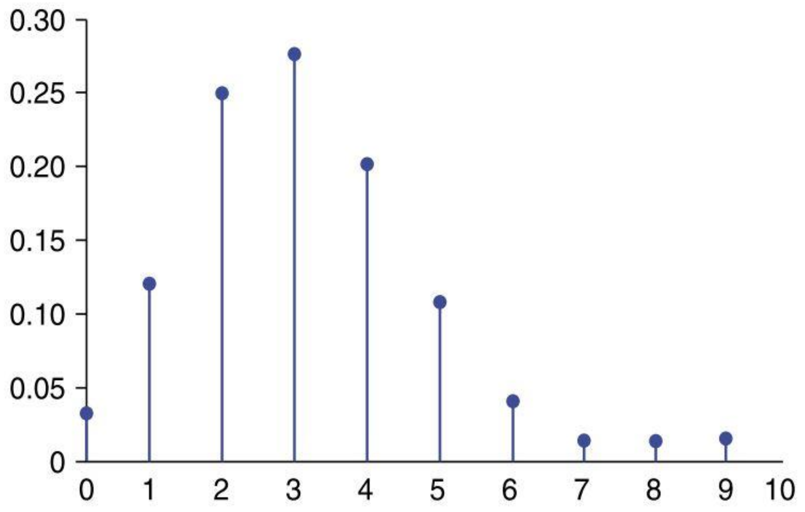
Olasılık ve İstatistik

Fırat İsmailoğlu, PhD

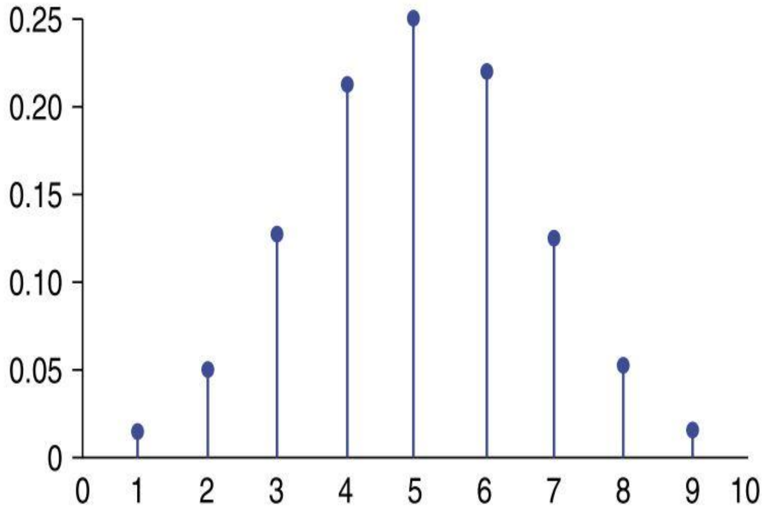
Sürekli Rastgele Değişkenler

Binomial Rastgele Değişken (devam)

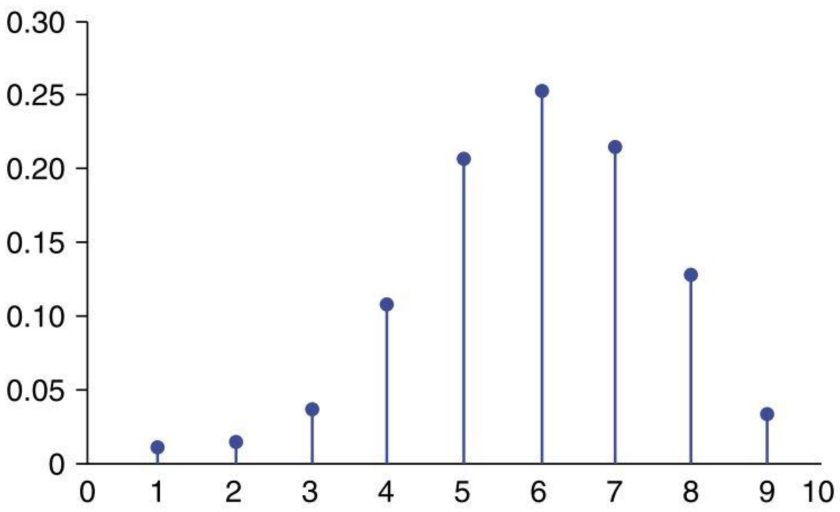
$n = 10$ iken (yani deneyin tekrarlanma sayısı 10 iken) farklı başarı olasılıkları (p değerleri) için olasılık dağılımları:



$p = 0.3$



$p = 0.5$



$p = 0.6$

$E[X] = 10 \cdot 0.3 = 3$
(10 atışta 3 başarı olması beklenir ve genel olarak toplam başarı sayısının az olması beklenir)

ör. Bir bozuk para 6 defa atılıyor.

i) En az 5 defa tura gelme olasılığı nedir?

ii) En fazla 5 defa tura gelme olasılığı nedir?

Çözüm.

X binomial rastgele değişkeni 6 atışta gelen toplam tura sayısını göstersin.

i) En az 5 defa tura gelme olayları $X = 5$ ve $X = 6$ ‘dır.

$$P(X=5)+P(X=6)=\binom{6}{5}0.5^5 0.5^1+\binom{6}{6}0.5^6 0.5^0=7\times 0.5^6$$

ii) En fazla 5 defa tura olayları $X = 0, X = 1, X = 2, X = 3, X = 4$ ve $X = 5$ ‘tir.

$$\begin{aligned} P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5) \\ = 1-P(X=6) \\ = 1-\binom{6}{6}0.5^6 0.5^0 \\ = 1-0.5^6 \\ = 0.984375 \end{aligned}$$

ör. Bir fabrikada üretilen ürünlerin birbirinden bağımsız olarak 0.01 olasılıkla bozuk olduğu biliniyor. 1000 adetlik bir kolide kaç ürünün bozuk olması beklenir?

