SAB101 OLASILIK

BÖLÜM 3
KESİKLİ OLASILIK DAĞILIMLARI
3.3 Hipergeometrik Dağılım
3.4 Poisson Dağılımı
3.5 Çokterimli Dağılım

Doç. Dr. Furkan BAŞER
Ankara Üniversitesi
Uygulamalı Bilimler Fakültesi

3.3 Hipergeometrik Dağılım
3.3.1 Hipergeometrik Dağılımın Tanımı (1/3)

- r nin belirli bir türü ifade ettiği N sayıdan oluşan bir topluluk düşünün.
- Öğelerden biri rasgele seçilirse, belirli özelliğe sahip bu olayların olasılığı;

$$p = \frac{r}{N}$$

- Sonuç olarak, eğer n madde yerine konularak rasgele seçildiyse, o zaman X'in dağılımı,

$$X \sim B\left(n,\, r/N\right)$$

3.3.1 Hipergeometrik Dağılımın Tanımı (2/3)

- Bununla birlikte, eğer n madde yerine konmadan rasgele seçilirse, X dağılımı hipergeometrik dağılımdır.
- Hipergeometrik dağılımın olasılık fonksiyonu,

$$P(X=x)=\frac{\binom{r}{x}\times\binom{N-r}{n-x}}{\binom{N}{n}} \qquad \max\{0,n+r-N\}\le x\le \min\{n,r\}.$$

3.3.1 Hipergeometrik Dağılımın Tanımı (3/3)

- Beklenen değer ve varyans,

$$E(X) = n\frac{r}{N}$$

$$Var(X) = \left(\frac{N-n}{N-1}\right) \times n \times \frac{r}{N} \times \left(1 - \frac{r}{N}\right)$$

- Bu türdeki r özelliğini içeren N büyüklüğündeki bir popülasyondan yerine konulmaksızın, n boyutunda rasgele bir örneklem içinde belirli bir türdeki maddelerin sayısının dağılımını temsil eder.

- X ve Y bağımsız, binom dağılımına sahip rasgele değişkenleri ise,

$$X \sim B(r, p), \;\; Y \sim B(N-r, p)$$

X+Y=n koşulu altında X in koşullu olasılık fonksiyonu;

$$P\{X = x \mid X+Y=n\} = \frac{P\{X=x, X+Y=n\}}{P\{X+Y=n\}} = \frac{P\{X=x, Y=n-x\}}{P\{X+Y=n\}}$$

$$\Rightarrow P\{X = x \mid X+Y=n\} = \frac{P\{X=x\}P\{Y=n-x\}}{P\{X+Y=n\}}$$

$$= \frac{\binom{n}{x} p^x (1-p)^{n-x} \binom{N-r}{n-x} p^{n-x} (1-p)^{N-r-n+x}}{\binom{N}{n} p^n (1-p)^{N-n}}$$

$$= \frac{\binom{n}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Yani X+Y verildiğinde X in koşullu dağılımı hipergeometrik dağılımı verir!

- X in beklenen değeri:
- Xi aşağıdaki gibi bir rasgele değişkeni olsun.

Xi=1 i. Seçim kabul ediliyorsa, yoksa 0.

O zaman,

$$P\{X_i = 1\} = \frac{r}{N}$$

X, (r,N.n) parametresiyle hipergeometrik dağılsın. O zaman,

$$X = \sum_{i=1}^{n} X_i$$

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} P\{X_i = 1\} = \frac{nr}{N}$$

$$Var(X) = Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n}\sum_{j>i}^{n} Cov(X_i, X_j)$$

Xi Bernoulli rasgele değişkeni olduğu için,

$$Var(X_i) = P\{X_i = 1\}(1 - P\{X_i = 1\}) = \frac{r}{N}\frac{N-r}{N}$$

i<j için,

$$Cov(X_i, X_j) = E[X_i X_j] - E[X_i]E[X_j]$$

$$E[X_i X_j] = P\{X_i X_j = 1\} = P\{X_i = 1, X_j = 1\}$$

$$P\{X_i = 1, X_j = 1\} = P\{X_i = 1\}P\{X_j = 1 \mid X_i = 1\} = \frac{r}{N}\frac{r-1}{N-1}$$

$$Cov(X_i, X_j) = \frac{r}{N}\frac{r-1}{N-1} - \left(\frac{r}{N}\right)^2 = \frac{-r(N-r)}{N^2(N-1)}$$

