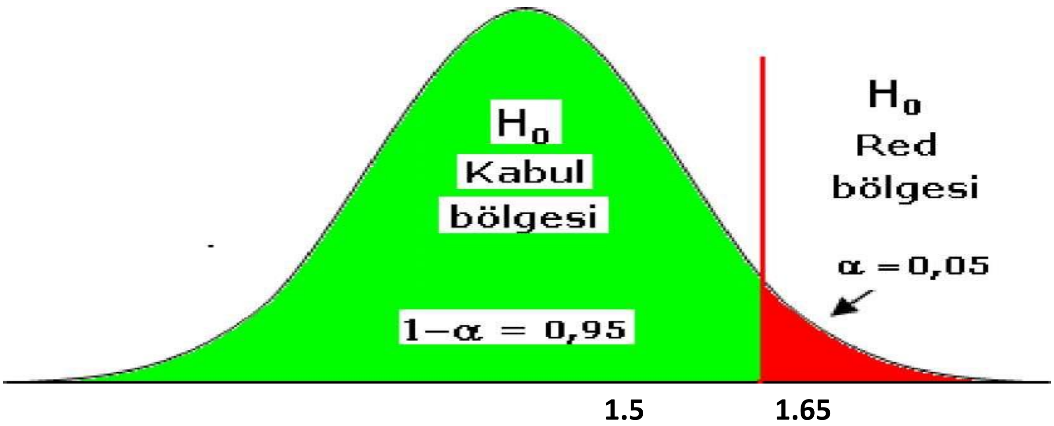
1. **Aşama:** Örneklemden yola çıkılarak Z_h değeri hesaplanır.

$$\mathbf{Z_h} = \frac{\mathbf{X - \mu}}{\boldsymbol{\sigma_{\bar{x}}}} = \frac{\mathbf{63 - 60}}{\frac{\mathbf{12}}{\mathbf{8}}} = \frac{\mathbf{3}}{\mathbf{2}} = \mathbf{1.5}$$

2. **Aşama:** Hesaplanan Z_h değeri Z_k değeri ile karşılaştırılır.

Z_h (1.5) > Z_k (1.65) olduğundan H_0 hipotezi reddedilemez.

Yorum: %95 güven düzeyinde ilaç ağrıyı en geç 60 dakika içinde geçirmektedir.



Örnek 3: Karatay diyetinin 3 ayda en az 10 kilo iddia edilmektedir. 144 kişi üzerinde 3 ay uygulanan diyetin ortalama 9 kilo zayıflattığı ve standart hatanın 4 kilo olduğu tesbit edilmiştir. 99 güven aralığında iddianın doğruluğunu test ediniz.

Hipotez testini kuralım.

$H_0 : \mu \geq 10$ kilo

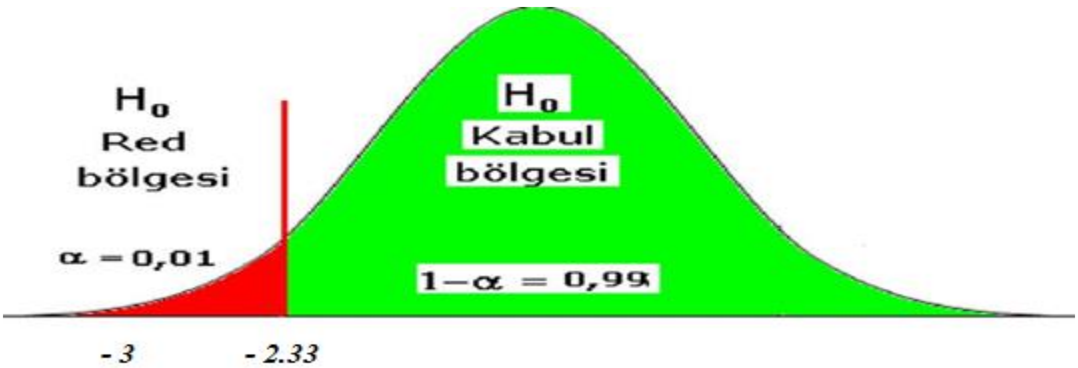
$H_0 : \mu < 10$ kilo

1. **Aşama:** Örneklemden yola çıkılarak Z_h değeri hesaplanır.

$$\mathbf{Z_h} = \frac{\mathbf{X} - \mu}{\sigma_{\bar{x}}} = \frac{\mathbf{9 - 10}}{\frac{\mathbf{4}}{\mathbf{12}}} = \frac{\mathbf{-1}}{\mathbf{0.33}} = \mathbf{-3}$$

2. **Aşama:** Hesaplanan Zh değeri Zk değeri ile karşılaştırılır.

Z_h (- 3) > Z_k (- 2.33) olduğundan H_0 hipotezi reddedilir.



Değerlendirme Soruları

1 – Bir ampul fabrikasının üretim sürecinde, 100 ampul rassal örneklem olarak seçilmiştir. Ampullerin ortalama ömrü 375 saat ve standart sapması s=15 saat olarak hesaplanmıştır. Üretilen ampullerin ortalama ömrünü % 95 güven düzeyiyle tahmin ediniz.

Çözüm:

n = 100 ampul

$\bar{X}$ = 375 saat

s = 15 saat

n > 30 olduğu için normal dağılıma sahiptir.

İlk olarak örneklem standart sapması bulunur.

$$\boldsymbol{s_x = \frac{s}{\sqrt{n}} = \frac{15}{\sqrt{100}} = 1.5\ saat}$$

% 95 olasılıkla

$$\bar{X} - Z_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$375 - (1.96)(1.5) \leq \mu \leq 375 + (1.96)(1.5)$$

$$372.06 \leq \mu \leq 377.94$$

Yukarıdaki sonucu % 95 güven düzeyinde ampullerin ömrü 372.06 ile 377.94 arasındadır.

2- SASKİ yöneticileri, Sakarya halkının ortalama aylık su tüketimini en az 20 litre olabileceğini düşünmektedir. Eğer yöneticilerin bu düşüncesi doğruysa Sakarya su sorunuyla karşı karşıya kalabilir. Bu amaçla, rassal olarak seçilen, 1000 kişiden veriler derlenmiş ve aylık ortalama su tüketiminin 22 litre ve standart sapmasının da 8 litre olduğu tespit edilmiştir. SASKİ yöneticilerinin kaygılanmalarının doğru olup olmadığını $\alpha = 0.01$ anlam düzeyini (% 99 güven düzeyi) kullanarak karar veriniz.

Çözüm:

Hipotez

$H_0 : \mu = 20$ litre

$H_1 : \mu > 20$ litre

$n = 1000$

$\bar{X} = 22$ litre

$\mu = 20$ litre

$s = 8$ litre

n > 30 olduğu için normal dağılıma sahiptir. Z değeri tablodan % 99 güven düzeyi için 2.33 bulunur. Bu bilgilerden sonra 2 aşama ile çözüme ulaşılır.

1. **Aşama:** Örneklemden yola çıkılarak Z_h degeri hesaplanır.

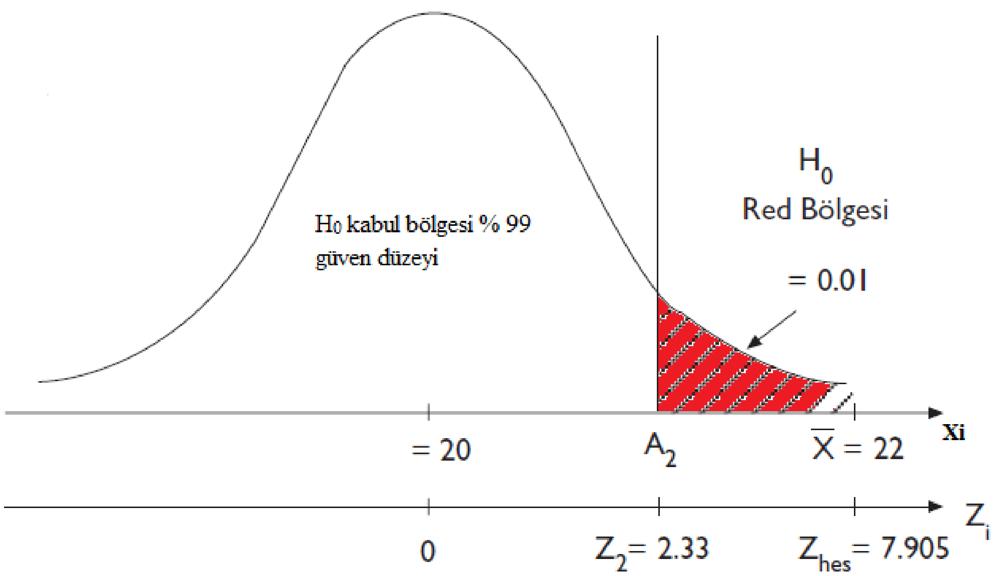
$$\mathbf{s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{8}{\sqrt{1000}} = 0.253}$$

$$\mathbf{Z_h = \frac{\bar{X} - \mu}{s_{\bar{x}}} = \frac{22 - 20}{0.253} = 7.9}$$

2. **Aşama:** Hesaplanan Z_h değeri Z_k değeri ile karşılaştırılır.

Z_h (7.9) > Z_k (2.33) oldugundan H_0 hipotezi reddedilir ve alternatif hipotez kabul edilir.

Yorum: Kişilerin aylık su tüketimi 20 litreden fazladır.



3- Bir çikolata firması 500 gramlık paketler halinde üretim yapmayı planlamaktadır. Üretimin planlandığı gibi gerçekleşip gerçekleşmediğini kontrol etmek için rassal olarak seçilen 100 paketin ortalama ağırlığı 495 gram ve standart sapma da 20 gram olarak bulunmuştur. Üretimin planlandığı gibi gerçekleşip gerçekleşmediğini α = 0.05 anlam düzeyini (% 95 güven düzeyi) kullanarak karar veriniz.

Çözüm:

Hipotez

H_0 : μ = 500 gram

H_1 : $\mu \neq$ 500 gram

n = 100

$\bar{X}$ = 495 gram

μ = 500 gram

s = 20 gram

n > 30 olduğu için normal dağılıma sahiptir. Z değeri tablodan % 95 güven düzeyi için 1.96 bulunur. Bu bilgilerden sonra 2 aşama ile çözüme ulaşılır.

1. **Aşama:** Örneklemden yola çıkılarak Z_h degeri hesaplanır.

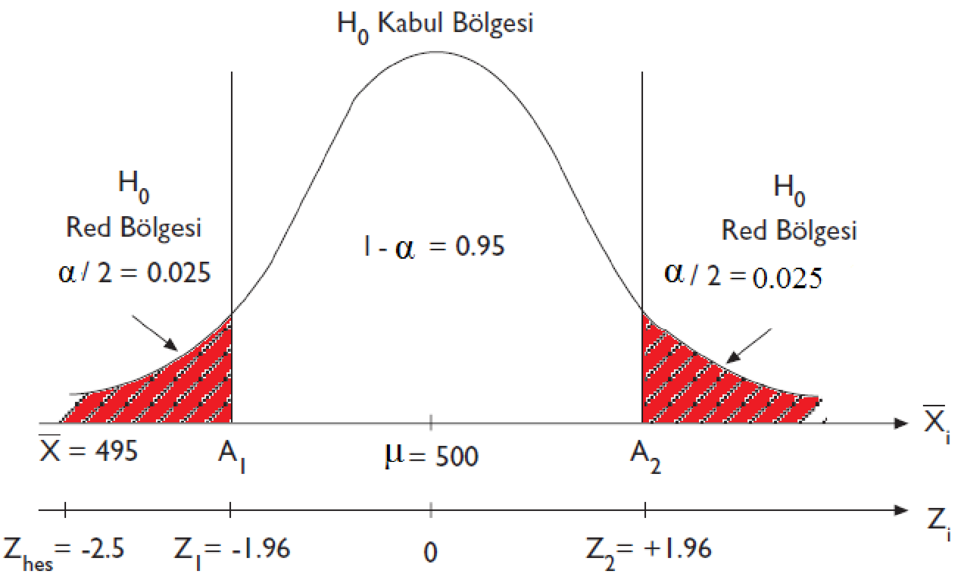
$$\mathbf{s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{20}{\sqrt{100}} = 2}$$

$$\mathbf{Z_h = \frac{\bar{X} - \mu}{s_{\bar{x}}} = \frac{495 - 500}{2} = -2.5}$$

2. **Aşama:** Hesaplanan Z_h değeri Z_k değeri ile karşılaştırılır.

Z_h (2.5) > Z_k (1.96) oldugundan H_0 hipotezi reddedilir ve alternatif hipotez kabul edilir.

Yorum: Paketlerin ağırlığı 500 gramdan farklıdır.



6. Eğri uydurma ve Regresyon

En küçük kareler yöntemi, birbirine bağlı olarak değişen iki fiziksel büyüklük arasındaki matematiksel bağlantıyı, mümkün olduğunca gerçeğe uygun bir denklem olarak yazmak için kullanılan, standart bir regresyon yöntemidir. Bir başka deyişle bu yöntem, ölçüm sonucu elde edilmiş veri noktalarına "mümkün olduğu kadar yakın" geçecek bir fonksiyon eğrisi bulmaya yarar. Gauss-Markov Teoremi'ne göre en küçük kareler yöntemi, regresyon için optimal yöntemdir.

The variance of $\{x_1, \ldots, x_N\}$, denoted by σ_x^2, is

$$\sigma_x^2 = \frac{1}{N}\sum_{n=1}^{N}(x_i - \overline{x})^2;$$

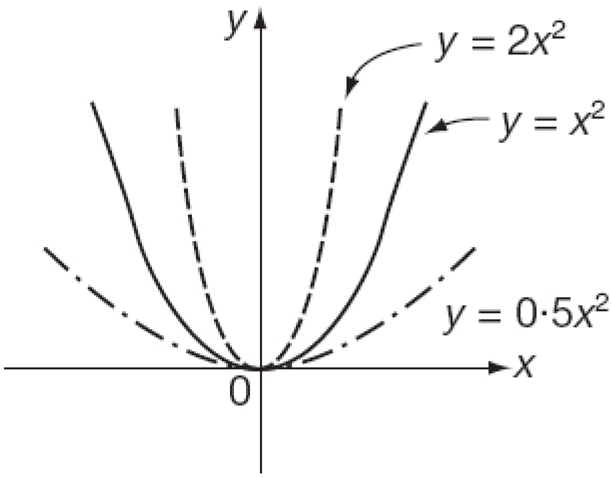
the standard deviation σ_x is the square root of the variance:

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_i - \overline{x})^2}.$$

Standart eğriler:

- *Düz çizgi*
- *İkinci dereceden eğriler*
- *Üçüncü dereceden eğriler*
- *Çember*
- *Elips*
- *Hiperbol*
- *Logaritmik eğriler*
- *Üstel eğriler*
- *Hiperbolik eğriler*
- *Trigonometrik eğriler*

Standard curves - *Second-degree curves*



The simplest second-degree curve is expressed by:

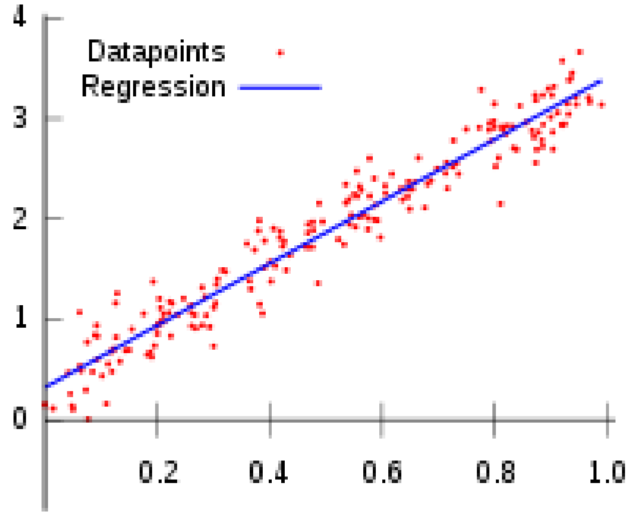
$$y = x^2$$

Its graph is a parabola, symmetrical about
The y-axis and existing only for $y \geq 0$. $y = ax^2$ gives a thinner parabola if $a > 1$ and a flatter parabola if $0 < a < 1$. The general second-degree curve is:

$$y = ax^2 + bx + c$$

where a, b and c determine the position,'width' and orientation of the parabola.

6.1. Doğrusal (Lineer) Regresyon



Kırmızı noktalar ölçümle elde edilmiş veri noktalarını, mavi çizgi ise en küçük kareler yöntemi ile bulunmuş teorik bağlantıyı ifade eder.

Çoğu zaman veri tablosuna tam olarak uyan bir fonksiyon bulmak mümkün olmaz; veri tablosuna en iyi uyan fonksiyon belirlenmeye çalışılır. Bir veri tablosuna en iyi uyan fonksiyonu bulma sürecine **regresyon analizi** denir. Regresyon analizi yaparken en çok kullanılan yöntemlerden biri en küçük kareler yöntemidir.

Belli ölçümler sonucunda i = 1, 2, . . . , n için (x_i , y_i) verileri elde edilmiş olsun. Burada, her bir y_i'nin x_i'ye bağlı olarak değiştiği varsayılmaktadır. Yapılan ölçümlerin doğası gereği, her i = 1, 2, . . . , n için $y_i = f(x_i)$ olacak biçimde bir fonksiyonun var olduğu, ölçümlerde yapılan hata nedeniyle bu eşitliklerin bazıları veya hepsinin sağlanmadığı kabul edilebilir. Bu düşünceyle, ölçülen y_i değeri $f(x_i)$ için yaklaşık değer kabul edilerek bu yaklaşımdaki hatanın minimum olduğu f fonksiyonu belirlenmeye çalışılır. Bu amacı gerçekleştirmek için f fonksiyonunun bir takım parametrelere bağlı bir ifadesi bulunduğu varsayılıp eldeki veriler yardımıyla bu parametreler belirlenmeye çalışılır. Örneğin, f fonksiyonu $y = f(x) = mx + b$ ifadesinde olduğu gibi bir doğrusal fonksiyon veya $y = f(x) = ax^2 + bx + c$ ifadesinde olduğu gibi bir karesel fonksiyon olabilir ki bu durumda belirlenmesi gereken parametereler a , b , c , m dir.

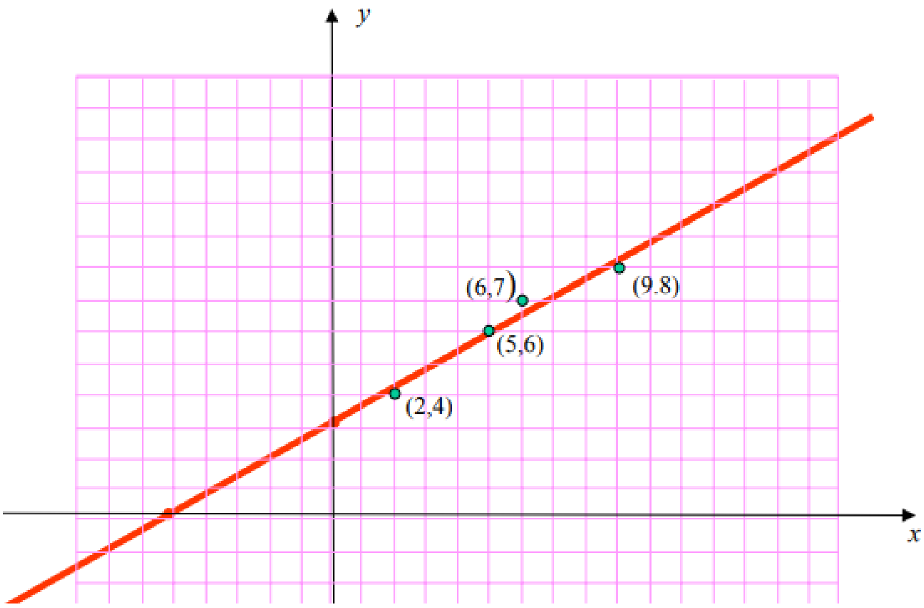
En küçük kareler yönteminde aranan fonksiyon, ya da onun parametreleri, tüm farkların kareleri toplamı olan

$$\sum_{i=1}^{n}(y_i - f(x_i))^2 = (y_1 - f(x_1))^2 + \cdots + (y_n - f(x_n))^2$$

ifadesini minimum yapacak şekilde belirlenir. Sözü edilen kareler toplamının minimum olması için her bir hatanın küçük olması gerekir.