Çözüm.

X binomial rastgele değişkeni 1000'lik kolideki toplam bozuk ürün sayısını göstersin.1000'lik kolideki beklenen bozuk ürün sayısı X'in beklenen değeridir: $E[X] = 1000 \times 0.01 = 10.$

ör. NBA'da son ikiye kalan takımlar birbirleriyle 7 maç yaparlar. Bu 7 maçın 4'ünü alan şampiyon olur. Son ikiye kalan takımlardan birincisinin ikinci takımı yenme olasılığı 0.7 olsun. Buna göre turnuvanın ilk 4 maçında birinci takımın ikinci takımı $3-1$ yenmesi ve daha sonra ikinci takımın kalan 3 maçı alarak şampiyon olması olasılığı nedir?

Çözüm.

Önce ilk 4 maçın $3-1$ olma olasılığını hesaplayalım.

X binomial rastgele değişkeni birinci takımın 4 maçta kazanacağı toplam galibiyet sayısını versin.

$X = 3$ ile ilgileniyoruz; çünkü $X = 3$ iken birinci takım ikinci takımı üç defa yenmiş ikinci takımda birinci takımı bir kere yenmiştir.

$$P(X = 3) = \binom{4}{3} 0.7^3 0.3^1 = 0.4116$$

Yukarıdaki duruma ilaveten kalan 3 maçın 3'ünü de ikinci takım kazansın istiyoruz.

Y binomial rastgele değişkeni ikinci takımın 4 maçta kazanacağı toplam galibiyet sayısını versin.

$$P(Y = 3) = \binom{3}{3} 0.3^3 0.7^0 = 0.027$$

Iki olay aynı anda olsun isityoruz. O halde bulduğumuz iki olasılığı çarparız.

$$P(X = 3) \times P(Y = 3) = 0.4116 \times 0.027 = 0.111$$

ör. Bir fabrikada üretilen malların birbirinden bağımsız olarak bozuk olma olasılığı $p = 0.05$ olsun. Bu mallar 10'luk kolilere konuluyor. Eğer bir kolide en az bir bozuk mal çıkarsa koli iade ediliyor.

i) Buna göre bir kolinin iade edilme olasılığı nedir?

ii) Üç koli alınırsa bu kolilerden birinin iade edilme olasılığı ne olur?

Çözüm.

i) X binomial rastgele değişkeni 10'luk kolideki toplam bozuk ürün sayısını göstersin.

$X \geq 1$ olması durumunda koli iade edilir. Bunun olma olasılığı:

$$\begin{aligned} P(X \geq 1) &= P(X = 1) + \cdots + P(X = 10) \\ &= 1 - P(X = 0) \\ &= 1 - \binom{10}{0} 0.05^0 (1 - 0.05)^{10} \\ &\approx 0.6 \end{aligned}$$

ii) Y binomial rastgele değişkeni alınan 3 koliden iade edilen toplam koli sayısını göstersin. Yukarıdaki hesapladık ki herhangi bir kolinin iade edilme olasılığı 0.6'dır.

$$P(Y = 1) = \binom{3}{1} 0.6^1 0.4^2 = 0.288$$

Sürekli Rastgele Değişken (Continuous Random Variables)

Şimdiye kadar ayrık rastgele değişken ile ilgilendik. Yani rastgele değişkenin alabileceği değerler sınırlı idi. Örneğin bir bozuk paranın üç kere atılma deneyinde, gelen tura sayılarını gösteren X rastgele değişkeninin alabileceği değerler yalnızca $0,1,2,$ yada 3'tü. Bunlardan başkası olamazdı.

Şimdi ise belirli bir aralıkta her değeri alabilen (yani sonsuz değer) alabilen rastgele değişken türü ile ilgilenceğiz. Bu rastgele değişken türüne sürekli rastgele değişken diyeceğiz.

Sürekli rastgele değişkenler genelde bir aletle ölçülen büyüklüklerdir:

- arabanın hızı
- kişinin kilosu, boyu
- ortamın sıcaklığı...

Olasılık Yoğunluk Fonksiyonu (Probability Density Function)(PDF)

X rastgele değişkeni ayrık (discrete) iken $P(X = x)$ olasılığı ile yani X, rastgele değişkeninin belirli bir x değerini alma olasılığı ile ilgilenmiştik. Bize bu olasılığı veren fonksiyona olasılık kütle fonksiyonu (pmf) demiş, p_X ile göstermiştik. $\left(p_X(x) = P(X = x)\right)$.

X rastgele değişkeni sürekli olduğunda, X'in belirli tek bir değeri alma olasılığı 0 olur. (Bir kişinin boyunun 188.921 olma olasılığı 0'dır gibi düşünülebilir).

