ASTROİSTATİSTİK
10. KONU

Hazırlayan: Doç. Dr. Tolgahan KILIÇOĞLU

10. OLASILIK DAĞILIMLARI

Bu bölümde olasılık ile istatistik arasında bir köprü kurmamızı sağlayan olasılık dağılımı kavramını göreceğiz. Olasılık dağılımları istatistikte sınıflanmış veriler için oluşturduğumuz frekans dağılımlarına benzerler. Olasılık dağılımının oluşturulabilmesi için örneklem uzayındaki her birimin olasılığı bir grafiğe aktarılır.

10.1 Tanımlar

Olay: Önceki bölümlerde de tanımladığımız gibi olay olasılığı hesaplanmak istenen bir durumdur. Örneğin, zar atıldığında 5 gelmesi veya bir kişinin hastalanması vb.

Deneme Sayısı (N): Bir işlemin kaç defa gerçekleştirildiğidir. Örneğin 100 kez zar atılması veya bir sınava 5 kez girilmesi deneme sayısıdır.

Başarı (p): İşlemin deneme sayısı kadar gerçekleştirilmesinin ardından olayın kaç defa gerçekleştiğinin deneme sayısına oranıdır. Bu bir olasılık değeri olup "başarı olasılığı" olarak adlandırılır. Örneğin, olay bir zar atıldığında 5 gelmesi ise, 100 kez zar atıldığında 18 kez 5 gelmesi durumunda $p=18/100=0.18$ olur. Başarı olasılığının mutlaka olumlu bir olay olması gerekmez. Örneğin, otobüs kazalarında ölüm oranlarının incelendiği bir istatistik çalışmada başarı parametresi hayatını kaybeden insan sayısı ile ilgili olacaktır. Buradaki "başarı" sözcüğünün gerçek hayattaki olayla doğrudan bir ilgisi yoktur. Sadece o olayın gerçekleşmesi ile ilgisi vardır.

Başarısızlık (q): İşlemin deneme sayısı kadar gerçekleştirilmesinin ardından olayın kaç defa gerçekleşmediğinin deneme sayısına oranıdır. Bu da bir olasılık değeridir ve "başarısızlık olasılığı" olarak adlandırılır. Bir olayın gerçekleşme olasılığı p, gerçekleşmeme olasılığı q ise $p+q=1$ olması gerektiği görülür. Bu durumda q nun belirli olmadığı durumlarda $q=1-p$ ifadesinden hesaplanabilmektedir.

Rassal Değişken (x): Bir olayın belirli sayıda denemede kaç defa gerçekleşeceğinin olasılığının p olduğunu belirtmiştik. Örneğin; yine olay bir zar atıldığında 5 gelmesi ise, $p=0.18$ olması durumunda 100 atış sonrasında 18 defa 5 gelmesi beklenir. Ancak bu bir olasılıktır ve gerçekleşmeyebilir. Bazen olayın deneme sayısı kadar tekrarda belirli bir sayıda gerçekleşmesinin olasılığını da bulmak isteyebiliriz. Örneğin zar 100 kez atıldığında 90 kez 5 gelmesinin olasılığı bizim için önemli olabilir. Buradaki 90 değeri rassal değişkendir. Bu anlamda, bir olasılık hesabı yapılırken olayın gerçekte rassal değişken kere gerçekleşmesinin olasılığı hesaplanmaktadır.

Olasılık: Bir olayın gerçekleşme sıklığının (f) deneme sıklığına (N) bölümüdür.

10.1 Bernoulli Dağılımı

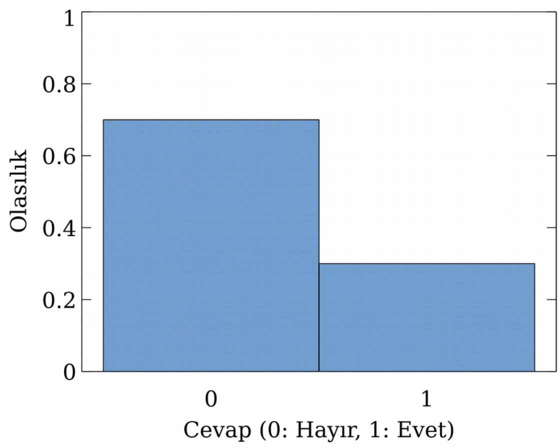
Bir işlemde rasgele değişken sadece 0 ve 1 gibi iki değer alabiliyorsa bu bir Bernoulli dağılımı oluşturur. Bozuk para atıldığında yazı veya tura gelmesi bu dağılıma en güzel örneklerdendir. Cevabı evet/hayır, doğru/yanlış, olumlu/olumsuz vb. olan sorulara verilen yanıtların da dağılımı bir Bernoulli dağılımı oluşturur. 0 sayısı negatif yanıtları temsil ederken 1 sayısı pozitif yanıtları temsil eder. Bernoulli dağılımı gerçekte ilerleyen bölümde göreceğimiz Binom dağılımında $N=1$ alındığı basit bir formdur.

Şimdi Bernoulli dağılımını daha kolay anlatmak için bir evcil hayvan anketini ele alalım. Bir mahallenin tüm sakinlerine bir anket uygulanarak evlerinde hayvan besleyip beslemedikleri soruluyor. Anket sonucunda popülasyonun %30'unun soruya **evet** cevabı verdiği görülüyor. Öncelikle p ve q değerlerini yazalım;

$p=0.30$ ve

$q=1-0.30=0.70$

Şekil 10.1'de yukarıda verdiğimiz örnek için Bernoulli dağılımı gösterilmektedir. Görüldüğü gibi Bernoulli dağılımı çok basit ve temel bir dağılımdır. Bu dağılımdan söz etmemizin temel nedeni dağılımın Binom dağılımına temel oluşturmasıdır. Şimdi bir Bernoulli dağılımında nasıl ortalama ve standart sapma bulunur onu görelim.