Örnek:
Bir makineden üretilen ürünlerin tablosu şu şekilde tutulmuştur,

2. ayda 4 adet
5. ayda 6 adet
6. ayda 7 adet
9.ayda 8 adet



f(x) = mx + b

$$\sum_{i=1}^{n}(y_i - f(x_i))^2 = (y_1 - f(x_1))^2 + \cdots + (y_n - f(x_n))^2$$

2. ayda 4 adet , y1-f(x1)=4-2m-b
5. ayda 6 adet, y2-f(x2)=6-5m-b
6. ayda 7 adet, y3-f(x3)=7-6m+b
9.ayda 8 adet, y4-f(x4)=8-9m+b

$$F(m,b)=(4-2m-b)^2+(6-5m-b)^2+(7-6m-b)^2+(8-9m-b)^2.$$

Eğimlerini bulabilmek için kısmi türevleri alınır ve sıfıra eşitlenir.

$$F_m(m,b)=2(4-2m-b)(-2)+2(6-5m-b)(-5)+2(7-6m-b)(-6)+2(8-9m-b)(-9)=0,$$

$$F_b(m,b)=2(4-2m-b)(-1)+2(6-5m-b)(-1)+2(7-6m-b)(-1)+2(8-9m-b)(-1)=0.$$

Sadeleştirmelerden sonra aşağıdaki denklemler elde edilir.
146m +22b=152
22m + 4b=25

Bu iki bilinmeyenli iki denklemin çözümünden m=0.58, b=3.06 bulunur.

f(x) = 0.58x + 3.06

m ve b nin doğrudan hesaplanmasını sağlayacak formüller farkların karelerinin toplamınde elde edilir.

$$F(m,b)=\sum_{i=1}^{n}(y_i-mx_i-b)^2=(y_1-mx_1-b)^2+\cdots+(y_n-mx_n-b)^2$$

Yukarıdaki denklemde m ve b ye göre kısmi türevler alınıp sıfıra eşitlenirse aşağıdaki denklemler elde edilir.

$$F_m(m,b)=\sum_{i=1}^{n}2(y_i-mx_i-b)(-x_i)=-2\left(\sum_{i=1}^{n}x_i^2\right)m-2\left(\sum_{i=1}^{n}x_i\right)b+2\left(\sum_{i=1}^{n}x_iy_i\right)=0$$

$$F_b(m,b)=\sum_{i=1}^{n}2(y_i-mx_i-b)(-1)=-2\left(\sum_{i=1}^{n}x_i\right)m-2\left(\sum_{i=1}^{n}1\right)b+2\left(\sum_{i=1}^{n}y_i\right)=0$$

ya da

$$\begin{cases}\left(\sum_{i=1}^{n}x_i^2\right)m+\left(\sum_{i=1}^{n}x_i\right)b=\sum_{i=1}^{n}x_iy_i\\ \left(\sum_{i=1}^{n}x_i\right)m+\quad nb\quad=\sum_{i=1}^{n}y_i\end{cases}$$

denklem sistemi çözülerek bulunur.

Bu denklem sisteminin daima tek bir çözümü bulunduğuna dikkat ediniz.

$$m = \frac{n(\sum_{k=1}^{n} x_k y_k) - (\sum_{k=1}^{n} x_k)(\sum_{k=1}^{n} y_k)}{n(\sum_{k=1}^{n} x_k^2) - (\sum_{k=1}^{n} x_k)^2} \quad , \quad b = \frac{\sum_{k=1}^{n} y_k - m(\sum_{k=1}^{n} x_k)}{n}.$$

Örnek:

(0 , 6.4), (1 , 2.6), (2 , 0.5), (3 , 0.6) ve (4 , 0.3) veri noktalarına en iyi uyan y=mx + b, doğru denklemini bulunuz.

$$\sum_{k=1}^{5} x_k = 0+1+2+3+4 = 10 \quad , \quad \sum_{k=1}^{5} y_k = 6.4+2.6+0.5+0.6+0.3 = 10.4$$

$$\sum_{k=1}^{5} x_k y_k = 0\cdot(6.4)+1\cdot(2.6)+2\cdot(0.5)+3\cdot(0.6)+4\cdot(0.3) = 2.6+1+1.8+1.2 = 6.6$$

$$\sum_{k=1}^{5} x_k^2 = 0+1+4+9+16 = 30 \quad , \quad (\sum_{k=1}^{5} x_k)(\sum_{k=1}^{5} y_k) = 10\cdot(10.4) = 104$$

$$m = \frac{n(\sum_{k=1}^{n} x_k y_k) - (\sum_{k=1}^{n} x_k)(\sum_{k=1}^{n} y_k)}{n(\sum_{k=1}^{n} x_k^2) - (\sum_{k=1}^{n} x_k)^2} = \frac{5\cdot(6.6)-104}{5\cdot 30-100} = \frac{33-104}{50} = \frac{-71}{50} = -1.42,$$

$$b = \frac{\sum_{k=1}^{n} y_k - m(\sum_{k=1}^{n} x_k)}{n} = \frac{10.4-(-1.42)\cdot 10}{5} = \frac{10.4+14.2}{5} = \frac{24.6}{5} = 4.92.$$

y=-1.42x+4.92

6.2. İkinci mertebe en küçük kareler yöntemi

$$E = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

f(x)=$a_2x^2 + a_1 x + a_0$

$$E = \sum_{i=1}^{N} (y_i - a_2x_i^2 - a_1bx_i - a_0)^2$$

$E_a = \sum_{i=1}^{N} 2(\, y_i - a_2x_i^2 - a_1x_i - a_0)(-x_i^2)$=0

$E_b = \sum_{i=1}^{N} 2(\, y_i - a_2x_i^2 - a_1x_i - a_0)(-x_i)$=0

$E_c = \sum_{i=1}^{N} 2(\, y_i - a_2x_i^2 - a_1x_i - a_0)(-1)$=0

$$\sum_{i=1}^{N} y_i x_i^2 = a_2 \sum_{i=1}^{N} x_i^4 + a_1 \sum_{i=1}^{N} x_i^3 + a_0 \sum_{i=1}^{N} x_i^2$$

$$\sum_{i=1}^{N} y_i x_i = a_2 \sum_{i=1}^{N} x_i^3 + a_1 \sum_{i=1}^{N} x_i^2 + a_0 \sum_{i=1}^{N} x_i$$

$$\sum_{i=1}^{N} y_i = a_2 \sum_{i=1}^{N} x_i^2 + a_1 \sum_{i=1}^{N} x_i + a_0 N$$

6.3. Genel polinom regresyon modeli

Genel polinom regresyon modeli, en küçük kareler yöntemi kullanılarak geliştirilebilir. En küçük kareler yöntemi, polinomdan tahmin edilen değerler ile veri kümesinden beklenen değerler arasındaki farkı en aza indirmeyi amaçlamaktadır.

The coefficients of the polynomial regression model ($a_k, a_{k-1}, \cdots, a_1, a_0$) may be determined by solving the following system of linear equations.

$$\begin{bmatrix} N & \sum_{i=1}^{N} x_i & \cdots & \sum_{i=1}^{N} x_i^k \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 & \cdots & \sum_{i=1}^{N} x_i^{k+1} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{N} x_i^k & \sum_{i=1}^{N} x_i^{k+1} & \cdots & \sum_{i=1}^{N} x_i^{2k} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i y_i \\ \vdots \\ \sum_{i=1}^{N} x_i^k y_i \end{bmatrix}$$

Örnek:
Aşağıdaki veri kümesine uygun 2. dereceden bir polinom eğrisinin nasıl geliştirileceğini gösterin.

x	-3	-2	-1	-0.2	1	3
y	0.9	0.8	0.4	0.2	0.1	0

Bu veri kümesinde N = 6 ve 2. dereceden bir polinom için k = 2 dir. En küçük kareler yönteminin uygulanması aşağıdaki doğrusal sistemi sağlar.

$$\begin{bmatrix} 6 & -2.2 & 24.04 \\ -2.2 & 24.04 & -8.008 \\ 24.04 & -8.008 & 180.0016 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2.4 \\ -4.64 \\ 11.808 \end{bmatrix}$$

Sistemi çözmek için Cramer kuralını kullanarak, M matrisini alarak ve sütun vektörü b'yi i sütununa yerleştirerek Mi matrislerinin her birini üretiyoruz, örneğin M0, M1, M2,:

$$M_0 = \begin{bmatrix} 2.4 & -2.2 & 24.04 \\ -4.64 & 24.04 & -8.008 \\ 11.808 & -8.008 & 180.0016 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} 6 & 2.44 & 24.04 \\ -2.2 & -4.64 & -8.008 \\ 24.04 & 11.808 & 180.0016 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 6 & 2.44 & 2.4 \\ -2.2 & -4.64 & -4.64 \\ 24.04 & 11.808 & 11.808 \end{bmatrix}$$

$$a_0 = \frac{det(M_0)}{det(M)} = \frac{2671.20}{11661.27} = 0.2291$$
$$a_1 = \frac{det(M_1)}{det(M)} = \frac{-1898.46}{11661.27} = -0.1628$$
$$a_2 = \frac{det(M_2)}{det(M)} = \frac{323.76}{11661.27} = 0.0278$$

$$y = 0.0278x^2 - 0.1628x + 0.2291$$

Örnek:

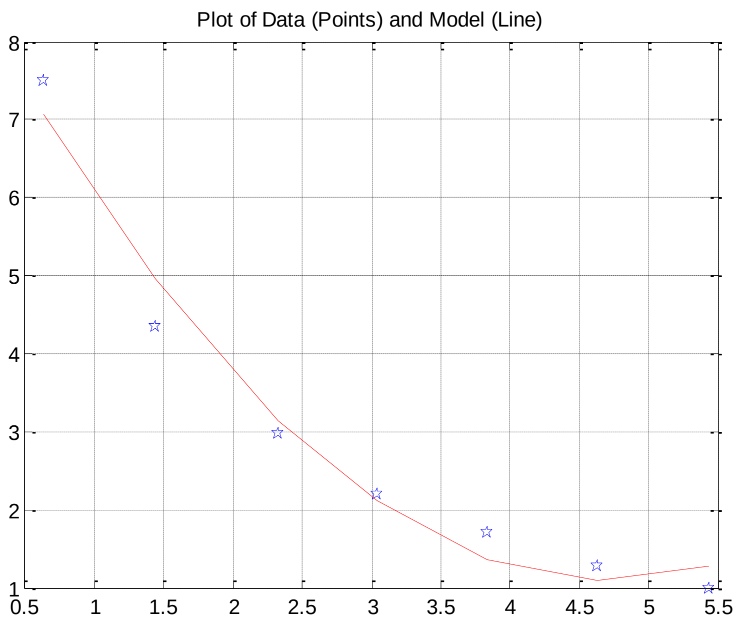
ikinci dereceden en küçük kareler ve Matlab ; p = polyfit(x,y,n), n=2

```
clear all
close all
x = [5.435    4.635    3.835    3.035    2.325    1.435    0.635];
y = [1.00      1.28      1.70      2.20      2.97      4.35      7.50   ];
p = polyfit(x, y, 2)                        % Quadratic Function Fit
v = polyval(p, x)                         % Evaluate
TSE = sum((v - y).^2)                % Total Squared Error
figure(1)
plot(x, y, 'bp')
hold on
plot(x, v, '-r')
hold off
grid
title('Plot of Data (Points) and Model (Line)')
```

p =   0.3582 -3.3833   9.0748

v =   1.2664   1.0877   1.3674   2.1056   3.1447   4.9573   7.0708

TSE =   0.8110



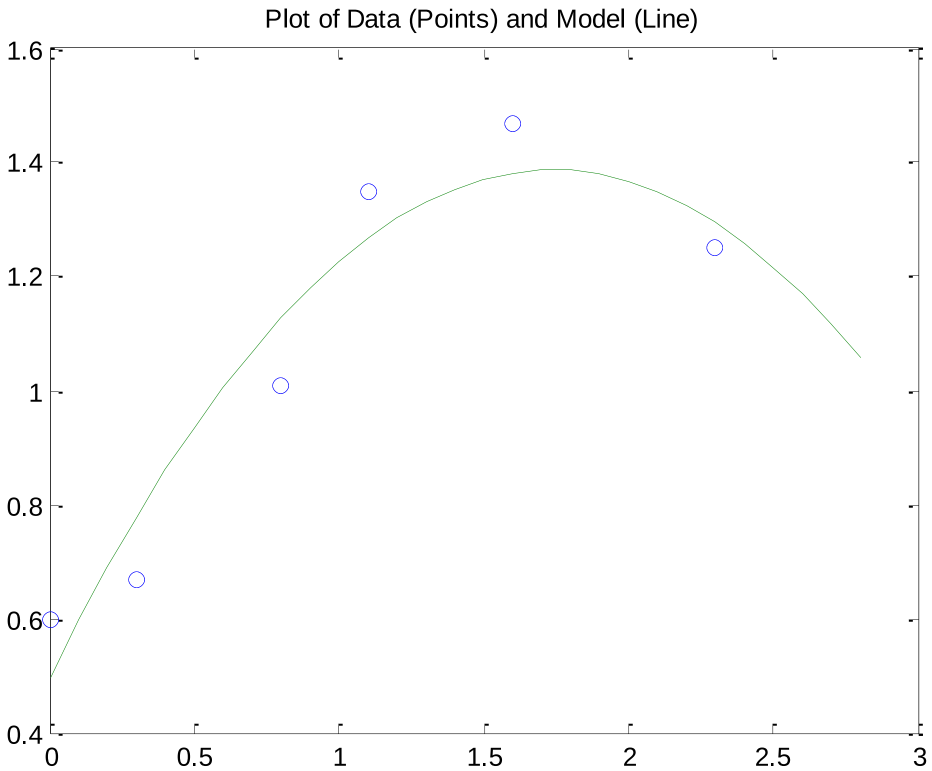
Örnek:

ikinci dereceden en küçük kareler ve Matlab ; p = polyfit(x,y,n), n=2

```
clear all
close all
t = [0 0.3 0.8 1.1 1.6 2.3];
y = [0.6 0.67 1.01 1.35 1.47 1.25];
figure, plot(t,y,'o')
title('Plot of y Versus t')
p = polyfit(t,y,2)
t2 = 0:0.1:2.8;
y2 = polyval(p,t2);
hold on
plot(t,y,'o',t2,y2)
title('Plot of Data (Points) and Model (Line)')
```

p = -0.2942 1.0231 0.4981

$f(x) = -0.2942x^2 + 1.0231x + 0.4981$



Örnek:

Uyum İşlevini Kullanarak Üstel Modellere Sığdırma

```
clear all
close all
x = (0:0.2:5)';
y = 2*exp(-0.2*x) + 0.1*randn(size(x));
f = fit(x,y,'exp1')
plot(f,x,y)
```

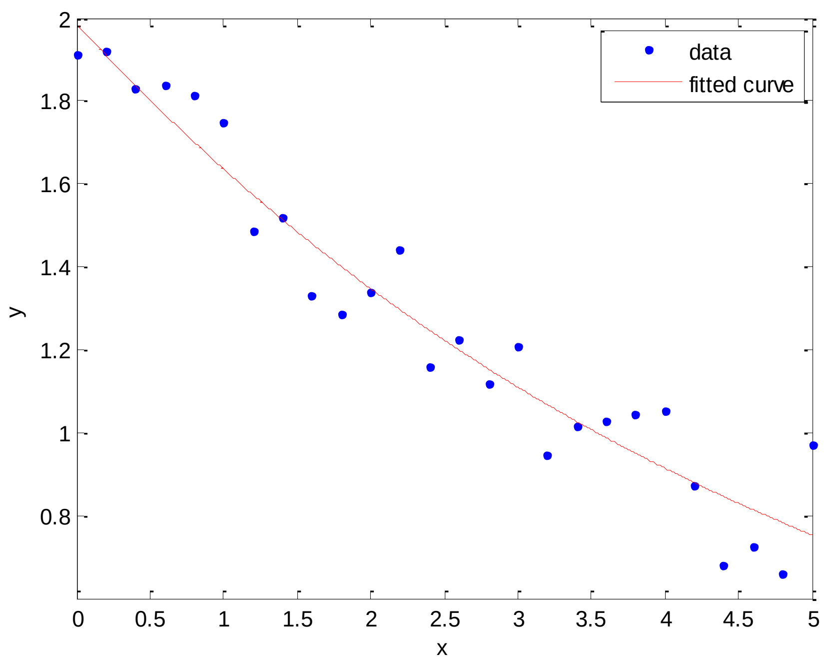
f = General model Exp1:

f(x) = a*exp(b*x)

Coefficients (with 95% confidence bounds):

a = 2.103 (1.999, 2.208)

b = -0.2222 (-0.2464, -0.1981)



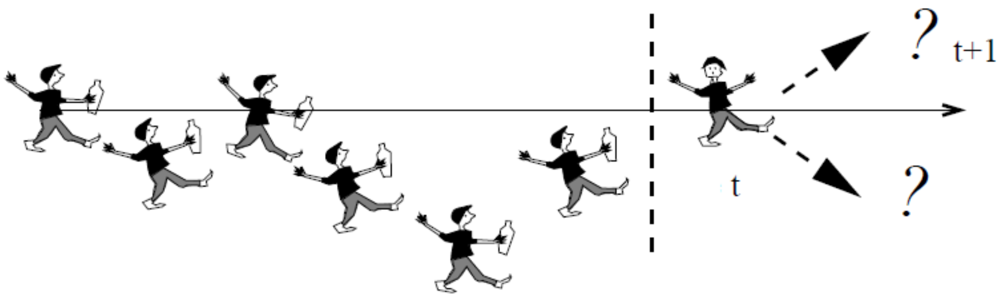
7. Markov Zincir Analizi

Stokastik (Rastgele - Rassal) süreç, Varlığı deneysel olarak kanıtlanmış zaman veya mekana göre değişen ya da evrilen olguları tanımlamak için kullanılan bir olasılık modelidir. Evrilmek: Bir biçimden başka bir biçime doğal olarak dönmek. Şu an hava Güneşli yarın nasıl olacak? Yağmurlu mu? Bulutlu mu? Güneşli mi?

Stokastik bir süreç, zamana bağlı olasılıklı bir şekilde gelişen matematiksel modeldir. Markov zinciri adı verilen özel bir stokastik süreç çalışılmasında, sistemin bir sonraki durumu önceki durumlara değil, yalnızca mevcut, şu andaki duruma bağlıdır. Yarının izleri bugündedir. Yarın için bugün karar vereksin. Stokastik süreçlerin analizinde Markov Zincirleri teorisi, ismini Rus matematikçi A.A. Markov'dan (1856–1922) almıştır. Herbir durum ancak ve ancak bir önceki durumun sonucudur, ondan önceki durumların sonucu değildir. Rassal değişkenlerin zamanla nasıl değiştiği stokastik süreçleri de içerir. Örneğin borsada bir hissenin fiyatının nasıl değiştiği veya bir firmanın piyasa payının nasıl değiştiği stokastik süreçle ilgilidir. Bir stokastik süreç örneği olan Markov Zincirleri eğitim, pazarlama, sağlık hizmetleri, muhasebe ve üretim alanları gibi alanlara uygulanmaktadır. Özellikle makine öğrenmesine yönelik geliştirilen matematiksel modellerin temelini oluşturmaktadır.

Markov zincirleri, ayrık zamanlı stokastik süreçlerin özel bir türüdür. Basit bir ifadeyle herhangi bir zamanda ayrık zamanlı stokastik süreç sonlu sayıda durumdan birinde olabilir. Ayrık zamanlı stokastik süreç aşağıdaki koşulu sağlıyorsa süreç Markov zinciridir.

Süreçler $\{X_0, X_1, X_2, . . .\}$, şeklinde yazılabilir, burada t= 0,1,2,… her bir durum için X_t, t zamanındaki durumdur. Durumların geçiş diyagramı üzerinden baktığımız süreçlerin ortak bir özelliği var: X_{t+1} durumu sadece X_t durumuna bağlıdır. $X_0, X_1, . . . , X_{t-1}$ durumlarına bağlı değildir.



Tanımlar:

Tanım-1: t zamanında bir Markov zincirinin durumu X_t değeridir.

Tanım-2: Bir Markov zincirinin durum (S) kümesi, her bir X_t'nin alabileceği değer kümesidir.

Örneğin, S = {1, 2, 3, 4, 5, 6, 7}. S kümesi N duruma sahiptir.

Tanım: Bir Markov zincirinin yörüngesi, X_0, X_1, X_2, için belirli değerler kümesidir. . ..

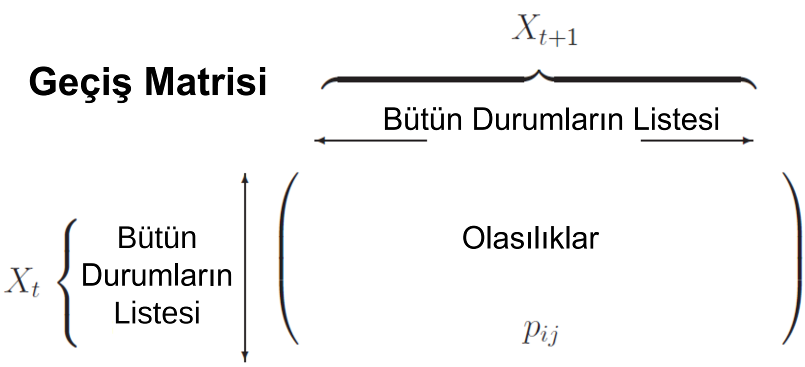
Daha genel olarak, eğer s0, s1, s2, s3, ... yörüngesini takip edersek X0 = s0, X1 = s1, X2 = s2, X3 = s3, değerlerini alır. "Yörünge" sadece "yol" anlamına gelen bir kelimedir.

