VERİ BİLİMİ İÇİN TEMEL İSTATİSTİK hafta-8

CEMİLE YILDIZÇAKAR

05.02.2021

Birikimli Dağılım Fonksiyonu (Cumulative Distribution Function cdf)

- X rassal değişkeninin *x* değerine eşit ya da daha küçük bir değer alma olasılığını veren fonksiyondur.
- Tüm rassal değişkenler için elde edilebilir.
- F(X) şeklinde gösterilir.

$$F(x) = P(X \leq x) \;\; for \;\; -\infty < x < \infty$$

Birikimli dağılım fonksiyonunun sağlaması gereken koşullar
Requirements for Cumulative Distribution Function

- $0 \leq F(x) \leq 1$ for all x
- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$
- $F(x)$ is a *nondecreasing* function of x

Teorem:

- X rassal değişkeni için $x_1 < x_2 < x_3 < \cdots < x_n$ ise;

$$\boxed{f(x_1) = F(x_1)}$$

$$\boxed{f(x_i) = F(x_i) - F(x_{i-1}) \text{ for } i = 2, 3, \dots, n}$$

Örnek

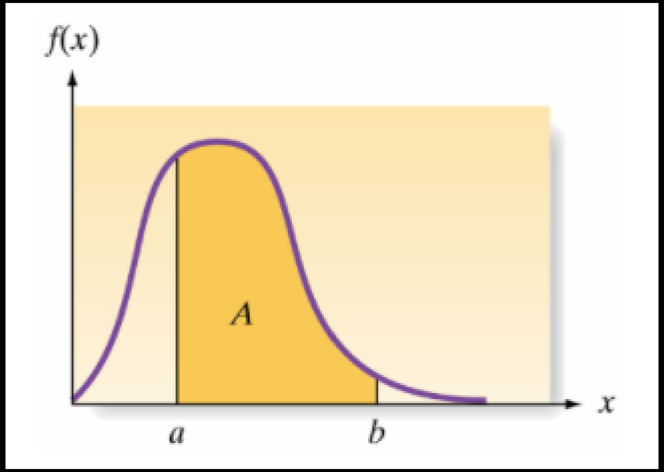
- Bir futbol takımının yaptığı maçlarda attığı gol sayısının fonksiyonu aşağıdaki gibi elde edilmiştir.

X	0	1	2	3	4	5
P(X)	0.1	0.2	0.4	0.15	0.1	0.05

- Bu fonksiyonun olasılık dağılımı olup olamayacağını inceleyiniz.
- Atılan gol sayısının birikimli dağılım fonksiyonunu oluşturunuz.

Sürekli Rassal Değişkenlerin Olasılık Dağılımları

- Sürekli rassal değişkenlerin olasılık dağılımları bir eğri oluştururlar. X'in bir fonksiyonu olan bu eğriye olasılık yoğunluk fonksiyonu (probability density function -pdf) denir.



- f(x) fonksiyonu hiçbir değerin olasılığını göstermez. Bu fonksiyonun integrali alınırsa olasılık elde edilebilir.
- Herhangi bir f(x) fonksiyonunun grafiğinin altında a ve b noktaları arasında kalan alan; P(a < X < b) olasılığını verir.

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\, dx$$

Sürekli olasılık dağılımının sağlaması gereken koşullar (Requirements for continuous probability distributions)

- f(x) > 0 for all possible intervals of x
- $\int_{-}^{+} f(x)dx = 1$

Birikimli Dağılım Fonksiyonu
Cumulative Distribution Function (cdf)

$$F(y_0) = P(Y \le y_0) = \int_{-\infty}^{y_0} f(y)dy$$

Birikimli Dağılım Fonksiyonunun Özellikleri

1. The CDF is non-negative: $F(x) \geq 0$. Probabilities are never negative.
2. The CDF goes to zero on the far left: $\lim_{x \to -\infty} F(x) = 0$. X is never less than $-\infty$.
3. The CDF goes to one on the far right: $\lim_{x \to \infty} F(x) = 1$. X is never more than ∞.
4. The CDF is non-decreasing: $F(b) \geq F(a)$ if $b \geq a$. If $b \geq a$, then the event $X \leq a$ is a sub-set of the event $X \leq b$, and sub-sets never have higher probabilities. (This was a problem in HW2.)

Teorem

- X sürekli bir rassal değişken, a ve b reel sayılar ise $(a \leq b)$;

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

Teorem

- X sürekli bir rassal değişken, a ve b reel sayılar ise ($a \leq b$);

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

Beklenen Değer (Expected Value)

- Bir rassal değişkenin beklenen değeri veya ortalaması; ilgili deney çok fazla sayıda tekrarlandığında gözlenmesi beklenen değerdir.
- E(X) şeklinde ifade edilir.

