İstatistik ve Olasılık

Hazırlayan: M. Kemal Güvenç

Öğretmen: Veysel Harun Şahin

Koşullu Olasılık

A ve B bir olay olsun.

$P(A|B) \Rightarrow$ B olayı gerçekleştiğine göre A'nın gerçekleşme olasılığına koşullu olasılık denir.

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

$$P(AB) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

$$\Rightarrow P(A_1 A_2 \ldots A_n) = \underbrace{\underbrace{P(A_1) \cdot P(A_2|A_1)}_{P(A_1 A_2)} \cdot P_3(A_3|A_1 A_2)}_{P(A_1 A_2 A_3)} \ldots P(A_n|A_1 A_2 \ldots A_{n-1})$$

Toplam Olasılık

E ve F iki olay olsun

Bu durumda $E = \underbrace{EF \cup EF'}_{E \cap (F \cup F')}$ yazılabilir.

Buradan hareketle

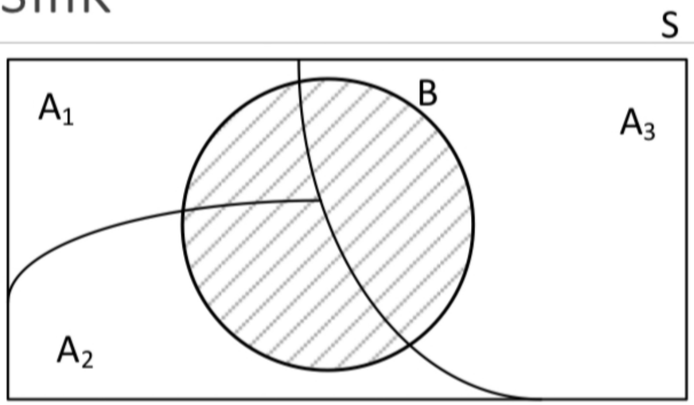
$$P(E) = P(EF) + P(EF')$$

$$P(E) = P(F)P(E|F) + P(F')P(E|F')$$

$$P(E) = P(F)P(E|F) + \big(1 - P(F)\big)P(E|F')$$

Toplam Olasılık

Üç olay durumu



$$S = A_1 \cup A_2 \cup A_3$$

$$P(S) = P(A_1) + P(A_2) + P(A_3) = 1$$

$$B = (A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B)$$

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$$

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)$$

Baye's Eşitliği

Bilinenler
- $P(A_i)$
- $P(B|A_i)$

İstenen
- $P(A_i|B)$

$$P(A_i|B)=\frac{P(A_i\cap B)}{P(B)}=\frac{P(A_i)\,P(B|A_i)}{\sum_{j=1}^{n}P(A_j)\cdot P(B|A_j)}$$

Bağımlı ve Bağımsız Olaylar

⇒ İki olaydan birinin olması diğerini etkiliyorsa bu olaylara bağımlı olaylar, etkilemiyorsa bağımsız olaylar denir.

⇒ Eğer iki olay bağımsız ise

$$P(A|B)=P(A)=\frac{P(AB)}{P(B)} \Rightarrow P(A)\cdot P(B)=P(AB)$$

Bağımsız Olaylar

3 adet olay da birbirinden bağımsız olabilir.

A, B ve C'nin 3 adet olay olduğunu düşünelim.

Bu olaylar aşağıdaki koşulları sağlıyorsa birbirinden bağımsızdır.

$$P(ABC) = P(A)P(B)P(C)$$

$$\left.\begin{array}{l} P(AB) = P(A)P(B) \\ P(AC) = P(A)P(C) \\ P(BC) = P(B)P(C) \end{array}\right\} \text{İkili bağımsızlık}$$

3 adet olaydan daha çok sayıda olay da birbirinden bağımsız olabilir.

Ayrık Olaylar

$\Rightarrow$ Aynı anda meydana gelmeyen olaylara ayrık olaylar denir

$\Rightarrow$ Ayrık olaylar bağımlı olaylardır

$\Rightarrow P(AB) = 0$

$\Rightarrow P(A \cup B) = P(A) + P(B)$

Rastgele Değişkeni:

⇒ Örnek uzayın elemanlarına karşılık gelen, sayısal değerlere sahip ilgilendiğimiz niceliklere rastgele değişken denir.

Gösterge Rastgele Değişkeni:

⇒ $I = \begin{cases} 1 & A \text{ gerçekleşir ise} \\ 0 & A \text{ gerçekleşmez ise} \end{cases}$

Kesikli Rastgele Değişkeni:

⇒ Mümkün değerler kümesi bir dizi olan rastgele değişkene kesiklidir denir.