Markov Özelliğini matematiksel gösterimle şu şekilde formüle ediyoruz:

$$\boxed{\mathbb{P}(X_{t+1} = s \,|\, X_t = s_t, X_{t-1} = s_{t-1}, \ldots, X_0 = s_0) \;=\; \mathbb{P}(X_{t+1} = s \,|\, X_t = s_t),}$$

tüm t = 1, 2, 3, . . . için ve tüm s0, s1,. . . , st, s durumları için.

Eğer {X_0, X_1, X_2, . . .} Markov özelliğini sağlar ise {X_0, X_1, X_2, ... } ayrık rasgele değişkenlerin bir sırası Markov zinciri olur.



$$p_{ij} = \mathbb{P}(X_{t+1} = j \,|\, X_t = i)$$

Geçiş matrisinin özellikleri:

1) Kare matristir, çünkü tüm olası durumların hem satır hem de sütun olarak kullanılmak zorundadır.
2) Tüm giriş verileri 0 ile 1 arasındadır; Bunun nedeni bütün girdilerin olasılıkları temsil edilmesidir.
3) Herhangi bir satırdaki verilerin olasılıkları toplamı 1 olmalıdır, çünkü satırdaki sayılar, bir durumun diğer durum olan tüm geçiş olasıklarını verir.

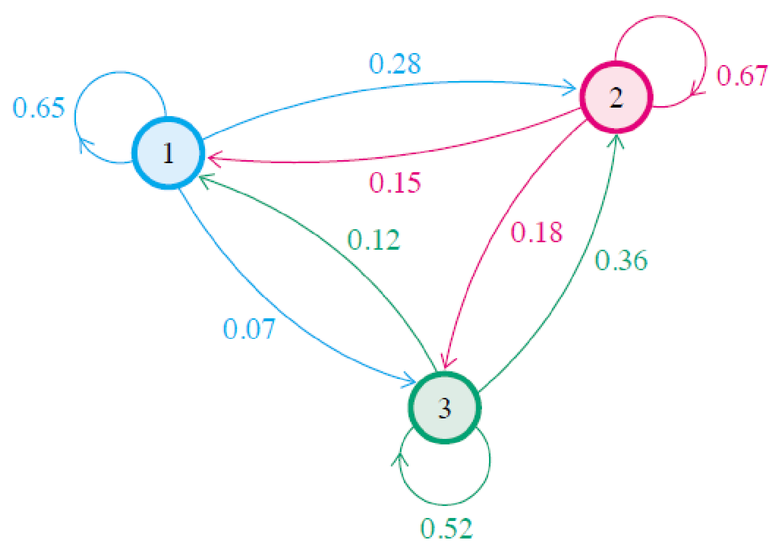
Bir deneyin denemelerinin bir sırası, bir Markov zinciri olabilmesi için

1) Her bir deneyin sonucu, belirli durumlardan bir tanesi olmalıdır,
2) Bir deneyin sonucu, sadece şu andaki duruma bağlı ve geçmiş herhangi bir duruma bağlı olmamalıdır. Olasılık kaysayıları, geçmişteki gözlemlerden ya da ölçümlerden elde edilir.

Durumların Geçiş Tablosu:

		Next Generation		
	State	**1**	**2**	**3**
Current Generation	**1**	0.65	0.28	0.07
	2	0.15	0.67	0.18
	3	0.12	0.36	0.52

Durumların Geçiş Diyagramı:



Bir geçiş matrisinde, durumlar yanda ve üstte gösterilir. P tablo için geçiş matrisini temsil ediyorsa, bir durum olduktan sonra diğer durumların olma olasılığının veren matrisdir.

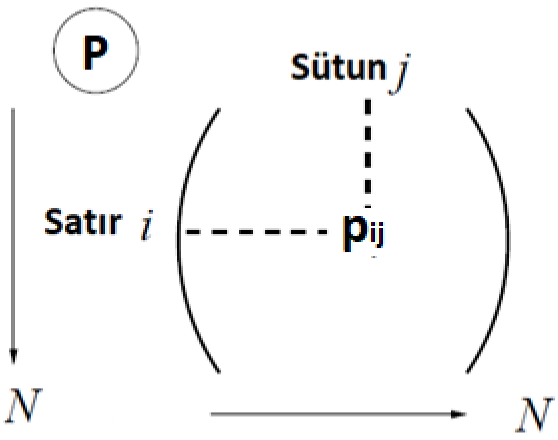
Geçiş Matrisi:

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.65 & 0.28 & 0.07 \\ 0.15 & 0.67 & 0.18 \\ 0.12 & 0.36 & 0.52 \end{bmatrix} \end{array} = P$$

p_{23} = 0.18 nedir? 2.durumdan bir sonraki adımda 3. Duruma gitme olasılığını verir.

Matris:

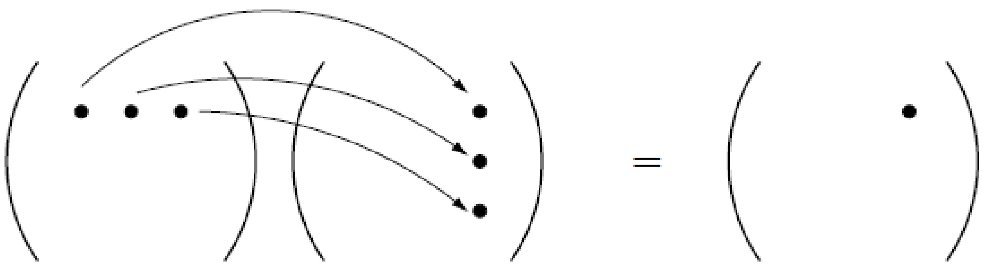
P matrisi N x N matrisi olsun. $P=(p_{ij})$ biçiminde yazılır. P'nin (i,j) elemanı hem p_{ij} hemde $(P)_{ij}$ olarak yazılır.



Matris çarpımı:

$A=(a_{ij})$ ve $B=(b_{ij})$ N x N boyutunda matrisler olsun. A ile B matrislerinin çarpımı A x B =AB dir.

$$(AB)_{ij} = \sum_{k=1}^{N} a_{ik} b_{kj}$$



Bir matrisin karesi:

$$(A^2)_{ij} = \sum_{k=1}^{N} (A)_{ik}(A)_{kj} = \sum_{k=1}^{N} a_{ik}a_{kj}.$$

7.1. Olasılık Vektörü

Geçiş matrisinin aşağıdaki koşulları sağlaması gerekir:

1) Her bir eleman olasılıktır ve değeri 0 ile 1 arasındadır. Olasılıklar negatif ve birden büyük olamaz.
2) Her bir satırdaki olasılıkların toplamı 1'e eşittir. Geçiş matrisinin bir satırdaki elemanları muhtemel olayların gerçeklenme olasılıklarından doğan sonuçları vermesi nedeni ile, olasılıklar toplamının bir olması açıktır.

Geçiş matrisi genel bir notasyon ile

$$0 \leq P_{ij} \leq 1,$$

$$\sum_{j=1}^{m} P_{iJ} = 1, (i = 1,2,....,m)$$

olmak üzere,

$$P=\begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1m} \\ \vdots & \vdots & \vdots & \vdots \\ P_{m1} & P_{m2} & \cdots & P_{mm} \end{bmatrix}$$

kare matrisi yazılır. i satırı temsil etmek üzere v^i satır vektörü olarak tanımlanır. v^i satır vektörlerinin birer olasılık vektörü olduğu unutulmamalıdır.

$v^2 = (P_{21}\ P_{22}\ \ldots\ P_{2m})$ olasılık vektörü geçiş matrisinin ikinci satırını temsil eder ve vektör elemanları sırasıyla durum-2'den diğer bütün durumlara geçme olasılıklarını verir. O halde P geçiş matrisi v^i olasılık vektörlerinden oluşur ve geçiş matrisi koşullarını sağlar.

Örnek:

Bir havanın üç durumunu sınıflandıralım,

Durum-1: Yağmurlu

Durum-2: Bulutlu

Durum-3: Güneşli

Varsayım: Yarınki hava durumu sadece bugünün hava durumuna bağlıdır! Hava durumuna ait Yağmurlu, Bulutlu ve Güneşli olma olasıkları aşağıdaki derlenmiş olsun.

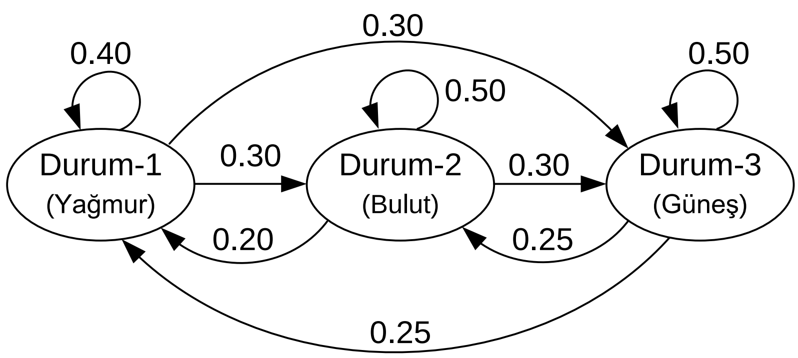
Şimdiki durum (n=0)	Bir sonraki durum (n=1)		
	Yağmur (%)	Bulut (%)	Güneş (%)
Yağmur	40	30	30
Bulut	20	50	30
Güneş	25	25	50

Tablo, şu anki durumu ve bir sonraki durumun olma olasılıklarını vermektedir. Tablodaki bilgiler P ile göstereceğimiz bir **geçiş matrisi** bulunur.

$$P = \begin{bmatrix} 0.40 & 0.30 & 0.30 \\ 0.20 & 0.50 & 0.30 \\ 0.25 & 0.25 & 0.50 \end{bmatrix}$$

Geçiş matrisinin her bir elemanı bir durumdan diğer bir duruma geçme olasılığını verir ve bu elemanlara P_{ij} denir. Bu ise, halen i. durumda olan sürecin bir sonraki adımda j. durumda olacağını gösteren şartlı olasılıktır.

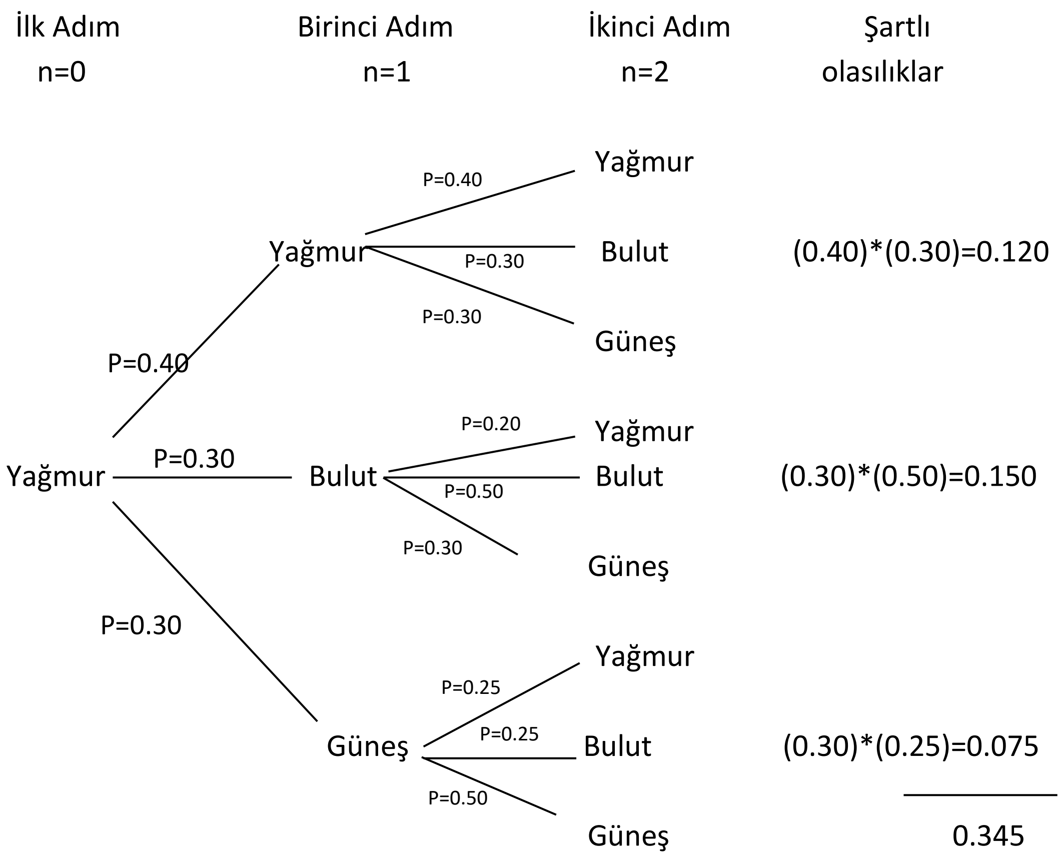
P_{13}=0.30 elemanının, şu an yağmur yağıyorsa bir sonraki gün güneşli olma olasılığını vermektedir. v^2 =(0.20 0.50 0.30) olasılık vektörü geçiş matrisinin ikinci satırını temsil eder ve vektör elemanları sırasıyla durum-2'den durum-1, durum-2 ve durum-3'e geçme olasılıklarını verir. O halde P geçiş matrisi v^i olasılık vektörlerinden oluşur ve geçiş matrisi koşullarını sağlar. Durumların geçiş diyagramı aşağıda verilmiştir.



7.2. Markov Zinciri ile Olasılık Analizi

Bir havanın n=0 da Yağmur olması, n=1 de Yağmur olması ve n=2 de Bulutlu olma olasılığı ne olacaktır? Soruyu olasılık teorisi teknikler ile cevaplamak olanağı vardır. Yağmur olan bir havanın ikinci adımda Bulut olana kadar olan değişik seçeneklerini teker teker ayırmak için bir diyagramla çalışmak kolaylık sağlar.

Şu an hava **Yağmur** ise ikinci adımda **Bulutlu** olma olasılığı nedir?



Şekil: n=2, ikinci adımda mümkün sonuçlar

Şekilde, 2. adımda Bulutlu olması için verilen olasılıkların toplanması ile 0,345 bulunur. Bu ise şartlı olasılıktır. Aşağıdaki tablo grafikteki işlemleri özetlemektedir ve n=0 adımında Şu an yağmurlu olan havanın n=2 adımda Bulutlu olma olasılığı 0,345 dir. Şu an yağmurlu iken yarın değil öbürgünün bulutlu olma olasılığı:%34.5 dır.

n=0	n=1 de	n=2 de	Şartlı Olasılık
Yağmur	Yağmur 0.40 Bulut 0.30 Güneş 0.30	Bulut 0.30 Bulut 0.50 Bulut 0.25	(0.40)*(0.30)=0.120 (0.30)*(0.50)=0.150 (0.30)*(0.25)=0.075 0.345

Bulunan sonuç Markov zincirlerinin sağladığı analiz tekniğiyle de elde edilebilir.

Önce V^1 olasılık vektörünün ne anlamda kullanıldığını açıklamaya gerek vardır. S^1 durumu için olasılık vektörü V^1= (0.40; 0.30; 0.30) dur. n=0 adımında Yağmur olan bir havanın ikinci adımdaki tüm tedarik olasılıklarını V^1 vektörü verecektir. V^1 vektörünün P matrisi ile çarpımı ile V^2 bulunur.

$$v^2 = v^1 P = [0.4\,, 0.3\,, 0.3]\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$$

= [0,295   0,345   0,360]

V^2 vektörü n=0 adımda Yağmur olan havanın n=1 adımında Yağmur iken n=2 adımında Yağmur, Bulutlu ve Güneşli olma olasıklarını vermektedir.

Markov zincirleri,olasılık problemlerinin özel halini formüle etmek için teknik ve analiz vermiştir.

V^2=P x P

$$V^2 = \begin{pmatrix} 0.2950 & 0.3450 & 0.3600 \\ 0.2550 & 0.3850 & 0.3600 \\ 0.2750 & 0.3250 & 0.4000 \end{pmatrix}$$

n: Yağmur iken
n=1'inci adımda [Yağmur, Bulutlu, Güneşli]=[0.4, 0.3, 0.3]
n=2'inci adımda [Yağmur, Bulutlu, Güneşli]=[0.295, 0.345, 0.360]
olarak bulunur.

Örnek:

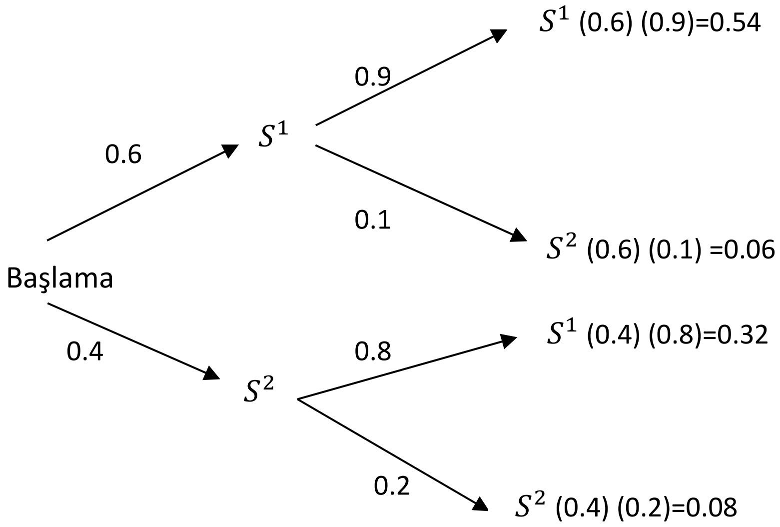
Bir markov sürecinin başlama vektörü $v^0 = (0.6 \quad 0.4)$ ve geçiş olasılık matrisi

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix}$$

Birinci adım için $v^1 = v^0 P$ olduğundan

$v^1 = (0.6 \;\; 0.4) \begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix} = (0.86 \;\; 0.14) = (v_1^1 \;\; v_2^1)$ bulunur.

O sürecin bir adım sonunda S^1 durumunda bulunma olasılığı 0.86 ve S^2 durumunda bulunma olasılığı 0.14 dir.Bu sonuşlar ağaç diyagramı ile de elde edilebilir.



Şekil: Başlama olasılıkları (0.6 0.4) olmak üzere bir adım sonra durum uzayı.

$v_1^1 = (0.6 \;\; 0.4) \begin{pmatrix} 0.9 \\ 0.8 \end{pmatrix}$=0.54+0.32=0.86

$v_2^1 = (0.6 \;\; 0.4) \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}$=0.06+0.08=0.14

$v^2 = v^1 P = (0.86 \;\; 0.14) \begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix} = (0.886 \;\; 0.114) = (v_1^2 \;\; v_2^2)$

veya

$v^2 = v^0 P^2 = (0.6 \;\; 0.4) \begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.8 & 0.2 \end{bmatrix} = (0.886 \;\; 0.114) = (v_1^2 \;\; v_2^2)$

7.3. Ergodik (Düzenli) Markov Zincirleri

Ergodik zincir, matematik olarak herhangi bir durumdan diğer bütün durumlara geçişin mümkün olduğu bir süreci tanımlar. Bu tanımda tam bir adımda bulunma şansı yoktur, ama başlama durumuna bakmadan bir duruma erişilmiş olmalıdır. Denge durumu koşullarına erişilmesini sağlamak için zincir mutlaka ergodik olmalıdır.

Ergodik markov zincirini sınırlayan hal düzenli (=reguler) zincirdir. Düzenli zincir, P geçiş olasılıkları matrisinin kuvvetlerinde bulunan elemanların sıfırdan farklı ve pozitif olmasını gerektirir. Ergodik zincirin kuvvetleri alınırsa veya matriste sıfır eleman kalmayıncaya kadar kuvvetler alınırsa ergodik zincirin düzenli olduğu görülür. Bu işlem aşağıda verilmiştir.Bütün düzenli zincirlerin ergodik olduğu doğrudur, ama tersininde doğru olmasına ihtiyaç yoktur. Bütün ergodik zincirlerin düzenli olmadığına dikkat etmelidir.

$$
\mathrm{P}=\begin{bmatrix} x & x & 0 & x & x \\ 0 & x & x & 0 & x \\ 0 & 0 & 0 & x & x \\ x & 0 & x & 0 & x \\ x & x & 0 & 0 & 0 \end{bmatrix}, \quad \mathrm{P}^2=\begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & 0 & x \\ x & x & 0 & x & x \\ x & x & x & x & x \end{bmatrix}, \mathrm{P}^4=\begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \end{bmatrix}
$$

Örnek: Aşağıdaki geçiş matrislerinin a) düzenli ve b) ergodik olmasını açıklayınız.

$$
\mathrm{P}= \begin{array}{c} 1 \\ 2 \\ 3 \end{array}\begin{bmatrix} x & x & 0 \\ x & 0 & x \\ 0 & x & x \end{bmatrix} \qquad \mathrm{P}^2=\begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \end{bmatrix}
$$

P geçiş matrisinin P^2 kuvvetinin elemanları pozitif veya sıfırdan farklı olduğundan P geçiş matrisi ile verilen Markov zinciri düzenlidir. Düzenli Markov zincirleri ise ergodiktir. Ayrıca 1 den 1 veya 2 ye doğrudan geçiş vardır,daha sonra da 2 den 3 e geçiş olanaklıdır.2 den 1e geçilebilir ve 3 den 2 ye ,1 e geçiş vardır.Dolayısıyla zincir ergodiktir,bütün durumlara geçiş olanağı vardır.

