

Olasılık ve İstatistik

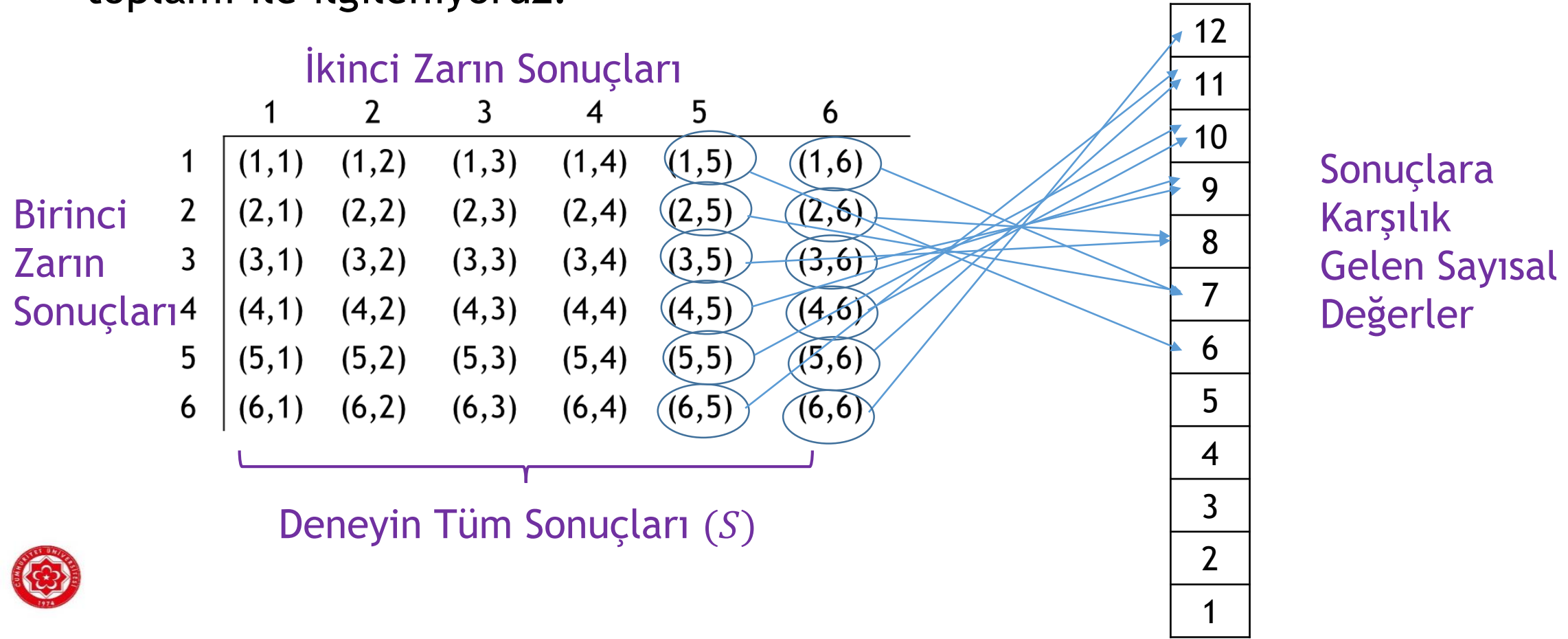
Fırat İsmailoğlu, PhD

Ayrık Rastgele Değişkenler

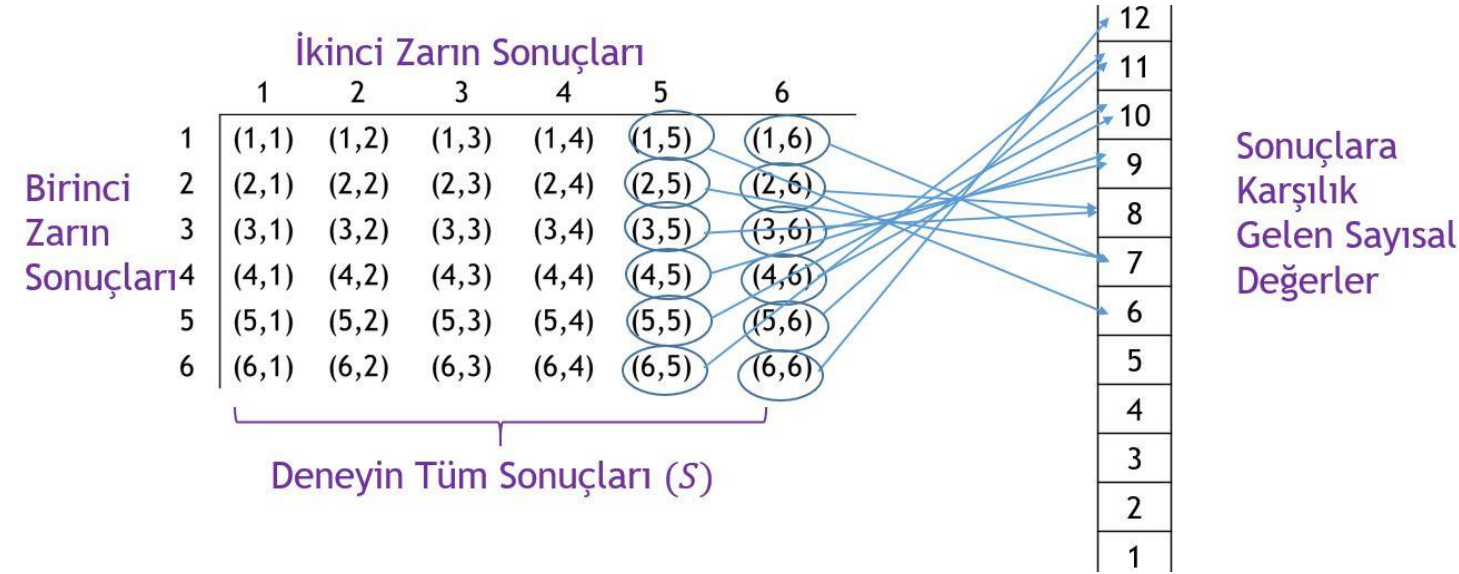
Rastgele Değişkenler (Random Variables)

