IST2084/ IST104.1/ IST104.2 Biyoistatistik (Biyoloji Bölümü)

Rastgele Değişkenlerin Beklenen Değeri ve Varyansı, Kesikli Olasılık Dağılımları

Fatih Kızılaslan

Marmara Üniversitesi

2019-2020 Bahar 10. Hafta

Beklenen Değerin Özellikleri

X rastgele değişkeninin beklenen değeri $E(X)$ rastgele değişkenin ortalama değerdir. Beklenen değer için genellikler kitle (popülasyon) ortalaması μ kullanırız, yani $\mu = E(X)$.

Özellikler:

- Sabitin beklenen değeri kendisine eşittir. $a \in \mathbb{R}$ için $E(a) = a$ dir.
- $a \in \mathbb{R}$ olmak üzere $E(aX) = aE(X)$ dir.
- $a, b \in \mathbb{R}$ olmak üzere $E(aX \pm b) = aE(X) \pm b$ dir.
- $g(X)$, X rastgele değişkeninin bir fonksiyonu olmak üzere

$$E(g(X)) = \begin{cases} \sum_x g(x)\ p(x) & , X \text{ kesikli r.d.} \\ \int_{-\infty}^{\infty} g(x)\ f(x)dx & , X \text{ sürekli r.d.} \end{cases}$$

biçiminde hesaplanır.

Örnek 1: Bir torbada 1'den 4'e kadar numaralı 4 top bulunmaktadır. Bu torbadan 2 top yerine koymama koşulu ile çekiliyor. X rastgele değişkeni "çekilen toplardaki numaraların toplamı" olarak tanımlansın. X rastgele değişkeninin olasılık fonksiyonunu ve beklenen değerini bulunuz.

Çözüm: Çekilen 2 topun üzerindeki sayıların toplamı $\{1, 2, 3, 4\}$ kullanarak $\{3, 4, 5, 6, 7\}$ olabilir. Böylece, örnek uzayımız $S = \{(1, 2), (2, 1), (1, 3), (3, 1), (2, 3), (3, 2), (4, 1), (1, 4), (2, 4), (4, 2), (3, 4), (4, 3)\}$ biçiminde 12 elemandan oluşur. X rastgele değişkeninin değer kümesi $\{3, 4, 5, 6, 7\}$ olmak üzere olasılık fonksiyonu

x	3	4	5	6	7
$p(x)$	$2/12$	$2/12$	$4/12$	$2/12$	$2/12$

biçiminde bulunur. Beklenen değer

$$\begin{aligned} E(X) &= \sum_{x=3}^{7} x\ p(x) = 3(2/12) + 4(2/12) + 5(4/12) + 6(2/12) + 7(2/12) \\ &= 60/12 = 5 \text{ bulunur.} \end{aligned}$$

Varyans ve Standat Sapma

Bir rastgele değişkeni tanımlayan iki değerden biri ortalama (beklenen değer) diğeri varyanstır.

Bir X rastgele değişkeninin varyansı $Var(X)$ veya σ^2 ile gösterilir ve

$$\begin{aligned} Var(X) &= E\left((X-\mu)^2\right) \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

olarak tanımlanır. Burada $\mu = E(X)$ ortalamadır.
Varyansı genellikle σ^2 ile gösteririz. Varyansın pozitif karekökü σ'ya **standart sapma** denir.

Özellikler:

- Sabitin varyansı 0 dır. Yani, $a \in \mathbb{R}$ için $Var(a) = 0$ dır.
- $a \in \mathbb{R}$ için $Var(aX) = a^2 Var(X)$ dir.
- $a, b \in \mathbb{R}$ olmak üzere $Var(aX \pm b) = a^2 Var(X)$ dir.

Örnek 2: Bir paranın 3 kez atılması deneyinde X rastgele değişkeni gelen turaların sayısı olmak üzere olasılık fonksiyonu

x	0	1	2	3
$p(x)$	$1/8$	$3/8$	$3/8$	$1/8$

biçiminde ve $E(X) = 3/2$ olarak bulmuştuk. $Var(X)$ bulalım.

Çözüm: $Var(X) = E(X^2) - (E(X))^2$ olduğundan öncelikle $E(X^2)$ hesaplayalım.

$$\begin{aligned} E(X^2) &= \sum_{x=0}^{3} x^2 \, p(x) = 0^2(1/8) + 1^2(3/8) + 2^2(3/8) + 3^2(1/8) \\ &= 24/8 = 3 \end{aligned}$$

bulunur. Böylece,

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 \\ &= 3 - (3/2)^2 = 3/4 \end{aligned}$$

bulunur.

Örnek 3: Örnek 1'de tanımlanan X rastgele değişkeninin varyansı ve standart sapmasını bulunuz.