Şekil 10.1 Örnekte verilen evcil hayvan anketi için Bernoulli dağılımı

10.1.1 Bernoulli Dağılımında Ortalama

Bernoulli dağılımı için ortalama bilinen şekilde alınır. Veriler olasılıklarına göre ağırlıklandırılır. Yani 0 değeri q ile, 1 değeri ise p değeri ile ağırlıklandırılarak toplanır ve buradan ilginç bir sonuç olarak

$\mu=q\cdot 0+p\cdot 1=p$

elde edilir. Görüldüğü gibi bir Bernoulli Dağılımı'nın ortalaması p değerine eşittir. Evcil hayvan anketi örneği için bu durumda ortalama;

$$\mu = p = 0.30$$

olur. Ancak bu kesirli sayının nasıl bir anlamı vardır? Çünkü anket sadece evet veya hayır cevabı içermektedir. Arada bir cevap olması mümkün değildir. Burada, ortalama değer şöyle düşünülebilir: böyle bir mahallede rastgele 10 kişiye evinde hayvan besleyip beslemediği sorulduğunda tahminen 3 kişi evet cevabı verecektir.

10.1.2 Bernoulli Dağılımında Varyans ve Standart Sapma

Daha önce de tanımladığımız gibi varyans ortalamadan olan sapmaların karelerinin toplamıdır. Bernoulli dağılımında sadece iki değer olduğu için 0 ve 1'in ortalamadan olan farkı alınır;

$$\sigma^2 = q \cdot (0-\mu)^2 + p \cdot (1-\mu)^2$$

Burada artık $\mu = p$ ve $q = 1-p$ olduğunu biliyoruz. Bu değerleri yerine yazarsak;

$$\sigma^2 = (1-p) \cdot (0-p)^2 + p \cdot (1-p)^2$$

$$\sigma^2 = (1-p) \cdot p^2 + p \cdot (1-2p+p^2)$$

$$\sigma^2 = p^2 - p^3 + p - 2p^2 + p^3$$

$$\sigma^2 = p - p^2$$

$$\sigma^2 = p(1-p) \quad \text{veya}$$

$$\sigma^2 = pq$$

olduğu görülür. Standart sapma (σ) da bu değerin karekökü olacaktır:

$$\sigma = \sqrt{p(1-p)} \quad \text{veya} \quad \sigma = \sqrt{pq}$$

Evcil hayvan anketi için $p = 0.30$ olarak elde etmiştik. Bu durumda anketin cevaplarının (0 veya 1) varyansı ve standart sapmasının

$$\sigma^2 = p(1-p) = 0.30(1-0.30) = 0.21 \quad \text{ve}$$

$$\sigma = \sqrt{p(1-p)} = \sqrt{0.21} = 0.46$$

olduğu görülür.

10.1.3 Bernoulli Dağılımında Çarpıklık ve Basıklık

Yukarıda yaptığımız işlemlere benzer şekilde çarpıklık için de bir çıkarım yapıldığında çarpıklığın ifadesinin

$$ç=\frac{q-p}{\sqrt{pq}}$$

olduğu görülecektir. Evcil hayvan anketi için hesaplandığında değerinin 0.87 olduğu görülecektir. Yani veriler pozitif (sağa) çarpıktır.

Basıklık değerinin bir Bernoulli dağılımında ne anlam ifade ettiği bir soru işaretidir. Ancak biz yine de ifadesini çıkarmadan alacağı değerler hakkında kısaca bir bilgi verelim. Eğer bir Bernoulli dağılımında negatif veya pozitif tercihlerden biri çok tercih edilmişse dağılım sivrileşecek ve basıklık değeri sonsuza doğru büyüyecektir. Eğer negatif ve pozitif değerlerin sayısı birbirlerine eşitse en basık durum olan kare dağılım ortaya çıkacak ve basıklığın değeri –1.2 olacaktır.

10.2 Binom Dağılımı

Bir yöntem N defa tekrarlandığında bir olayın x = {x_0, x_1, x_2, ...} defa gerçekleşmesinin olasılıkları bir binom dağılım oluşturabilir. Hemen bunu bir örnekle anlatalım.

Örnek 10.1 Hilesiz bir para 4 defa atıldığında 0, 1, 2, 3 ve 4 kere tura gelmesinin olasılıklarını bir histogram üzerinde gösteriniz.

Cevap 10.1 Burada rassal değişken (x) 0, 1, 2, 3 ve 4 değerlerini almaktadır. Şimdi bir para 4 defa atıldığında oluşabilecek tüm olası durumları (örneklem uzayı) yazalım;

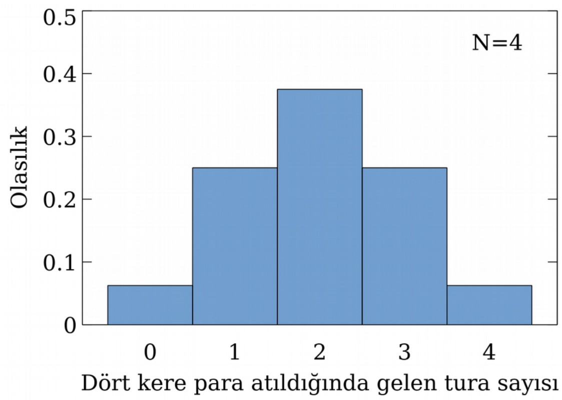
Örneklem uzayı: {TTTT, TTTY, TTYT, TTYY, TYTT, TYTY, TYYT, TYYY, YTTT, YTTY, YTYT, YTYY, YYTT, YYTY, YYYT, YYYY}

Yukarıdaki toplam 16 (2^4) durum bulunmaktadır. Şimdi aşağıdaki tabloda x'in alabileceği farklı değerleri için olasılıkların dağılımını Çizelge 10.1'de gösterelim:

Çizelge 10.1 4 kere madeni para atıldığında farklı sayıda tura gelme olasılıklarının dağılımı

	x **kere tura gelme olasılığı**
$x=0$	1/16
$x=1$	4/16
$x=2$	6/16
$x=3$	4/16
$x=4$	1/16

Şimdi bulduğumuz bu değerleri Şekil 10.2'de bir histogram üzerinde gösterelim;