Bir deneyin önceden bilemediğimiz sonuçları olduğunu ve olabilecek tüm sonuçların örnek uzayı (S) oluşturduğunu söylemiştik. Genelde bu sonuçlara karşılık gelen sayısal değerlerle ilgileniriz.

İki zarın atılma deneyinde tüm sonuçlar şunlardır. Diyelimki biz zarların toplamı ile ilgileniyoruz.



Rastgele Değişkenler (Random Variables)



Bu şekilde her sonuca karşılık gelen bir sayısal değer vardır (burada aslında her sonuç için bir ok vardır ama çok karmaşık gözükmemesin diye her oku çizmedim)

İki zar atılma deneyinde sonuçların sahip olduğu başka değerlerle ilgilenebiliriz:

- zarlardaki büyük olan değer
- zarlardaki küçük olan değer
- zarların toplamının asal olup olmadığı (asal :1, asal değil: 0)
- zarlarda kaç defa 4 görüldüğü (0, 1 yada 2)
- zarların farkının karesi

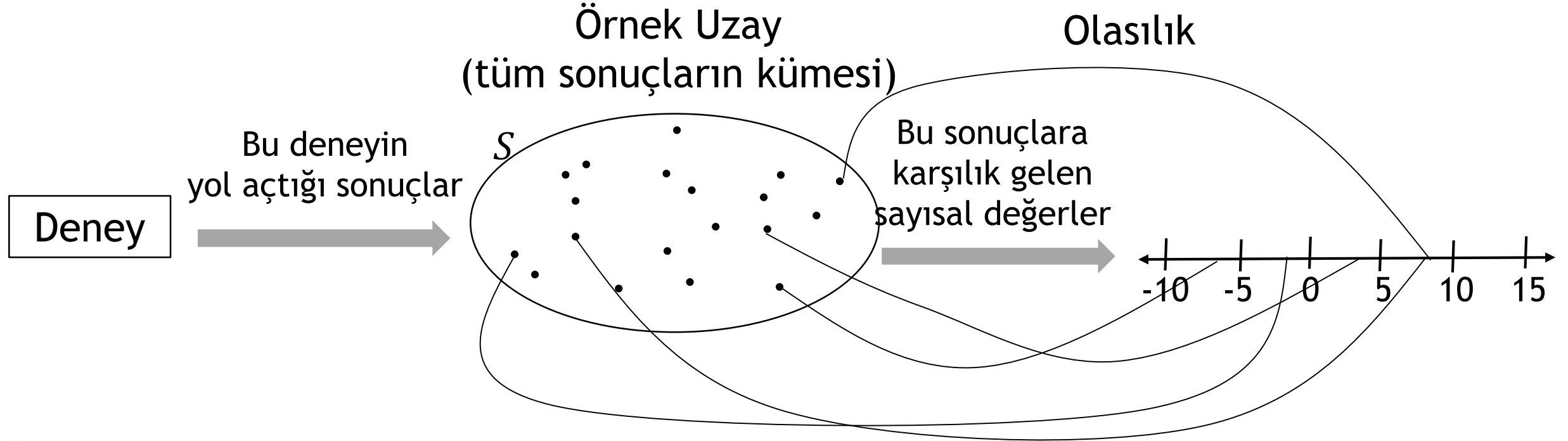


Rastgele Değişkenler (Random Variables)

Yada bir bozuk paranın 5 defa atılma deneyinde gelen tura sayısı: 0,1,2,3,4,5

Rastgele 10 kişi seçilme deneyinde, bu 10 kişiden araba sürebilenlerin sayısı.