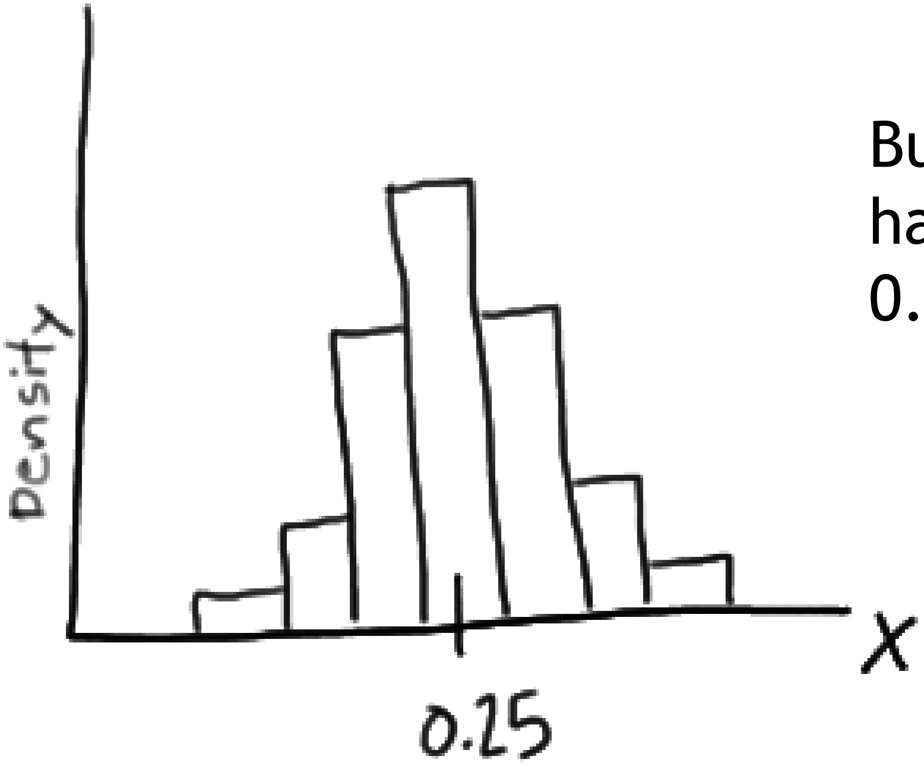
Onun yerine bir (a, b) aralığı için $P(a < X < b)$ olasılğı ile ilgileneceğiz. (Bir kişinin boyunun 176 ile 184 cm arasında olma olasılığı nedir gibi). Bunu olasılık yoğunluk fonksiyonu yardımıyla bulacağız.

ör. Bir hamburgercide hamburgerdeki et ağırlığının 0.25 kg olduğu iddia ediliyor. Fakat aldığınız herhangi iki hamburgerdeki et ağırlıkları 0.23 kg ve 0.28 kg olabilir. Yani herhangi bir kesinlik yok. Bu durumda alacağınız bir hamburgerin örneğin 0.20 ile 0.30 kg arasında olma olasılığı ne olur?

X rastgele değişkeni bir hamburgerin ağırılığı olmak üzere $P(0.20 < X < 0.30) =?$

Olasılık Yoğunluk Fonksiyonu (Probability Density Function)(PDF)

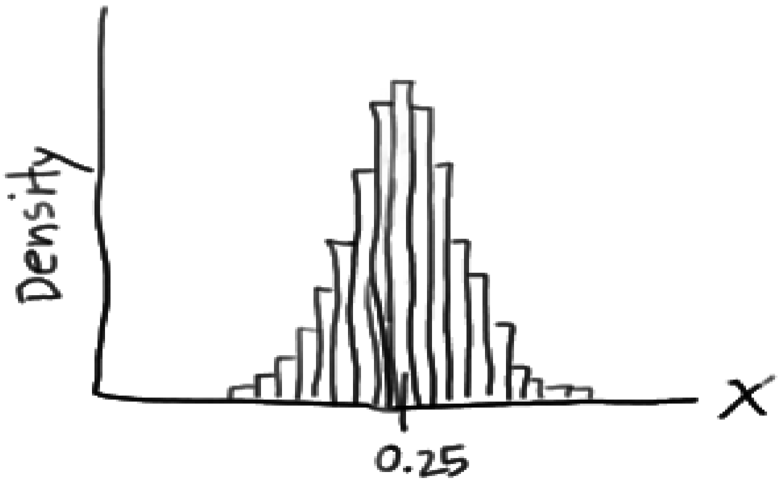
Diyelimki 100 tane hamburger aldınız, bunların içindeki et miktarlarını ölçtünüz ve aşağıdaki histogramı oluşturdunuz.



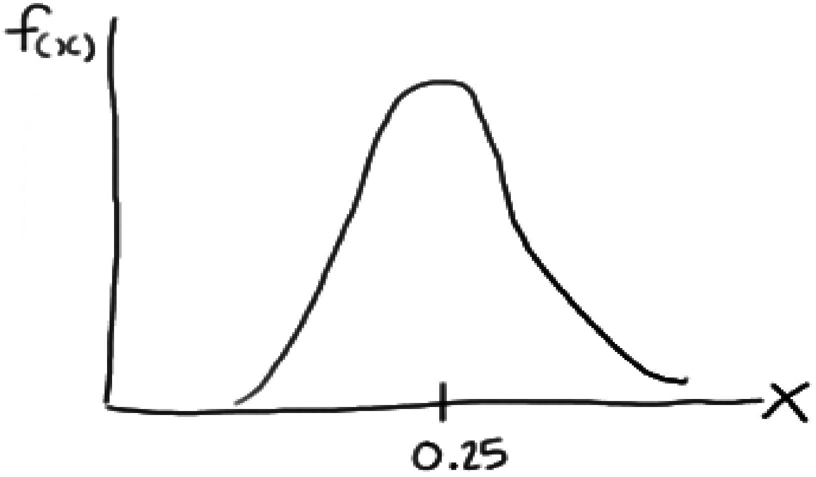
Buradan görülüyorki gerçekten hamburgerlerin en sık görülen ağırlığı 0.25 kg.