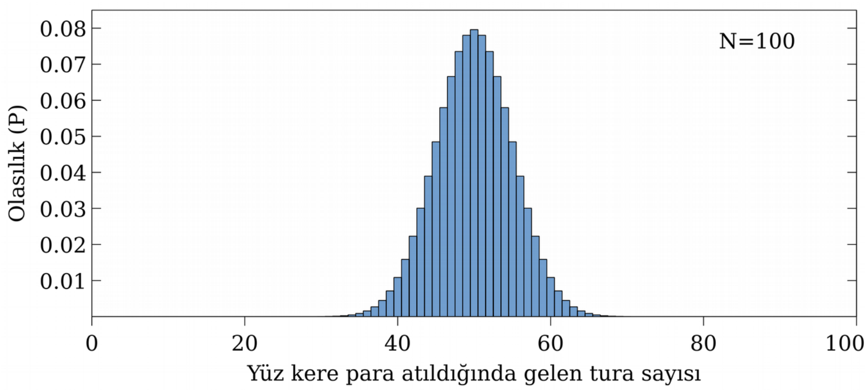


Şekil 10.2 Dört kere para atıldığında 0, 1, 2, 3 ve 4 kere tura gelme olasıkları (olasılık dağılımı)

Görüldüğü gibi bir madeni para 4 kere atıldığında 2 kere tura gelme şansı oldukça fazla (%37.5) iken 4 kere tura gelme veya hiç tura gelmeme şansları ise oldukça düşüktür (%6.25). Dağılım bir anlamda normal dağılımı andırmaktadır; ancak x-ekseninde çok az sayıda (5 adet) rastgele değişken değeri bulunduğundan dağılımın şekli tam olarak anlaşılamamaktadır. Buna karşın dağılımın simetrik olduğu rahatlıkla söylenebilir. Bu tür bir dağılım bir binom dağılıma örnektir.

Örnek 10.2 Hilesiz bir para 100 defa atıldığında tura gelme sayılarının olasılık dağılımını bir histogram üzerinde gösteriniz.

Cevap 10.2 Bir önceki soruda kullanılan yöntemle bu sorunun çözülmesi olanaksızdır. Çünkü bu soruda 100 tane madeni paranın oluşturacağı farklı durumların sayısı 2^{100} ($>10^{30}$) adettir! Bu durumları tek tek yazarak aralarından aradıklarımızı seçmek mümkün değildir. Birazdan bu hesaplamaların daha kolay şekilde nasıl yapılacağını göreceğiz. Şimdilik bu soruyu çözdüğümüzü ve 0 ile 100 arasında tura gelme olasıklarını Şekil 10.2'de gösterdiğimizi düşünelim.



Şekil 10.2 Yüz kere para atıldığında farklı sayılarda tura gelme olasıkları (olasılık dağılımı)

Şekil 10.2'de elde ettiğimiz binom dağılım gerçekten de normal dağılıma (çan eğrisine) çok benzemektedir. Normal dağılımdan en belirgin farkı verilerin kesikli olmasıdır (deneme sayısı N kadardır). Buradan normal dağılımın kesikli halinin çarpıklık sergilemeyen bir binom dağılım olduğunu ve N arttıkça binom dağılımın normal dağılıma yaklaştığını rahatlıkla söyleyebiliriz.

10.2.1 Binom Dağılımında Ortalama

Sınıflanmış veriler için verdiğimiz klasik ortalama ifadesini hatırlayalım:

$$\mu=\frac{\sum f_i x_i}{N}$$

Bu ifade aşağıdaki şekilde de yazılabilir;

$$\mu=\sum \frac{f_i}{N} x_i$$

Burada karşımıza çıkan $\frac{f_i}{N}$ teriminin daha önce istatistikte **oransal frekans** olarak tanımlamıştık. Olasılık hesabında ise bu orana ilgili i. rassal değişkenin **olasılığı** denir. Dolayısıyla bir binom dağılındaki değerlerin ortalaması değerlerin kendisi ile olasılıklarının çarpımının toplamıdır:

$$\mu=\sum P_i x_i$$

Herhangi bir dağılım için yukarıda verilen bu ifade ile ortalama bulunabilir. Ancak binom dağılımı N defa tekrarkanan bir Bernoulli dağılımı olarak düşünülebilir. Böylece ortalama değer Bernoulli dağılımına benzer çekilde aşağıdaki ifade ile bulunabilir:

$$\mu=Np$$

Burada p yine başarı olasılığıdır.

Örnek 10.3 100 kere madeni para atıldığında ortalama kaç defa tura gelir?

Cevap 10.3 Burada $N=100$ ve $p=\frac{1}{2}$ olduğuna göre;

$$\mu=Np=100\cdot\frac{1}{2}=50$$

olarak elde edilir. Bu değerin doğru olup olmadığını Şekil 10.2'yi inceleyerek görsel olarak kontrol ediniz.

10.2.2 Binom Dağılımında Varyans ve Standart Sapma

Binom dağılımında da varyans yine Bernoulli dağılımına benzer şekilde;

$\sigma^2=Npq$ şeklinde hesaplanır. Buradan standart sapma;

$$\sigma=\sqrt{Npq}$$

olacaktır.

Örnek 10.4 100 kere madeni para atıldığında gelecek tahmini tura sayısının standart sapmasını hesaplayınız.

Cevap 10.4

$$\sigma=\sqrt{Npq}=\sqrt{100\cdot\frac{1}{2}\cdot\frac{1}{2}}=5$$

olur. Bu durumda 100 kere para atıldığında Örnek 10.3'te elde ettiğimiz ortalama tura sayısı standart sapmasıyla birlikte;

$$\mu=50\pm5$$

olarak ifade edilebilir.

10.1.3 Binom Dağılımında Çarpıklık

Çarpıklık ifadesi de Bernoulli dağılımına benzer olup sadece paydaya N çarpan olarak gelmektedir;

$$ç=\frac{q-p}{\sqrt{Npq}}$$

Burada binom dağılımın normal dağılıma göre ikinci bir farkı ortaya çıkar. Bir yöntemin başarısının (p) %50 olmaması durumunda binom dağılım çarpıklık sergileyebilmektedir. Başarı oranı 0.5'ten büyükse dağılım negatif (sola) çarpık, 0.5'ten küçükse de pozitif (sağa) çarpık olur.