Örnek:

$$
P=\begin{array}{c} \\ 1 \\ \\ 2 \\ 3 \end{array}
\begin{array}{c}
\begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\
\begin{bmatrix} x & 0 & x & 0 \\ 0 & x & 0 & x \\ x & 0 & x & 0 \\ 0 & x & 0 & x \end{bmatrix}
\end{array}
\quad P^2=\begin{bmatrix} x & 0 & x & 0 \\ 0 & x & 0 & x \\ x & 0 & x & 0 \\ 0 & x & 0 & x \end{bmatrix}
\quad P^4\begin{bmatrix} x & 0 & x & 0 \\ 0 & x & 0 & x \\ x & 0 & x & 0 \\ 0 & x & 0 & x \end{bmatrix}
$$

P geçiş matrisinin kuvvetleri P matrisini tekrar vermektedir.Dolayısıyle P stokastik matrisi düzenli bir zincir değildir.Zira ilk matriste sıfır elemanları vardır ve kuvvetlerde sıfır elemanlar aynen kalmıştır. Ayrıca 1 den 1 veya 3 e,ve 3 den 3 veya 1 e geçiş vardır. Dolayısıyle 1. durumdan 2. duruma veya 4. duruma geçiş olanağı yoktur, ve zincir ergodik değildir.

Decide whether the following transition matrices are regular.

(a) $A=\begin{bmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.5 & 0.5 \\ 0.6 & 0.4 & 0 \end{bmatrix}$

Solution Square A.

$$A^2=\begin{bmatrix} 0.5625 & 0.3125 & 0.125 \\ 0.3 & 0.45 & 0.25 \\ 0.45 & 0.35 & 0.2 \end{bmatrix}$$

Since all entries in A^2 are positive, matrix A is regular.

(b) $B=\begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Solution Find various powers of B.

$$B^2=\begin{bmatrix} 0.25 & 0 & 0.75 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix};\ B^3=\begin{bmatrix} 0.125 & 0 & 0.875 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix};\ B^4=\begin{bmatrix} 0.0625 & 0 & 0.9375 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Further powers of B will still give the same zero entries, so no power of matrix B contains all positive entries. For this reason, B is not regular.

Örnek:

Aşağıda verilen Markov zincirinin ergodik ve düzenli olmasını araştırınız.

$$\text{P}=\begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix} \qquad \text{P}^2=\begin{bmatrix} 1 & 0 \\ 3/4 & 1/4 \end{bmatrix}$$

P matrisinin kuvvetleri alınırsa (1.2) elemanı daima (0) olacaktır. Dolayısıyla verilen Markov zincir düzenli ve ergodik değildir.

Örnek:

$$\text{P}=\begin{bmatrix} 0 & 1 \\ 1/3 & 2/3 \end{bmatrix}$$

Düzenli Markov zinciri olduğunu gösteriniz.

$$\text{P}^2 = P * P = \begin{bmatrix} 0 & 1 \\ 1/3 & 2/3 \end{bmatrix} * \begin{bmatrix} 0 & 1 \\ 1/3 & 2/3 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/3 \\ 2/9 & 7/9 \end{bmatrix}$$

P^2 nin bütün elemanları pozitif olduğundan P düzenli Markov zinciridir.

Örnek:

Aşağıdaki matrisin reguler (düzenli) olup olmadığını bulunuz.

$$P = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.5 & 0.5 \\ 0.6 & 0.4 & 0 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0.5625 & 0.3125 & 0.125 \\ 0.3 & 0.45 & 0.25 \\ 0.45 & 0.35 & 0.2 \end{bmatrix}$$

P^2 matrisindeki bütün değerler pozitif olduğunda P matrisi düzenlidir.

Örnek:

Aşağıdaki matrisin reguler (düzenli) olup olmadığını bulunuz.

$$P = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0.25 & 0 & 0.75 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0.125 & 0 & 0.875 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$P^4 = \begin{bmatrix} 0.0625 & 0 & 0.9375 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

P^2 matrisindeki bütün değerler pozitif olmadığından P matrisi düzenli değildir.

7.4. Denge Durumu Koşulları

Düzenli ergodik zincirlerde denge durumu koşullarının varlığını göstermek üzere P^n 'nin hesaplanması ile gösterilir. P nin kuvvetlerinde izlenebileceği gibi n büyüdükçe Pij değerleri sabit bir sayıya veya limite yaklaşmaktadır. Geçiş matrisi denge durumuna yaklaşmaktadır.

Örnek:

P nin n. kuvvetlerine göre n=1 den n=8 e kadar değerler için ilişkin hesaplar verilmiştir.

$$P=\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.25 & 0.25 & 0.5 \end{bmatrix} \qquad P^2=\begin{bmatrix} 0.295 & 0.345 & 0.360 \\ 0.255 & 0.385 & 0.360 \\ 0.275 & 0.325 & 0.400 \end{bmatrix}$$

$$P^3=\begin{bmatrix} 0.277 & 0.351 & 0.372 \\ 0.269 & 0.359 & 0.372 \\ 0.275 & 0.345 & 0.380 \end{bmatrix} \qquad P^4=\begin{bmatrix} 0.2740 & 0.3516 & 0.3744 \\ 0.2724 & 0.3744 & 0.3744 \\ 0.2740 & 0.3500 & 0.3760 \end{bmatrix}$$

$$P^5=\begin{bmatrix} 0.27352 & 0.35160 & 0.37488 \\ 0.27320 & 0.35192 & 0.37488 \\ 0.27360 & 0.35120 & 0.37520 \end{bmatrix} \quad P^6\begin{bmatrix} 0.273448 & 0.351576 & 0.374976 \\ 0.273384 & 0.356400 & 0.374976 \\ 0.273480 & 0.351400 & 0.375040 \end{bmatrix}$$

$$P^7=\begin{bmatrix} 0.2734384 & 0.3515664 & 0.3749952 \\ 0.2734256 & 0.3515792 & 0.3749952 \\ 0.2734480 & 0.3515440 & 0.3750080 \end{bmatrix}$$

$$P^8=\begin{bmatrix} 0.27343744 & 0.35156352 & 0.37499904 \\ 0.27343488 & 0.35156608 & 0.3749904 \\ 0.27244000 & 0.35155840 & 0.37500160 \end{bmatrix}$$

Denge durum koşullarının hesaplanması:

Denge durumunda her bir v_i^n olasılık vektörlerinin bütün değerleri için eşit olmaya meyletmektedir. Dolayısıyle aşağıdaki iki kural yazılır:

1) n nin yeterince daha büyük değerleri için, v_i^n olasılık vektörleri bütün o değerleri için aynıdır ve değişmez.
2) $v_i^{n+1} = V_i P$ ve $v_i^{n+1} = V_i^n$ olduğundan V*P=V eşitliğini sağlayan bir V denge vektörü vardır.
3) $Vi, \text{Ö}zde\check{g}erleri\ verir.$

V vektörü denge durumu koşullarını içeren olasılıkları kapsar.

$\sum_{j=1}^{m} P_i^j = 1$ bağıntısı bu değerleri elde etmek için analitik yöntemi sağlar. V bir olasılık vektörü olduğundan

$$\sum_{j=1}^{m} v_i^j = 1$$

bağıntısı geçerlidir.(2) nolu koşuldan [v $_1$ v $_2$ v $_3$ ......v $_m$]*P=[v $_1$ v $_2$ v $_3$ .....v $_m$] yazılır.

Sol tarafta bulunan satır vektörü ile P geçiş matrisi çarpımından yine bir satır vektörü bulunarak eşitliğin sağ tarafındaki vektör elemanlara her bir eleman eşit olması yazılarak (m) adet denklem elde edilir. Olasılıkların toplamıın 1 e eşit olma şartı ile de (m) bilinmeyen, (m+1) denklemden çözülebilir. Fakat (m+1) adet denklemden biri elimine edilerek denklem takımına katılmaz.

Örnek:

$$P=\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$$

Örneğimize belirlenen kuralları uygulayalım.

$\sum_{j=1}^{3} v_1 = 1$ ....$v_1 + v_2 + v_3 = 1$

$[v_1 \ \ v_2 \ \ v_3] * P = [v_1 \ \ v_2 \ \ v_3]$

$0.4v_1 + 0.2v_2 + 0.25v_3 = v_1$

$0.3v_1 + 0.5v_2 + 0.25v_3 = v_2$

$0.3v_1 + 0.3v_2 + 0.5v_3 = v_3$

Fakat (3) adet denklemden biri elimine edilir.
Üç bilinmeyenli denklem sisteminin çözümü ile; $v_1 = 0.273$, $v_2 = 0.352$, $v_3 = 0.375$ bulunur.

Örnek:

Aşağıdaki geçiş matrisinden hareketle denge durumu koşullarını bulunuz.

$$P=\begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.1 & 0.6 & 0.3 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

V=[v_1 v_2 v_3] denge vektörünün bulunması istenmektedir.O halde,

$$v_1+v_2+v_3=1$$

$$0.2v_1+0.1v_2+0.5v_3=v_1$$
$$0.5v_1+0.6v_2+(0)v_3=v_2$$
$$0.3v_1+0.3v_2+0.5v_3=v_3$$

bağıntıları yazılır. İkinci eşitlik elimine edilerek

$$v_1+v_2+v_3=1$$
$$0.5v_1-0.4v_2=0$$
$$0.3v_1+0.3v_2-0.5v_3=0$$

yazılır ve çözüm

$$v_1=0.282$$
$$v_2=0.352$$
$$v_3=0.366$$

olarak bulunur. Dolayısiyle denge vektörü: V=[0.282 0.352 0.366] olur.

Örnek:

Bir imalat tezgahında standart bir mamul imalinden sonra, takip eden mamulünde aynı kalitede olma olasılığı 0.9 ve hatalı bir mamul imalinden sonra takip eden mamulün standart olma olasılığı 0.8 dir. Üretim sürecinde her kademe sonuçları bir ön sonuca bağlı olacağı varsayımı ile geçiş matrisini kurunuz ve süreci irdeleyiniz.

Standart bir mamul üretimi S_1 , hatalı bir mamul üretimi S_2 ile gösterilerek

$$P=\begin{bmatrix}0.9 & 0.1\\0.8 & 0.2\end{bmatrix}$$

Geçiş matrisi bulunur.Geçiş matrisi kuvvetleri alınarak

$$P^2=\begin{bmatrix}0.9 & 0.1\\0.8 & 0.2\end{bmatrix}*\begin{bmatrix}0.9 & 0.1\\0.8 & 0.2\end{bmatrix}=\begin{bmatrix}0.89 & 0.11\\0.88 & 0.12\end{bmatrix}$$

$$P^3=P^2P=\begin{bmatrix}0.89 & 0.11\\0.88 & 0.12\end{bmatrix}*\begin{bmatrix}0.9 & 0.1\\0.8 & 0.2\end{bmatrix}=\begin{bmatrix}0.889 & 0.111\\0.888 & 0.112\end{bmatrix}$$

$$P^4=P^3P=\begin{bmatrix}0.889 & 0.111\\0.888 & 0.112\end{bmatrix}*\begin{bmatrix}0.9 & 0.1\\0.8 & 0.2\end{bmatrix}=\begin{bmatrix}0.8889 & 0.1111\\0.8888 & 0.1112\end{bmatrix}$$

Bulunur ve hesaplara devam ederek sürecin denge vektörünün V=(8/9,1/9) olduğu gösterilebilir. O halde verilen süreç düzenli bir Markov zinciridir.

Sürecin uzun bir dönemde S_1 durumuna gitme olasılığı 8/9 ve S_2 durumuna girme olasılığı 1/9 dur. Diğer bir deyişle uzun bir dönemde üretimin 8/9 u standart mamul, 1/9 u hatalı mamul olacaktır. Örneğin 999 adetlik bir parti üretiminde 888 mamul standart, 111 mamul ise hatalı olur.

Katı Stokastik Matris:

$$P=\begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

(m) durum sayısını göstermek üzere P geçiş matrisi

$$\sum_{i=1}^{m} P_{ij} = \sum_{j=1}^{m} P_{ij} = 1$$

bağıntısını sağlarsa katı stokastik matris adını alır.Katı stokastik matrislerde denge vektörü bileşenleri bütün matrislerde j değerleri için $v_j = 1/m$ dir.Örnekte denge vektörü, V=[v_1 v_2 v_3], $v_1 = v_2 = v_3 = 1$,V=(1/3 1/3 1/3) olur.

Örnek:

$$P=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

P geçiş matrisinin denge vektörü, VP=V bağıntısından V=(1/2 1/2) olarak bulunur. P nin kuvvetleri alınırsa

n=tek sayı için P^n =P

n=çift sayı için $P^n = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

bulunur ve P nin kuvvetlerinde sıfır bulunduğu için zincir düzenli değildir.

7.5. Emici Markov Zincirleri

Tüm Markov zincirleri düzenli değildir. Aslında, Markov zincirlerinin en önemli yaşam bilimleri uygulamalarından bazıları düzenli olan geçiş matrislerini içermez. Markov zincirlerinin fikirlerini canlı organizmaları modellemek için kullandığımızda, ortak bir durum ölümdür. Organizma bu duruma girdiğinde, ayrılması mümkün değildir. Yaşam bilimlerinde yaygın olarak kullanılan bir Markov zinciri türü, bir emici Markov zinciri olarak adlandırılır.

Not: Elektronik ve bilgisayar kontrollu sistemlerin kararlı hallerinde ise hangi durumuda nereden başlarsa başlasın aynı noktaya gider ve orada çalışmasını sürdürür.

Geçiş matrisinin aşağıdaki koşulları sağlaması gerekir:

3) Her bir eleman olasılıktır ve değeri 0 ile 1 arasındadır. Olasılıklar negatif ve birden büyük olamaz.
4) Her bir satırdaki olasılıkların toplamı 1'e eşittir. Geçiş matrisinin bir satırdaki elemanları muhtemel olayların gerçeklenme olasılıklarından doğan sonuçları vermesi nedeni ile, olasılıklar toplamının bir olması açıktır.

Geçiş matrisi genel bir notasyon ile

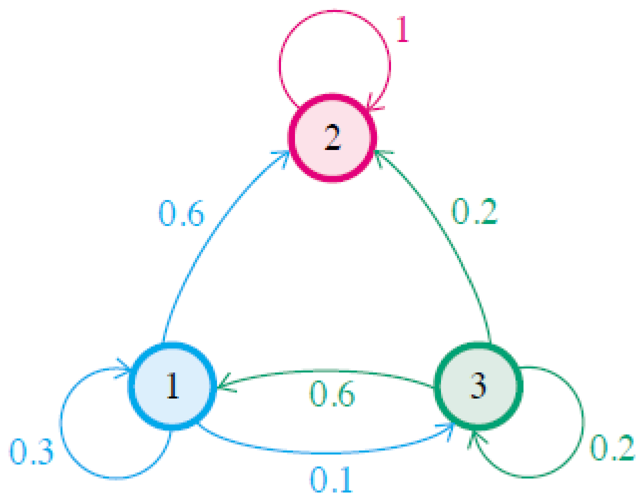
$$0 \leq P_{ij} \leq 1,$$

$$\sum_{j=1}^{m} P_{iJ} = 1, (i = 1,2,\ldots,m)$$

For example, suppose a Markov chain has transition matrix

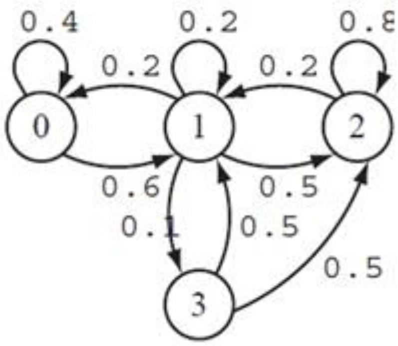
$$\begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0 & 1 & 0 \\ 0.6 & 0.2 & 0.2 \end{bmatrix} = P. \end{array}$$

The matrix shows that p_{12}, the probability of going from state 1 to state 2, is 0.6, and that p_{22}, the probability of staying in state 2, is 1. Thus, once state 2 is entered, it is impossible to leave. For this reason, state 2 is called an *absorbing state*. Figure 4 shows a transition diagram for this matrix. The diagram shows that it is not possible to leave state 2.

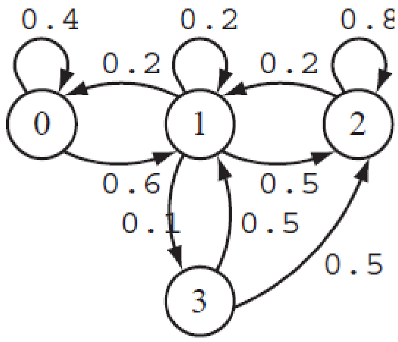


Örnek:

a) Geçiş diyagramında Markov geçiş matrisini yazınız.



Find the state transition matrix **P** for the Markov chain below.



$$\mathbf{P} \square \begin{bmatrix} 0.4 & 0.6 & 0 & 0 \\ 0.2 & 0.2 & 0.5 & 0.1 \\ 0 & 0.2 & 0.8 & 0 \\ 0 & 0.5 & 0.5 & 0 \end{bmatrix}$$

b) Geçiş matrisi düzenli mi?

P geçiş matrisinin P^2 den P^4 kuvvetinin elemanları pozitif veya sıfırdan farklı olduğundan P geçiş matrisi ile verilen Markov zinciri düzenlidir. Düzenli Markov zincirleri ise ergodiktir.

Ergodik zincir, matematik olarak herhangi bir durumdan diğer bütün durumlara geçişin mümkün olduğu bir süreci tanımlar. Bu tanımda tam bir adımda bulunma şansı yoktur, ama başlama durumuna bakmadan bir duruma erişilmiş olmalıdır.

Ergodik markov zincirini sınırlayan hal düzenli (=reguler) zincirdir. Düzenli zincir, P geçiş olasılıkları matrisinin kuvvetlerinde bulunan elemanların sıfırdan farklı ve pozitif olmasını gerektirir. Ergodik zincirin kuvvetleri alınırsa veya matriste sıfır eleman kalmayıncaya kadar kuvvetler alınırsa ergodik zincirin düzenli olduğu görülür.

Bulduğunuz geçiş matrisi a) düzenli mi ve b) ergodik mi beirleyiniz.

c) Aşağıdaki geçiş matrisi a) düzenli mi ve b) ergodik mi beirleyiniz.

$$B = \begin{bmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

d) Varsayalık ki, Markov zinciri aşağıdaki geçiş matrisine sahip olsun. Geçiş diyagramını çiziniz. Emici bir Markov zinciri ise hangi durum ölümcüldür.

$$\begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0 & 1 & 0 \\ 0.6 & 0.2 & 0.2 \end{bmatrix} \end{array} = P.$$

Tüm Markov zincirleri düzenli değildir. Aslında, Markov zincirlerinin en önemli yaşam bilimleri uygulamalarından bazıları düzenli olan geçiş matrislerini içermez. Yaşam bilimlerinde yaygın olarak kullanılan bir Markov zinciri türü, bir emici Markov zinciri olarak adlandırılır. Markov zincirlerinin fikirlerini canlı organizmaları modellemek için kullandığımızda, ortak bir durum ölümdür. Organizma bu duruma girdiğinde, ayrılması mümkün değildir.

7.6. Yörünge olasılığı

Bir yörüngenin $X_0, X_1, \ldots, X_n$ için bir değerler dizisi olduğunu hatırlayın. Markov özelliğinden dolayı, herhangi bir yörüngenin olasılığını başlangıç olasılığını ve sonraki tüm tek aşamalı olasılıkları çarparak bulabiliriz. Tüm durumların başlangış anındaki olasılık değerleri toplamı bire eşittir.

Yöründe herbir durum için başlangıç olasılık değeri ve yörüngeler verilir. Yörüngeler sürekli olmak zorundadır. Kopuk yörünge olamaz. Eğer yörünge olasılığı verilmiş ise başlangıç olasılık değeri hesaplanır.

Markov Model: Sequence Prob.