Bu tür rastgele değişkenlere kesikli rastgele değişken denir.

1) Olasılık Kitle Fonksiyonu:

⇒ Kesikli rastgele değişkenlerin olasılık kitle fonksiyonu mevcuttur. $p(a)$, kesikli X rastgele değişkeninin olasılık kitle fonksiyonu olsun.

$$p(a) = P(X = a)$$

⇒ X rastgele değişkeni $x_1, x_2, x_3 \ldots$ değerlerini alıyor olsun.

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

2) Birikimli Dağılım Fonksiyonu:

Bir X rastgele değişkeninin birikimli dağılım fonksiyonu veya kısaca dağılım fonksiyonu F, herhangi bir x gerçek sayısı için aşağıdaki şekilde tanımlanır.

$$F(x) = P(X \leq x)$$

$X \sim F$ gösterimi, F'nin, X'in dağılım fonksiyonu olduğunu ifade eder.

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b)$$

$\Rightarrow$ Eğer $a < b$ ise

$$P(a < X \leq b) = F(b) - F(a)$$

Kesikli rastgele değişkenin birikimli dağılım fonksiyonu, olasılık kitle fonksiyonu kullanılarak aşağıdaki gibi ifade edilebilir.

$$F(a) = \sum_{\text{tüm } x \leq a} p(x)$$

$\Rightarrow$ X'in değerleri $x_1 < x_2 < x_3 \ldots$ şeklinde olsun.

Bu durumda F bir basamak fonksiyondur.

$$P(a < X \leq b) = F(b) - F(a)$$

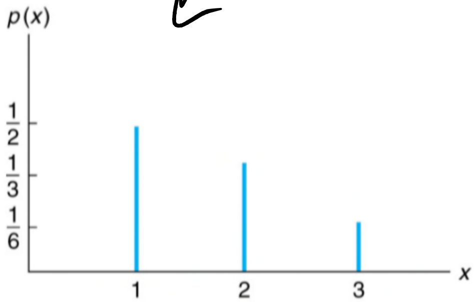
$$\begin{aligned} P(a \leq X \leq b) &= P(X = a) + P(a < X \leq b) \\ &= P(X = a) + F(b) - F(a) \end{aligned}$$

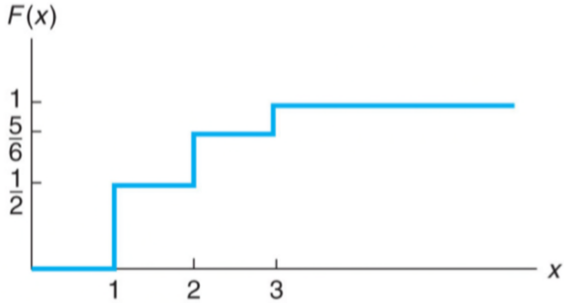
$$\begin{aligned} P(a < X < b) &= P(a < X \leq b) - P(X = b) \\ &= F(b) - F(a) - P(X = b) \end{aligned}$$

$$\begin{aligned} P(a \leq X < b) &= P(X = a) + P(a < X \leq b) - P(X = b) \\ &= P(X = a) + F(b) - F(a) - P(X = b) \end{aligned}$$

$$P(X > a) = 1 - P(X \leq a) = 1 - F(a)$$

$p(x)$ ve $F(x)$ fonksiyonlarının grafiği aşağıdaki gibidir.





3) Beklenti:

X rastgele değişkeninin beklentisi (beklenen değeri) $E[X]$ ile gösterilir ve aşağıdaki şekilde tanımlanır.

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

Beklenti, X'in alabildiği mümkün değerlerin ağırlıklı ortalamasıdır.

X'in her bir değeri varsayılan olasılığı ile ağırlıklandırılır.

$$I = \begin{cases} 0 & A \text{ olayı gerçekleşmemiş ise} \\ 1 & A \text{ olayı gerçekleşmiş ise} \end{cases}$$

$$E[I] = 0P(A') + 1P(A)$$

$$E[I] = P(A)$$

Gösterge Rastgele değişkeninin beklentisi

X, $p(x)$ olasılık kitle fonksiyonuna sahip kesikli bir rastgele değişken olsun. Bu durumda gerçek değerli herhangi bir g fonksiyonu için

$$E[g(X)] = \sum_x g(x) \cdot p(x)$$

Beklentinin Özellikleri :

a ve b sabitse

$$E[aX + b] = aE[X] + b$$

Bir sabitin beklentisi kendisidir.

$a = 0$ ise $E[b] = b$

Sabitle çarpılan bir rastgele değişkenin beklentisi nedir?