10.1.4 Binom Dağılımında Olasılık

Binom dağılımda bir olayın x (rassal değişken) kez gerçekleşme olasılığı aşağıdaki ifade ile elde edilir;

$$P(x)=\binom{N}{x}\cdot p^x\cdot q^{(N-x)}$$

Hatırlatma: Burada $\binom{N}{x}$ ifadesi N elemanlı bir kümeden seçilen x elemanlı kombinasyonların toplamı olup aşağıdaki ifade ile hesaplanır:

$$\binom{N}{x}=\frac{N!}{x!\cdot(N-x)!}$$

Artık bu öğrendiklerimizi kullanarak Örnek 10.2'yi tüm olasılıkları tek tek yazmadan da çözebiliriz. Şimdi konuya ilişkin benzer bir başka bir örnek yapalım.

Örnek 10.5 Hilesiz bir madeni para 10 defa atıldığında 7 kere yazı gelmesinin olasılığı nedir?

Cevap 10.5

Olay **yazı gelmesi**dir. Bu nedenle başarı olasılığı $p=\frac{1}{2}$ 'dir. Buradan başarızlık olasılığı;

$$q=1-p=1-\frac{1}{2}=\frac{1}{2}$$

olarak elde edilir. $N=10$ ve $x=7$ olduğuna göre;

$$P(7)=\binom{10}{7}\cdot p^7\cdot q^{(10-7)}$$

$$P(7)=\left(\frac{10!}{7!\cdot(10-7)!}\right)\cdot\left(\frac{1}{2}\right)^7\cdot\left(\frac{1}{2}\right)^{(10-7)}$$

$$P(7)=\left(\frac{10!}{7!\cdot(3)!}\right)\cdot\left(\frac{1}{2}\right)^7\cdot\left(\frac{1}{2}\right)^3$$

$$P(7)=0.1171875$$

olarak elde edilir. Yani 10 kez madeni para atıldığında 7 kez yazı gelme olasılığı yaklaşık %12'dir.

Görüldüğü gibi binom dağılımının hesaplanması biraz zahmetlidir. Özellikle, N büyük olduğu zaman, örneğin $N=100$ gibi, hesaplamalarda sayıların üstleri oldukça büyümektedir. Ancak, N nin büyük olması ve p nin çok küçük olması durumunda binom dağılımı hesaplanması biraz daha esnek olan Poisson dağılımına yaklaşır. N sonsuza yaklaştığında ise binom dağılım normal dağılıma yakınsamaktadır. Şimdi bu iki diğer dağılım türünü de irdeleyelim.

10.3 Poisson Dağılımı

Eğer bir yöntem çok defa tekrarlandığında olay çok nadir gerçekleşiyorsa, başka bir deyişle, başarı (p) oldukça düşükse, olayın kaç defa gerçekleşeceğine ilişkin binom olasılık dağılımı bir Poisson dağılımına yakınsar. Binom dağılımının Poisson dağılımına yakınsaması aşağıda yer alan eşitsizliğin sağlanması durumunda söz konusu olmaktadır:

$$Np \lesssim 5$$

Şimdi binom dağılım ifadesini kullanarak Poisson dağılımını elde edelim. Öncelikle binom dağılımın ifadesini hatırlayalım;

$$P(x)=\binom{N}{x}\cdot p^x\cdot q^{(N-x)}$$

Binom dağılımda μ olarak gösterdiğimiz ortalama değer Poisson dağılımında λ ile gösterilir; ancak hesabında bir farklılık söz konusu değildir:

$$\lambda = Np$$

Bu ifadeden p çekilirse;

$$p = \frac{\lambda}{N}$$

ve aynı zamanda,

$$q = 1 - \frac{\lambda}{N}$$

olacaktır.

Şimdi, bu p ve q ifadelerini binom ifadesinde yerine koyalım ve kombinasyon ifadesini açalım;

$$P(x) = \left(\frac{N!}{x!(N-x)!}\right)\left(\frac{\lambda}{N}\right)^{x}\left(1 - \frac{\lambda}{N}\right)^{(N-x)}$$

En sağdaki parantez üstlere göre ikiye ayrılırsa;

$$P(x) = \left(\frac{N!}{x!(N-x)!}\right)\left(\frac{\lambda}{N}\right)^{x}\left(1 - \frac{\lambda}{N}\right)^{N}\left(1 - \frac{\lambda}{N}\right)^{-x}$$

Şimdi N nin sonsuza yaklaştığı limit durumu ele alalım;

$$\lim_{N\to\infty} P(x) = \lim_{N\to\infty}\left(\left(\frac{N!}{x!(N-x)!}\right)\left(\frac{\lambda}{N}\right)^{x}\left(1 - \frac{\lambda}{N}\right)^{N}\left(1 - \frac{\lambda}{N}\right)^{-x}\right)$$

Burada, λ ve x sabitler olduklarından eşitliğin sol tarafına alınabilir;

$$\lim_{N\to\infty} P(x) = \frac{\lambda^{x}}{x!}\lim_{N\to\infty}\left(\left(\frac{N!}{(N-x)!}\right)\left(\frac{1}{N^{x}}\right)\left(1 - \frac{\lambda}{N}\right)^{N}\left(1 - \frac{\lambda}{N}\right)^{-x}\right)$$

