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

Kesikli Olasılık Dağılımları

Fatih Kızılaslan

Marmara Üniversitesi

2019-2020 Bahar 11. Hafta

Hatırlatma

- **Bernoulli Dağılımı**
  X rastgele değişkeni Benoulli dağılımına sahip ise 0 ve 1 olmak üzere iki sonucu vardır. Olasılık fonskiyonu
  $$p(x) = P(X = x) = p^x(1-p)^{1-x},\ x = 0, 1$$
  ve $E(X) = p$, $Var(X) = p(1-p)$ dir.
- **Binom Dağılımı**
  n tane birbirinden bağımsız aynı p başarı olasılıklı Bernoulli deneyinin tekrarlansın. X r.d. bu n denemedeki başarılı deneylerin sayısı olsun. Bu durumda $X \sim Binom(n, p)$ olur. Olasılık fonskiyonu
  $$p(x) = P(X = x) = \binom{n}{x} p^x(1-p)^{n-x},\ x = 0, 1, 2, ..., n$$
  ve $E(X) = np$ ve $Var(X) = np(1-p)$ dir.
- **Geometrik Dağılım** (Bugün inceleyeceğiz!)

Binom Dağılımı Örnekler

Örnek 1: Bir topluluğun %40'ı kadındır. Bu topluluktan rastgele seçilen 10 kişiden
a) 6 tanesinin kadın olma olasılığını bulunuz.
b) En az 2 tanesinin kadın olasılığını bulunuz.

Çözüm: İki sonucu olan bir durum $n = 10$ kez tekrar edilmiştir. Bu nedenle Binom dağılımını kullanabiliriz. İlgilendiğimiz durumun (olayın) olasılığı yani başarı olasılığı $p = P(\text{Kadın olma}) = 0.4$ dir. Böylece, bu 10 kişi içindeki toplam kadın sayısını X r.d. ile gösterirsek
$X \sim Binom(10, 0.4)$ ve olasılık fonksiyonu

$$p(x) = P(X = x) = \binom{10}{x} (0.4)^x \, (0.6)^{10-x}, \;\; x = 0, 1, 2, ..., 10$$

olur.

a) $P(X = 6) = \binom{10}{6} (0.4)^6 \, (0.6)^{10-6} = 0.1114767$

b)

$$\begin{aligned}
P(\text{en az 2 kadın olması}) &= P(X \geq 2) \\
&= 1 - P(\text{2'den az kadın olması}) \\
&= 1 - P(X < 2) \\
&= 1 - [P(X = 0) + P(X = 1)] \\
&= 1 - \left[0.6^{10} + 10\ 0.4\ 0.6^{9}\right] \\
&= 1 - 0.0463574 = 0.9536426
\end{aligned}$$

Örnek 2: Bir basketbol oyuncusunun herhangi bir atışında sayı yapma olasılığı %30 dur. Her atışı diğerinden bağımsız olduğu bir maçta bu oyuncu 4 atış yaptığında bu atışların
a) bir tanesinde sayı yapma olasılığını bulunuz.
b) en az bir tanesinde sayı yapma olasılığını bulunuz.
c) Ortalama basket sayısını ve standart sapmasını bulunuz.

Çözüm: Başarı olasılığı sayı yapma olasılığı $p = 0.3$ olur. $n = 4$ olduğundan 4 atışta yaptığı sayıların sayısını X r.d. ile gösterelim. $X \sim Binom(4, 0.3)$ olur.

a) $P(X=1) = \binom{4}{1} 0.3^1 (1-0.3)^{4-1} = 0.4116$

b)

$$\begin{aligned} P(\text{en az bir tanesinde sayı}) &= P(X \geq 1) = 1 - P(X < 1) \\ &= 1 - P(X=0) = 1 - 0.7^4 = 0.7599 \end{aligned}$$

c) Ortalama başarılı durum yani basket sayısı: $E(X) = np = 4(0.3) = 1.2$ ve varyansı $Var(X) = np(1-p) = 4\ 0.3\ 0.7 = 0.84$, standart sapma $\sigma = \sqrt{0.84} = 0.916$ bulunur.

Örnek 3: Hilesiz bir zar 2400 kez atılıyor. Kaç atışta zarın üst yüzünde 1 gelmesini beklersiniz. Bu duruma ilişkin varyans nedir ?