Bu nedenle,

$$Var(X) = \frac{r(N-r)}{N^2} - \frac{n(n-1)N(N-r)}{N^2(N-1)} = n\frac{r}{N}\left(1 - \frac{r}{N}\right)\frac{N-n}{N-1}$$

$$p = r/N$$

$E[X] = np$ ve $Var(X) = np(1-p)\dfrac{N-n}{N-1}$

N sonsuza gittiğinde,

$$Var(X) = np(1-p)$$

Örnek: Süt Kabı İçeriği(1/3)

- Sütün, 16 süt kabı bulunan kutulara perakende satış noktalarına gönderildiğini varsayalım. Altı kilonun altında ağırlığa sahip kapların bulunduğu belirli bir kutu denetleme için açılır ve beş kap rasgele seçilir.

- Denetmen tarafından seçilen örnekte düşük kilolu süt kaplarının sayısının dağılımı → $N=16$, $r=6$, and $n=5$ parametreleriyle hipergeometrik dağılım.

Örnek: Süt Kabı İçeriği (2/3)

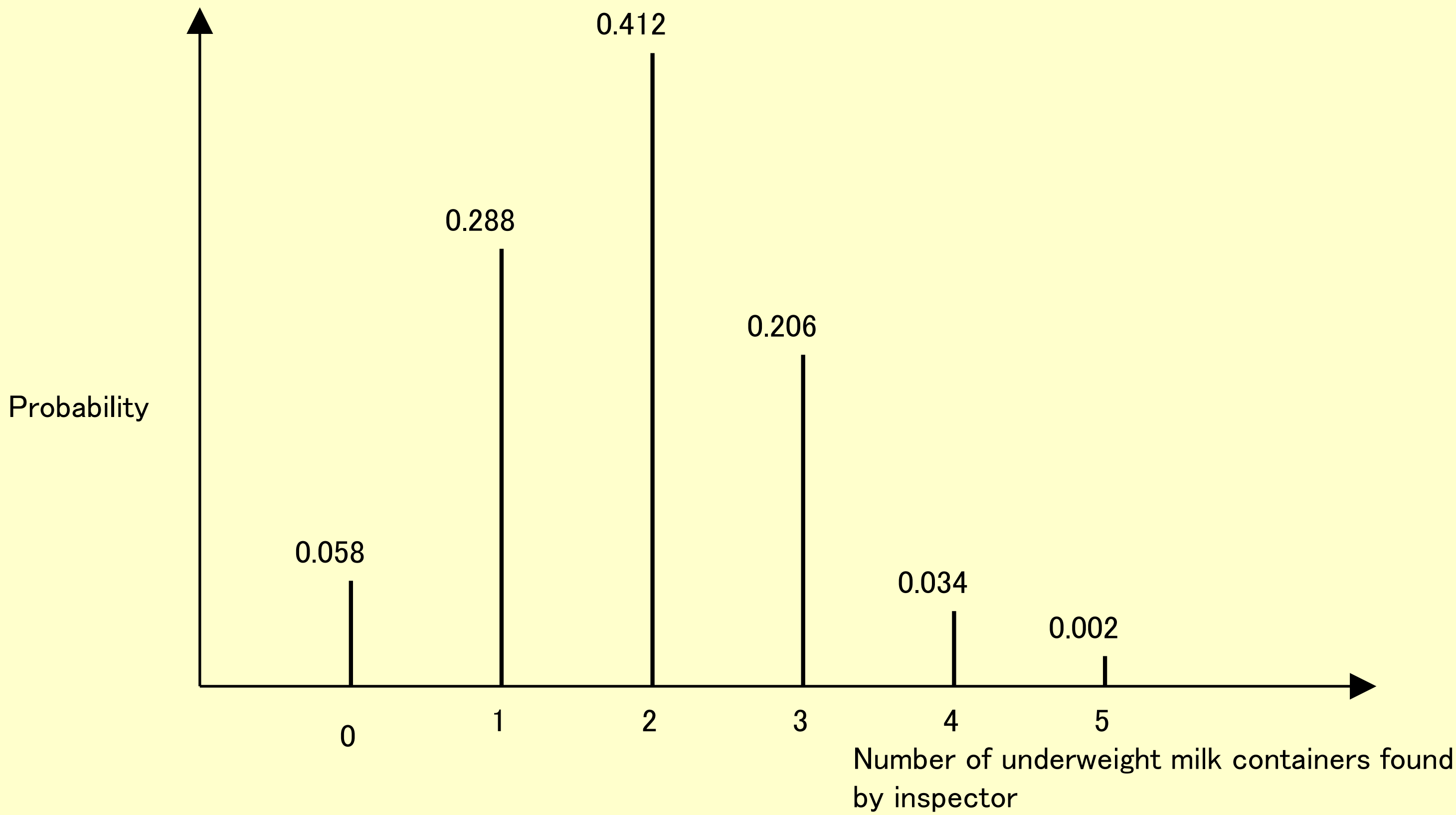
- Denetmenin tam olarak iki tane az kilolu kutu seçmesi olasılığı