$b = 0$ ise $E[aX] = aE[X]$

n adet rastgele değişkenin toplamlarının beklentisini aşağıdaki gibi yazabiliriz.

$$E[X_1 + X_2 + X_3 + \cdots + X_n] = E[X_1] + E[X_2] + E[X_3] + \cdots + E[X_n]$$

$\Rightarrow$ X rastgele değişkeninin beklenen değeri olan $E[X]$'e aynı zamanda X'in ortalaması ya da birinci momenti de denir.

$(n \geq 1)$ olmak üzere $E[X^n]$'e ise X'in n'inci momenti denir.

$$E[X^n] = \sum_x x^n \cdot p(x)$$ $\Rightarrow$ X kesikli ise

4) Varyans:

Bir X rastgele değişkeninin değişimi, ortalamasından ne kadar uzakta olduğuna bakılarak ölçülebilir.

X, μ ortalamaya sahip ($E(X) = \mu$) bir rastgele değişken ise bu rastgele değişkenin varyansını aşağıdaki şekilde hesaplayabiliriz.

$$Var(X) = E\left[(X-\mu)^2\right]$$

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

$$I = \begin{cases} 0 & A \text{ olayı gerçekleşmemiş ise} \\ 1 & A \text{ olayı gerçekleşmiş ise} \end{cases}$$

$$Var(I) = E[I^2] - (E[I])^2$$

$$E[I^2] = 0^2 P(A') + 1^2 P(A) = P(A)$$

$$E[I] = 0P(A') + 1P(A) = P(A)$$

$$\begin{aligned}
Var(I) &= E[I^2] - (E[I])^2 \\
&= P(A) - (P(A))^2 \\
&= P(A)[1 - P(A)] \\
&= P(A)P(A')
\end{aligned}$$

Varyansın Özellikleri :

a ve b sabitse

$$Var(aX + b) = a^2 Var(X)$$

n adet rastgele değişken bağımsız ise toplamlarının varyansını aşağıdaki gibi yazabiliriz.

$$Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n)$$

Kesikli Rastgele Değişkenler

1) Bernoulli Rastgele Değişkeni :

Bir deneyin başarılı olup olmadığını gösteren rastgele değişkendir. Eğer deney başarılı ise 1, değilse 0 değerini alır.

X ⇒ Bir Bernoulli rastgele değişkeni

$P(X=0) = 1-p \quad P(X=1) = p$

$E[X] = 1 \cdot \overbrace{P(X=1)}^{p} + 0 \cdot \overbrace{P(X=0)}^{1-p} = p$

$Var(X) = p \cdot (1-p)$

2) Binom Rastgele Değişkeni :

p olasılıkla başarılı ve $1-p$ olasılıkla başarısız olan birbirinden bağımsız n adet deney gerçekleştirdiğimizi düşünelim. X rastgele değişkeni, bu n adet deneyin başarılı sonuç sayısını gösteren rastgele değişken olsun. Bu durumda X rastgele değişkenine $(\boldsymbol{n}, \boldsymbol{p})$ **parametreli binom rastgele değişkeni** denir ve olasılık kitle fonksiyonu aşağıdaki gibidir.

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \qquad i = 0,1,2, \dots, n$$

Bernoulli rastgele değişkeni $(1, p)$ parametreli binom rastgele değişkenidir.

$E[X] = n \cdot p$ $\qquad$ $E[X^k] = n \cdot p \cdot E[(Y+1)^{k-1}]$

$\rightarrow$ Y, $(n-1, p)$ parametreli binom rastgele değişkendir.

$Var(X) = n \cdot p \cdot (1-p)$

$$F(i) = \sum_{k=0}^{i} \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

$$P(X=k+1) = \frac{p}{1-p} \cdot \frac{n-k}{k+1} \cdot P(X=k)$$

3) Poisson Rastgele Değişkeni :

(n, p) parametreli bir binom rastgele değişkende, n değeri büyük ve p değeri küçük ise bunu $\lambda = np$ parametreli bir poisson rastgele değişken olarak da düşünebiliriz.

Poisson rastgele değişkene yaklaşım yapabileceğimiz durumları şöyle sıralayabiliriz;

- Bir sayfadaki hatalı karakter sayısı
- Bir ülkede 100 yaşının üzerindeki kişi sayısı
- Bir günde çevrilen yanlış telefon numarası sayısı
- İlk kullanıldığı günde arızalanan akıllı telefon sayısı

$X \Rightarrow$ Bir poisson rastgele değişkeni olmak üzere

$$P(X=i) = e^{-\lambda} \cdot \frac{\lambda^i}{i!} \qquad \frac{P(X=i+1)}{P(X=i)} = \frac{\lambda}{i+1}$$

$$F(a) = P(X \leq a) = \sum_{i \leq a} P(X=i)$$

$E[X] = \lambda$ $\qquad$ $Var[X] = \lambda$

4) Geometrik Rastgele Değişkeni:

Her birisinin başarılı olma olasılığı p olan bağımsız deneyler yapılsın. Deney başarılı olana kadar gerçekleştirilen tekrar sayısını gösteren rastgele değişkene geometrik rastgele değişken denir.