Sonuç olarak deneylerin sonuçlarının sahip olduğu sayısal değere rastgele değişken (random variable) diyeceğiz, büyük harfle göstereceğiz: X, Y, Z, \dots



Ayrık Rastgele Değişkenler (Discrete Random Variables)

Eğer bir rastgele değişkenin alabileceği değerler sonlu sayıda elemanı olan bir kümeden seçiliyorsa (yani bu rastgele değişken sonlu tane değer alabiliyorsa) bu rastgele değişkene ayrık (discrete) rastgele değişken diyeceğiz.

ör. Bir ailenin 3 çocuğunun kaçının kız olduğu ile ilgilenelim.

Örnek uzay: $S = \{(k, k, k), (k, k, e), (k, e, k), (e, k, k), (k, e, e), (e, e, k), (e, k, e), (e, e, e)\}$

Burada rastgele değişken 3 olabilir: (k, k, k) sonucu için

Rastgele değişken 2 olabilir: $(k, k, e), (k, e, k), (e, k, k)$ sonuçları için

Rastgele değişken 1 olabilir: $(k, e, e), (e, e, k), (e, k, e)$ sonuçları için

Rastgele değişken 0 olabilir: (e, e, e) sonucu için.

Sonuç olarak burada rastgele değişkenin alabileceği değerler sonlu tanedir:

$$X = 0, X = 1, X = 2, X = 3$$

olabilir. O halde bu rastgele değişken ayrıktır.



Ayrık Rastgele Değişkenler (Discrete Random Variables)

ör. 110 kişinin aldığı olasılık dersinden kaç kişinin geçeceği ile ilgilenelim. Eğer X rastgele değişken bu sayıyı gösterirse, X ; 0-110 arası bir tam sayı alır (yani 111 tamsayıdan biridir). Bu ayrık bir rastgele değişkendir.

ör. Bir bozuk paranın 5 kez atılma deneyinde biz paraların kaç tanesinin Tura geldiği ile ilgilenelim. X rastgele değişkeni burada 0-5 rası bire tam sayı değeri alır, sonlu olduğundan bir ayrık rastgele değişkenidir.

ör. Bir kişinin başvurduğu bir iş deneyinde, olası sonuçlar iki tanedir: işe alınma, işe alınmama. Örnek uzayın yalnızca iki elemanı var. Diyelimki X rastgele değişkeni işe alınma sonucu için 1, işe alınmama sonucu için 0 değerini alsın. Bu durumda da X rastgele değişkeni iki değerden birini alabilceği için yani sonlu olduğu için bir ayrık rastgele değişken olur.



Olasılık Dağılımı (Probability Distribution)

Bir deneyde rastgele değişkenin alabileceği her değere karşılık bir olasılık vardır. Diyelimki bir deneyde X rastgele değişkeni x_1, x_2, \dots, x_n değerlerinden birini alsın. Bu değerlerden birini alma olasılığını $P(X = x_i)$ şeklinde göstereceğiz.

Örneğin bir önceki örnekte, X rastgele değişkeni (e, e, e) sonucu için 0 değerini alır. O halde $P(X = 0) = \frac{1}{8}$

X rastgele değişkeni $(k, e, e), (e, e, k), (e, k, e)$ sonuçları için 1 değerini alır. O halde $P(X = 1) = \frac{3}{8}$

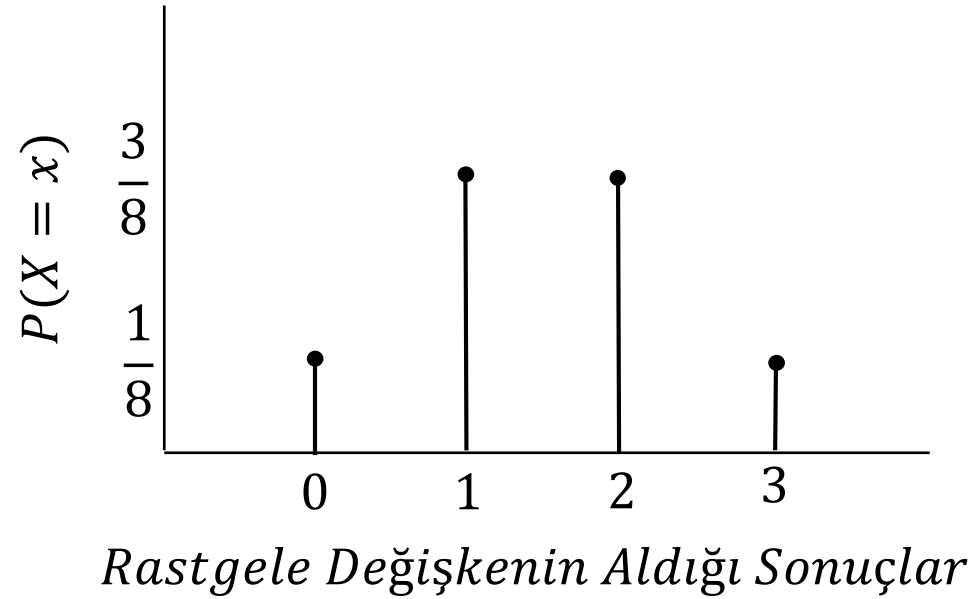
X rastgele değişkeni $(k, k, e), (k, e, k), (e, k, k)$ sonuçları için 2 değerini alır. O halde $P(X = 2) = \frac{3}{8}$

X rastgele değişkeni (k, k, k) , sonucu için 3 değerini alır. O halde $P(X = 3) = \frac{1}{8}$



Olasılık Dağılımı (Probability Distribution)

x	$P(X = x)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$
$\frac{+}{1}$	



$P(X = x_1), P(X = x_2), \dots, P(X = x_n)$ olasılıklarının hepsine birden X ayrık rastgele değişkeninin **olasılık dağılımı** diyeceğiz.