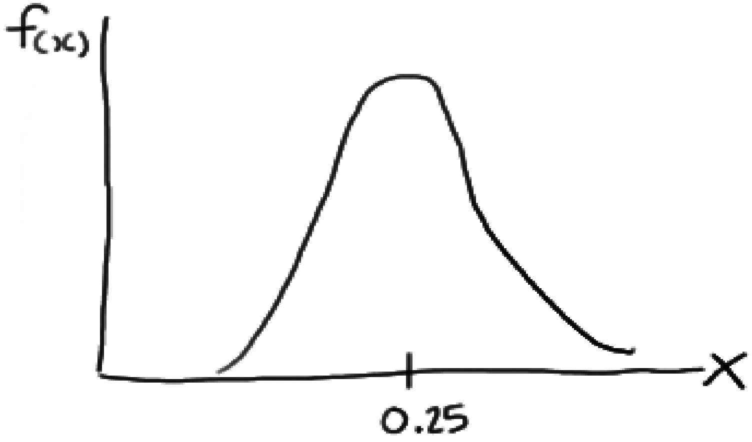
Olasılık Yoğunluk Fonksiyonu (Probability Density Function)(PDF)

Şimdi histogramdaki aralık büyüklüğünü küçültelim.



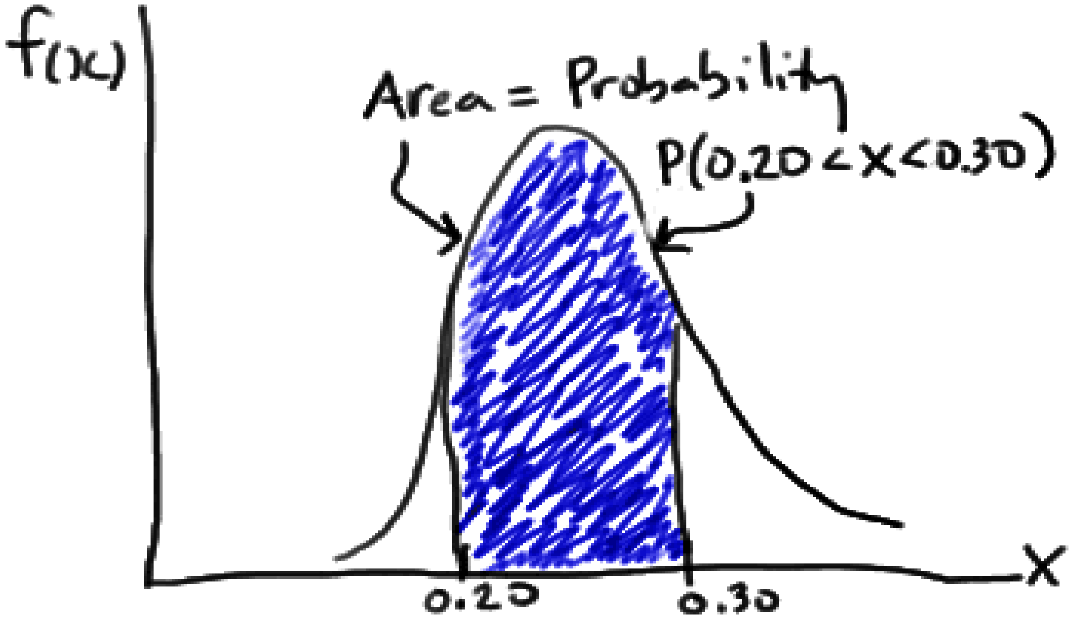
Histogramdaki aralık büyüklüğünü çok daha küçültürsek,örneğin 0 yaparsak aralıklar yalnızca bir değer alır. Herbir değerin olasılığı bir nokta olur. Bu noktalar birleştirilirse aşağıdaki gibi bir eğri elde edilir:





Ortaya çıkan bu eğriye olasılık yoğunluk fonksiyonu denir, $f(x)$ ile gösterilir.
Bu eğrinin altında kalan alanın büyüklüğü 1'dir.
Bir sürekli rastgele değişkenin belirli bir aralıkta olma olasılığı, eğrinin altında bu aralık tarafından sınırlanan alan olur.

Rastgele seçilen bir hamburgerdeki et ağırılığının 0.20 ile 0.30 arasında olma olasılığı: $P(0.20 < X < 0.30) =$

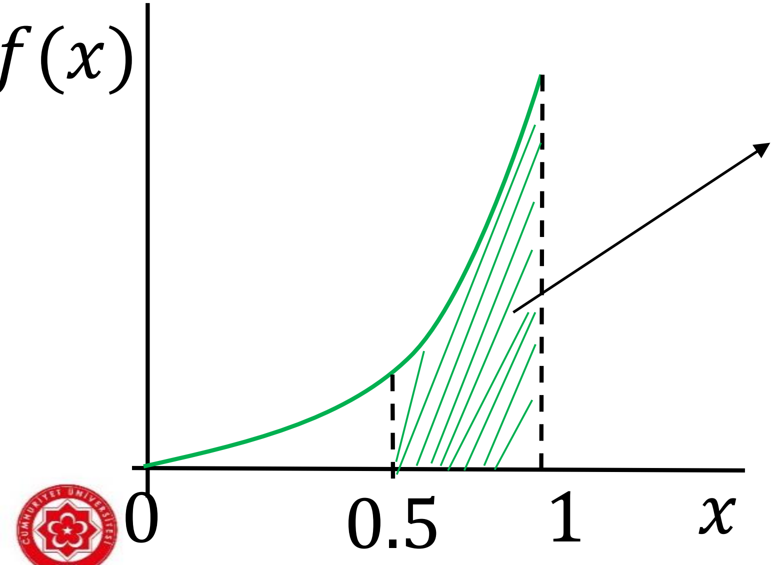


Sonuç olarak X bir sürekli rastgele değişken, $f(x)$ de onun olasılık yoğunluk fonksiyonu iken X'in bir (a,b) aralığında olma olasılığı:

$$\boxed{P(a < X < b) = \int_a^b f(x)dx}$$

ör. X rastgele değişkenin olsaılık yoğunluk fonksiyonu $f(x) = 3x^2$; ve X'in alabilceği değerler [0,1] aralığında olsun. Bu durumda $P(0.5 < X < 1)$ olasılığı ne olur:

Çözüm:



$$P(0.5 < X < 1) = \int_{0.5}^{1} 3x^2 dx = x^3|_{0.5}^{1} = 1^3 - 0.5^3 = \frac{7}{8}$$

Bir önceki örnek için $P(X = 0.5) = ?$

$$P(X = 0.5) = \int_{0.5}^{0.5} 3x^2 dx = x^3|_{0.5}^{0.5} = 0.5^3 - 0.5^3 = 0$$

Sonuç X bir sürekli rastgele değişken iken X'in belirli tek bir değeri alma olasılığı her zaman 0'dır.

Normal Rastgele Değişken

En önemli sürekli rastgele değişken türüdür.

X normal rastgele değişkeninin olasılık yoğunluk fonksiyonu:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Burada μ, X'in beklenen değeridir: $\mu = E[X]$. σ, X'in standart sapmasıdır.

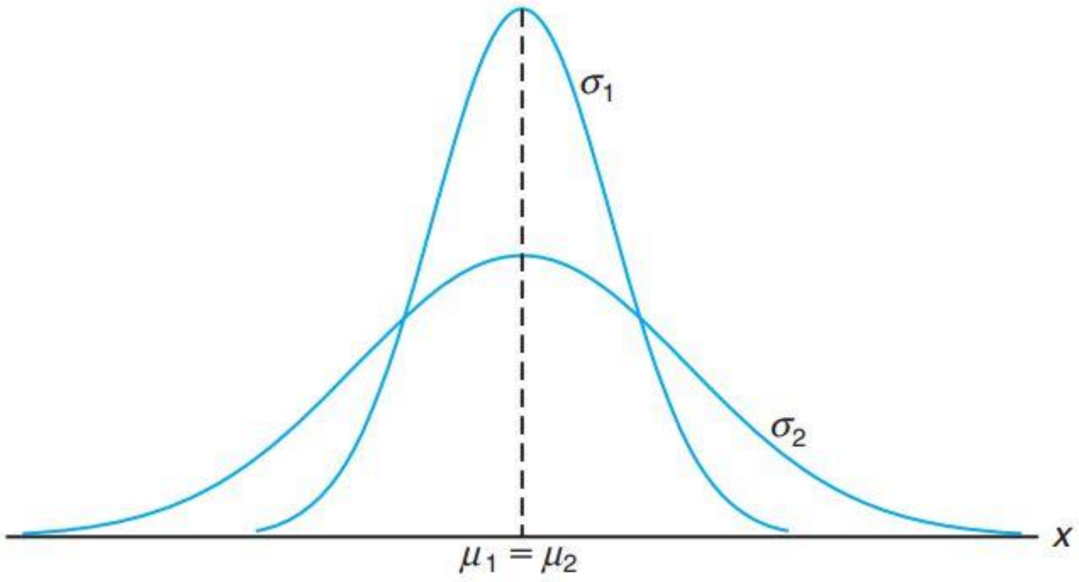
Bu fonksiyonun integrali alınamayacağından biz bu fonksiyonu direkt kullanmayacağız.

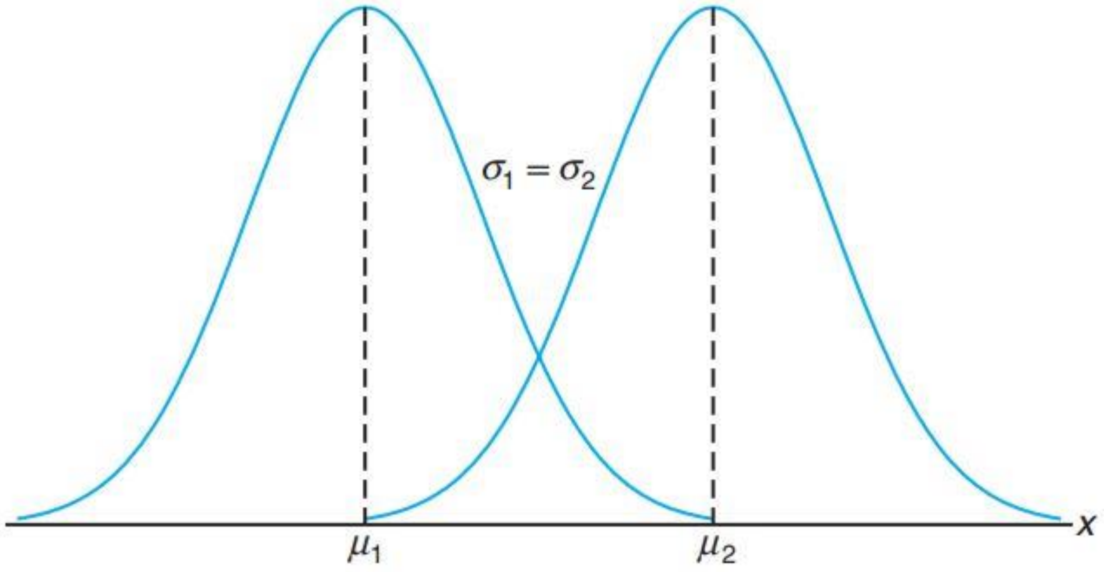
Normal Rastgele Değişken

X'rastgele değişkeni ortalaması (μ) ve standart sapması (σ) ile karakterize olur.