Örneğin X bir geometrik rastgele değişken ise $P(X=3)$ demek, ilk iki deneyin başarısız ve 3. deneyin başarılı olma olasılığı demektir.

$$P(X=i) = (1-p)^{i-1} \cdot p$$

$$\boxed{\text{Not: } \sum_{i=0}^{n} a^i = \frac{1-a^{n+1}}{1-a}}$$

5) Negatif Binom Rastgele Değişkeni:

Her birisinin başarılı olma olasılığı p olan bağımsız deneyler yapılsın. r. başarının i. denemede ortaya çıkma durumunu gösteren rastgele değişkene negatif binom rastgele değişken denir.

Örneğin X bir negatif binom rastgele değişken ise $r=3$ için $P(X=10)$ ifadesi, 3. başarılı deneyin 10. tekrarda olma olasılığı demektir.

$$P(X=i) = \binom{i-1}{r-1} \cdot p^r \cdot (1-p)^{i-r}$$

Sürekli Rastgele Değişkenler

Mümkün değerler kümesi bir aralık olan rastgele değişkene süreklidir denir.

Bu tür rastgele değişkenlere sürekli rastgele değişken denir.

Örneğin bir akıllı telefonun pilinin ömrü 3 ile 5 yıl arasındadır dersek burada akıllı telefon pilinin ömrünü gösteren rastgele değişken süreklidir ve değeri (3, 5) aralığında herhangi bir değerdir.

Sürekli rastgele değişkenlerin olasılık yoğunluk fonksiyonu vardır.

$X \Rightarrow$ Bir sürekli rastgele değişken

$f(x) \Rightarrow X$'in olasılık yoğunluk fonksiyonu

$$B = [a, b], \quad P(X \in B) = P(a \leq X \leq b) = \int_a^b f(x).dx$$

$a = b \Rightarrow \int_a^a f(x).dx = 0 \Rightarrow$ Bu yüzdendir ki sürekli rastgele değişkenin bir değere eşit olma olasılığı sıfırdır.

$$B = (-\infty, +\infty), \quad P(X \in B) = \int_{-\infty}^{+\infty} f(x).dx = 1$$

$F(a)$, X'in birikimli dağılım fonksiyonu olmak üzere,

$$F(a) = P(X < a) = P(X \leq a) = P(X \in (-\infty, a]) = \int_{-\infty}^{a} f(x).dx$$

$\Rightarrow \frac{d}{da} F(a) = f(a) \Rightarrow$ rastgele değişkenin a değeri çevresinde olma ihtimalidir.

Sürekli rastgele değişkeninin beklentisi şudur:

$$E[X] = \int_{-\infty}^{+\infty} x \cdot f(x) \cdot dx$$

$\Rightarrow$ $g(x)$, x'in bir fonksiyonu olmak üzere

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x) \cdot f(x) \cdot dx$$

$$\Rightarrow E[aX+b] = a \cdot E[X] + b$$

$$\Rightarrow E[X^n] = \int_{-\infty}^{+\infty} x^n \cdot f(x) \cdot dx$$

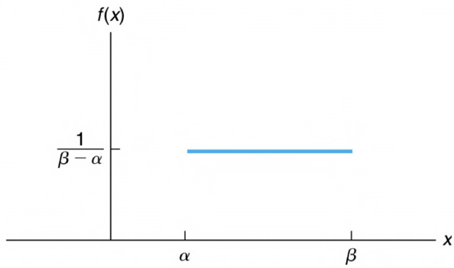
$$\Rightarrow E[X_1 + X_2 + X_3 + \dots + X_n] = E[X_1] + E[X_2] + E[X_3] + \dots + E[X_n]$$

$$\Rightarrow Var(X) = E[X^2] - (E[X])^2$$

1) Düzgün Rastgele Değişkeni :

Bir X rastgele değişkeninin olasılık yoğunluk fonksiyonu aşağıdaki gibi ise bu rastgele değişkene $[\alpha, \beta]$ aralığında düzgün dağılmıştır denir.