Doğal olarak $\sum_{i=1}^n x_i = 1$.

ör. Diyelimki bir kasede 7 mavi 3 beyaz top var. Deney olarak bu kaseden iki top rastgele çekiliyor. X rastgele değişkeni çekilen toplardaki beyaz top sayısı olsun. Bu durumda X rastgele değişkenin olasılık dağılımı ne olur?

Çözüm.

- Çekilen iki topun hiçbiri beyaz olmayabilir. Bu durumda $X = 0$ olur. Bu durumda çekilen iki top da mavidir. Bunun olasılığını hesaplayalım.

A : ilk topun mavi olma olayı, B : ikinci topun mavi olma olayı olsun. Biz bu olayların ikisi de olsun istiyoruz: $A \cap B$.

Çarpma kuralından: $P(A \cap B) = P(A)P(B|A)$

$P(A)$: ilk topun mavi olma olasılığı: $\frac{7}{10}$

$P(B|A)$: çekilen ilk topun mavi olduğu bilinirken ikincinin mavi olma olasılığı. ilk top mavi olduğundan geriye kalan toplam 9 topun 6'sı mavidir: $P(B|A) = \frac{6}{9}$

O halde $P(X = 0) = \frac{7}{10} \times \frac{6}{9} = \frac{7}{15}$



- Çekilen iki topun biri beyaz biri mavi olabilir. Bu durumda toplam beyaz top sayısı 1 olur: $X = 1$.

İlk top mavi ikinci top beyaz veya ilk top beyaz ikinci top mavi olabilir.

İlk top mavi olma olasılığı: $\frac{7}{10}$

İlk top mavi iken ikincisinin beyaz olma olasılığı: $\frac{3}{9}$

İlk top beyaz olma olasılığı: $\frac{3}{10}$

İlk top beyaz iken ikincisinin mavi olma olasılığı: $\frac{7}{9}$

$$P(X = 1) = \frac{7}{10} \times \frac{3}{9} + \frac{3}{10} \times \frac{7}{9} = \frac{7}{15}$$

- Çekilen topların ikisi birden beyaz olabilir. Bu durumda toplam beyaz sayısı 2 olur. Yani $X = 2$.

İlkinin beyaz olma olasılığı: $\frac{3}{10}$

İlk top beyaz iken ikincinin de beyaz olma olasılığı: $\frac{2}{9}$

$$P(X = 2) = \frac{3}{10} \times \frac{2}{9} = \frac{1}{15}$$

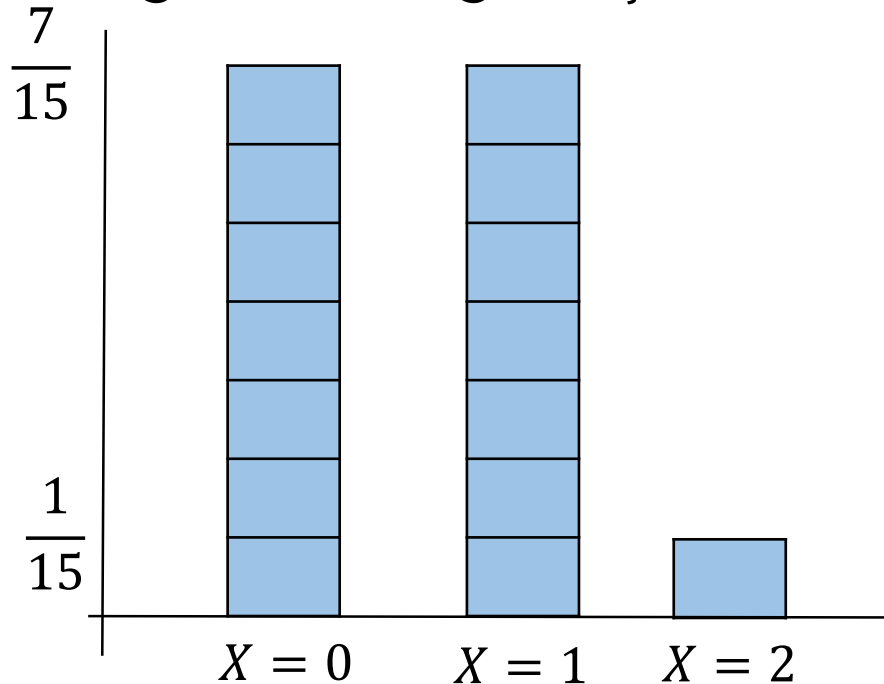


Olasılık dağılımı:

	$X = 0$	$X = 1$	$X = 2$	Toplam
$P(X = x)$	$\frac{7}{15}$	$\frac{7}{15}$	$\frac{1}{15}$	1

Eğer bu toplam 1 çıkmazsa, bir yerlerde hata yapmışsınız demektir.