Çözüm: Olasılık fonksiyonu

x	3	4	5	6	7
$p(x)$	$2/12$	$2/12$	$4/12$	$2/12$	$2/12$

biçiminde ve $E(X) = 5$ bulmuştuk. $E(X^2)$ hesaplayalım.

$$\begin{aligned} E(X^2) &= \sum_{x=3}^{7} x^2 p(x) = 3^2 \tfrac{2}{12} + 4^2 \tfrac{2}{12} + 5^2 \tfrac{4}{12} + 6^2 \tfrac{2}{12} + 7^2 \tfrac{2}{12} \\ &= 320/12 = 80/3 \end{aligned}$$

bulunur. Böylece,

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 \\ &= (80/3) - 5^2 = 5/3 \end{aligned}$$

ve standart sapma $\sigma = \sqrt{5/3} = 1.29$ bulunur.

Örnek 4: Bir yazı tura oyununda bir oyuncu bir parayı 2 kez attığında 1 yazı gelirse 1TL, 2 yazı gelirse 2TL ve hiç yazı gelmezse 5 TL kazanacaktır. Bu oyuncunun bu oyunun sonundaki kazancının ortalamasını ve standart sapmasını bulunuz.

Çözüm: X rastgele değişkenini bu oyuncunun 2 para atışı sonucunda kazandığı para miktarı olarak tanımlayalım. Böylece, X r.d.nin alabilecek olduğu değerler $\{1, 2, 5\}$ dir. Bu değerler için olasılıkları bulalım. 2 para atışı için örnek uzay $S = \{YY, YT, TY, TT\}$ biçiminde olur. Böylece, olasılık fonksiyonu

x	1TL (1 Y)	2TL (2Y)	5TL (0Y)
$p(x)$	2/4	1/4	1/4

biçiminde bulunur. Böylece, ortalama kazanç

$$\mu = E(X) = \sum_{x=1,2,5} x\ p(x) = 1(2/4) + 2(1/4) + 5(1/4) = 9/4$$

bulunur.

Ayrıca,

$$E(X^2) = \sum_{x=1,2,5} x^2 \ p(x) = 1^2(2/4) + 2^2(1/4) + 5^2(1/4) = 31/4$$

bulunur. Böylece, $Var(X) = (31/4) - (9/4)^2 = 43/16$ ve $\sigma = \sqrt{43/16} = 1.64$ bulunur.

Örnek 5: X r.d.nin olasılık fonksiyonu

$$p(x) = \begin{cases} k \ x & ,x = 1,2,3,4 \\ 0 & , \text{diğer durumlar} \end{cases}$$

olarak verilmiş olsun. **a)** k sabitini bulunuz. **b)** $E(X) =?$, **c)** $Var(X) =?$, **d)** $P(1 < X \leq 3) =?$, **e)** $P(X \leq 3) =?$

Çözüm: **a)** $\sum_{x=1}^{4} p(x) = 1$ olması gerektiğinden

$$1 = \sum_{x=1}^{4} p(x) = \sum_{x=1}^{4} k \ x = k(1+2+3+4) = 10k$$

bulunur. Böylece, $k = 1/10$ bulunur.

Böylece,

$$p(x) = \begin{cases} \frac{x}{10} & ,x = 1,2,3,4 \\ 0 & , \text{diğer durumlar} \end{cases}$$

olur.

b)

$E(X) = \sum_{x=1}^{4} xp(x) = 1(1/10)+2(2/10)+3(3/10)+4(4/10) = 30/10 = 3$ bulunur.

c) Bulunuz. Cevap: $Var(X) = 1$

d) Bulunuz. Cevap: $P(1 < X \leq 3) = 5/10$

e) Bulunuz. Cevap: $P(X \leq 3) = 6/10$

Olasılık Dağılımları

- Rastgele olayları belirleyen rastgele değişkenin olasılık dağılımının bilinmesi bu olaylar ile ilgilenen araştırmacılar için önemlidir. Bu nedenle olasılık teorisinde bazı özel olay veya durumları ifade eden olasılık dağılımları oluşturulmuştur.
- Bu olasılık dağılımlarını kesikli ve sürekli olmak üzere 2 gruba ayırırız.
- İlk olarak bazı kesikli olasılık dağılımlarını inceleyeceğiz. Bu dağılımlar: **Bernoulli, Binom, Geometrik** ve **Poisson** dağılımlarıdır.

Bernoulli Dağılımı: Bir rastgele deneyin iki sonucu (iyi-kötü, başarılı-başarısız, olumlu-olumsuz, evet-hayır, çalışıyor-bozuk gibi) olması durumunda kullanılır. Bu tip deneyler **Bernoulli deneyi** olarak adlandırılır.

Örneğin, bir elektronik cihazın çalşıyor olması veya bozuk olması durumu, bir para atışını sonucu, bir dersten başarılı olma veya olmama durumu gibi olaylar.