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \alpha \le x \le \beta \\ 0 & diger\ durumlarda \end{cases}$$



$$P(a < X < b) = \int_a^b f(x) \cdot dx = \frac{1}{\beta-\alpha} \int_a^b dx = \frac{b-a}{\beta-\alpha}$$

$$E[X] = \frac{\alpha+\beta}{2} \qquad E[X^2] = \frac{\beta^2+\alpha\beta+\alpha^2}{3} \qquad Var(X) = \frac{(\beta-\alpha)^2}{12}$$

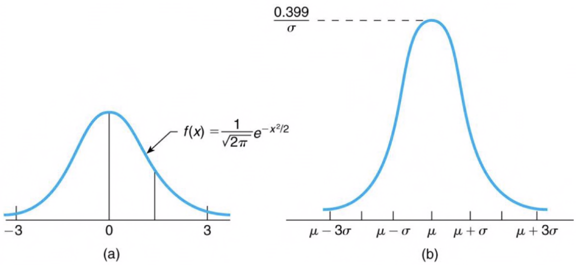
2) Normal (Gaussian) Rastgele Değişkeni :

Bir X rastgele değişkeninin olasılık yoğunluk fonksiyonu aşağıdaki gibi ise bu rastgele değişkene μ ve σ^2 parametreleri ile normal dağılmıştır denir ve $X \sim N(\mu, \sigma^2)$ şeklinde gösterilir.

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

$f(x)$ fonksiyonu, μ etrafında simetrik olan ve $x = \mu$ olduğunda $\frac{1}{\sigma\sqrt{2\pi}} \approx \frac{0{,}399}{\sigma}$ maksimum değerine ulaşan çan biçimli bir eğridir.

Olasılık yoğunluk fonksiyonu altında kalan alan olasılık bilgisini verir.



$$P(X > x_1) = \int_{x_1}^{+\infty} f(x)dx$$

$$P(X < x_2) = \int_{-\infty}^{x_2} f(x)dx$$

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx$$

Örnekler :

- Bir kişinin boyu
- Gaz içerisindeki herhangi bir molekülün herhangi bir doğrultudaki hızı
- Fiziksel bir niceliği ölçerken yapılan hatalar

$E[X] = \mu$

$Var(X) = \sigma^2$

3) Standart Normal Rastgele Değişkeni :

X, parametreleri μ ve σ^2 olan bir normal dağılıma sahip bir rastgele değişken olsun.

Bu durumda, $Z = \frac{X-\mu}{\sigma}$ rastgele değişkeni 0 ortalamalı ve 1 varyanslı bir normal dağılıma sahiptir.

Z rastgele değişkenine, standart veya birim normal dağılıma sahiptir denir.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-y^2}{2}} dy$$

$\Phi(x)$'in negatif olamayan x için alabileceği değerler, hazır tablolarda verilir.

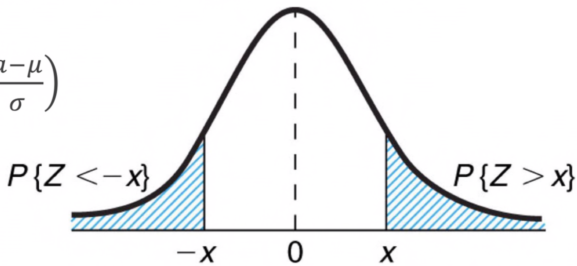
$$\Phi(-x) = 1 - \Phi(x)$$

Bu durumda, X rastgele değişkeninin birikimli dağılım fonksiyonu aşağıdaki gibi hesaplanabilir.

$$F_X(a) = P(X \leq a) = P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right)$$

$$F_X(a) = P\left(Z \leq \frac{a-\mu}{\sigma}\right)$$

$$F_X(a) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$



Bu aşamadan sonra $\Phi(x)$'in değeri tablodan bulunur.

Standart normal dağılım tablosu tablo başlıklarında verilen değerlerden küçük olma olasılıklarını verir.

$a < b$ için

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

$$= P\left(Z < \frac{b-\mu}{\sigma}\right) - P\left(Z < \frac{a-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Z = Başlık sütun değeri + Başlık Satır Değeri

4) Üstel Rastgele Değişkeni:

Bir $\lambda > 0$ sabiti için olasılık yoğunluk fonksiyonu aşağıdaki gibi olan rastgele değişkene λ parametreli üstel rastgele değişken denir.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Bu rastgele değişkenin birikimli dağılım fonksiyonu aşağıdaki gibidir.

$$F(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x}, \quad x \geq 0$$

Üstel rastgele değişken sıklıkla belli bir olayın gerçekleşmesine kadar geçen zamanın dağılımı olarak görünür.
Örneğin;

- Bir deprem olana kadar geçen süre
- Gelen ilk yanlış aramaya kadar geçen süre

$E[X] = \frac{1}{\lambda}$ $\qquad$ $Var(X) = \frac{1}{\lambda^2}$

Üstel rastgele değişken ve poisson rastgele değişken arasında ilişki mevcuttur.