Standart sapma (σ) , X'in merkezi μ'den ne kadar uzağa yayılıdığını gösterir.

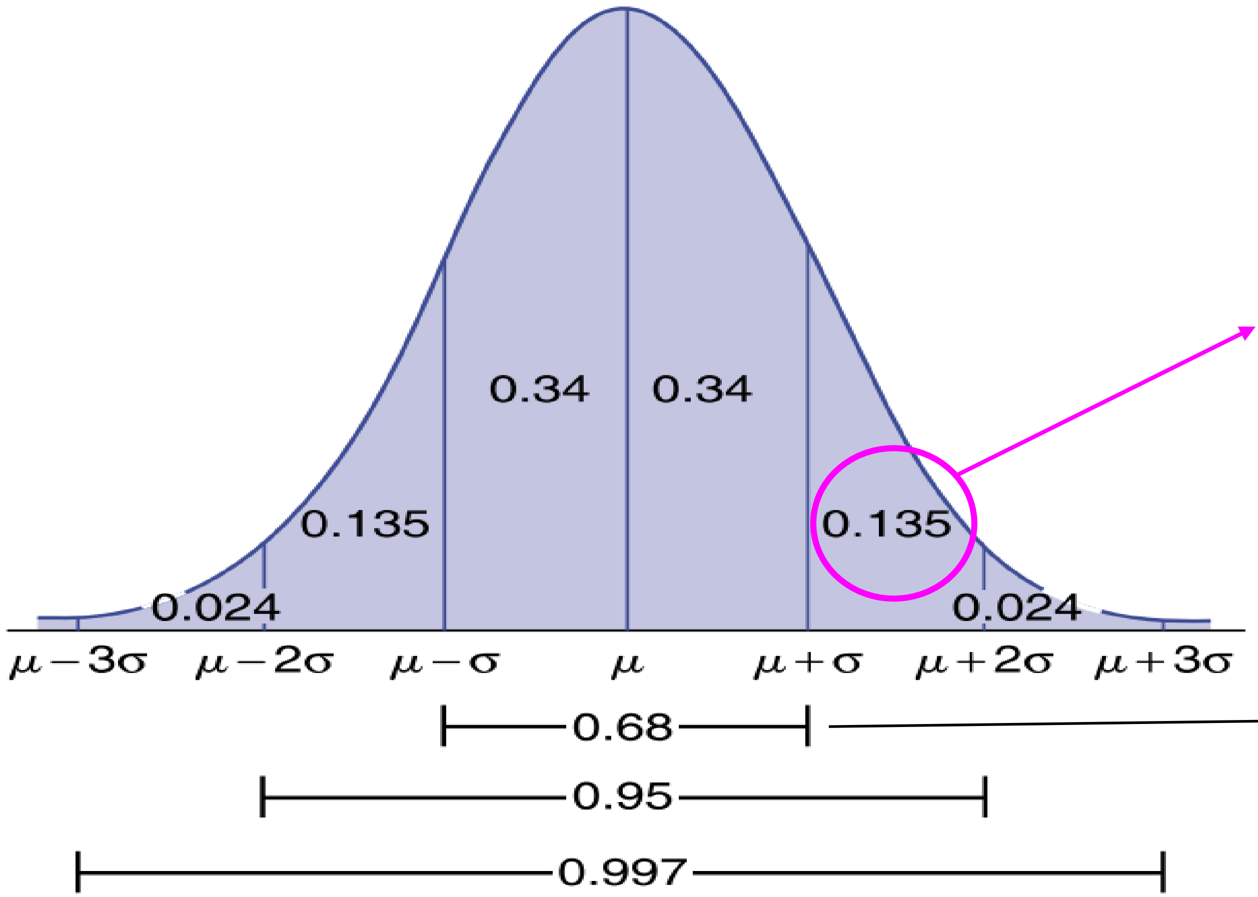
X ve Y ortalamalari aynı, fakat standart sapmaları farklı farklı olsun. Y'nin standart sapması σ_2, X'nin standart sapması σ_1'den büyük olsun: $\sigma_1 < \sigma_2$. Bu durumda aşağıda soldaki grafik elde edilir. Eğer standart sapmları ayni, ortalamaları $\mu_1 < \mu_2$ olacak şekilde farklı ise sağdaki grafik elde edilir.





Normal Eğrinin Altında Kalan Alan

X normal sürekli rastgele değişkeninin olasılık yoğunluk fonksiyonun grafiğine ***normal eğri*** diyeceğiz. Bu eğrinin altında kalan alanı kullanarak, normal dağılıma sahip bir rastgele değişkenin belirli bir aralıkta değer almasının olasılığını

