Kovaryans ve Korelasyon

IST 108 Olasılık ve İstatistik
Bahar 2016

Yrd. Doç. Dr. Ferhat Dikbıyık

Kovaryans

- Nasıl beklenti ve varyans bize tek bir rastgele değişken hakkında bir bilgi veriyorsa, aynı şekilde kovaryans da iki rastgele değişkenin arasındaki ilişkiyi tanımlar.
- X ve Y arasındaki kovaryans Cov(X, Y) ile gösterilir ve şöyle tanımlanır.

$$
\begin{aligned}
Cov(X, Y) &= E[(X - E[X])(Y - E[Y])] \\
&= E[XY - E[X]Y - XE[Y] + E[X]E[Y]] \\
&= E[XY] - E[X]E[Y]
\end{aligned}
$$

Kovaryans

- Eğer X ve Y rastgele değişkenleri birbirinden bağımsızsa herhangi bir *g* ve *h* fonksiyonu için;

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

- Bu durumda birbirinden bağımsız iki rastgele değişken arasındaki kovaryans;

$$\begin{aligned} Cov(X,Y) &= E[XY] - E[X]E[Y] \\ &= E[X]E[Y] - E[X]E[Y] \\ &= 0 \end{aligned}$$

- Ama tersi her zaman doğru değildir. Yani kovaryans 0 ise bu rastgele değişkenler mutlaka birbirinden bağımsız demek değildir.

Kovaryans

- Örneğin, X rastgele değişeni 0, 1, ve -1 değerlerini eşit olasılıkla alsın ve Y rastgele değişkeni X sıfıra eşit olduğunda 1, aksi halde 0 olsun.
- Bu durumda Y'nin de tanımından anlaşılacağı üzere X ve Y birbirinden bağımsız değillerdir.
- Ama kovaryanslarını hesapladığımızda sıfır çıkacaktır. Çünkü XY her zaman sıfıra eşittir, yani E[XY]=0 ve E[X] de sıfıra eşittir. Bu durumda birbirlerinden bağımsız olmadıkları halde kovaryansları sıfıra eşit olur.

Kovaryans özellikleri

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(X,X) = Var(X)$
- $Cov(aX,Y) = aCov(X,Y)$
- $Cov(\sum_{i=1}^{n} X_i, \sum_{j=1}^{m} Y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} Cov(X_i, Y_j)$

Rastgele değişkenlerin toplamlarının varyansı

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

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j)$$

$$= \sum_{i=1}^{n} Var(X_i) + \underbrace{\sum_{i \neq j}\sum Cov(X_i, X_j)}_{=0}$$

Eğer ikili olarak birbirlerinden bağımsızlar ise

$$= \sum_{i=1}^{n} Var(X_i)$$

Bimomial Rastgele Değişkenin Varyansı

- *n* ve *p* parametrelerine sahip bir Binomial Rastgele Değişken X'in varyansını hesaplayın.
- $X = X_1 + X_2 + \ldots X_n$
- $X_i = 1$, deney başarılı olduğunda ve $X_i = 0$ deney başarısız olduğunda.
- Bütün X_i değerleri birbirinden bağımsız olduğuna göre;
- $Var(X) = Var(X_1) + Var(X_2) + \ldots Var(X_n)$
- $Var(X_i) = E[Xi^2] – (E[X_i])^2 = E[X_i] – (E[X_i])^2 = p - p^2$
- $Var(X) = np(1-p)$

Korelasyon

- İki rastgele değişken arasındaki korelasyon ρ(X, Y) ile gösterilir ve şöyle tanımlanmıştır.

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

- İki rastgele değişken arasındaki korelasyon -1 ile 1 arasındadır.

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

Korelasyon bize ne anlatır?