- Conditional probability

$$P(A, B) = P(A \mid B)P(B)$$

- Sequence probability of Markov model

Chain rule

$$\begin{aligned} &P(q_1, q_2, \cdots, q_T) \\ &= P(q_1)P(q_2 \mid q_1)\cdots P(q_{T-1} \mid q_1, \cdots, q_{T-2})P(q_T \mid q_1, \cdots, q_{T-1}) \\ &= P(q_1)P(q_2 \mid q_1)\cdots P(q_{T-1} \mid q_{T-2})P(q_T \mid q_{T-1}) \end{aligned}$$

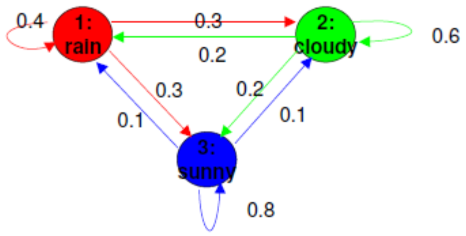
1st order Markov assumption

Örnek:

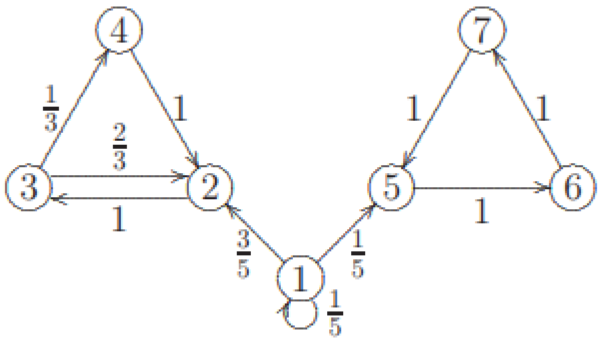
What is the probability that the weather for the next 7 days will be “sun-sun-rain-rain-sun-cloudy-sun” when today is sunny?

$S_1 : rain, \quad S_2 : cloudy, \quad S_3 : sunny$

$$
\begin{aligned}
P(O \mid \text{model}) &= P(S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3 \mid \text{model}) \\
&= P(S_3) \cdot P(S_3 \mid S_3) \cdot P(S_3 \mid S_3) \cdot P(S_1 \mid S_3) \\
&\quad \cdot P(S_1 \mid S_1) P(S_3 \mid S_1) P(S_2 \mid S_3) P(S_3 \mid S_2) \\
&= \pi_3 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23} \\
&= 1 \cdot (0.8)(0.8)(0.1)(0.4)(0.3)(0.1)(0.2) \\
&= 1.536 \times 10^{-4}
\end{aligned}
$$



Örnek:



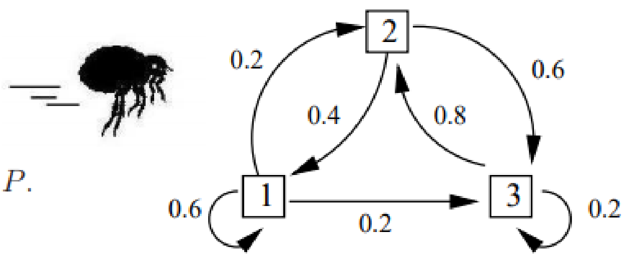
X0 ~ (3 /4 , 0, 1/ 4 , 0, 0, 0, 0).   1, 2, 3, 2, 3, 4 yörüngesinin olasılığı nedir?

P(1,2,3,2,3, 4)=P(X_0 = 1) x p12 x p23 x p32 x p23 x p34

=3/4 x 3/5 x 1 x 2/3 x 1 x 1/3=1/10

Worked Example: distribution of X_t and trajectory probabilities

Purpose-flea zooms around the vertices of the transition diagram opposite. Let X_t be Purpose-flea's state at time t $(t = 0, 1, \ldots)$.



(a) Find the transition matrix, P.

$$\textit{Answer: } P = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.4 & 0 & 0.6 \\ 0 & 0.8 & 0.2 \end{pmatrix}$$

(b) Find $\mathbb{P}(X_2 = 3 \mid X_0 = 1)$.

$$\mathbb{P}(X_2 = 3 \mid X_0 = 1) = \left(P^2\right)_{13} = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} \cdot & \cdot & 0.2 \\ \cdot & \cdot & 0.6 \\ \cdot & \cdot & 0.2 \end{pmatrix}$$

$$= 0.6 \times 0.2 + 0.2 \times 0.6 + 0.2 \times 0.2$$

$$= 0.28.$$

Note: we only need one element of the matrix P^2, so don't lose exam time by finding the whole matrix.

Şu an X_0=1 durumdan, iki sonraki durum olan X_2=3 ise $(P^2)_{13}$ hesaplanır.

İkі adımda 1 durumundan 3 durumuna nasıl gidilir? 3 yoldan gidilir.

S11 * S13 = 0.6*0.2

S13 * S33 =0.2*0.2

S12 * S23 =0.2*06

Toplam=0.28

Örnek:

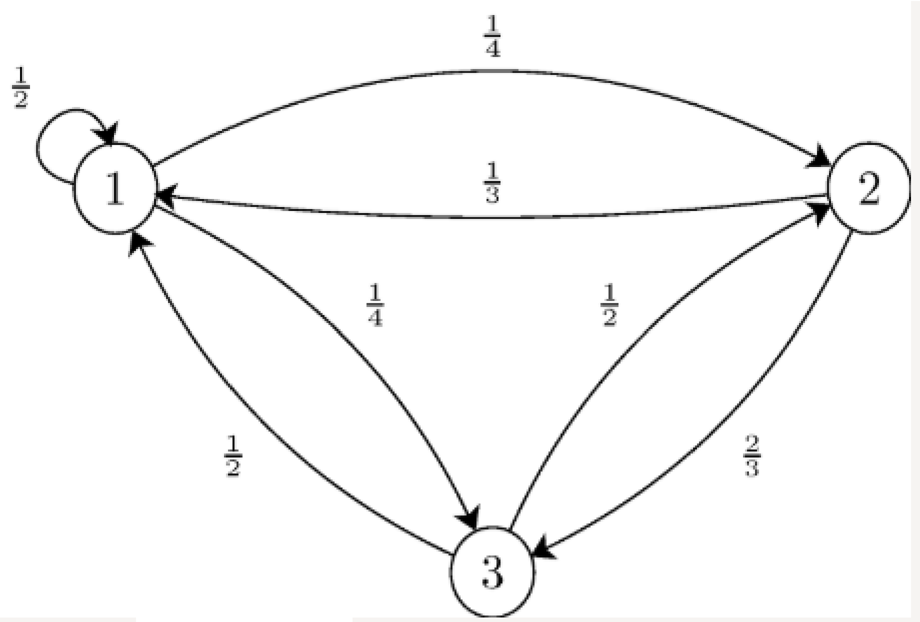
Consider the Markov chain with three states, $S = \{1, 2, 3\}$, that has the following transition matrix

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}.$$

a. Draw the state transition diagram for this chain.
b. If we know $P(X_1 = 1) = P(X_1 = 2) = \frac{1}{4}$, find $P(X_1 = 3, X_2 = 2, X_3 = 1)$.

X_1 adımında S3 durumunda iken, X_2 adımında S2 durumuna gidecek, sonunda X_3 adımında S1 durumuna gidilecek.

a)



A state transition diagram.

b) Yörüngesi S3, S2, S1 olduğuna göre S1 ve S2 durumları için başlangıç koşulu bilindiğinden önce S3 için başlangıç koşulu hesaplanır. Tüm durumların başlangış anındaki olasılık değerleri toplamı bire eşittir.

$$\begin{aligned} P(X_1 = 3) &= 1 - P(X_1 = 1) - P(X_1 = 2) \\ &= 1 - \frac{1}{4} - \frac{1}{4} \\ &= \frac{1}{2}. \end{aligned}$$

Durum-3, ardından durum-2 ardından durum-1 gelme olasılığının hesaplanması,

$$\begin{aligned} P(X_1 = 3, X_2 = 2, X_3 = 1) &= P(X_1 = 3) \cdot p_{32} \cdot p_{21} \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3} \\ &= \frac{1}{12}. \end{aligned}$$

Örnek:

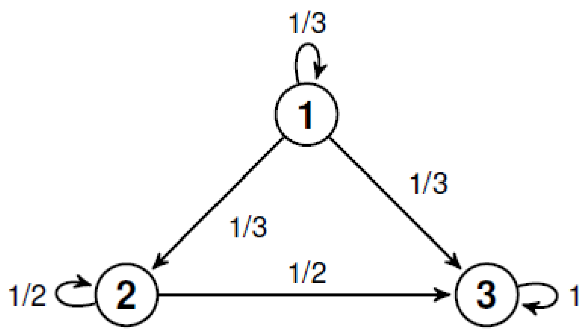
Şekildeki geçiş diyagramına karşılık gelen üç durumlu ayrık zamanlı Markov zincirini düşünün. X (0) 'ın başlangıç dağılımının f (1) = f (2) = 1/2 ile verildiğini varsayın.

Compute the following

1. P(X(0) = 1, X(1) = 2, X(2) = 3).

2. P(X(2) = i), for i = 1, 2, 3.

3. P(T3 = 2) where



P(X(0) = 1, X(1) = 2, X(2) = 3) = f(1) * P12 * P23 $= \frac{1}{2} * \frac{1}{3} * \frac{1}{2} = \frac{1}{12}$

7.7. Saklı Markov Modeli

Markov zinciri modellerinde sistemin, bir durumdan diğer durumlardan birine geçişi söz konusudur. Sistemin bulunabileceği durumlar ve durumlar arası geçişler açık bir şekilde gözlemlenebilir durumdadır. Saklı Markov Modelinde ise durumlara göre gizli isteklerin olma durumları söz konusudur.

Tarihçesi

Saklı Markov Modelinin teorisi 1970'li yıllarda Baum ve Eagon (1967), Petrie (1969) ve Baum (1972) tarafından geliştirilmiştir. Aslında Saklı Markov Modeli 1940'lı yıllarda çalışılmış fakat teori tam olarak geliştirilemediğinden dolayı uygulaması yapılamamıştır.

Saklı Markov Modeli Nedir?

Saklı Markov Modeli sinyal işlemede classifier(sınıflandırıcı) olarak kullanılan, çoklukla konuşma tanıma ve genetik kodların çözülmesinde faydalanılan stokastik bir modeldir. Bir başka deyişle, mevcut durumun açıklaması, sürecin gelecekteki evrimini etkileyebilecek tüm bilgiyi kapsar. Markov Analizi, gelecekteki davranışları tahmin etmek için belirsizlik ortamında kullanılan güçlü bir tahmin etme tekniğidir.

Saklı Markov Modeli ve Tahmin Algoritmaları (Hidden Markov Model and Estimation Algorithms)

Günümüzde, belirsizlik altında karar alma ve ileriye yönelik tahminde bulunma birey, firma ve devlet gibi iktisadi karar birimlerinin karşı karşıya olduğu en temel sorunlardandır. Bir sistemin belirsizliği konu alması, sözü edilen sistemin tamamen kontrol altına alınamamasından kaynaklanır. Saklı Markov Modelindeki durum dizisini bulabilmek için üç önemli algoritma kullanılır:

- Tanıma Problemi - İleri Yön Algoritması
- Durum Dizisinin Bulunması Problemi - Viterbi Algoritması
- Model Parametrelerinin Öğrenilmesi - Baum Welch Algoritması

İleri Algoritması: Modeldeki durumların sırasını bulmada kullanılıyor. Ortaya çıkabilecek tüm durum sıralarının olasılıkları toplanıyor.

Viterbi Algoritması: İleri Algoritması'ndaki gibi tüm olasılıkları toplamak yerine, Viterbi algoritmasında her durum sıralarından olasılık vektörleriyle en iyi örtüşeni seçiliyor. Böylece daha sağlıklı bir sonuç elde ediliyor.

Baum-Welch Algoritması: Gözlem dizisini baştan sona ve tekrar sondan başa geçerek gözlem olasılıklarını hesaplar. Böylece daha kesin sonuçlar bulur.

Saklı Markov Modeli Parametreler

Yandaki şekilde görülen Saklı Markov Modeli'nin olasılık parametreleri şu şekildedir:

x — durumlar

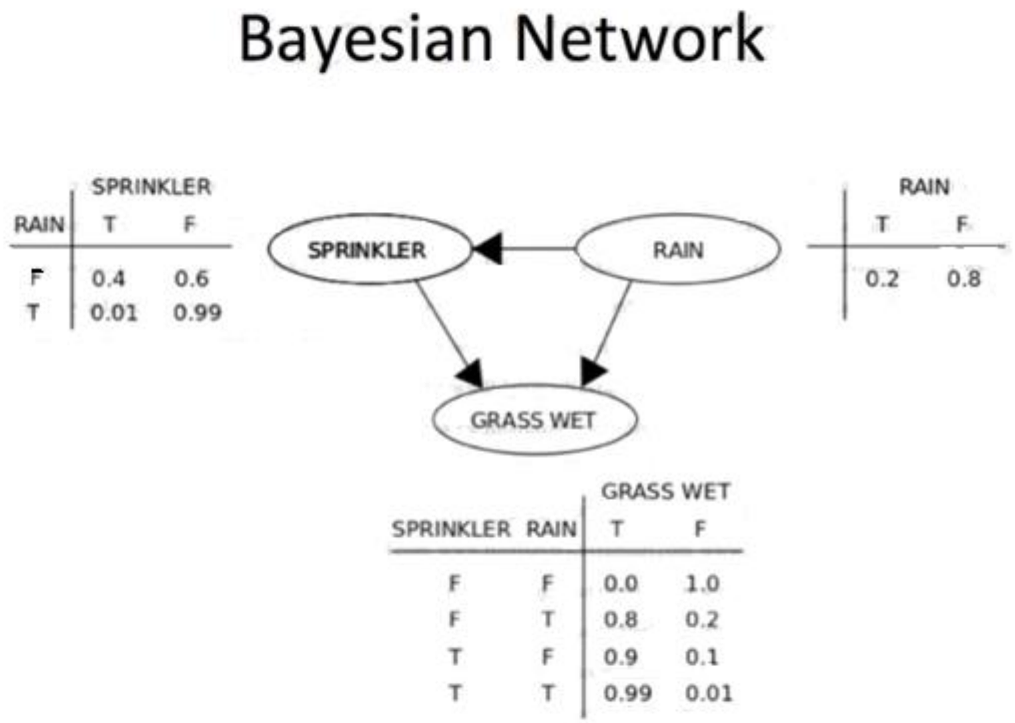
y— olası gözlemler

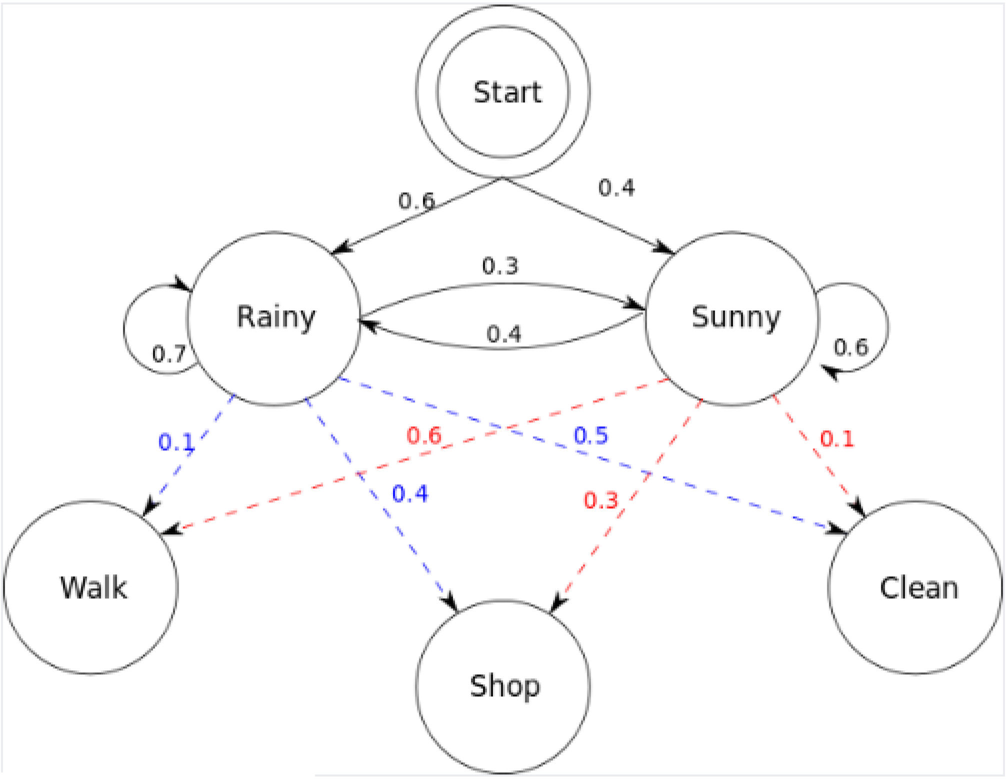
a— geçiş olasılıkları durumu

b— çıkış olasılıkları

Normal Markov Model'inde, durumlar, gözlemci için görünebilirdir ve bu yüzden tek parametre, durum geçiş olasılıklarıdır. Saklı Markov Modelinde, durum, direkt olarak görünebilir değildir, ama duruma bağlı çıkışlar görünürdür.

Uzakta yaşayan iki arkadaş Kemal ve Feyza , her gün telefonla o gün ne yaptıkları hakkında konuşurlar. Kemal sadece 3 faaliyetle ilgilenir; parkta yürümek, alışveriş ve evini temizlemek. Sadece o günkü hava durumuna göre yapılacak olan iş belirlenmektedir Feyza, Kemal'in yaşadığı yerdeki hava durumu hakkında kesin bir bilgiye saip değildir, Fakat genel yönelmeleri bilmektedir. Kemal'in her gün ne yaptığına dayanarak, Feyza oradaki hava durumunu tahmin etmeye çalışır.





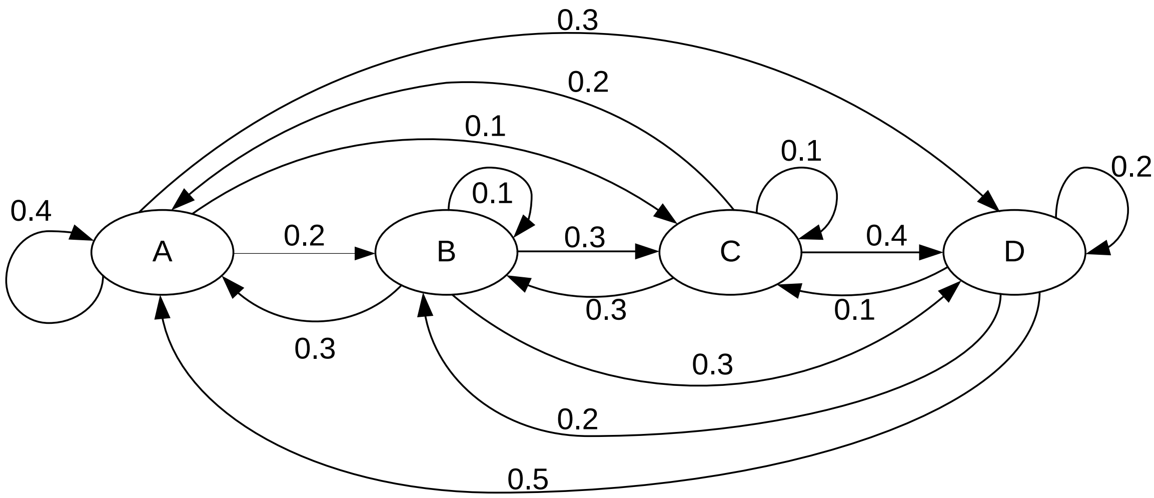
			Walk	
Start	Rainy	Sunny	True	False
0	0	0	k	1-k
0	0	1	k	1-k
0	1	0	k	1-k
0	1	1	k	1-k
1	0	0	k	1-k
1	0	1	0.4	0.6
1	1	0	0.1	0.9
1	1	1	k	1-k

Markov zincir modeli ile Bayesian metotunun bütünleştirilmesi:

Örnek: Bir fabrika üç makine bulunmaktadır. Bir makinede A, B, C ürünlerini üretmektedir. Markov Zinciri durum geçiş diyagaramı aşağıda verilmiştir.

	Bir sonraki durum			
Şu anki durum	A	B	C	D
A	0.4	0.2	0.1	0.3
B	0.3	0.1	0.3	0.3
C	0.2	0.3	0.1	0.4
D	0.5	0.2	0.1	0.2

Durum geçiş diyagramı:

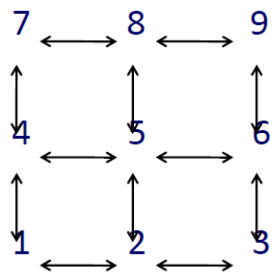


Ürünlerin hatalı ve sağlam oranları aşağıda verilmiştir.