X_i 'nin alabileceği tüm mümkün değerleri $X_1, X_2, \ldots\ldots, X_k$ ve olasılıkları sırasıyla $p_1, p_2, \ldots, p_k$ olan bir tesadüfi değişken olsun. Bu durumda X_i'nin beklenen değeri:

$E(X_i) = \sum_i^k X_i p_i$ şeklinde tanımlanır.

n örnek hacmi olmak üzere $\bar{x} = \frac{1}{n}\sum_i^k x_i$ şeklinde ifade edilen örnek ortalamasının beklenen değerini bulmak istediğimizde $\bar{x}$'nın tüm mümkün değerlerinin bulunması gerekir. $\bar{x}$'nın mümkün bütün değerlerini bulabilmek için N hacimli bir populasyondan seçilebilecek n hacimli tüm mümkün örneklerin bulunması gerekir. Bu durumda $\bar{x}$'in beklenen değeri:

$E(\bar{x}) = \sum_i^k \bar{x}_i p_i$ şeklinde tanımlanır.

Sürekli rassal değişkenler için :

$$E(X) = \int_{\text{all } x} x\, f(x)\, dx$$

Beklenen değerin özellikleri

- $E(c) = c$
- $E(cX)=cE(X)$
- $E(cX)=cE(X)$
- $E(X+Y)= E(X) + E(Y)$

Örnek

- Hilesiz bir zar atılsın.
  - Tek sayı gelirse gelen sayı kadar para kazanılıyor.
  - Çift sayı gelirse, 4 lira kaybediliyor.
- Bu oyunu oynarsanız ne kadar kazanç beklersiniz?

Let X = your earnings

X=1	P(X=1) = P({1}) =1/6
X=3	P(X=1) = P({3}) =1/6
X=5	P(X=1) = P({5}) =1/6
X=-4	P(X=1) = P({2,4,6}) =3/6

E(X) = 1*1/6 + 3*1/6 + 5*1/6 + (-4)*1/2 = 1/6 + 3/6 +5/6 – 2= -1/2

Örnek

For the variable X with p.d.f.

$$f(x) = \begin{cases} \frac{1}{2}x, & 0 \le x \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

find $E(X)$

$$E(X) = \int_0^2 \frac{1}{2} x.x \, dx = \left[\frac{1}{6}x^3\right]_0^2 = \frac{8}{6} = \frac{4}{3}$$

Örnek

- Aşağıda pdf'i verilen X rassal değişkeninin ortalamasını hesaplayınız.

$$f(x) = \begin{cases} 2(1-x) & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$

Çözüm

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_0^1 x\left[2(1-x)\right] dx \\ &= 2\int_0^1 \left(x - x^2\right) dx \\ &= 2\left(\frac{x^2}{2} - \frac{x^3}{3}\right)\Big|_0^1 \\ &= 1/3 \end{aligned}$$

Varyans (Variance)

$$\sigma_X^2 = E[(X - \mu_X)^2]$$

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

Kesikli rassal değişkenler için:

$$Var(X) = \sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 P(x_i)$$

Sürekli rassal değişkenler için :

$$Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x_i - \mu)^2 f(x_i)\, dx$$

X ve Y herhangi rassal değişkenler, c sabit bir sayı ise;

- $\mathrm{Var}(c) = 0$
- $\mathrm{Var}(c+X) = \mathrm{Var}(X)$
- $\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X)$
- $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ ⇨ ***X ve Y bağımsız ise!!!!***

Varyansın Özellikleri

$\forall\, a, b, c \in \mathbb{R}$ için,

1) $Var(a) = 0$

2) $Var(bX + c) = b^2 Var(X)$

3) X ve Y tesadüfi değişkenler ise,

$$Var(aX \pm bY) = a^2 Var(X) \pm 2abCov(X,Y) + b^2 Var(Y)$$

KOVARYANS

- Kovaryans iki değişken arasındaki doğrusal ilişkinin değişkenliğini ölçen bir kavramdır. Betimsel istatistiktir. Yani var olan bir şeyi bize söyler. Ortada tahmin yoktur. Sonucun pozitif olması artan bir doğrusal ilişkiyi, negatif olması azalan bir doğrusal ilişkiyi ve sıfır civarında olması ilişkinin olmadığını gösterir.

Teşekkür Ederim

LinkedIn
https://www.linkedin.com/in/cemile-yıldızçakar-34782248/

Email
yildizcakar.cemile@gmail.com

Cemile YILDIZÇAKAR