Çözüm: Bir zar atışında 6 farklı sonuç vardır. Ancak, burada ilgilenilen durum zarın üst yüzünde 1 olması veya olmaması olduğundan 2 sonucumuz vardır.
$n = 2400$ tekrar yapılıyor ve $p = P(1 \text{ gelme durumu}) = 1/6$ olur. Böylece, ortalama başarılı durum sayısı $E(X) = np = 2400(1/6) = 400$ ve $Var(X) = 2400(1/6)(5/6) = 333.33$ bulunur.

Soru 1: 2019 yılında Biyoistatistik dersini alan 77 öğrenciden 36 öğrencinin derste başarılı olduğu biliniyor. 77 öğrenciden rastgele seçilen 5 öğrenciden
a) Hepsinin Biyoistatistik dersinde başarılı olma olasılığını,
b) En 2 tanesinin Biyoistatistik dersinde başarılı olma olasılığını bulunuz.
c) En çok 3 tanesinin Biyoistatistik dersinde başarılı olma olasılığını bulunuz.

Geometrik Dağılım

- Bağımsız tekrar edilen Bernoulli deneylerinde ilk başarılı deneme elde edilinceye kadar gerçekleşen deneme sayısını X ile gösterirsek. X'in alabilecek olduğu değerler $x = 1, 2, 3, ...$ olur.
- Bu durumda X r.d. **Geometrik dağılıma** sahiptir denir ve $X \sim Geo(p)$ biçiminde gösterilir. Olasılık fonksiyonu

$$p(x) = P(X = x) = p(1-p)^{x-1}, \; x = 1, 2, 3, ...$$

  biçimindedir. Burada p başarı olasılığıdır ve $0 < p < 1$ dir.
- Ayrıca, $E(X) = 1/p$ ve $Var(X) = (1-p)/p^2$ dir.

Örneğin,

- bir para atışı deneyinde ilk defa yazı gelinceye kadar yapılan denemelerin sayısı,
- bir atıcının ilk başarılı atışı gerçekleştirene kadar yaptığı atışların sayısı,

birer geometrik rastgele değişkendir.

Örnek 4: Herhangi bir satış temsilcisinin aylık satış hedefine ulaşma olasılığı %70 dir. Yeni işe başlayan bir temsilcinin
a) aylık hedefine ilk kez 3. ayda ulaşması olasılığını bulunuz.
b) aylık hedefine en çok 3. ayda ulaşması olasılığını bulunuz.
c) aylık hedefe ulaşıncaya kadar ortalama kaç ay geçer ?

Çözüm: Başarı olasılığı hedefe ulaşması $p = 0.7$ dir. İlk başarı elde edinceye kadar yapılan denemelerin sayısı için Geometrik dağılımı kullanırız. X r.d. $Geo(0.7)$ olmak üzere olasılık fonksiyonu

$$p(x) = P(X = x) = 0.7(1 - 0.7)^{x-1}, \; x = 1, 2, 3, \ldots$$

olur. **a)**

$$\begin{aligned} P(X &= 3) = P(\text{ilk 2 ayda hedefe ulaşamıyor, 3.de ulaşıyor}) \\ &= 0.7(0.3)^{3-1} = 0.063 \end{aligned}$$

b)

$$\begin{aligned} P(X &\leq 3) = P(X = 1) + P(X = 2) + P(X = 3) \\ &= 0.7(0.3)^0 + 0.7(0.3)^1 + 0.7(0.3)^2 = 0.973 \end{aligned}$$

c) $E(X) = 1/p = 1/0.7 = 1.428571$ ay

Örnek 5: Bir hastalığın riskini arttırdığı bilinen bir gen insanların %15'inde bulunduğu biliniyor. Bir klinik araştırmada bu genin tespiti için test yapılmaktadır. Bu geni taşıyan ilk kişiyi bulana kadar toplam 5 test yapma olasılığını bulunuz.