Üstel rastgele değişken

- Belli bir olay meydana gelene kadar geçen süre

Poisson rastgele değişken

- Olay sayısı

Bir olay meydana gelene kadar geçen süre - Belli bir sürede gerçekleşen olay sayısı

$\Rightarrow$ $\lambda_1 \Rightarrow$ Poisson rastgele değişkenine ait

$\lambda_2 \Rightarrow$ Üstel rastgele değişkenine ait

$$\lambda_1 = \lambda_2$$

Kovaryans:

Kovaryans, iki rastgele değişken arasındaki ilişkinin bir ifade biçimidir.

X ve Y iki rastgele değişken olsun.

μ_x ve μ_y sırasıyla X ve Y rastgele değişkenlerinin ortalamaları olsun.

X ve Y rastgele değişkenlerinin kovaryansı aşağıdaki şekilde tanımlanır.

$$Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)]$$

Eşitliğin sağ tarafının açılımını yaparsak aşağıdaki ifadeyi elde ederiz.

$$Cov(X,Y) = E[XY] - E[X]E[Y]$$

Özellikler:

$$Cov(X,Y) = Cov(Y,X)$$

$$Cov(X,X) = Var(X)$$

$$Cov(aX,Y) = aCov(X,Y)$$

$$Cov(X+Z,Y) = Cov(X,Y) + Cov(Z,Y)$$

$$Cov\left(\sum_{i=1}^{n} X_i, Y\right) = \sum_{i=1}^{n} Cov(X_i, Y)$$

Bağımlı rastgele değişkenlerin varyanslarının toplamı, kovaryansları kullanılarak aşağıdaki şekilde ifade edilebilir.

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} Cov(X_i, X_j)$$

İki rastgele değişken için

$$Var(X+Y) = Var(X) + Var(Y) + Cov(X,Y) + Cov(Y,X)$$

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

X ve Y bağımsız rastgele değişkenlerse kovaryansları 0'a eşit olur.

$$Cov(X,Y) = 0$$

Neden?

X ve Y bağımsız rastgele değişkenlerse

$$E[XY] = E[X]E[Y]$$

⇒ Önemli !!

Dolayısıyla

$$Cov(X,Y) = E[XY] - E[X]E[Y]$$

$$Cov(X,Y) = E[XY] - E[XY] = 0$$

X ve Y bağımsız rastgele değişkenlerse kovaryansları 0'a eşittir. Ancak bunun tersi her zaman doğru değildir.

İki rastgele değişkenin kovaryansı 0'a eşit ise bu iki rastgele değişken bağımsızdır diyemeyiz.

Bağımlı rastgele değişkenlerin de kovaryansları da 0'a eşit olabilir.

$Cov(X,Y)$'nin değeri pozitif ise

- X arttıkça Y'nin artma eğiliminde olduğunun göstergesi

$Cov(X,Y)$'nin değeri negatif ise

- X arttıkça Y'nin azalma eğiliminde olduğunun göstergesi

Korelasyon :

X ve Y iki rastgele değişken olsun.

Bu iki rastgele değişken arasındaki ilişkinin gücü korelasyon ile ifade edilebilir.

Korelasyon aşağıdaki şekilde tanımlanır.

$$Corr(X,Y) = \rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Korelasyon her zaman -1 ile $+1$ arasında değer alır.

$$-1 \leq Corr(X,Y) \leq 1$$

$Corr(X,Y) = 1 \Rightarrow Y = aX + b \text{ } ve \text{ } a > 0$

$Corr(X,Y) = -1 \Rightarrow Y = aX + b \text{ } ve \text{ } a < 0$

$Corr(X,Y)$'nin değeri pozitif ise

- X arttıkça Y'nin artma eğiliminde olduğunun göstergesi

$Corr(X,Y)$'nin değeri negatif ise

- X arttıkça Y'nin azalma eğiliminde olduğunun göstergesi

$Corr(X,Y) = 0$ ise X ve Y rastgele değişkenleri **birbiri ile ilişkisizdir** denir.

Korelasyon, iki rastgele değişkenin arasındaki doğrusal ilişkiye dair bilgi verir.