- Bernoulli deneyinde ilgilenilien sonucu "**başarı**" diğer sonucu "**başarısızlık**" olarak tanımlayalım. Ayrıca, matematiksel olarak başarılı olmayı 1 ile başarısızlığı 0 ile gösterelim.
- Bernoulli deneyinde başarı sonucu için 1 değerini, başarısızlık sonucu için 0 değerini alan X rastgele değişkenine **Bernoulli rastgele değişkeni** denir.
- Başarılı olma olasılığı $0 \leq p \leq 1$ olmak üzere p ile gösterilirse bu X rastgele değişkeninin olasılık fonksiyonu

$$p(x) = \begin{cases} p & ,x = 1 \text{ (Başarılı)} \\ 1 - p & ,x = 0 \text{ (Başarısız)} \\ 0 & , \text{diğer durumlar} \end{cases}$$

  biçiminde olur.
- Bu olasılık fonksiyonunu

$$p(x) = P(X = x) = p^x (1 - p)^{1-x}, \; x = 0, 1$$

  biçiminde de yazabiliriz.

X rastgele değişkeni p başarı olasılıklı Bernoulli dağılımına sahip ise olasılık fonksiyonu

$$p(x) = P(X = x) = p^x (1-p)^{1-x}, \; x = 0, 1$$

biçiminde verilir. Bernoulli dağılımının parametresi p'dir.
Ayrıca,

$$\begin{aligned} E(X) &= \sum_{x=0}^{1} x \; p(x) = 0 + 1(p) = p, \\ E(X^2) &= \sum_{x=0}^{1} x \; p(x) = 0 + 1^2(p) = p \text{ ve } Var(X) = p - p^2 = p(1-p) \end{aligned}$$

olarak bulunur.

Bernoulli deneyinde deney bir kez yapılır. Bernoulli deneyinin n kez ve birbirinden bağımsız olarak tekrarlanmasıyla Binom dağılımı elde edilir.

Binom Dağılımı

- İki sonucu olan bir deney (Bernoulli deneyi) aynı koşullar altında n kez tekrarlanıyor.
- Deneyler birbirinden bağımsızdır.
- Bu deneyde ilgilenilen sonucu ifade eden başarı olasılığı p her deneyde aynıdır.

Bu koşulların sağlandığı n bağımsız Bernoulli deneyi sonucunda elde edilecek **toplam başarılı olma sayısını** X rastgele değişkeni ile gösterelim.

- Bu durumda X rastgele değişkeninin alacak olduğu değerler $\{0, 1, 2, ..., n\}$'dir.
- Bu X'in olasılık dağılımına **binom dağılımı** denir.
- X rastgele değişkeninin olasılık fonksiyonu

$$p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \ x = 0, 1, 2, ..., n$$

biçiminde olur.

- Binom dağılımın iki parametresi vardır: n deneyin tekrar sayısı ve p başarılı olma olasılığı
- Eğer X r.d. n ve p parametreli Binom dağılımına sahip ise genellikle $X \sim Binom(n, p)$ biçiminde gösteririz.
- $X \sim Binom(n, p)$ r.d. için $E(X) = np$ ve $Var(X) = np(1 - p)$ olarak bulunur.

Örnek 6: Bir atıcının hedefi vurma olasılığı $2/3$'tür. X r.d. birbirinden bağımsız olarak 8 atış yapıldığında toplam hedefi vurma sayısı olmak üzere
a) Hedefi tam 3 kez vurma,
b) Hedefi en az 1 kez vurma,
c) Hedefi en çok 1 kez vurma olasılıklarını bulunuz.

Çözüm: Burada tanımlanan X r.d. başarı olasılığı (ilgilenilen olay: hedefi vurma) $2/3$ ve tekrar sayısı 8 olan Binom dağılımına sahiptir. Yani, $X \sim Binom(8, 2/3)$ olur. Dolayısıyla olasılık fonksiyonu,

$$p(x) = P(X = x) = \binom{8}{x} \left(\frac{2}{3}\right)^x (1 - \frac{2}{3})^{8-x}, \; x = 0, 1, 2, ..., 8$$

olur.

a)

$$P(X = 3) = \binom{8}{3} \left(\frac{2}{3}\right)^3 (1 - \frac{2}{3})^{8-3} = \frac{448}{6561} \simeq 0.068$$

b)

$$\begin{aligned} P(X \geq 1) &= 1 - P(X < 1) = 1 - P(X = 0) \\ &= 1 - \binom{8}{0} \left(\frac{2}{3}\right)^0 (1 - \frac{2}{3})^{8-0} \\ &= 1 - \left(\frac{1}{3}\right)^8 = 0.9998476 \end{aligned}$$

c)

$$\begin{aligned} P(X &\leq 1) = P(X=0) + P(X=1) \\ &= \binom{8}{0}\left(\frac{2}{3}\right)^{0}(1-\frac{2}{3})^{8-0} + \binom{8}{1}\left(\frac{2}{3}\right)^{1}(1-\frac{2}{3})^{8-1} \\ &= \left(\frac{1}{3}\right)^{8} + 8\frac{2}{3}\left(\frac{1}{3}\right)^{7} = \left(\frac{1}{3}\right)^{7}\left(\frac{17}{3}\right) = 0.00259 \end{aligned}$$