Bu olasılık dağılımı histogram şeklinde de gösterilebilir:



ör. Diyelimki bir sigorta şirketi bireysel hayat sigortası yaptıran bir müşterisi öldüğünde yakınlarına 100,000 TL ödeme yapıyor. Diyelimki bu şirketin iki müşterisi var ve bu müşterilerin bu sene ölme olasılıkları 0.1 ve 0.05 olsun; ayrıca bu kişilerin ölme olasılıkları birbirinden bağımsız olsun. X rastgele değişkeni, şirketin bu sene ölüm olma durumunda yapacağı toplam ödemeyi gösterebilir. Bu durumda X rastgele değişkeninin olasılık dağılımı ne olur?

Çözüm.

Yani Müşterilerin hiçbiri ölmez. $X = 0$ olma durumu. $P(X = 0) = 0.9 \times 0.95 = 0.855$

Müşterilerden biri ölebilir: Bu durumda müşterilerin biri ölür, diğeri ölmez. $X = 100,000$ olma durumu.

$$P(X = 100,000) = 0.1 \times 0.95 + 0.05 \times 0.9 = 0.14$$

Müşterilerin ikisi de ölebilir. $X = 200,000$ olma durumu

$$P(X = 200,000) = 0.1 \times 0.05 = 0.005$$

	$X = 0$	$X = 100,000$	$X = 200,000$
$P(X = x)$	0.855	0.14	0.005

Olasılık Dağılımı



ör. Bir fırında sabahları 3 tane özel yapım pasta çıkıyor olsun. Bu pastalara günlük talep aşağıdaki gibi olsun.

0	0.15
1	0.20
2	0.35
3	0.15
4	0.10
5 yada daha fazla	0.05

X rastgele değişkeni günün sonunda satılmadan kalan pasta sayısını gösterebilir. Bu durumda X rastgele değişkeninin olasılık dağılımı ne olur?

Çözüm.

Pasta hiç satılmayabilir. $X = 3$ olur. Bu durumda hiç talep gelmemiştir: $P(X = 3) = 0.15$

1 pasta satılmış olabilir. $X = 2$ olur. Bu durumda 1 talep gelmiştir: $P(X = 2) = 0.20$

2 pasta satılmış olabilir. $X = 1$ olur. Bu durumda 2 talep gelmiştir: $P(X = 1) = 0.35$

3 pasta satılmış olabilir. $X = 0$ olur. Bu durumda 3, 4, 5 yada daha fazla talep gelmiştir: $P(X = 0) = 0.15 + 0.10 + 0.05 = 0.3$.



Olasılık Kütle Fonksiyonu (Probability Mass Function (PMF))

Olasılık kütle fonksiyonu aldığı ayrık rastgele değişkeni, onun olasılığına götüren fonksiyonun adıdır. p_X ile gösterilir.

$$p_X: X \rightarrow [0,1]$$

Bu fonksiyon aslında daha önce gösterdiğimiz $P(X = x)$ in kendisidir, yani X rastgele değişkenin x 'değerini alma olasılığını gösterir:

$$p_X(x) = P(X = x)$$

ör. X rastgele değişkeni, bir bozuk parayı üç kere atma deneyinde, gelen toplam tura sayısını gösterebilir. Bu rastgele değişkenin alabileceği her değer için olasılık kütle fonksiyonunu hesaplayınız.

$$p_X(0) = P(X = 0) = 1/8$$

$$p_X(1) = P(X = 1) = 3/8$$

$$p_X(2) = P(X = 2) = 3/8$$

$$p_X(3) = P(X = 3) = 1/8$$



Beklenen Değer (Expected Value)

Olasılık teorisindeki bir başka anahtar konu rastgele değişkenin beklenen değeridir.

X 'in beklenen değeri X 'in aldığı değerlerin ağırlıklı ortalamasıdır. Burada ağırlıklar değerlerin olasılıklarıdır (yani olasılık kütle fonksiyonudur).

X rastgele değişkenin beklenen değeri $E[X]$ ile gösterilir ve şöyle hesaplanır:

$$E[X] = \sum_{i=1}^n x_i P(X = x_i)$$

x_1, x_2, \dots, x_n değerleri X rastgele değişkenin alabileceği değerler ve $P(X = x_i)$, x_i rastgele değişkeninin görülme olasılığıdır.

Beklenen Değer (Expected Value)

ör. Bir zar atma deneyinde X rastgele değişkeni gelen zarı gösterson. Bu durumda X ; 1,2,3,4,5,6 değerlerinden her birini $\frac{1}{6}$ olasılıkla alır.