	Sağlam	Bozuk
A	%90	%10
B	%80	%20
C	%70	%30
D	%60	%40

7.8. Markov Zinciri Uygulama

Örnek : Aşağıdaki şekildeki sayılar köşe noktaları veya dönüşleri belirleyen kavşakları ve aradaki çizgiler de yolları belirlemektedir. Bir arabanın dönüş veya doğrudan gitmesini eş olasılıkla varsayarak köşelerde bulunmak isteğini geçiş olasılıkları matrisi ile gösteriniz.



$$
\begin{array}{c}
\\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9
\end{array}
\begin{array}{c}
\begin{array}{ccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9
\end{array} \\
\begin{bmatrix}
0 & 1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 \\
0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\
1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\
0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 & 1/4 & 0 \\
0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 1/3 \\
0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 \\
0 & 0 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 1/3 \\
0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 & 0
\end{bmatrix}
\end{array}
$$

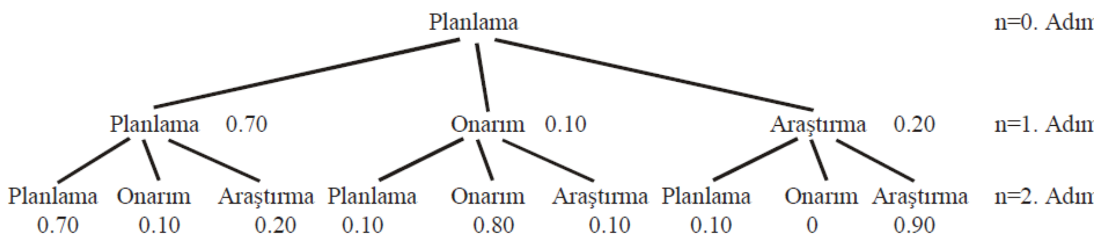
2 nolu köşede bulunması halinde 1, 3 veya 5 köşelerinde bulunma olasılığı 1/3 olacaktır. 5 nolu köşede ise takiben 2, 4, 6 veya 8 köşelerine 1/4 olasılıkla gidebilir v.s.

Geçiş matrisi aşağıdaki gibidir ve mevcut herhangi bir durumdan, verilen herhangi bir duruma geçilir. Dolayısıyla süreç ergodiktir.

Örnek:

Şu anda (n=0) Planlamada olan bir mühendisin iki yıl sonra (n=2) Onarım Bölümünde olma ihtimali nedir?

	P	O	A
P	0,7	0,1	0,2
O	0,1	0,8	0,1
A	0,1	0	0,9



P->P->O = 0,7*0,1=**0,07**; P->O->O = 0,1*0,8=**0,08**; P->A->O = 0,2*0=**0**

Planlama Bölümünde çalışan mühendisin ikinci yılda Planlama, Onarım ve Araştırma Bölümlerine atanma olasılıkları:

	P	O	A
P	0,7	0,1	0,2
O	0,1	0,8	0,1
A	0,1	0	0,9

$$V_i^n = V_i^{n-1}.P$$

$$V_1^2 = V_1^1.P = (0.7\ \ 0.1\ \ 0.2)\,.\begin{vmatrix} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0 & 0.9 \end{vmatrix} = (0.52\ \ 0.15\ \ 0.33)$$

Daha genel olarak bu problemde, n=2 yıl sonraki bütün geçiş ihtimallerini bilmek istersek P matrisinin karesi alınır:

	P	O	A
P	0,7	0,1	0,2
O	0,1	0,8	0,1
A	0,1	0	0,9

$$P^2 = \begin{vmatrix} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0 & 0.9 \end{vmatrix} . \begin{vmatrix} 0.7 & 0.1 & 0.2 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0 & 0.9 \end{vmatrix} = \begin{vmatrix} 0.52 & 0.15 & 0.33 \\ 0.16 & 0.65 & 0.19 \\ 0.16 & 0.01 & 0.83 \end{vmatrix}$$

n. Adım Sonunda Her Bir Grupta Kaç Kişi Bulunur?

$$m = n.P^n$$

n=(n_1, n_2, ...): dönem başı mevcutlar vektörü

m=(m_1, m_2, ...): dönem sonu mevcutlar vektörü

n. Adım Sonunda Her Bir Grupta Kaç Kişi Bulunur?

Dönem başı personel durum mevcutları n=(100, 80, 120) vektörü ile verilirse 2. yıl sonunda gruplar arasındaki dağılım şöyle bulunabilir:

$$m = (100\ 80\ 120) . \begin{vmatrix} 0.52 & 0.15 & 0.33 \\ 0.16 & 0.65 & 0.19 \\ 0.16 & 0.01 & 0.83 \end{vmatrix} = (84\ \ 68\ \ 148)$$

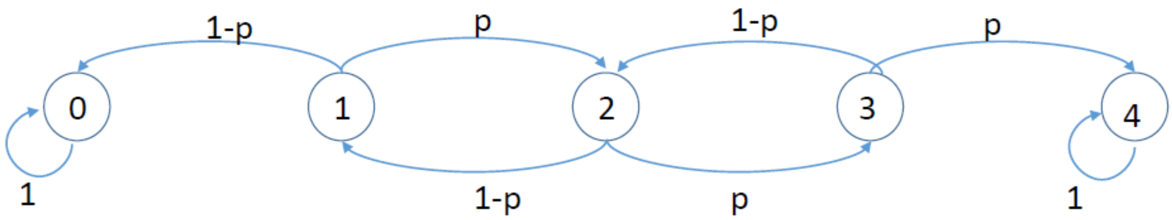
Kumarbazın İflası Problemi

Oyuna başlagıç anında kumarbaz 2 TL'ye sahiptir. 1,2,… zamanlarında kumarbaz oyun oynar ve 1TL bahse girer ve p olasılıkla oyunu kazanır ve (1-p) olasılıkla oyunu kaybeder. Burada hedef 4 TL sahibi olunca oyunu bitirmektir. Dikkat edilirse elde 0 TL kalınca da oyun bitmektedir.

$X_0, X_1, \ldots, X_t$ ayrık zamanlı stokastik süreç olarak ortaya çıkar. $X_0 = 2$ bilinmektedir ve sabittir. Fakat X_1 ve sonra X_t değerleri rassaldır.

Örneğin p olasılıkla $X_1=3$ ve (1-p) olasılıkla $X_1=1$ olur. Eğer X_t , 0 veya 4 değerlerine eşit değilse p olasılıkla $X_{t+1}=X_t + 1$ ve 1-p olasılıkla $X_{t+1}=X_t-1$ olur.
Eğer $X_t=0$ ise X_{t+1} ve daha sonraki X_t değerleri 0'a eşittir
Eğer $X_t=4$ ise X_{t+1} ve daha sonraki X_t değerleri 4'e eşittir

t+1'deki para t zamanına kadar birikmiş paraya (t zamanındaki paraya) bağlı olduğundan bu süreç bir Markov zinciridir. Oyunun kuralları zamanla değişmediği için bu aynı zamanda stasyoner(sabit) Markov zinciridir.



Geçiş matrisi

$$
\begin{array}{cc|ccccc|}
 & & \multicolumn{5}{c}{\text{Durum}} \\
 & & \$0 & \$1 & \$2 & \$3 & \$4 \\
 & 0 & 1 & 0 & 0 & 0 & 0 \\
 & 1 & 1-p & 0 & p & 0 & 0 \\
P = & 2 & 0 & 1-p & 0 & p & 0 \\
 & 3 & 0 & 0 & 1-p & 0 & p \\
 & 4 & 0 & 0 & 0 & 0 & 1
\end{array}
$$

8. Algoritmik Olasılık - Kolmogorov

Gözlemlerden ya da deneylerden elde edilen verilerin tablolar halinde yazılması bir teori yaratmıyor. Söz konusu ham verilerin yorumlanarak, herkesin anlayacağı kısa bir dille anlatılması gerekiyor. O da yetmiyor, teorinin gelecekte olacaklar hakkında bilgi içermesi gerekiyor. Gezegenin yörüngesini biliyorsam, onun ne zaman nerede olacağını hesaplayabiliyorum. Bu iş, kehanetten çok farklı bir şeydir. Bunlar olduğunda, ham veriler bir teoriye dönüşmüş oluyor. Elbette, toplanan verilerin duyarlığı, kullanılan gözlem/deney aletlerinin gelişmişliğine bağlı olduğu gibi, verilerin yorumlanması da bilim adamının bilgi ve yetenekleriyle sınırlıdır. Şu anda, bir teorinin doğru ya da yanlışlığı amacımız için önem taşımıyor. Yanlış teoriler, nasıl olsa, bir gün bilimsel bilgilerin biriktiği ambardan atılacaktır. 8 Bilimin gücü burada yatar. Bilimin bilgi ambarı çok dinamiktir, yanlış olduğu kanıtlanan teoriler hemen yerlerini yeni teorilere kendiliğinden bırakırlar. Şimdi, bir teori kurma olgusunu algoritmik seçkisizlik kavramıyla ifade edeceğiz. Algoritmik seçkisizlik tanımını vermeden önce, Solomonoff‟un bilimsel teoriyi açıklamak için kullandığı "inductive inference" yönteminden söz etmeliyiz. Bilim adamı bir sürü deney/gözlem yapar. Bunları bitlerden oluşan bir dizi (mesaj) olarak düşünelim. Bilim adamı bu mesajı iletmek istemektedir. Bu mesajı gönderen en az bir tane algoritma vardır ve o da dizinin kendisidir. Bundan başka algoritmalar da olabilir. Bilim adamı mesajı gönderen bir algoritma kurmuş olsun. Algoritma, ilettiği mesajdan kısa değilse bir teori olamaz. Algoritma ilettiği mesajdan daha kısa ise, diziyi aynen iletmekle kalmayıp gelecek gözlemler için de öngörü yapıyorsa, bu algoritma bir teoridir. Bu koşulu sağlayan birden çok algoritma varsa, daima en kısa (bit sayısı en az) olan algoritma tercih edilir. Bu tercih Occam's razor diye bilinir: Aynı işi yapan teoriler arasından en basiti tercih edilmelidir.

Algoritmik Seçkisizlik: Yukarıdaki örneklerden hereketle, Chaitin ve Kolmogorov, seçkisiz diziyi şöyle tanımladılar:
Kendisinden daha kısa bir algoritma ile yazılamayan dizi seçkisizdir.

Bu tanım, sezgisel kavrama dayalı olasılık kuramını sağlam bir temel üzerine taşımaktadır. Elbette, seçkisizliğin bu yeni tanımı, olasılık kuramının aksiyomlarını yoketmiyor, olasılığın hiç bir uygulamasını değiştirmiyor, yalnızca temeli sağlamlaştırıyor. Olasılık açısından peşinde olduğumuz şey, gözlemlerden çıkan seçkisiz bir $x1$, $x2$, $x3$, ... , xn dizisi verilmişken, bir sonraki terimin, yani $xn+1$ teriminin ne olacağını öngörebilmektir. $x1$, $x2$, $x3$, ... , xn dizisini ileten ve $xn+1$ teriminin ne olacağını öngören ve diziden kısa olan algoritma bir teoridir.

Kolmogorov Karmaşıklığı (complexity):
Verilen bir diziyi ileten sonsuz sayıda algoritma kurulabilir. Örneğin, "233 e 1 ekle", "235 ten 1 çıkar", 117 yi 2 ile çarp" gibi algoritmaların hepsi 234 dizisini iletir. Bu tür algoritmalardan sonsuz sayıda yazılabileceği açıktır. Bizim için ilginç olanı en küçük olanıdır. Aynı diziyi ileten algoritmalar arasında en kısa olana minimal algoritma diyeceğiz. Bir dizi için bir tane minimal algoritma olabileceği gibi, bir çok minimal algoritma da olabilir.

İlettiği dizi ister seçkili, ister seçkisiz olsun minimal bir algoritmanın kendisi daima seçkisiz olmak zorundadır. Aksi taktirde, onu ileten daha kısa bir algoritma var olur ve dolayısıyla söz konusu algoritma minimal olamaz.

Karakterlerden (harf ve simgeler) oluşan bir s stringi düşünelim. s stringini yazdıran bir P programına s stringini iletiyor diyelim. P nin uzunluğu, P içindeki karakterlerin sayısıdır.

s stringini ileten en kısa P programının uzunluğuna S stringinin karmaşıklığı denir.

s stringini, yukarıdakiler gibi bitlerden oluşan bir dizi olarak düşünürsek, bu dizinin karmaşıklığı o diziyi ileten minimal algoritmanın uzunluğuna eşit olur. Buradan, seçkisizlik için şu denk tanımı elde ederiz:

Bit sayısı Kolmogorov karmaşıklığına eşit olan dizi seçkisizdir. Tabii, buradaki eşitlik yaklaşıklık anlamındadır. Dizilerin bit sayıları çok çok büyüdüğünde, aradaki farkın önemi kalmamaktadır. Kolmogorov karmaşası, seçkisizliği tanımlamakla kalmıyor, seçkisizliğin ölçümünü de veriyor.

Seçkisiz Sayıların Çokluğu:
Klasik anlamda, olasılık ile seçkisizlik (randomness) eşanlamlıdır. Seçkisiz bir süreçte, olayların (çıktıların) olma olasılıkları birbirlerine eşittir. Başka bir deyişle, bir süreçte seçkisizlik "amaç, neden, sıra ve öngörü yokluğu" diye tanımlanabilir. Bu nedenle, seçkisiz süreç, çıktısı öngörülebilir bir biçime (pattern) sahip olmayan ardışık oluşumlar zinciridir. Özel olarak, istatistikte, yanlı (bias) olmayan ya da bağlılaşımlı (correlated) olmayan olayları belirlemek için kullanılır. İstatistiksel seçkisizlik, daha sonraları informasyon kuramında informasyon entropisi kavramı içine alınmıştır.

Seçkisizlik kavramı, başlangıçta şans oyunlarından çıkmıştır. Örneğin zar atma, rulet oyunu, oyun kartlarını karma vb. Daha sonra yapılan elektronik kumar makinalarında da seçkisiz sayı üretimi işin esasıdır. Ama bu işte çok hile yapılabileceği için, bu tür oyun makinaları bir çok ülkede ya yasaklıdır ya da devletin sıkı denetimi altındadır. Bizde olduğu gibi, bazı ülkelerde, seçkisiz sayı üretimine dayalı piyangolar ülke genelinde serbestçe oynanabilir. Bundan farklı

olarak, çıktısı önceden öngörülemeyecek spor karşılaşmaları, at yarışları vb. oyunlar da seçkisizliğin (olasılığın) ilgi alanındadır.
Olasılık kavramının geçtiği her yerde, olabilecek olayları sayılarla ifade etmek mümkündür. Dolayısıyla, konu, esasında seçkisiz sayı üretimine dayalıdır. Stephan Wolfram' a göre, seçkisiz sayı üretme işi üç ayrı sınıfa ayrılabilir. Bu üç sınıfta üretilen sayıların nitelikleri birbirlerinden farklıdır.

1. *Çevreden gelen seçkisizlik*. Örneğin, çoğalmayı açıklayan hareket (Brownian motion).
2. *Başlangıç koşullarına hassas bağlı seçkisizlik.* Örneğin, kaos.
3. *Sözde seçkisizlik.* Bu sayılar tasarlanan bir sistem tarafından üretilir. Örneğin, bir bilgisayarla üretilen seçkisiz sayılar... Bu tür sayılar belli bir algoritma ile üretilir. Algoritma çok ağır hesaplamalara dayandırılarak, çıktı hiç kimsenin öngöremeyeceği duruma kolayca getirilebilir. Ama üretilen sayılar gerçek anlamıyla seçkisiz sayılamaz.

1960 yılında *Solomonoff*, bilimsel teorinin basit bir açıklamasını vermeye uğraşırken, *algoritmik olasılık* kavramını ilk ortaya atan kişidir. Bundan 5 yıl sonra, Solomonoff'dan ve birbirlerinden habersiz olarak *Kolmogorov* ve *Chaitin* aynı algoritmik seçkisizlik kavramını ortaya koydular. Kolmogorov o zamanının en ünlü matematikçilerinden birisidir, Chaitin ise henüz üniversitede matematik bölümü son sınıf öğrencisidir. Bu öğrencinin, daha sonra yaptığı çalışmalar olasılığa ve informasyon teorisine büyük katkılar sağlayacaktır.

***"Algoritmik seçkisizlik"* ya da *"algoritmik olasılık"* kavramı:**
Örnek 1. Dünyadaki Uzay Merkezi (UM) çok uzaktaki bir gezegene bir araştırmacı göndermiştir. Dalgın araştırmacımız, yapacağı hesaplar için kendisine mutlaka gerekli olan trigonometri cetvelini yanına almayı unutmuştur ve onun bir iletişim aracıyla kendisine acele gönderilmesini istemektedir. Bu uzak gezegenle telgraf, telefon, faks vb iletişim araçlarıyla iletilen mesajların çok pahalıdır. UM, pahalı iletişim ücretini ödemeyi göze alarak, *sin, cos, tan, cot, sec, cosec* fonksiyonları için hazırlanmış, yirmi haneli geniş bir trigonometri cetvelini bir iletişim aracıyla ile göndermek zorunda kalmıştır. Ama UM'de bir matematikçi varsa, işi çok ucuza getirebilir. Koca bir kitap olan trigonometri cetvelini göndermek yerine, *exp(ix) =cosx + isinx* formülünü göndermesi sorunu çözecektir. Bu kısa mesaj, araştırmacının istediği bütün bilgiyi içermektedir.

Örnek 2. Aradan binlerce yıl geçmiş olsun. Bilginimiz yorulmuştur ve hobilerine biraz zaman ayırmak için geçmiş yıllara ait basketbol maçlarını, skorları ve kimin hangi maçta kaç sayı yaptığını bilmek istemektedir. UM bu isteği çok haklı görmüş ve istenen bilgilerin gönderilmesini emretmiştir. Bu kez, matematikçiler de dahil olmak üzere, hiç kimse istenen maçlarla ilgili bilgileri tamamen içeren daha kısa bir mesaj (formül) yazamamıştır. Çaresiz, yüksek ücretler ödenerek, istenen bilgi gönderilecektir.

Bu iki örnekten çıkardığımız sonuç şudur. Bazı mesajları, anlamını aynen koruyarak, kısaltabiliriz. Bazı mesajları asla kısaltamayız.

Algoritmik Seçkisizlik tanımı, yukarıda verilen tanımda olduğu gibi insan sezgisine dayalı olmasın diye bilgisayar terminolojisine dayandırılacaktır. Günümüz bilgisayarları ikili (binary) sayıtlama dizgesine dayanır. İkili (binary) sayıtlama dizgesinde yalnızca 0 ve 1 sayakları (digit) vardır. Bilgisayar terminolojisinde ikili sayı sistemindeki hanelere *bit* denir. Bir bit'te (hanede) ya 0 ya da 1 sayağı yer alır.

İkili sayı dizgesini kullanarak her bilgiyi (mesajı) karşı tarafa gönderebiliriz. Başka bir deyişle, 0 ile 1 lerden oluşan dizilerle istediğimiz her bilgiyi yazabiliriz. Bunun için, örneğin, bir dildeki harfleri, kelimeleri, cümleleri, kavramları,... vb 0 ile 1 lerin belirli bir sırada sıralanmasıyla oluşan birer diziye karşılık getirmek yetecektir. Diziler sonlu ya da sonsuz olabilir. Mesajın ne kadar uzağa gideceğinin ve mesajın anlamının, şu andaki hedefimiz için bir önemi yoktur. O nedenle, mesajları 0 ile 1 lerden oluşan diziler olarak, uzak gezegeni de bilgisayarın çıktısı olarak düşüneceğiz. Amacımız, mesajın (dizinin) bilgisayar çıktısı olarak elde edilmesidir. O zaman mesajı yerine iletilmiş varsayacağız. Mesajı iletmek için, bilgisayara komutlar vermeliyiz. Verilecek komutlar herhangi bir bilgisayar dilinde yazılmış bir programdır. Biz buna *algoritma* diyeceğiz. Fiziksel kısıtlamaları yok sayıp, mesajın gönderilmesi için gerekli zamanın olduğunu ve algoritma doğru ise mesajın daima yerine ulaşığını varsayalım.

Bilgilerimize ya da sezgilerimize dayalı olarak bir dizi hakkında vereceğimiz *seçkili/seçkisiz* kararlarımızın ne kadar yanıltıcı olabileceğine bir çok örnek gösterebiliriz. Örneğin, 3,1451... dizisini gören iki kişi düşünelim. Bunlardan birisi *pi* sayısını biliyor olsun, ötekisi bilmiyor olsun. Birinci kişi bu diziyi istençle yazılmış (seçkili) bir dizi olarak, yani *pi* sayısı olarak algılarken, ikinci kişi bunu tamamen rasgele dizilmiş (seçkisiz) bir dizi olarak görebilir.

Kolmogorov, seçkisiz diziyi şöyle tanımladılar:
Kendisinden daha kısa bir algoritma ile yazılamayan dizi seçkisizdir.
Bu tanım, sezgisel kavrama dayalı olasılık kuramını sağlam bir temel üzerine taşımaktadır. Elbette, seçkisizliğin bu yeni tanımı, olasılık kuramının aksiyomlarını yoketmiyor, olasılığın hiç bir uygulamasını değiştirmiyor, yalnızca temeli sağlamlaştırıyor. Olasılık açısından peşinde olduğumuz şey, gözlemlerden çıkan seçkisiz bir x_1 , x_2 , x_3 , ... , x_n dizisi verilmişken, bir sonraki terimin, yani x_{n+1} teriminin ne olacağını öngörebilmektir. x_1 , x_2 , x_3 , ... , x_n dizisini ileten ve x_{n+1} teriminin ne olacağını öngören ve diziden kısa olan algoritma bir teoridir.