- $\rho(X, Y) = 1$, $Y = aX+b$ olduğunu ve burada $b = \sigma_y / \sigma_x > 0$ olduğunu gösterir.
- $\rho(X, Y) = -1$ ise, $Y = aX+b$ olduğunu ve burada $b = -\sigma_y / \sigma_x < 0$ olduğunu gösterir.
- Korelasyon katsayısı iki rastgele değişken arasında lineerliğin bir ölçüsüdür.
  - $\rho(X, Y)$ değeri +1 veya -1 civarında ise bu iki rastgele değişken arasında yüksek derecede bir lineerlikten söz edilebilirken,
  - $\rho(X, Y)$ değeri 0 civarında ise bu iki rastgele değişken arasındaki ilişkinin lineerlikten yoksun olduğunu söylemek mümkündür.
- Pozitif bir $\rho(X, Y)$ değeri X artarken Y'nin de artmaya meyilli olduğunu ve negatif bir $\rho(X, Y)$ değeri ise X artarken Y'nin azalmaya meyilli olduğunu gösterir.
- Eğer $\rho(X, Y) = 0$ ise, X ve Y için «birbiri ile ilişkisizdir» denir.

Örnek 1

- X bir rastgele değişkendir.
- $Y = aX + b$ ise, X ve Y arasındaki kovaryans ve korelasyonu hesaplayınız.

Örnek 1

- X bir rastgele değişkendir.
- $Y = aX + b$ ise, X ve Y arasındaki kovaryans ve korelasyonu hesaplayınız.
  - $$\begin{aligned} Cov(X,Y) &= E[XY] - E[X]E[Y] \\ &= E[X(aX+b)] - E[X](aE[X]+b) \\ &= E[aX^2 + bX] - aE[X]^2 - bE[X] \\ &= aE[X^2] + bE[X] - aE[X]^2 - bE[X] \\ &= aE[X^2] - aE[X]^2 = aVar(X) \end{aligned}$$

Örnek 1

- X bir rastgele değişkendir.
- $Y = aX + b$ ise, X ve Y arasındaki kovaryans ve korelasyonu hesaplayınız.
  - $Cov(X,Y) = aVar(X)$
  - $\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$

$$= \frac{aVar(X)}{\sqrt{Var(X)a^2Var(X)}}$$

$$= \frac{a}{|a|} = \begin{cases} 1 & a > 0 \\ -1 & a < 0 \end{cases}$$

Örnek 2

- X, rastgele değişkeni (-1, 1) arasında düzgün (uniform) olarak dağıtılmıştır.
- $Y = X^n$ ise, X ve Y arasındaki kovaryans ve korelasyonu hesaplayınız.

Örnek 2

- X, rastgele değişkeni (-1, 1) arasında düzgün (uniform) olarak dağıtılmıştır.
- $Y = X^n$ ise, X ve Y arasındaki kovaryans ve korelasyonu hesaplayınız.

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

$$= E[XX^n] - \underbrace{E[X]}_{=0} E[X^n]$$

$$= E[X^{n+1}] = \int_{-1}^{1} \frac{x^{n+1}}{2} dx = \left.\frac{x^{n+2}}{2(n+2)}\right|_{-1}^{1} = \frac{(1)^{n+2} - (-1)^{n+2}}{2(n+2)}$$

$$= \begin{cases} \dfrac{1}{n+2} & n \text{ tek ise} \\ 0 & n \text{ çift ise} \end{cases}$$

Örnek 2

- X, rastgele değişkeni (-1, 1) arasında düzgün (uniform) olarak dağıtılmıştır.
- $Y = X^n$ ise, X ve Y arasındaki kovaryans ve korelasyonu hesaplayınız.
  - $Cov(X,Y) = \begin{cases} \frac{1}{n+2} & n \text{ tek ise} \\ 0 & n \text{ çift ise} \end{cases}$
  - n çift ise $\rho = 0$ olur. n tek ise
    - $\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{\frac{1}{n+2}}{\sqrt{\left(\frac{1}{3}\right)\left(\frac{1}{2n+1}\right)}} = \frac{\sqrt{3(2n+1)}}{n+2}$