O halde X 'in beklenen değeri: $\frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6 = 3.5$

ör. Diyelimki çalıştığınız işte her ay 0.6 olasılıkla 1000 TL bonus, 0.3 olasılıkla 500TL bonus, 0.1 olasılıkla hiç bonus almıyorsunuz. Bu işte bir ayda beklediğiniz bonus miktarı ne kadar olur?

Çözüm.

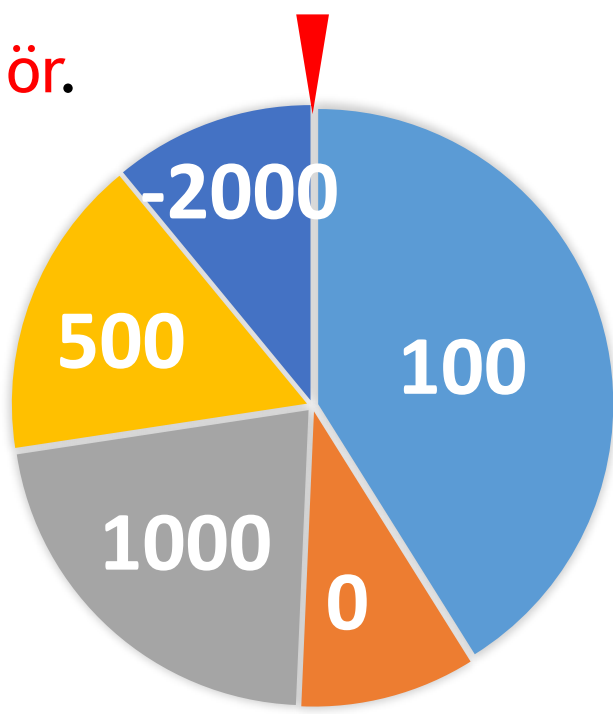
X rastgele değişkeni alacağınız bonus olsun. $P(X = 0) = 0.1, P(X = 500) = 0.3, P(X = 1000) = 0.6$

$$E[X] = 0 \times 0.1 + 500 \times 0.3 + 1000 \times 0.6 = 750$$

olduğundan ayda 750TL bonus almak beklenir.



ör.



Yandaki şeklin bir tür çarkıfelek olduğunu ve 30 defa çevrildiğini varsayalım. Aşağıdaki tablo her bir dilimin kaç defa geldiğini gösterebilir.

Gelen Dilim	Kaç Defa Geldiği
100	12
0	3
1000	7
500	6
–2000	2

Buna göre çarkıfeleği bir defa döndürmekle ne kadar kazanmayı bekleriz?

ör. X rastgele değişkeni bir çevrimde sonucunda gelebilecek miktarı gösterebilir. Bu durumda X 'in alabileceği değerler: 100, 0, 1000, 500, -2000'dir.

$$P(X = 100) = \frac{12}{30}, P(X = 0) = \frac{3}{30}, P(X = 1000) = \frac{7}{30}, P(X = 500) = \frac{6}{30}, P(X = -2000) = \frac{2}{30}$$



Beklenen Değer ~ Ortalama

Beklenen değer ile ortalama aslında birbirine çok yakın iki kavramdır. Eğer bir kümede her bir elemanın seçilme şansı aynı ise, bu kümenin ortalaması ile kümede seçilen elemanı gösteren rastgele değişkenin beklenen değeri aynı olur.

ör. Diyelimki 5 kişilik bir arkadaş grubundaki kişilerin boyları cm cinsinden 178, 184, 169, 190, 164 olsun.

Bu kişilerden biri rastgele seçilsin. X rastgele değişkeni seçilen kişinin boyunu gösterecek. X 'in alacağı değerler 178, 184, 169, 190, 164 olur ve her bir değeri alma olasılığı aynıdır: $P(X = x_i) = \frac{1}{5} (\forall i \in \{1,2,3,4,5\})$.

$$\text{O halde } E[X] = 178 \times \frac{1}{5} + 184 \times \frac{1}{5} + 169 \times \frac{1}{5} + 190 \times \frac{1}{5} + 164 \times \frac{1}{5} = 177$$

Bu değer aynı zamanda gruptaki kişilerin boy ortalamasıdır:

$$\bar{x} = \frac{178 + 184 + 169 + 190 + 164}{5} = 177$$



$E[X]$ 'in Yorumlanması

$E[X]$, X rastgele değişkenin, deneyin çok fazla kez tekrar etmesi durumunda alması beklenen değerdir (yaklaştığı değerdir).

Bir bozuk para atma deneyinde tura gelme olasılığı 0.5 idi. Fakat bu, para iki defa atıldığında birnin tura gelmesini beklemek yada 10 defa atıldığında 5'inin tura gelmesini beklemek değildi. Bu, uzun dönemde görülen sıklıktı. Bu olasılığın yakalanması için paranın çok çok fazla kez atılması (yani deneyin çok çok fazla kez tekrarlanması) gerekir. Yani örneğin para 100.000 defa atılırsa yaklaşık 50.000 kez tura görebiliriz.