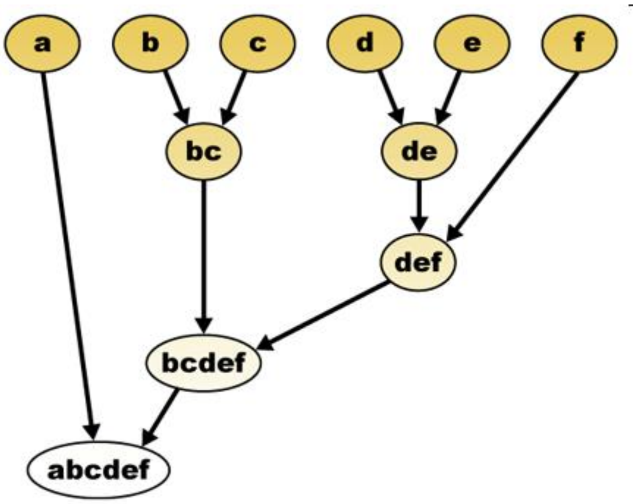
Bazı dizilerde tekrarlanan patternler olabilir. *d* dizisi *s* içinde periyodik olarak tekrarlanan bir altdizi olsun. Örneğin, *s = 01010101010101010101* dizisi *d = 01* altdizisinin *10* kez tekrarlanmasıyla oluşmuştur. *P ="n kez d yaz"* algoritması *s* dizisini ileten algoritmalardan birisidir. Ohalde, *s* dizisinin algoritmik karmaşıklığı, *P* algoritmasının uzunluğundan büyük olamaz. Asıl amacımız, *P* nin uzunluğu ile *s* dizisinin bit uzunluğunu karşılaştırmaktır. Bu karşılaştırma bize, *s* dizisinin seçkisiz olup olmadığı konusunda bir ölçü verecektir.

Önce *P* algoritmasının uzunluğunu irdeleyelim. *n* sayısının algoritmik karmaşıklığı yaklaşık olarak log_2n dir. Yaklaşık diyoruz, çünkü algoritmanın gerçek uzunluğu kullanılan makina diline bağlıdır. Yeterince büyük *n* sayıları için log_2n sayısı *n* sayısından çok küçüktür, dolayısıyla algoritmanın uzunluğu *s* dizisinin uzunluğu ile karşılaştırılırken göreli olarak ihmal edilebilir. Geriye kalan *"kez"* ve *"yaz"* stringlerinin uzunluğu zaten yok denilecek kadardır, onlar da ihmal edilebilir. Ohalde, *P ="n kez d yaz"* algoritmasının uzunluğunu, *s* dizisinin uzunluğu ile karşılaştırırken belirleyici olan tek etmenin *d* dizisinin uzunluğu (bit sayısı) olduğu sonucuna varırız.

Buradan yola çıkarak *n* bit uzunluğundaki dizilerin algoritmik karmaşıklıklarını *n-1, n-10, n-100, n-1000, ...* gibi sınıflara ayırabiliriz. Artık *P* nin uzunluğu ile *d* nin uzunluğunu yaklaşık eşit sayarak, aşağıdaki inductive yöntemi uygulayabiliriz.

Uzunluğu *1* olmak üzere *n* bitlik dizi ileten kaç tane algoritma vardır? *"n kez 0 yaz"* ve *"n kez 1 yaz"* algoritmaları bu işi yapan iki algoritmadır. Birincisi *000...0* dizisini, ikincisi ise *111...1* dizisini iletir. Bu algoritmaların ilkinde *d* dizisi yalnızca *'0'* dan, ikincisinde ise yalnızca *'1'* den ibarettir. Her ikisinin de uzunluğu *1* bittir. Bir bitlik başka algoritma yoktur. 1 bitlik algoritmaların sayısını 2^1 biçiminde gösterebiliriz. Benzer olarak, 0 ile 1 sayaklarından elde edilecek 2 bitlik dizilerin sayısı 4 dür: 00, 10, 11, 10. Ohalde, İki bitlik algoritmaların sayısı 2^2 dir. Benzer düşünüşle, üç bitlik algoritmaların sayısı 2^3 *, ... , n-11* bitlik algoritmaların sayısı 2^{2-11} olacaktır. Bunların toplamı $(2^1 + 2^2 + 2^3 + ... 2^{n-11}) = 2^{n-10} -2$ dir. Demek ki, uzunluğu *n-10* dan az olan algoritmaların sayısı 2^{n-10} dan daha azdır. Öte yandan *n* bit uzunluğundaki dizilerin sayısı 2^n dir. Görüldüğü gibi, bunlar arasında ancak 2^{n-10} tanesinin algoritmik karmaşıklığı *n-10* dan küçüktür. $2^{n-10} / 2^n = 1 / 1024$ olduğuna göre, *1024* diziden ancak *1* tanesinin algoritmik karmaşası *n-10* dan küçüktür. Bundan anlaşılıyor ki seçkili sayılar çok seyrektir, sayıların çoğunluğu seçkisizdir.

Yukarıda yaptıklarımızdan şu sonuç çıkmaktadır: Bir dizi verildiğinde onun seçkili olduğunu göstermek için, diziyi ileten ve diziden daha kısa olan bir algoritma olduğunu göstermek yetecektir. Bulunacak bu algoritmanın minimal olması gerekmiyor. Ama bir dizinin seçkisiz olduğunu göstermek için onu ileten daha kısa bir algoritmanın var olmadığını göstermek gerekir.



Intuitions about probability:

i. Since $0 \le f_n(A) \le n$ we have $0 \le r_n(A) \le 1$. Thus the probability of A should be in [0, 1].

ii. $f_n(\emptyset) = 0$ and $f_n(Everything) = n$. Thus the probability of $\emptyset$ should be 0 and the probability of *Everything* should be 1.

iii. Let B be *Everything* except A. Then $f_n(A) + f_n(B) = n$ and $r_n(A) + r_n(B) = 1$. Thus the probability of A plus the probability of B should be 1.

iv. Let $A \subseteq B$. Then $r_n(A) \le r_n(B)$ and thus the probability of A should be no bigger than that of B.

v. Let $A \cap B = \emptyset$ and $C = A \cup B$. Then $r_n(C) = r_n(A) + r_n(B)$. Thus the probability of C should be the probability of A plus the probability of B.

vi. Let $C = A \cup B$. Then $f_n(C) \le f_n(A) + f_n(B)$ and $r_n(C) \le r_n(A) + r_n(B)$. Thus the probability of C should be at most the sum of the probabilities of A and B.

vii. Let $C = A \cup B$ and $D = A \cap B$. Then $f_n(C) = f_n(A) + f_n(B) - f_n(D)$ and thus the probability of C should be the probability of A plus the probability of B minus the probability of D.

The probability space:

Definition

A probability space is a tuple $(\Omega, \mathcal{F}, P)$ where

- Ω is the sample space or set of all elementary events
- $\mathcal{F}$ is the set of events (for our purposes, we can consider $\mathcal{F} = \mathcal{P}(\Omega)$)
- $P : \mathcal{F} \to \mathbb{R}$ is the probability function

The probability (P) of some event (E), denoted P(E), is defined with respect to a "universe" or sample space (Ω) of all possible elementary events in such a way that P must satisfy the Kolmogorov Axioms.

- **First axiom**

  For any set *E,* the probability of an event set is represented by a real number between 0 and 1.

- **Second axiom**

  The probability that some elementary event in the entire sample set will occur is 1, or certainty. More specifically, there are no elementary events outside the sample set.

- **Third axiom**

  The probability of an event set which is the union of other disjoint subsets is the sum of the probabilities of those subsets. This is called ***σ-additivity***. If there is any overlap among the subsets this relation does not hold.

Kolmogorov's axioms

Kolmogorov formulated three axioms that the probability function P must satisfy. The rest of probability theory can be built from these axioms.

A1 For any $A \in \mathcal{F}$, there is a nonnegative real number $P(A)$

A2 $P(\Omega) = 1$

A3 Let $\{A_n \mid 1 \leq n\}$ be a collection of pairwise disjoint events. Let $A = \bigcup_{n=1}^{\infty} A_n$ be their union. Then

$$P(A) = \Sigma_{n=1}^{\infty} P(A_n)$$

9. Durumsal Mantık Ugulamaları

Digital Logic Basics:

Gate	Symbol	Truth-Table	Expression
NAND		see below	$Z = \overline{X \bullet Y}$
AND		see below	$Z = X \bullet Y$
NOR		see below	$Z = \overline{X + Y}$
OR		see below	$Z = X + Y$

NAND

X	Y	Z
0	0	1
0	1	1
1	0	1
1	1	0

AND

X	Y	Z
0	0	0
0	1	0
1	0	0
1	1	1

NOR

X	Y	Z
0	0	1
0	1	0
1	0	0
1	1	0

OR

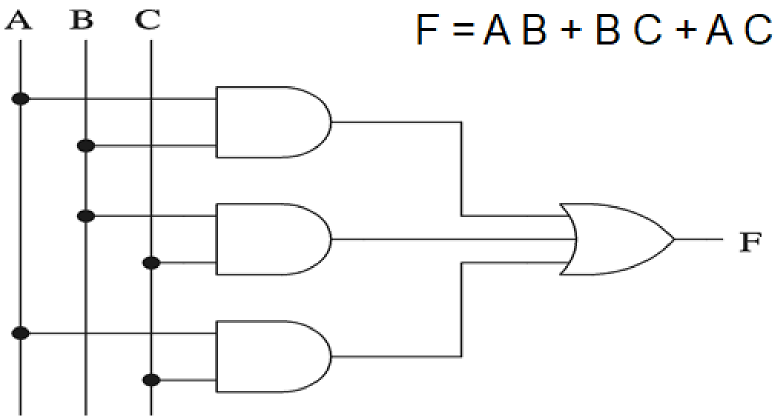
X	Y	Z
0	0	0
0	1	1
1	0	1
1	1	1

Logical functions can be expressed in several ways:

- Truth table
- Logical expressions
- Graphical form

3-input majority function

A	B	C	F
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	1



- Logical expression form

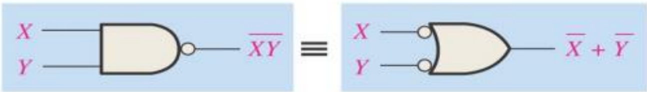
  $F = A\,B + B\,C + A\,C$

Laws and Rules of Boolean Algebra

- Laws of Boolean Algebra
  - Commutative Law
    - Commutative Law of Addition: $A + B = B + A$
    - Commutative Law of Multiplication: $AB = BA$
  - Associative Law
    - Associative Law of Addition: $A + (B + C) = (A + B) + C$
    - Associative Law of Multiplication: $A(BC) = (AB)C$
  - Distributive Law
    - $A(B + C) = AB + AC$

Demorgan's Theorems

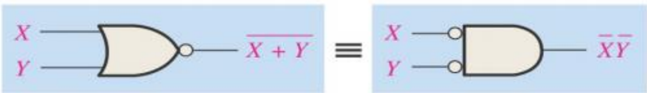
$\overline{XY} = \overline{X} + \overline{Y}$



NAND — Negative-OR

Inputs		Output	
X	Y	$\overline{XY}$	$\overline{X} + \overline{Y}$
0	0	1	1
0	1	1	1
1	0	1	1
1	1	0	0

$\overline{X + Y} = \overline{X}\,\overline{Y}$



NOR — Negative-AND

Inputs		Output	
X	Y	$\overline{X + Y}$	$\overline{X}\,\overline{Y}$
0	0	1	1
0	1	0	0
1	0	0	0
1	1	0	0

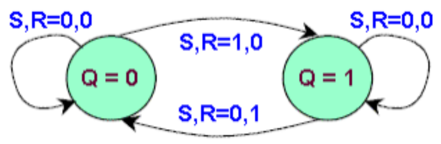
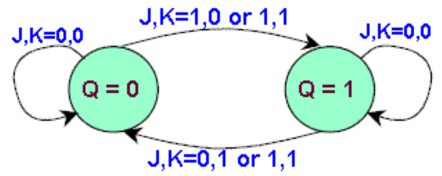
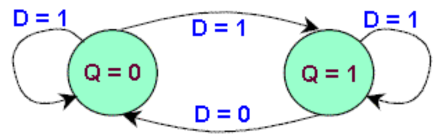
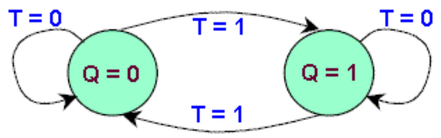
Laws and Rules of Boolean Algebra

- Laws of Boolean Algebra
  - The 12 Rules of Boolean Algebra
    - $A + 0 = A$
    - $A + 1 = 1$
    - $A \cdot 0 = 0$
    - $A \cdot 1 = A$
    - $A + A = A$
    - $A + \overline{A} = 1$
    - $A \cdot A = A$
    - $A \cdot \overline{A} = 0$
    - $\overline{\overline{A}} = A$
    - $A + AB = A$
    - $A + \overline{A}B = A + B$
    - $(A + B)(A + C) = A + BC$

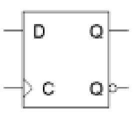
Sequential Logic:

Has memory; the circuit stores the result of the previous set of inputs. The current output depends on inputs in the past as well as present inputs. The basic element in sequential logic is the bistable latch or flip-flop, which acts as a memory element for one bit of data.

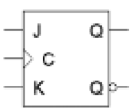
Flip Flop:

NAME	STATE DIAGRAM
SR	
JK	
D	
T	

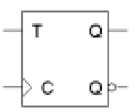
Tüm Flip-Flop(FF)lar için karakteristik tablolar



Q(t)	Q(t+1)	D	İşlem
0	0	0	Reset
0	1	1	Set
1	0	0	Reset
1	1	1	Set



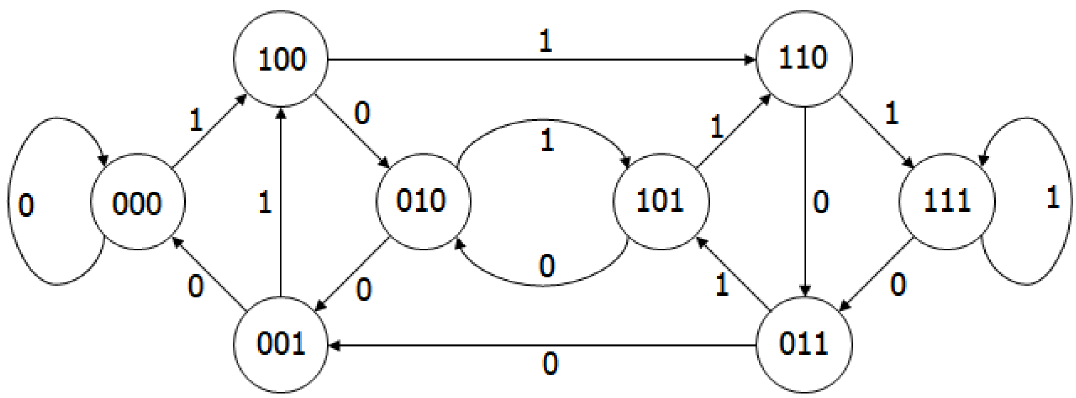
Q(t)	Q(t+1)	J	K	İşlem
0	0	0	x	Değişmez/reset
0	1	1	x	Set/tümleyen
1	0	x	1	Reset/tümleyen
1	1	x	0	Değişmez/set



Q(t)	Q(t+1)	T	Operation
0	0	0	Değişmez
0	1	1	Tümleyen
1	0	1	Tümleyen
1	1	0	Değişmez

State Table:

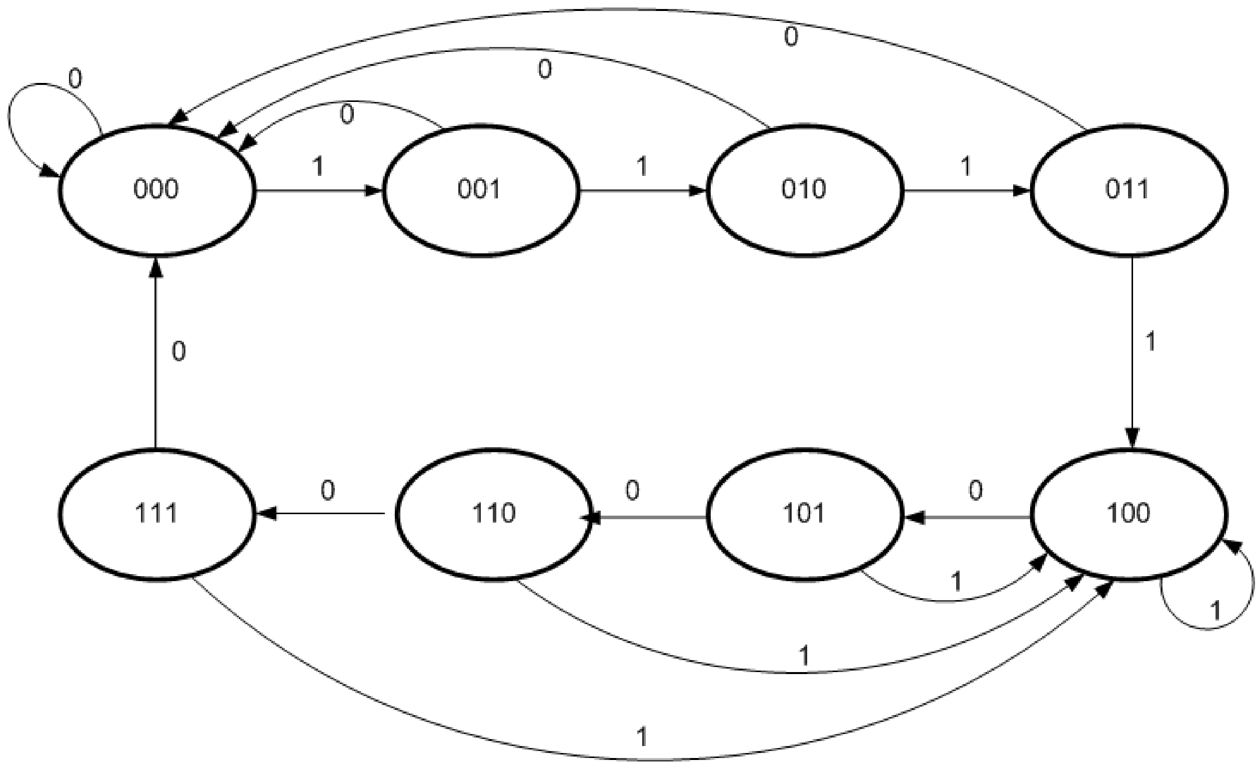
- The sequential circuit function can be represented in graphical form as a state diagram with the following components:
  - A circle with the state name in it for each state
  - A directed arc from the Present State to the Next State for each state transition
  - A label on each directed arc with the Input values which causes the state transition, and
  - A label:
    - On each circle with the output value produced, or
    - On each directed arc with the output value produced.



Örnek:

Durum diyagramı verilen devrenin

- Şuanki ve birsonraki durumlara göre D -ikili devre sayısını bulun
- Durum Tablosunu oluşturun
- Karnaugh Diyagramı ile indirgeyerek çıkış denklemlerini bulun
- Devreyi çizin
- Yorumlayın



Şu anki ve bir sonraki durumlara göre ikili devre sayısını bulunması:

- İkili flip – flop sayısı=3 adettir. Çünkü Durum diyagramında tüm durumlar 0 ile 7 arasında değişmektedir. Toplam durum sayısı=8=2^3 dür.
- Şu anki durumlar D-ikili devresini Q çıkışlarında bulunmaktadır. Bir sonraki durum ise D-ikili devresinin D girişlerinde bulunmaktır.
- Clok'un yükselen kenarı ile D-ikili devresi tetiklendiğinde Q-çıkışları D-girişlerine eşit olur.