Şimdi bu ifadeyi üç parçaya ayıralım ve her bir parçanın limit durumuna bakalım:

$$\lim_{N\to\infty} P(x) = \frac{\lambda^{x}}{x!}\lim_{N\to\infty}\left(\underbrace{\left(\frac{N!}{(N-x)!}\right)\left(\frac{1}{N^{x}}\right)}_{\text{i.}}\underbrace{\left(1 - \frac{\lambda}{N}\right)^{N}}_{\text{ii.}}\underbrace{\left(1 - \frac{\lambda}{N}\right)^{-x}}_{\text{iii.}}\right)$$

i. İlk ifadede faktöriyeller açık olarak yazılırsa;

$$\lim_{N\to\infty}\left(\frac{N!}{(N-x)!}\right)\cdot\left(\frac{1}{N^x}\right)=\lim_{N\to\infty}\left(\frac{(N)(N-1)(N-2)\ldots(N-x+1)(N-x)(N-x-1)\ldots(1)}{N^x(N-x)!}\right)$$

ve buradan,

$$\lim_{N\to\infty}\left(\frac{(N)(N-1)(N-2)\ldots(N-x+1)(N-x)!}{N^x(N-x)!}\right)$$

görüldüğü gibi $(N-x)!$ ifadesi hem pay hem de payda da karşımıza çıktı. Bu ifade sadeleştirilirse;

$$\lim_{N\to\infty}\left(\frac{(N)(N-1)(N-2)\ldots(N-x+1)}{N^x}\right) \quad \text{veya} \quad \lim_{N\to\infty}\left(\left(\frac{N}{N}\right)\left(\frac{N-1}{N}\right)\left(\frac{N-2}{N}\right)\ldots\left(\frac{N-x+1}{N}\right)\right)$$

elde edilir. Görüldüğü gibi sağ tarafta yer alan her oran n sonsuza doğru gittikçe 1'e yaklaşacaktır. Bu durumda denklemdeki ilk ifade için aşağıdaki sonuç yazılır:

$$\lim_{N\to\infty}\left(\frac{N!}{(N-x)!}\right)\cdot\left(\frac{1}{N^x}\right)=1$$

ii. Şimdi ikinci ifadeyi inceleyelim:

$$\lim_{N\to\infty}\left(1-\frac{\lambda}{N}\right)^N$$

Bu bölümün başında $Np=\lambda \lesssim 5$ olduğunu söylemiştik. Bu nedenle λ nın küçük bir sayı olması gerektiğini unutmayalım. Limit parantezi içinde kalan $\left(1-\frac{\lambda}{N}\right)^N$ ifadesinde λ/N çok küçük bir sayı olsa da, bu ifadenin N. kuvvetinin alınıyor olması nedeniyle limit 1'e yakınsamaz. Matematikte aşağıdaki durum bilinmektedir:

$$\lim_{c\to\infty}\left(1+\frac{1}{c}\right)^c=e$$

Bu ifadeyi yukarıdaki ifadeye benzetmek için $c=-\frac{N}{\lambda}$ olduğunu kabul edelim. Böylece;

$$\lim_{N\to\infty}\left(1-\frac{\lambda}{N}\right)^N=\lim_{N\to\infty}\left(1+\frac{1}{c}\right)^{c(-\lambda)}=e^{-\lambda}$$

elde edilir.

iii. Şimdi geriye kalan son ifadeye bakalım:

$$\lim_{c\to\infty}\left(1-\frac{\lambda}{N}\right)^{(-x)}$$

Bu ifadede N sonsuza yaklaştıkça sonuç $(1)^{(-x)}$ e sadeleşir ve limit 1 e yakınsar.

i., ii. ve iii. parçalarda bulduğumuz sonuçlar birleştirildiğinde Poisson dağılımının olasılık ifadesi elde edilir:

$$P_\lambda(x)=\frac{e^{-\lambda}\lambda^x}{x!}$$

Görüldüğü gibi Poisson dağılımı ifadesi binom dağılımından çok daha pratik şekilde hesaplanabilmektedir.

10.3.1 Poisson Dağılımında Varyans ve Standart Sapma

Binom dağılımda varyansı $\sigma^2=Npq$ ifadesi ile elde ediyorduk. Poisson dağılımında p çok küçük bir değere sahip olduğundan q değeri 1'e çok yakın olur. Bu nedenle Poisson dağılımında varyans $\sigma^2=Np$ olacaktır. Poisson dağılımında, ilginç bir şekilde, varyans ile ortalama değer birbirlerinin aynısıdır:

$$\sigma^2=\lambda=Np$$

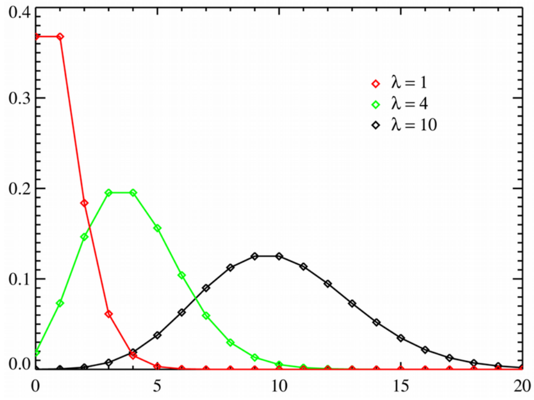
Buradan eğer standart sapma hesaplanmak istenirse;

$$\sigma=\sqrt{Np}$$

ifadesi kullanılabilir.

10.3.2 Poisson Dağılımında Çarpıklık

Poisson dağılımında *p* çok küçük bir değer olduğundan Binom dağılım için verilen çarpıklık ifadesinden her zaman oldukça sağa çarpık bir eğri elde edileceği kolaylıkla söylenebilir. Şekil 10.3'te λ = 1, 4 ve 10 olan dağılımlar gösterilmektedir. Bu grafiklerde λ = 1 ve 4 olanların sağa çarpık Poisson dağılımına yakınsadıkları, λ = 10 olanın ise yine sağa çarpık olan bir binom dağılım sergilediği görülür.



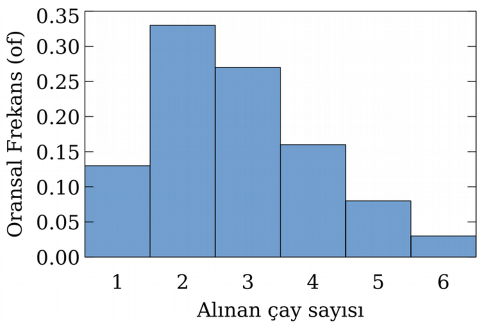
Şekil 10.3 λ = 1, 4 ve 10 değerleri için Poisson/binom dağılımlar.