Benzer şekilde, $E[X]$ de deney çok çok fazla kez tekrarlanırsa en sık göreceğimiz X değeri olur.

ör. Bir bozuk paranın iki defa atılma deneyini ele alalım. X rastgele değişkeni toplam gelen tura sayısını gösterebilir.

Bu durumda X 'in alacağı değerler 0, 1, ve 2 olur. Bunların olasılıkları

$$P(X = 0) = 1/4$$

$$P(X = 1) = 1/2$$

$$P(X = 2) = 1/4$$



X 'in beklenen değeri:

$$E[X] = 0 \times \frac{1}{4} + 1 \times \frac{1}{2} + 2 \times \frac{1}{4} = 1$$

Bir parayı iki defa atma deneyi çok çok fazla kez tekrar ederse, bu deneylerde genel olarak iki atıştan birinin tura olduğunu görürüz.

ör. Diyelimki bir davada avukat ya dava sonucuna bakmaksızın sabit ücret 1200 lira alacak, yada davanın kazanılması durumunda 5000 lira alacak davanın kaybedilmesi durumunda hic bir sey olmayacak. Aşağıda verilen davanın kazanılması olasılıkları için hangi durumlarda avukat standart ücret almayı seçmelidir?

a) $\frac{1}{2}$ b) $\frac{1}{5}$

Çözüm.

X rastgele değişkeni, avukatın standart ücreti almadığı durumdaki kazanacağı parayı gösterebilir. Bu durumda X , 0 yada 5000 olabilir.

a) $P(X = 5000) = 0.5, P(X = 0) = 0.5$



$$E[X] = 5000 \times 0.5 + 0 \times 0.5 = 2500$$

Yani davayı kazanma olasılığı 0.5 iken eğer avukat standart ücreti seçmesse, 2500 lira kazanması beklenir. Bu değer standart ücretten fazla olduğu için bu durumda avukat standart ücreti tercih etmemelidir.

$$\text{b) } P(X = 5000) = \frac{1}{5}, \quad P(X = 0) = \frac{4}{5}$$

$$E[X] = 5000 \times \frac{1}{5} + 0 \times \frac{4}{5} = 1000$$

Davayı kazanma olasılığı 0.2 iken, eğer avukat standart ücreti seçmesse, 1000 lira kazanması beklenir. Bu durumda standart ücreti seçmelidir.

ör. Bir yazılım firması bir tam gün çalışmazsa 4000 lira kaybediyor. Bu firma bilgisayarlar çalışırken elektrik kesilirse 12000 lira zarara uğruyor. Eğer yarın elektriklerin kesilme olasılığı 0.25 ise, bu firma kaybı minimize etmesi için, yarın çalışmalı mıdır, yoksa çalışmamalı mıdır?

Çözüm.

X rastgele değişkeni çalışılan gündeki zarar miktarı olsun. Elektrikler kesilmez ise zarar 0'dır; elektrikler kesilirse zarar 12000'dir. O halde X , 0 yada 12000 değerini alır.



Ertesi gün çalışılırsa zararın 0 olma olasılığı 0.75; zararın 12000 olma olasılığı 0.25'tir. O halde X 'in beklenen değeri:

$$E[X] = 0 \times 0.75 + 12000 \times 0.25 = 3000$$

Çalışılan günde beklenen kayıp 3000 liradır; ki bu kayıp hiç çalışılmama kaybı olan 4000 liradan azdır. O halde şirket çalışmalıdır.

ör. Olasılık sınavına arkadaşınızla beraber çalışırsanız 90, tek başına çalışırsanız 60 alacağınızı varsayalım. Yağmur yağarsa arkadaşınıza gitmeyeceksiniz, yağmaz ise arkadaşınıza gidip ders çalışacaksınız. Yağmurun yağma ihtimali yüzde yirmi ise, olasılık sınavından kaç almanız beklenmektedir?

Çözüm.

X rastgele değişkeni olasılık sınavından alacağınız puan olsun. X , 90 yada 60 değerini alabilir. Yağmurun yağma olasılığı 0.2, yağmama olasılığı 0.8 olduğuna göre 90 alma olasılığınız 0.8; 60 alma olasılığınız 0.2'dir. O halde sınavdan alacağınız beklenen değer:

$$E[X] = 90 \times 0.8 + 60 \times 0.2 = 84$$



ör. Eğer bir şirket 0.4 olasılıkla 50000 lira kazanması, yada 0.6 olasılıkla 35.000 zarar etmesi bekleniyorsa; bu şirket kar mı eder; yoksa zarar mı eder?