- $Corr(X,Y) = 1$ ise veya $Corr(X,Y) = -1$ ise X ve Y rastgele değişkenleri arasında tam doğrusal ilişki var demektir.
- $Corr(X,Y)$, 1 veya -1 civarında ise örneğin $0{,}8$ veya $-0{,}7$ gibi değer ise X ve Y rastgele değişkenleri arasında göreli olarak güçlü bir doğrusal ilişki var demektir.
- $Corr(X,Y)$, 0 civarında ise örneğin $0{,}3$ gibi değer ise X ve Y rastgele değişkenleri arasında göreli olarak zayıf bir doğrusal ilişki var demektir.

$Corr(X,Y) = 0{,}5$ ise bir şey diyemeyiz.

Tablolar

Sıklık (frekans) tablosu

Grafikler

Çizgi grafiği
Sütun grafiği
Sıklık poligonu
Göreli sıklık grafiği (pasta dilimi)
Histogram

Kök - Yaprak gösterimi:

12, 26, 37, 38, 40, 42, 43, 44, 67, 80 $\Rightarrow$

1	2
2	6
3	7; 8
4	0; 2; 3; 4
6	7
8	0

Veri kümesinin ortalaması:

Bütün verileri topla, toplam adede böl.

Veri kümesinin ortancası:

Veri kümesi artan sıra ile dizildiğinde tam ortadaki değerdir.

$n \Rightarrow$ Toplam veri adedi

$n =$ tek $\Rightarrow$ Ortanca $= \frac{n+1}{2}$. değerdir.

$n =$ çift $\Rightarrow$ Ortanca $= \frac{n}{2}$. ve $\frac{n}{2}+1$. değerlerin toplamının ortalamasıdır.

Veri kümesinin tepedeğeri:

Frekansı en fazla olan değer(ler)dir

Veri kümesinin varyansı:

$$s^2 = \frac{\sum_{i=0}^{n} (x_i - \bar{x})^2}{n-1} = \frac{\left(\sum_{i=0}^{n} x_i^2\right) - (n \cdot \bar{x}^2)}{n-1}$$

$s \rightarrow$ standart sapma

Chebyshev Eşitsizliği :

Ortalaması ve standart sapması verilmiş bir veri kümesi için Chebyshev eşitsizliği şunu söyler:

Herhangi bir $k \geq 1$ değeri için,

verinin en az $\%100\left(1-\frac{1}{k^2}\right)$'si

$\bar{x}-ks$ ile $\bar{x}+ks$ arasındadır.

Örnek Korelasyon Katsayısı :

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{(n-1)s_x s_y}$$

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$r>0$ olduğunda örnek veri çiftleri pozitif ilişkilidir denir.

$r<0$ olduğunda örnek veri çiftleri negatif ilişkilidir denir.

r, $x\,i$ ve $y\,i$ (i = 1 , … , n) veri çiftlerinin örnek korelasyon katsayısı ise a ve c 'nin her ikisi de pozitif veya her ikisi de negatif olmak koşulu ile

$axi + b$ ve $c\,yi + d$ (i = 1 , … , n)

veri çiftlerinin örnek korelasyon katsayısı da r'dir.

$X_1, X_2, \ldots, X_n$ bir yığından alınan n adet örneği göstersin.

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$E[\bar{X}] = \mu$ $\quad$ $Var[\bar{X}] = \frac{\sigma^2}{n}$ $\Rightarrow$ X'in varyansı

X'in ortalaması

Merkezi Limit Teoremi;

Beklentisi μ ve varyansı σ^2 olan $X_1, X_2 \ldots X_n$ olan rastgele değişkenlerinin toplamı **Normal Dağılım'a** yakınsanabilir.

$$X = X_1 + X_2 + \cdots + X_n \Rightarrow X \sim N(n.\mu, n.\sigma^2)$$

$$E[X] = n.\mu \qquad Var(X) = n.\sigma^2$$

$$\Rightarrow P(X < a) \cong P\left(z < \frac{a - n.\mu}{\sigma.\sqrt{n}}\right)$$

Merkezi Limit Teoremi ve Binom rastgele değişkeni

X_i = Bir Bernoulli rastgele değişkeni

$X = X_1 + X_2 + \cdots + X_n \Rightarrow X, (n.p)$ parametreli binom rastgele değişkeni

$$P(X < a) = P\left(z < \frac{a - np}{\sqrt{n.p.1-p}}\right)$$

Not: Süreklilik düzeltmesini unutma.

Not: Sınırlı yığında örnekleme konusunda da binom rastgele değişkeni kullanılıyor.

1) Varyans bilindiğinde hipotez testi:

$\alpha \Rightarrow$ önem seviyesi $\Rightarrow$ Soruda veriliyor.