Şimdi Poisson dağılımına ilişkin verdiğimiz tüm bilgileri kullanarak birkaç alıştırma yapalım.

Örnek 10.6 Bir kafe sahibi gelen 100 müşterisi için kişi başı kaç çay içildiğini aşağıdaki gibi kaydediyor. Bu verinin dağılımı için uygun bir istatistiksel bir model seçiniz ve dağılımın bu modele uyduğunu varsayarak gelen bir müşterinin 7 çay içme ve hiç çay içmeme olasılıklarını hesaplayınız.

İçilen çay sayısı	Müşteri sayısı
1	13
2	33
3	27
4	16
5	8
6	3

Cevap 10.6 Öncelikle müşteri sayılarını toplam müşteri sayısına bölerek oransal frekansları elde edelim ve veriyi bir histogram üzerinde gösterelim;



Dağılımın sağa çarpık olduğu görülmektedir ve dağılım bir Poisson dağılımını andırmaktadır. Poisson yaklaşımını kullanabilmemiz için $\lambda \lesssim 5$ koşulu sağlanmalıdır. Yukarıdaki verinin ortalaması hesaplandığında 2.82 olduğu görülecektir. Bu değer 5'ten az olduğundan Poisson yaklaşımı bu veri için kullanılabilir. Ancak, verinin varyansı 5.6 civarında olup ortalama değere eşit değildir. Bu durumda Poisson dağılımı bu veriyi çok iyi modellemeyebilir. Biz yine de bir yaklaşım olarak soruda Poisson dağılımını kullanacağız.

$\lambda = 2.82$ olduğuna göre Poisson dağılımı ifadesi;

$$P_\lambda(x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-2.82}(2.82)^x}{x!}$$

Bu durumda 7 çay içilmesi olasılığı:

$$P_\lambda(7) = \frac{e^{-2.82}(2.82)^7}{7!} = 0.02$$

yani olasılığın %2 olduğu görülür.

Hiç çay içilmemesi olasılığı ise;

$$P_\lambda(0) = \frac{e^{-2.82}(2.82)^0}{0!} = 0.06$$

yani %6 olarak elde edilir.

Frekans dağılımı incelendiğinde müşterilerin %27'sinin 3 bardak çay içtiği görülmektedir. Şimdi, bir karşılaştırma olması bakımından 3 bardak çay içme olasılığını bir de Poisson dağılımı ile hesaplayalım;

$$P_\lambda(3) = \frac{e^{-2.82}(2.82)^3}{3!} = 0.22$$

Poisson dağılımının %22'lik oranla oldukça yakın bir olasılık verdiği görülür.

Poisson dağılımının kullanılabilmesi için sadece ortalama değerin bilinmesi yeterli olmaktadır. Yani, yukarıdaki soruda ortalama değer verilir ve dağılımın bir Poisson dağılımı olduğu söylenirse başka bir veriye ihtiyaç duyulmaksızın hangi rassal değişkenin hangi olasılığa sahip olduğu hemen belirlenebilmektedir. Poisson dağılımının en yaygın kullanıldığı durumlar uzun zaman aralıklarında gerçekleşen nadir olaylardır. Şimdi, bu durumlara ilişkin bir örnek çözelim.

Örnek 10.7 Ankara Üniversitesi Kreiken Rasathanesi'ndeki T40 teleskobu yılda ortalama 4 kez arızalanmaktadır. Buna göre;

i. Bir yılda teleskobun hiç arızalanmama olasılığı nedir?
ii. 2 aylık bir sürede teleskobun hiç arızalanmama olasılığı nedir?
iii. 4 aylık bir gözlem döneminde teleskobun 2 kere arızalanma olasılığı nedir?
iv. 6 ayda teleskobun 2 veya 2'den çok arızalanma olasılığı nedir?

Cevap 10.7

i. Bir yılda teleskobun hiç arızalanmaması aşağıdaki değerlerle temsil edilir:

$\lambda=4$ (yani 1 yıl için ortalama arızalanma) ve $x=0$ kez arızalanma durumu. Buradan,

$$P_\lambda(0)=\frac{e^{-4}(4)^0}{0!}=0.02$$

Yani, T40 teleskobunun bir yılda hiç arızalanmama olasılığı %2 dir.

ii. 2 ayda teleskobun hiç arızalanmaması için öncelikle λ nın indirgenmesi gerekir: λ değeri 1 yıldaki (12 aydaki) ortalama arızalanma miktarıdır. 2 aylık süre için λ nın yeni değeri

$$\lambda'=\lambda\cdot\frac{2}{12}=4\cdot\frac{2}{12}=0.667$$

olur. Hiç arızalanmama olasılığı sorulduğundan yine $x=0$ olmalıdır. Buna göre;

$$P_\lambda(0)=\frac{e^{-0.667}(0.667)^0}{0!}=0.51$$

Yani, T40 teleskobunun 2 ayda hiç arızalanmama olasılığı %51 dir.

iii. 4 ayda teleskobun 2 defa arızalanması aşağıdaki değerlerle temsil edilecektir:

$\lambda'=\lambda\cdot\frac{4}{12}=4\cdot\frac{4}{12}=1.333$ ve $x=2$

Buradan,

$$P_\lambda(2)=\frac{e^{-1.333}(1.333)^2}{2!}=0.23$$

elde edilir. Yani, T40 teleskobunun 4 ayda 2 kez arızalanma olasılığı %23 tür.

iv. Bu şık diğer şıklardan farklıdır. 6 ayda 2 veya 2'den fazla arızalanma oranının bulunması için 6 ayda 2, 3, 4, 5 ... (sonsuza kadar) kez arızalanma olasılıklarının hesaplanması ve daha sonra bu olasılıkların toplanması gerekmektedir. Ancak, böyle bir işlemi yapmamız pratikte mümkün değildir. Burada farklı bir strateji takip edeceğiz: <u>tüm olasılıkların toplamının 1 olması</u>. Dikkat edilirse bu sorudaki koşulu sağlamayan sadece iki durum mevcuttur: teleskobun 6 ayda hiç arızalanmaması veya 1 kez arızalanması. Eğer bu iki olasılığı hesaplar, toplar ve çıkan sonucu 1'den çıkarırsak aradığımız olasılığı elde etmiş oluruz.