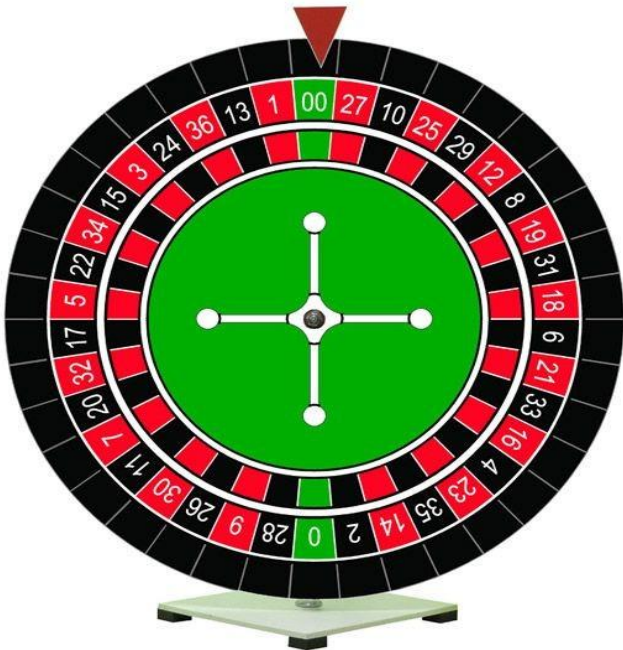
Çözüm.

X şirketin kazancı olsun. Bu durumda X , 50000 veya -35000 olabilir.

$$E[X] = 50000 \times 0.4 - 35000 \times 0.6 = -1000$$

şirketin 1000 lira zarar etmesi beklenmektedir.

ör.



Bir rulet çarkında 18 kırmızı, 18 siyah 2 yeşil numara vardır. 1000 liranızı kırmızıya koyarsanız ve kırmızı gelirse 1000 lira kar edersiniz, kırmızı gelmez ise 1000 lira zarar edersiniz. Bu durumda 1000 lira yatırdığınızda ne kadar kazanmayı beklersiniz?

Çözüm:

X , rastgele değişkeni kazancı gösterebilir. X , 1000 yada -1000 olabilir.

$$E[X] = 1000 \times \left(\frac{18}{38}\right) - 1000 \times \left(\frac{20}{38}\right) \approx -52$$

52 lira kaybetmeniz beklenir.



Rastgele Değişkenin Varyansı (Variance of Random Variables)

Değişkenleri özetlerken ortalama ve varyanstan faydalanıyorduk. Yani bir değişkenin ortalaması ve varyansı o değişkeni tanımlamaya, özetlemeye yetiyordu.

Rastgele değişkenleri de özetlerken, beklenen değerini ve varyansını göstereceğiz. Beklenen değer, onun ortalaması gibi düşünebiliriz. Fakat rastgele değişkenin aldığı değerlerin ortalamadan yani beklenen değerden ne kadar saptığını da bilmemiz gerekir. Örnek olarak aşağıdaki rastgele değişkenlerin beklenen değerleri aynıdır.

ör. $X = 0, P(X = 0) = 1, E[X] = 0.$

ör. $W = -10, P(W = -10) = 0.5; W = 10, P(W = 10) = 0.5, E[W] = 0.$

Bu örneklerde X ve W rastgele değişkenlerinin aldıkları değerler çok farklı olmasına rağmen beklenen değerleri aynıdır. Bu rastgele değişkenlerin beklenen değerlerinden ortalama ne kadar uzaklıkta olduğunu da hesaplamamız gerekir.



Rastgele Değişkenin Varyansı (Variance of Random Variables)

X rastgele değişkeni beklenen değerinden ne kadar sapması beklenir:

$(X - E[X])^2$ 'nin beklenen değeri nedir, yani $E[(X - E[X])^2]$?

$$\begin{aligned} E[(X - E[X])^2] &= E[(X^2 + (E[X])^2 - 2E[X]X)] \\ &= E[X^2] + E[E[X]^2] - E[2E[X]X] \\ &= E[X^2] + E[X]^2 - 2E[X]E[X] \\ &= E[X^2] + E[X]^2 - 2E[X]^2 \\ &= E[X^2] - E[X]^2 \end{aligned}$$

O halde bir X rastgele değişkeninin varyansı:

$$\boxed{Var(X) = E[X^2] - (E[X])^2}$$

ör. Bir bahise yatırılan 10 TL, 0.7 olasılıkla 10 lira kaybettirsin, 0.2 olasılıkla 40 lira kazandırısın ve 0.1 olasılıkla 80 TL kazandırısın. X rastgele değişkeni geliri göstermek üzere $Var(X)=?$

Çözüm:

$$E[X] = -10 \times 0.7 + 40 \times 0.2 + 80 \times 0.1 = 9$$



$$E[X] = -10 \times 0.7 + 40 \times 0.2 + 80 \times 0.1 = 9$$

olduğundan bu bahisten 9 lira kazanmamız beklenir.

X^2 değerleri: 100, 1600, 6400 olur.

$$E[X^2] = 100 \times 0.7 + 1600 \times 0.2 + 6400 \times 0.1 = 1030$$

$$Var(X) = 1030 - 9^2$$

$$= 949$$