Varyansın Bilindiği Durum $\Phi(z_{\alpha/2}) = 1-\frac{\alpha}{2}$ $Z = \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0)$ $\Phi(z_\alpha) = 1-\alpha$		
Çift Yanlı Hipotez Testi H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$	Tek Yanlı Hipotez Testi H_0: $\mu = \mu_0$ veya $(\mu \leq \mu_0)$ H_1: $\mu > \mu_0$	Tek Yanlı Hipotez Testi H_0: $\mu = \mu_0$ veya $(\mu \geq \mu_0)$ H_1: $\mu < \mu_0$
H_0 kabul: $-z_{\alpha/2} \leq Z \leq z_{\alpha/2}$	H_0 kabul: $Z \leq z_\alpha$	H_0 kabul: $Z \geq -z_\alpha$
H_0 ret: $Z > z_{\alpha/2}$ veya $Z < -z_{\alpha/2}$	H_0 ret: $Z > z_\alpha$	H_0 ret: $Z < -z_\alpha$

- İlk önce H_0 ve H_1'i yaz.
- $z_{\alpha/2}$ veya z_α'yı hesapla.
- Z'yi hesapla.
- Son olarak doğruluğunu test et.

Not: Burada kabul ya da ret ettiğin şeyin H_0 olduğunu unutma!

Hata tipi II:

$$\beta(\mu) = P\left(\frac{\mu_0-\mu}{\sigma/\sqrt{n}} - z_{\alpha/2} \leq Z \leq \frac{\mu_0-\mu}{\sigma/\sqrt{n}} + z_{\alpha/2}\right)$$

$\beta(\mu)$ fonksiyonu, **işletim karakteristiği (İK) eğrisi** olarak adlandırılır ve gerçek beklenti μ iken H_0'ın kabul edilme ihtimalini verir.

$1 - \beta(\mu)$ ise **güç fonksiyonu** olarak adlandırılır. Verilen bir μ değeri için, testin gücü, μ gerçek değeri iken reddedilme ihtimalidir.

2) Varyans bilinmediğinde hipotez testi :

$\alpha \Rightarrow$ önem seviyesi $\Rightarrow$ Soruda veriliyor.

Varyansın Bilinmediği Durum - t Testi		
$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$		$T = \frac{\sqrt{n}}{S}(\bar{X} - \mu_0)$
Çift Yanlı t Testi $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	Tek Yanlı t Testi $H_0: \mu = \mu_0$ veya $(\mu \leq \mu_0)$ $H_1: \mu > \mu_0$	Tek Yanlı t Testi $H_0: \mu = \mu_0$ veya $(\mu \geq \mu_0)$ $H_1: \mu < \mu_0$
H_0 kabul: $-t_{\alpha/2,n-1} \leq T \leq t_{\alpha/2,n-1}$	H_0 kabul: $T \leq t_{\alpha,n-1}$	H_0 kabul: $T \geq -t_{\alpha,n-1}$
H_0 ret: $T > t_{\alpha/2,n-1}$ veya $T < -t_{\alpha/2,n-1}$	H_0 ret: $T > t_{\alpha,n-1}$	H_0 ret: $T < -t_{\alpha,n-1}$

- İlk önce H_0 ve H_1'i yaz.
- $t_{\alpha/2,n-1}$'i veya $t_{\alpha,n-1}$'i 't tablosu'na bakarak hesapla.
- T değerini hesapla.
- Son olarak doğruluğunu test et.

Not: Burada kabul ya da ret ettiğin şeyin H_0 olduğunu unutma.

3) İki normal yığının ortalamalarının testi :

$\alpha \Rightarrow$ önem seviyesi $\Rightarrow$ Soruda veriliyor.

İki Normal Yığının Ortalamalarının Testi

$\Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2}$

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

$\Phi(z_\alpha) = 1 - \alpha$

Çift Yanlı Test	Tek Yanlı Test
H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$	H_0: $\mu = \mu_0$ veya $(\mu \leq \mu_0)$ H_1: $\mu > \mu_0$
H_0 kabul: $-z_{\alpha/2} \leq Z \leq z_{\alpha/2}$	H_0 kabul: $Z \leq z_\alpha$
H_0 ret: $Z > z_{\alpha/2}$ veya $Z < -z_{\alpha/2}$	H_0 ret: $Z > z_\alpha$

- İlk önce H_0 ve H_1'i yaz.
- $z_{\alpha/2}$ veya z_α'yı hesapla.
- Z'yi hesapla.
- Son olarak doğruluğunu test et.

Not: Burada kabul ya da ret ettiğin şeyin H_0 olduğunu unutma.

Hipotez testleri

Basit ve bileşik hipotez ?

Kritik bölge ?

Hata tipleri ?

Serbestlik derecesi $t_{\alpha,(n-1)}$