Öncelikle λ değerini indirgersek;

$\lambda'=\lambda\cdot\frac{6}{12}=4\cdot\frac{6}{12}=2$ elde edilir.

Teleskobun 6 ayda hiç arızalanmama olasılığı ($x=0$):

$$P_{\lambda}(0)=\frac{e^{-2}(2)^{0}}{0!}=0.1353$$

Teleskobun 6 ayda 1 kez arızalanma olasılığı ($x=1$):

$$P_{\lambda}(1)=\frac{e^{-2}(2)^{1}}{1!}=0.2707$$

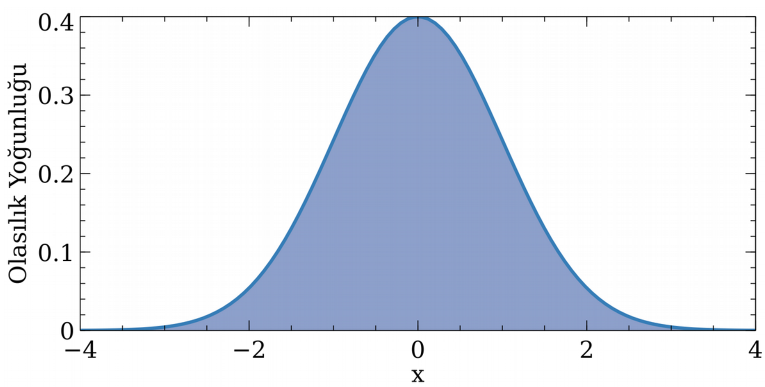
Bu iki değer toplanır ve sonuç 1'den çıkarılırsa sorunun cevabı elde edilmiş olur;

$$1-(0.1353+0.2707)=0.594$$

Bu sonuç bize 6 ay içerisinde teleskobun 2 veya 2'den fazla bozulma ihtimalinin %59 civarında olduğunu işaret etmektedir.

10.4 Normal Dağılım

Dersimizin başından beri sıkça söz ettiğimiz normal dağılım (Gauss dağılımı veya çan eğrisi), gerçekte $N=\infty$ olan bir binom dağılımı olarak düşünülebilir. Normal dağılımın burada sözünü ettiğimiz diğer dağılımlardan en önemli farkı **sürekli olması**dır. Normal dağılımın matematiksel bir fonksiyonu bulunmaktadır ve bu fonksiyonda diğer dağılımlarda olduğu gibi faktöriyel içeren ifadeler **yer almaz**. Bu sayede rassal değişken kesikli veri olmak zorunda değildir (Şekil 10.4).



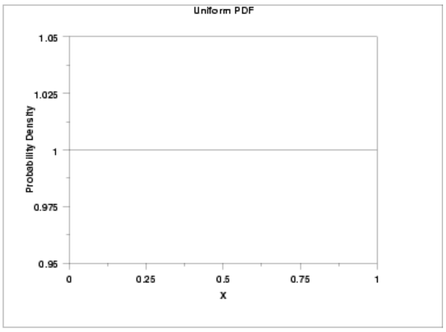
Şekil 10.4 Normal dağılım

N sayısının yüksek olduğu bir veri eğer Poisson dağılımına yakınsamıyorsa böyle bir veri üzerinde binom dağılımı ile hesaplama yapmak mümkün veya pratik olmayabilir. Bu durumda veri üzerinde Normal dağılım (Gauss fonksiyonu) kullanılarak hesaplama yapmak her zaman daha pratik ve oldukça güvenilir olacaktır.

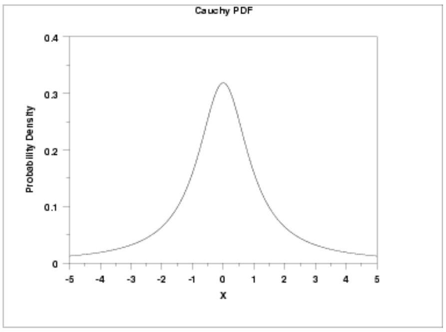
Bir sonraki dersimizde normal dağılıma ilişkin hesaplamaların nasıl yapılacağını göreceğiz. Ayrıntıları daha sonra irdelemek üzere şimdilik bu konuyu burada bırakıyoruz.

10.5 Diğer Dağılım Türleri

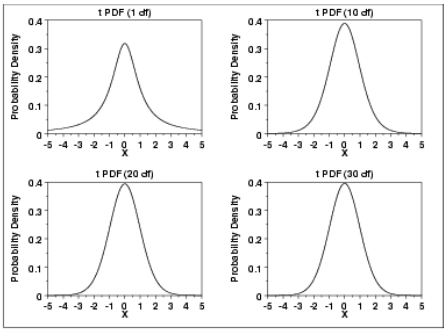
İstatistikte farklı amaçlar için kullanılabilecek çok sayıda farklı olasılık dağılımı türleri bulunmaktadır. Bu dağılımların bir bölümü aşağıda gösterilmektedir.