Çözüm: İstenilen durum: Toplam 5 test yapılıyor, ilk 4 testte başarısız sonuç, 5. de başarılı sonuç.
Böylece, bu durumun gerçekleşmesi olasılığı: $(0.85)^4(0.15) = 0.07830094$ dir.
Ayrıca, Geometrik dağılım da kullanabiliriz. X r.d. geni taşıyan ilk kişiyi bulana kadar yapılan testlerin sayısı olsun. Bu durumda $P(X = 5) = 0.15(1 - 0.15)^4 = 0.07830094$ olur.

Poisson Dağılımı

İstatistikte en sık kullanılan
kesikli olasılık dağılımlarından biri de Poisson dağılımıdır. Fransız Matematikçi Simeon Denis Poisson (1781-1840) tarafından tanımlanmıştır.

- **Poisson dağılımı** belli bir aralıkta rastgele ortaya çıkan nadir olayların sayısının olasılık dağılımıdır.
- Bu aralık bir zaman aralığı, bir bölge veya alan olabilir.

Poisson dağılımını,

- Bir çağrı merkezine bir saatte gelen çağrıların sayısı,
- Bir havaalanına her saat inen uçakların sayısı,
- Bir kitabın her sayfasındaki hataların sayısı,
- Bir fabrikada bir günde üretilen hatalı ürünlerin sayısı,
- Bir köprüde bir günde meydana gelen kazaların sayısı,
  gibi durumlarda kullanabilriz.

Belli bir zaman aralığında ilgilenilen olayın ortaya çıkma sayısı olan X r.d. aşağıdaki koşulları sağlıyorsa X Poisson rastgele değişenidir.

1. Verilen aralıkta ilgilenilen olayın ortaya çıkma olasılığı sabittir ve aralığın büyüklüğü ile orantılıdır.
2. İki ayrık zaman aralığında ortaya çıkan olaylar birbirinden bağımsızdır.

X Poisson rastgele değişkeninin olasılık fonksiyonu

$$p(x) = P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \; x = 0, 1, 2, ..., \; \lambda > 0$$

biçimindedir. $(e = 2.718282)$
λ : Poisson dağılımının parametresi olup, gerçekleşen ortalama olay sayısını gösterir.

- Beklenen değer $E(X) = \lambda$ ve varyans $Var(X) = \lambda$ dır.

Örnek 6: Geçmişte hava koşulları nedeniyle İstanbul'da 1 yılda okullar ortalama 3 gün kapalı kalmıştır. Gelecek yıl okulların 4 gün kapalı kalması olasılığı nedir?

Çözüm: X r.d. Poisson dağılımına sahiptir, zaman aralığımız 1 yıl olup, 1 yılda ortalama gerçekleşen olay sayısı 3 olduğundan $\lambda = 3$ olur. Böylece, X r.d. için olasılık fonksiyonu

$$p(x) = P(X = x) = \frac{e^{-3}3^x}{x!}, \; x = 0, 1, 2, ...$$

dır. İstenilen 1 yılda okulların 4 gün kapalı olması olasılığı

$$P(X = 4) = \frac{e^{-3}3^4}{4!} = 0.1680314$$

olarak bulunur.

Örnek 7: İşlek bir caddede bir gün içinde meydana gelen trafik kazalarının sayısının ortalama 1.5 olduğu görülmektedir.
a) Bu caddede herhangi bir günde kaza olmaması ve 1 kaza olması olasılığını,
b) Bir haftalık zaman süresinde kaza olmaması olasılığını bulunuz.

Çözüm: 1 gündeki ortalama kaza sayısı 1.5 olduğundan $\lambda = 1.5$ olur ve

$$p(x) = P(X = x) = \frac{e^{-1.5}(1.5)^x}{x!}, \; x = 0, 1, 2, ...$$

a) $P(X = 0) = \dfrac{e^{-1.5}(1.5)^0}{0!} = e^{-1.5} = 0.2231302$ ve

$P(X = 1) = \dfrac{e^{-1.5}(1.5)^1}{1!} = e^{-1.5}(1.5) = 0.3346952$

b) Bir haftalık sürede meydana gelen kazaların sayısı için Y r.d. kullanalım. Bir haftalık zaman süresindeki ortalama kaza sayısı $7(1.5) = 10.5$ olduğundan $\lambda = 10.5$ olur. Böylece,

$$P(Y = 0) = \frac{e^{-10.5}(10.5)^0}{0!} = e^{-10.5} = 0.000028 \text{ bulunur.}$$