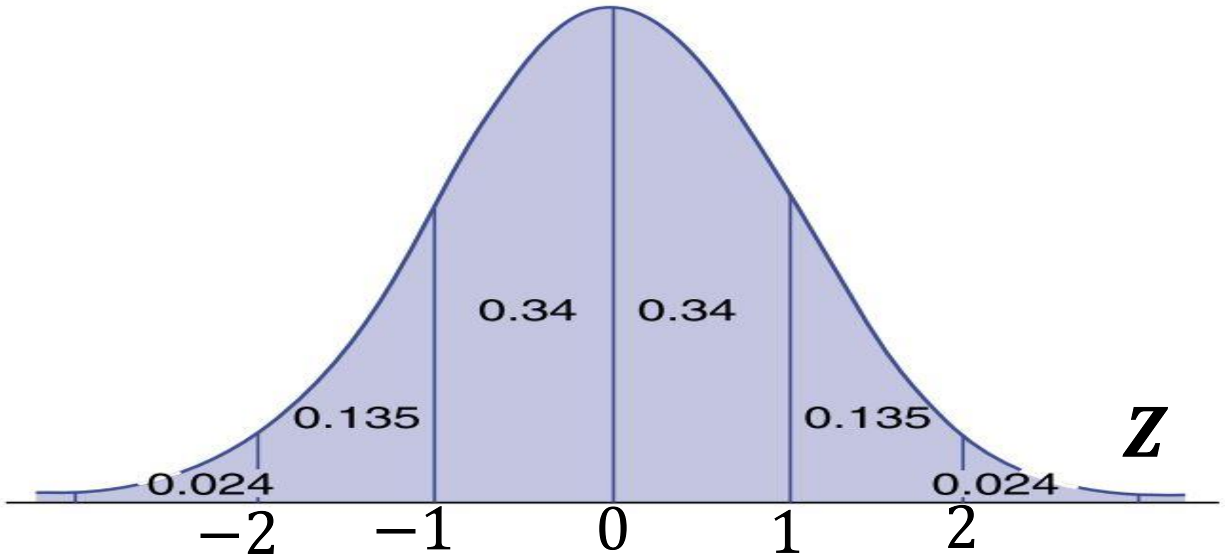


X'in %13.5'i $[\mu+\sigma,\ \mu+2\sigma]$ aralığındadır.

X'in %68'i $[\mu-\sigma,\ \mu+\sigma]$ aralığındadır.

ör. Bir sınıftaki kişilerin IQ skorlarının normal dağılıma sahip olduğunu, bu kişilerin IQ ortalamalarının 100 olduğunu ve standart sapmanın 16 olduğunu varsayalım.

Bu durumda diyebilirizki sınıftaki kişilerin %68'inin IQ ortalaması [84, 116] arasındadır. Sınıftaki kişilerin %95'inin IQ ortalaması $[100 - 2 \times 16, 100 + 2 \times$



Bir X normal rastgele değişkeni, standart normal rastgele dönüştürülürken, X’den önce μ çıkartılır ardından bölüm satandart sapmaya σ bölünür.

$$\boxed{Z \leftarrow \frac{X-\mu}{\sigma}}$$

Bir önceki örnekte sınıftan seçilen rastgele birinin IQ’su 120’den fazla olma olasılığı nedir?

$$Z \leftarrow \frac{120-100}{16}$$

olup,120’ye karşılık gelen Z değeri 1.25’tir.

O halde $P(X > 120) = P(Z > 1.25) = 1 - P(Z \leq 1.25) = 1 - 0.8944 = 0.1056$

Burada $P(Z \leq 1.25)$ olasılığını Standart Normal Olasılıklar Tablosu’ndan bulduk.

Bu tablo her zaman $P(Z \leq z)$ $(z \in [-4, 4]\,)$ olasılıklarını verir.

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

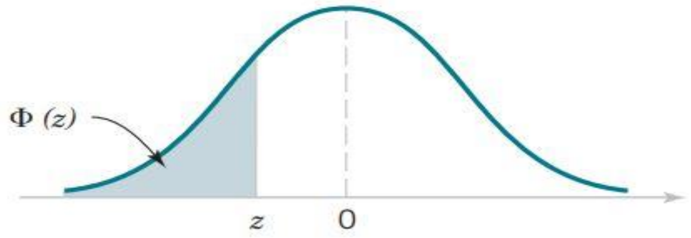


TABLE III Cumulative Standard Normal Distribution

z	−0.09	−0.08	−0.07	−0.06	−0.05	−0.04	−0.03	−0.02	−0.01	−0.00
−3.9	0.000033	0.000034	0.000036	0.000037	0.000039	0.000041	0.000042	0.000044	0.000046	0.000048
−3.8	0.000050	0.000052	0.000054	0.000057	0.000059	0.000062	0.000064	0.000067	0.000069	0.000072
−3.7	0.000075	0.000078	0.000082	0.000085	0.000088	0.000092	0.000096	0.000100	0.000104	0.000108
−3.6	0.000112	0.000117	0.000121	0.000126	0.000131	0.000136	0.000142	0.000147	0.000153	0.000159
−3.5	0.000165	0.000172	0.000179	0.000185	0.000193	0.000200	0.000208	0.000216	0.000224	0.000233
−3.4	0.000242	0.000251	0.000260	0.000270	0.000280	0.000291	0.000302	0.000313	0.000325	0.000337
−3.3	0.000350	0.000362	0.000376	0.000390	0.000404	0.000419	0.000434	0.000450	0.000467	0.000483
−3.2	0.000501	0.000519	0.000538	0.000557	0.000577	0.000598	0.000619	0.000641	0.000664	0.000687
−3.1	0.000711	0.000736	0.000762	0.000789	0.000816	0.000845	0.000874	0.000904	0.000935	0.000968
−3.0	0.001001	0.001035	0.001070	0.001107	0.001144	0.001183	0.001223	0.001264	0.001306	0.001350
−2.9	0.001395	0.001441	0.001489	0.001538	0.001589	0.001641	0.001695	0.001750	0.001807	0.001866
−2.8	0.001926	0.001988	0.002052	0.002118	0.002186	0.002256	0.002327	0.002401	0.002477	0.002555
−2.7	0.002635	0.002718	0.002803	0.002890	0.002980	0.003072	0.003167	0.003264	0.003364	0.003467
−2.6	0.003573	0.003681	0.003793	0.003907	0.004025	0.004145	0.004269	0.004396	0.004527	0.004661