Durum Tablosunun oluşturulması ve Karnaugh diyagramı yardımıyla indirgnmesi:

Şu anki durum			Giriş	Bir sonraki durum		
Q2	Q1	Q0	C	D2	D1	D0
0	0	0	0	0	0	0
0	0	0	1	0	0	1
0	0	1	0	0	0	0
0	0	1	1	0	1	0
0	1	0	0	0	0	0
0	1	0	1	0	1	1
0	1	1	0	0	0	0
0	1	1	1	1	0	0
1	0	0	0	1	0	1
1	0	0	1	1	0	0
1	0	1	0	1	1	0
1	0	1	1	1	0	0
1	1	0	0	1	1	1
1	1	0	1	1	0	0
1	1	1	0	0	0	0
1	1	1	1	1	0	0

Denklemler:

- D2 = Q2Q1’ + Q2Q0’ + Q1Q0C
- D1 = Q2Q1Q0’C’ + Q2’Q1Q0’C + Q2’Q1’Q0C + Q2Q1’Q0C’
- D0 = Q2’Q0’C’ + Q2Q0’C

10. Ekler

10.1. Matematiksel Kavramlar

Properties of Exponential and Logarithmic Equations

Let a be a positive real number such that $a \neq 1$, and let x and y be real numbers. Then the following properties are true:

1. $a^x = a^y$ if and only if $x = y$
2. $\log_a x = \log_a y$ if and only if $x = y$ $(x > 0,\ y > 0)$

Inverse Properties of Exponents and Logarithms

	Base a	*Natural Base e*
1.	$\log_a(a^x) = x$	$\ln(e^x) = x$
2.	$a^{(\log_a x)} = x$	$e^{(\ln x)} = x$

$4^{x+2} = 64$	Original Equation
$4^{x+2} = 4^3$	Rewrite with like bases
$x + 2 = 3$	Property of exponential equations
$x = 1$	Subtract 2 from both sides

The solution is 1. Check this in the original equation.

$\ln(2x - 3) = \ln 11$	Original Equation
$2x - 3 = 11$	Property of logarithmic equations
$2x = 14$	Add 3 to both sides
$x = 7$	Divide both sides by 2

The solution is 7. Check this in the original equation.

$5 + e^{x+1} = 20$ Original Equation

$e^{x+1} = 15$ Subtract 5 from both sides

$\ln e^{x+1} = \ln 15$ Take the logarithm of both sides

$x + 1 = \ln 15$ Inverse Property

$x = -1 + \ln 15 \approx 1.708$ Subtract 1 from both sides

$5 + e^{x+1} = 20$ Original Equation

$5 + e^{1.708+1} \stackrel{?}{=} 20$ Substitute 1.708 for x

$5 + e^{2.708} \stackrel{?}{=} 20$ Simplify

$5 + 14.999 \approx 20$ Solution checks ✓

$2^x = 7$ Original Equation

$\log 2^x = \log 7$ Take the logarithm of both sides

$x(\log 2) = \log 7$ Property of Logarithms

$x = \frac{\log 7}{\log 2} \approx 2.807$ Solve for x

$2^x = 7$ Original Equation

$\log_2 2^x = \log_2 7$ Take the logarithm of both sides

$x = \log_2 7$ Inverse Property

$x = \frac{\log 7}{\log 2} \approx 2.807$ Change of Base Formula

$4^{x-3} = 9$ Original Equation

$\log 4^{x-3} = \log 9$ Take the logarithm of both sides

$(x-3)\log 4 = \log 9$ Property of Logarithms

$x - 3 = \frac{\log 9}{\log 4}$ Divide both sides by $\log 4$

$x = 3 + \frac{\log 9}{\log 4} \approx 4.585$ Solve for x

$4^{x-3} = 9$ Original Equation

$\log_4 4^{x-3} = \log_4 9$ Take the logarithm of both sides

$x - 3 = \log_4 9$ Inverse Property

$x - 3 = \frac{\log 9}{\log 4}$ Change of Base Formula

$x = 3 + \frac{\log 9}{\log 4} \approx 4.585$ Solve for x

$2\log_4 x = 5$ Original Equation

$\log_4 x = \frac{5}{2}$ Divide both sides by 2

$4^{5/2} = x$ Change to exponential form

$x = 32$ Simplify

$20 \ln 0.2x = 30$ Original Equation

$\ln 0.2x = 1.5$ Divide both sides by 20

$0.2x = e^{1.5}$ Change to exponential form

$x = 5e^{1.5} \approx 22.408$ Divide both sides by 0.2

$3\log x = 6$ Original Equation

$\log x = 2$ Divide both sides by 3

$10^2 = x$ Change to exponential form

$x = 100$ Simplify

$\log_3 2x - \log_3(x-3) = 1$ Original Equation

$\log_3 \frac{2x}{x-3} = 1$ Condense the left side

$3^{\log_3 \frac{2x}{x-3}} = 3^1$ Exponentiate both sides

$\frac{2x}{x-3} = 3$ Inverse Property

$2x = 3x - 9$ Multiply both sides by $x - 3$

$x = 9$ Solve for x

1. $3^{x-1} = 81$	1. 5	22. $3^{x-2} = 81$	22. 6
2. $8^x = 4$	2. $\frac{2}{3}$	23. $\log_3 x = 5$	23. 243
3. $e^x = 5$	3. 1.609	24. $\log_4 x = 3$	24. 64
4. $-14 + 3e^x = 11$	4. 2.120	25. $\log_2 2x = \log_2 100$	25. 50
5. $-6 + \ln 3x = 0$	5. 134.476	26. $\ln(x+4) = \ln 7$	26. 3
6. $\log(3x+1) = 2$	6. 33	27. $\log_3(2x+1) = 2$	27. 4
7. $\ln x - \ln 3 = 4$	7. 163.794	28. $\log_5(x-10) = 2$	28. 35
8. $2\ln 3x = 4$	8. 2.463	29. $3^x = 500$	29. 5.66
9. $5^{x+2} = 4$	9. -1.139	30. $8^x = 1000$	30. 3.32
10. $\ln(x+2)^2 = 6$	10. 18.086, -22.086	31. $\ln x = 7.25$	31. 1408.10
11. $4^{-3x} = 0.25$	11. $\frac{1}{3}$	32. $\ln x = -0.5$	32. 0.61
12. $2e^{2x} - 5e^x - 3 = 0$	12. 1.099	33. $2e^{0.5x} = 45$	33. 6.23
13. $\log_7 3 + \log_7 x = \log_7 32$	13. $\frac{32}{3}$	34. $100e^{-0.6x} = 20$	34. 2.68

14. $2\log_6 4x = 0$	14. $\frac{1}{4}$	35. $12(1-4^x) = 18$	35. No Solution
15. $\log_2 x + \log_2(x-3) = 2$	15. 4	36. $25(1-e^t) = 12$	36. -0.65
16. $\log_2(x+5) - \log_2(x-2) = 3$	16. 3	37. $\log 2x = 1.5$	37. 15.81
17. $4\ln(2x+3) = 11$	17. 6.321	38. $\log_2 2x = -0.65$	38. 0.32
18. $\log x - \log 6 = 2\log 4$	18. 96	39. $\frac{1}{3}\log_2 x + 5 = 7$	39. 64
19. $2^x = 64$	19. 6	40. $4\log_5(x+1) = 4.8$	40. 5.90
20. $5^x = 25$	20. 2	41. $\log_2 x + \log_2 3 = 3$	41. $\frac{8}{3}$
21. $4^{x-3} = \frac{1}{16}$	21. 1	42. $2\log_4 x - \log_4(x-1) = 1$	42. 2

Binomial expansion

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots$$

If n is a positive integer the series terminates and is valid for all x: the term in x^r is ${}^nC_r x^r$ or $\binom{n}{r}$ where ${}^nC_r \equiv \dfrac{n!}{r!(n-r)!}$ is the number of different ways in which an unordered sample of r objects can be selected from a set of n objects without replacement. When n is not a positive integer, the series does not terminate: the infinite series is convergent for $|x| < 1$.

Taylor and Maclaurin Series

If $y(x)$ is well-behaved in the vicinity of $x = a$ then it has a Taylor series,

$$y(x) = y(a+u) = y(a) + u\frac{\mathrm{d}y}{\mathrm{d}x} + \frac{u^2}{2!}\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + \frac{u^3}{3!}\frac{\mathrm{d}^3y}{\mathrm{d}x^3} + \cdots$$

where $u = x - a$ and the differential coefficients are evaluated at $x = a$. A Maclaurin series is a Taylor series with $a = 0$,

$$y(x) = y(0) + x\frac{\mathrm{d}y}{\mathrm{d}x} + \frac{x^2}{2!}\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + \frac{x^3}{3!}\frac{\mathrm{d}^3y}{\mathrm{d}x^3} + \cdots$$

Power series with real variables

e^x	$= 1 + x + \dfrac{x^2}{2!} + \cdots + \dfrac{x^n}{n!} + \cdots$	valid for all x
$\ln(1+x)$	$= x - \dfrac{x^2}{2} + \dfrac{x^3}{3} + \cdots + (-1)^{n+1}\dfrac{x^n}{n} + \cdots$	valid for $-1 < x \le 1$
$\cos x$	$= \dfrac{\mathrm{e}^{\mathrm{i}x} + \mathrm{e}^{-\mathrm{i}x}}{2} = 1 - \dfrac{x^2}{2!} + \dfrac{x^4}{4!} - \dfrac{x^6}{6!} + \cdots$	valid for all values of x
$\sin x$	$= \dfrac{\mathrm{e}^{\mathrm{i}x} - \mathrm{e}^{-\mathrm{i}x}}{2\mathrm{i}} = x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!} + \cdots$	valid for all values of x
$\tan x$	$= x + \dfrac{1}{3}x^3 + \dfrac{2}{15}x^5 + \cdots$	valid for $-\dfrac{\pi}{2} < x < \dfrac{\pi}{2}$
$\tan^{-1} x$	$= x - \dfrac{x^3}{3} + \dfrac{x^5}{5} - \cdots$	valid for $-1 \le x \le 1$
$\sin^{-1} x$	$= x + \dfrac{1}{2}\dfrac{x^3}{3} + \dfrac{1.3}{2.4}\dfrac{x^5}{5} + \cdots$	valid for $-1 < x < 1$

Integer series

$$\sum_1^N n \;= 1 + 2 + 3 + \cdots + N = \frac{N(N+1)}{2}$$

$$\sum_1^N n^2 = 1^2 + 2^2 + 3^2 + \cdots + N^2 = \frac{N(N+1)(2N+1)}{6}$$

$$\sum_1^N n^3 = 1^3 + 2^3 + 3^3 + \cdots + N^3 = [1 + 2 + 3 + \cdots N]^2 = \frac{N^2(N+1)^2}{4}$$

$$\sum_1^\infty \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = \ln 2$$

$$\sum_1^\infty \frac{(-1)^{n+1}}{2n-1} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = \frac{\pi}{4}$$

$$\sum_1^\infty \frac{1}{n^2} = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots = \frac{\pi^2}{6}$$

$$\sum_1^N n(n+1)(n+2) = 1.2.3 + 2.3.4 + \cdots + N(N+1)(N+2) = \frac{N(N+1)(N+2)(N+3)}{4}$$

This last result is a special case of the more general formula,

$$\sum_1^N n(n+1)(n+2)\dots(n+r) = \frac{N(N+1)(N+2)\dots(N+r)(N+r+1)}{r+2}.$$

10.2. Türev - İnteral

$$\int x^n \, \mathrm{d}x = \frac{x^{n+1}}{n+1} + c$$

$$\int \frac{1}{x} \, \mathrm{d}x \quad = \ln x + c$$

$$\int \mathrm{e}^{ax} \, \mathrm{d}x \quad = \frac{1}{a} \mathrm{e}^{ax} + c$$

$$\int \ln x \, \mathrm{d}x \ = x(\ln x - 1) + c$$

$$\int x \, \mathrm{e}^{ax} \, \mathrm{d}x = \mathrm{e}^{ax} \left(\frac{x}{a} - \frac{1}{a^2} \right) + c$$

$$(uv)' = u'v + uv', \quad \left(\frac{u}{v} \right)' = \frac{u'v - uv'}{v^2}$$

$$\int_a^b u \, \mathrm{d}v = uv \Big|_a^b - \int_a^b v \, \mathrm{d}u$$

$$\int \sin x \, \mathrm{d}x \quad = -\cos x + c$$

$$\int \cos x \, \mathrm{d}x \quad = \sin x + c$$

$$\int e^x dx = e^x$$

$$\int e^{ax} dx = \frac{1}{a} e^{ax}$$

$$\int b^{ax} dx = \frac{1}{a \ln(b)} b^{ax} \qquad ;\mathrm{b} > 0$$

$$\int \ln(x) \, dx = x \ln(x) - x$$

$$\int a^x \ln(a) dx = a^x \qquad ;\mathrm{a} > 0$$

Dirac δ-'function'

$$\delta(t-\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\exp[i\omega(t-\tau)]\,d\omega.$$

If $f(t)$ is an arbitrary function of t then $\int_{-\infty}^{\infty}\delta(t-\tau)f(t)\,dt=f(\tau)$.

$\delta(t)=0$ if $t\neq 0$, also $\int_{-\infty}^{\infty}\delta(t)\,dt=1$

Türev:

1. $\frac{d}{dx}(c) \quad = \quad 0,$ where c is a constant

2. $\frac{d}{dx}(x^n) \quad = \quad nx^{n-1},$ where n is any real number

3. $\frac{d}{dx}(e^x) \quad = \quad e^x,$ $\quad \frac{d}{dx}(e^{cx}) = c\,e^{cx}$

4. $\frac{d}{dx}(\ln x) \quad = \quad \frac{1}{x},$ for $x>0$

5. $\frac{d}{dx}(\sin x) \quad = \quad \cos x$

6. $\frac{d}{dx}(\cos x) \quad = \quad -\sin x$

Differentiation

$$(uv)' = u'v + uv', \quad \left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$$

$$(uv)^{(n)} = u^{(n)}v + nu^{(n-1)}v^{(1)} + \cdots + {}^nC_r u^{(n-r)}v^{(r)} + \cdots + uv^{(n)}$$

$$\text{where } {}^nC_r \equiv \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\sin x) = \cos x \qquad \frac{\mathrm{d}}{\mathrm{d}x}(\sinh x) = \cosh x$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\cos x) = -\sin x \qquad \frac{\mathrm{d}}{\mathrm{d}x}(\cosh x) = \sinh x$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\tan x) = \sec^2 x \qquad \frac{\mathrm{d}}{\mathrm{d}x}(\tanh x) = \mathrm{sech}^2\, x$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\sec x) = \sec x \tan x \qquad \frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{sech}\, x) = -\mathrm{sech}\, x \tanh x$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\cot x) = -\mathrm{cosec}^2\, x \qquad \frac{\mathrm{d}}{\mathrm{d}x}(\coth x) = -\mathrm{cosech}^2\, x$$

$$\frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{cosec}\, x) = -\mathrm{cosec}\, x \cot x \qquad \frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{cosech}\, x) = -\mathrm{cosech}\, x \coth x$$

10.3. Trigeometri

Trigonometric Formula

$$\cos^2 A + \sin^2 A = 1 \qquad \sec^2 A - \tan^2 A = 1 \qquad \operatorname{cosec}^2 A - \cot^2 A = 1$$

$$\sin 2A = 2 \sin A \cos A \qquad \cos 2A = \cos^2 A - \sin^2 A \qquad \tan 2A = \frac{2 \tan A}{1 - \tan^2 A}.$$

$$\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B \qquad \cos A \cos B = \frac{\cos(A+B) + \cos(A-B)}{2}$$

$$\cos(A \pm B) = \cos A \cos B \mp \sin A \sin B \qquad \sin A \sin B = \frac{\cos(A-B) - \cos(A+B)}{2}$$

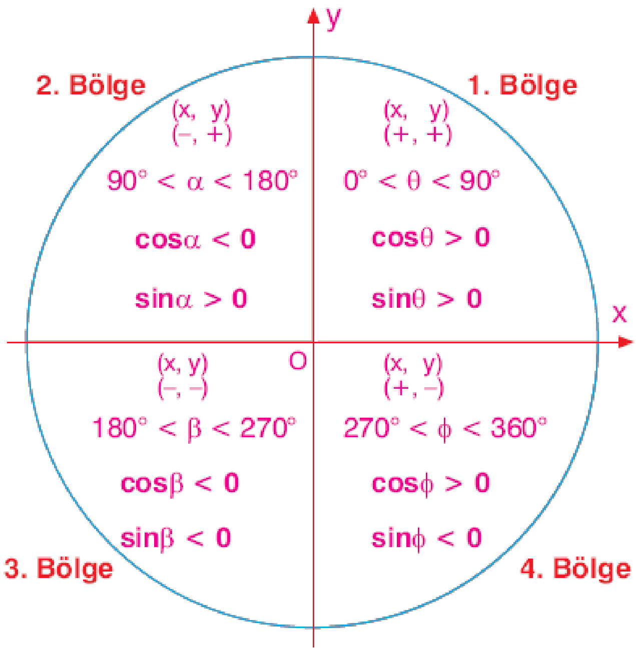
$$\tan(A \pm B) = \frac{\tan A \pm \tan B}{1 \mp \tan A \tan B} \qquad \sin A \cos B = \frac{\sin(A+B) + \sin(A-B)}{2}$$

$$\sin A + \sin B = 2 \sin \frac{A+B}{2} \cos \frac{A-B}{2} \qquad \cos^2 A = \frac{1 + \cos 2A}{2}$$

$$\sin A - \sin B = 2 \cos \frac{A+B}{2} \sin \frac{A-B}{2} \qquad \sin^2 A = \frac{1 - \cos 2A}{2}$$

$$\cos A + \cos B = 2 \cos \frac{A+B}{2} \cos \frac{A-B}{2} \qquad \cos^3 A = \frac{3 \cos A + \cos 3A}{4}$$

$$\cos A - \cos B = -2 \sin \frac{A+B}{2} \sin \frac{A-B}{2} \qquad \sin^3 A = \frac{3 \sin A - \sin 3A}{4}$$



$$\sin(\pi-\theta)=\sin\theta$$
$$\sin(\pi+\theta)=-\sin\theta$$
$$\cos(\pi-\theta)=-\cos\theta$$
$$\cos(\pi+\theta)=-\cos\theta$$

$$\sin\left(\frac{\pi}{2}-\theta\right)=\cos\theta$$
$$\sin\left(\frac{\pi}{2}+\theta\right)=\cos\theta$$
$$\cos\left(\frac{\pi}{2}-\theta\right)=\sin\theta$$
$$\cos\left(\frac{\pi}{2}+\theta\right)=-\sin\theta$$

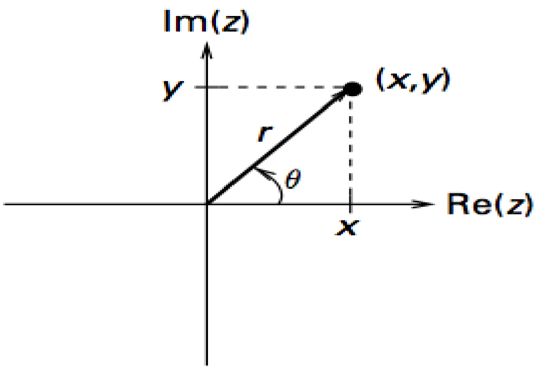
$$\cos x=\frac{1}{2}\left(e^{ix}+e^{-ix}\right)$$
$$\sin x=\frac{1}{2i}\left(e^{ix}-e^{-ix}\right)$$

	0°	30°	45°	60°	90°
sin	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1
cos	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0
tan	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$	Tanımsız

10.4. Complex

Complex number $z = x + jy$ (x and y real-valued $j = \sqrt{-1}$.)

$$r = |z| = \sqrt{x^2 + y^2},$$
$$\theta = \arg(z) = \tan^{-1}\frac{y}{x}$$



$$e^{j\theta} = \cos\theta + j\sin\theta$$

Complex Numbers

- A complex number **x** is of the form:

$$x = a + jb, \text{ where } j = \sqrt{-1}$$

a: **real part**, b: **imaginary** part

- Addition $(a + jb) + (c + jd) = (a + c) + j(b + d)$

- Multiplication $(a + jb) . (c + jd) = (ac - bd) + j(ad + bc)$

11. Kaynaklar

1. https://tr.wikipedia.org
2. https://www.wiley.com/legacy/Australia/Landing_Pages/c12ContinuousProbabilityDistributions_web.pdf
3. Olasılık ve İstatistik, Aydın Üstün, 2014.
4. İslamoğlu, A.H.(2009). Sosyal Bilimlerde Araştırma Yöntemleri (SPSS Uygulamalı), Beta Basım Yayım Dağıtım A.Ş., İstanbul.
5. Tütek, H., Gümüşoğlu, Ş., 2008, İ letme İstatistiği, Beta Basım Yayım Dağıtım A.Ş., İstanbul.
6. Olasılığın Matematiksel Temelleri , Timur KARAÇAY, Başkent Üniversitesi, Fen-Edebiyat Fakültesi
7. Olasılığın Temelleri, Timur Karaçay, Başkent Üniversitesi. Mantık, Matematik ve Felsefe IV.Ulusal Sempozyumu Foça, 5-8 Eylül 2006.
8. https://www.probabilitycourse.com/chapter11/11_2_7_solved_probs.php
9. http://en.wikipedia.org/wiki/Table_of_mathematical_symbols
10. https://cdn-acikogretim.istanbul.edu.tr/
11. Yılmaz Özkan, Uygulamalı İstatistik 1, Sakarya Kitapevi, 2008.
12. Özer Serper, Uygulamalı İstatistik 1, Filiz Kitapevi, 1996.
13. Meriç Öztürkcan, İstatistik Ders notları, YTÜ.
14. https://tr.wikipedia.org/wiki/Hipotez_testi