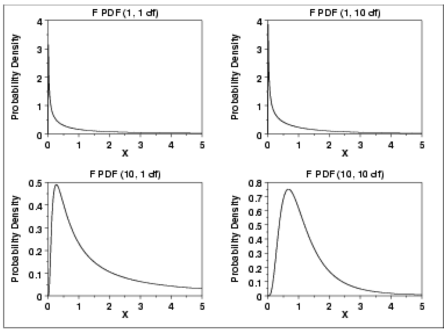
Tekdüze Dağılım



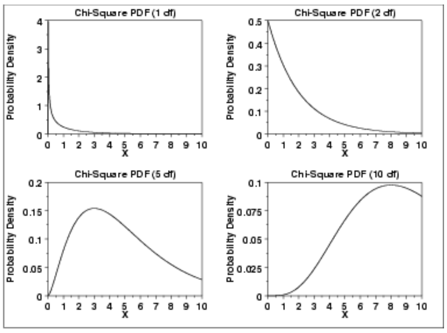
Cauchy Dağılımı



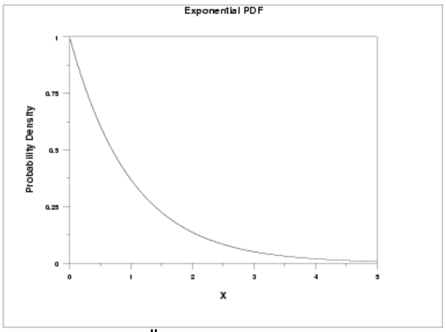
t Dağılım



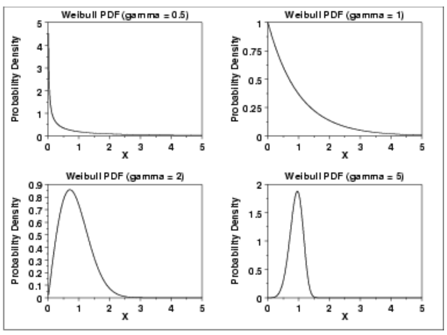
F Dağılımı



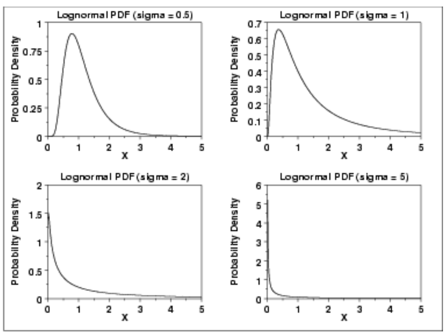
Chi-Kare Dağılımı



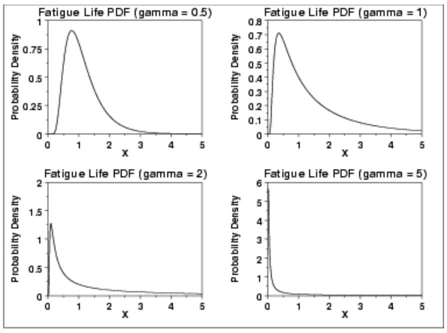
Üstel Dağılım



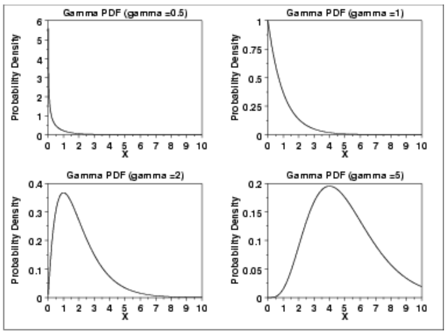
Weibull Dağılımı



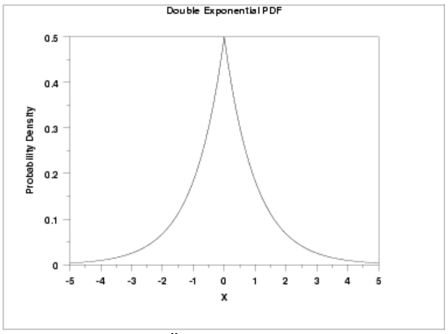
Lognormal Dağılım



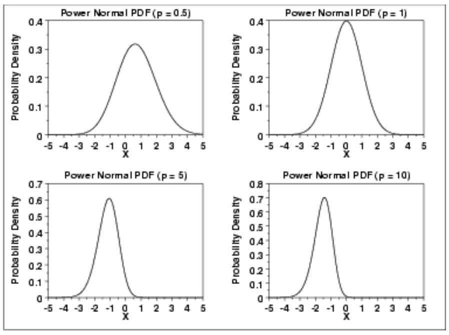
Birnbaum Saunders
(Yorulma Ömrü)
Dağılımı



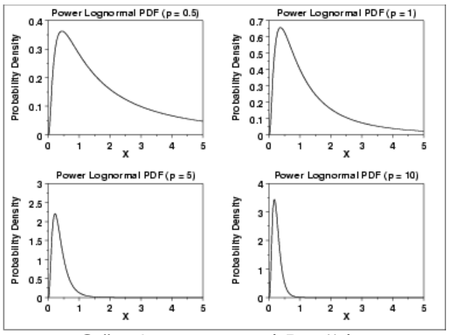
Gamma Dağılımı



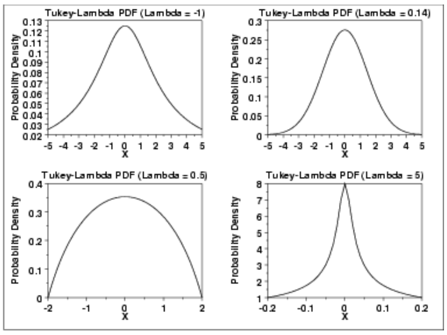
Çift Üstel Dağılım



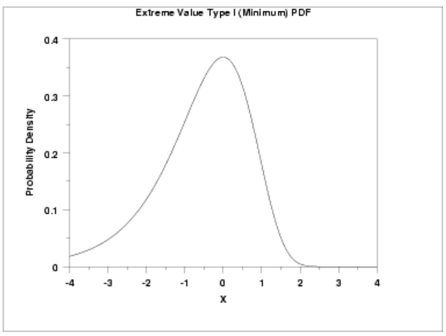
Güç Normal Dağılımı



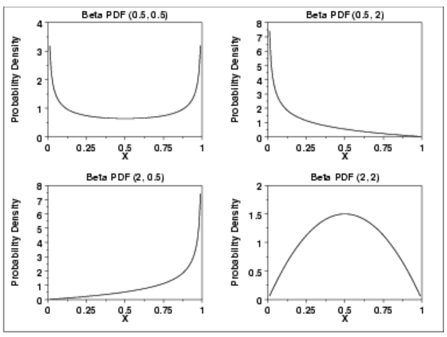
Güç Lognormal Dağılımı



Tukey-Lambda Dağılımı



Uç Değer Tip I Dağılımı



Beta Dağılımı

Bölüm 10.5'deki grafiklerin kaynağı: *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/, Tarih: 22.04.2019