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du$$

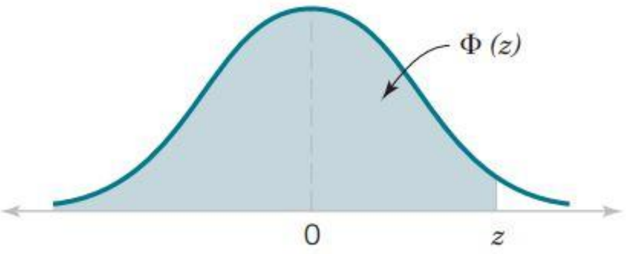
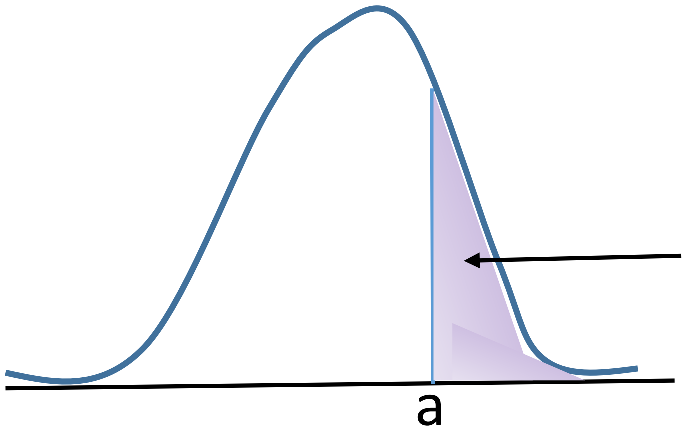


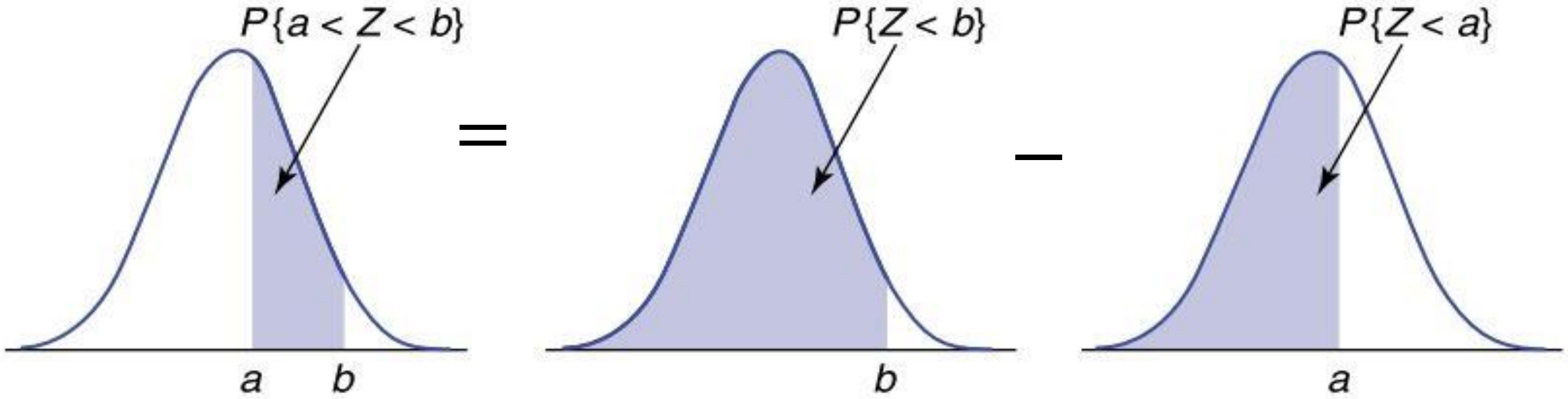
TABLE III **Cumulative Standard Normal Distribution (*continued*)**

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736

$$\boxed{P(Z \le 1.18) = 0.881000}$$



$$P(Z > a) = 1 - P(Z \leq a)$$



$$P(a < Z < b) = P(Z < b) - P(Z < a)$$

ör. Belirli bir otomobil lastiğinin ömrü, ortalaması 35000 mil, standart sapması 5000 mil sahip olacak şekilde normal dağılıma sahip olsun.

i) Bir lastiğin ömrünün 30000 ile 40000 arasında olma olasılığı nedir?

ii) Bir lastiğin ömrünün en az 43000 mil olma olasılığı nedir?

Çözüm.

$X = 30000$ sürekli rastgele değişkenine karşılık gelen Z değeri:

$$Z = \frac{30000 - 35000}{5000} = -1$$

$X = 40000$ sürekli rastgele değişkenine karşılık gelen Z değeri:

$$Z = \frac{40000 - 35000}{5000} = 1$$

$$P(30000 < X < 40000) = P(-1 < Z < 1) = P(Z < 1) - P(Z < -1)$$
$$= 0.8413 - 0.1378 = 0.735$$

ii) $P(X \geq 43000) = P(Z \geq (43000 - 35000)/5000) = P(Z \geq 1.6) =$
$$= 1 - P(Z < 1.6) = 1 - 0.9452 = 0.0548$$

ör. Gençlerin bir dakikadaki kalp atış sayısı, ortalama 72 ve 9.5 standart sapmaya sahip olacak şekilde normal dağılıma sahip olsun. Ayrıca eğer bir kişinin bir dakikadaki kalp sayış sayısı 95 üzeri ise askerlikten muaf sayılıyor olsun. Buna göre gençlerin yüzde kaçı fazla kalp atış yüzünden askere alınmaz?

Çözüm.

X rastgele seçilen bir kişinin kalp atış sayısı olsun. Biz $P(X > 95)$ ile ilgileniyoruz.

$$P(X > 95) = P(Z > (95 - 72)/9.5) = P(Z > 2.42) = 1 - P(Z \leq 2.42) = 1 - 0.9922 = 0.0078$$

Bu olasılık ise $\%0.78$'e eşittir.