$$P(X=2)=\frac{\binom{6}{2}\times\binom{10}{3}}{\binom{16}{5}}=\frac{\left(\frac{6!}{2!4!}\right)\times\left(\frac{10!}{3!7!}\right)}{\left(\frac{16!}{5!11!}\right)}=0.412$$

- Denetmen tarafından seçilen düşük kilolu kutuların beklenen sayısı

$$E(X)=\frac{nr}{N}=\frac{5\times 6}{16}=1.875$$

Örnek: Süt Kabı İçeriği (3/3)



3.4 Poisson Dağılımı

3.4.1 Poisson Dağılımı Tanımı (1/3)

- Bir öğedeki kusur sayısı
- Bir madde tarafından yayılan radyoaktif parçacıkların sayısı
- Belirli bir süre sınırı olan bir operatör tarafından alınan telefon çağrılarının sayısı gibi olayların dağılımını inceler.

- Yani, belirli sınırlar içinde meydana gelen "olayların" sayısıdır.

3.4.1 Poisson Dağılımı Tanımı (2/3)

- X rasgele değişkeni, λ parametresiyle Poisson dağılıyorsa

$$X \sim P(\lambda)$$

Olasılık fonksiyonu;

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

3.4.1 Poisson Dağılımı Tanımı (3/3)

- Poisson dağılımı, belirli bir olayın zaman birimi, mesafe veya hacim başına kaç kez oluştuğunu modellemek için genellikle yararlıdır ve ortalama ve varyansının her ikisi de λ parametre değerine eşittir.

- Beklenen değer ve varyans,

$$E(X) = Var(X) = \lambda$$

- Poisson dağılımı, n büyük ve p küçük olduğunda (n, p) parametreleriyle binom dağılımına yakınsar.
- X, (n,p) parametreleriyle binom dağılsın.

O zaman

$$\lambda = np$$

$$P\{X = i\} = \frac{n!}{(n-i)!\,i!} p^i (1-p)^{n-i} = \frac{n!}{(n-i)!\,i!} \left(\frac{\lambda}{n}\right)^i \left(1-\frac{\lambda}{n}\right)^{n-i}$$

Yani,

$$P\{X = i\} = \frac{n(n-1)\cdots(n-i+1)}{n^i} \frac{\lambda^i}{i!} \frac{(1-\lambda/n)^n}{(1-\lambda/n)^i}$$

n büyük, p küçük olduğunda,

$$\left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \frac{n(n-1)\cdots(n-i+1)}{n^i} \approx 1, \left(1-\frac{\lambda}{n}\right)^i \approx 1$$

Bu yüzden,

$$P\{X = i\} \approx e^{-\lambda} \frac{\lambda^i}{i!}$$

Örnek: Yazılım Hataları (1/3)

- Bir yazılım parçasındaki hata sayısının λ = 3 parametresine sahip bir Poisson dağılımı olduğunu varsayalım.

[Beklenen hata sayısı] = [hata sayısının varyansı] = 3.

- Hatasız yazılım parçalarının olasılığı,

$$P\left(X=0\right)=\frac{e^{-3}\times 3^{0}}{0!}=e^{-3}=0.050$$

Örnek: Yazılım Hataları (2/3)

- Yazılım parçalarında 3 ve daha fazla hata olma olasılığı,

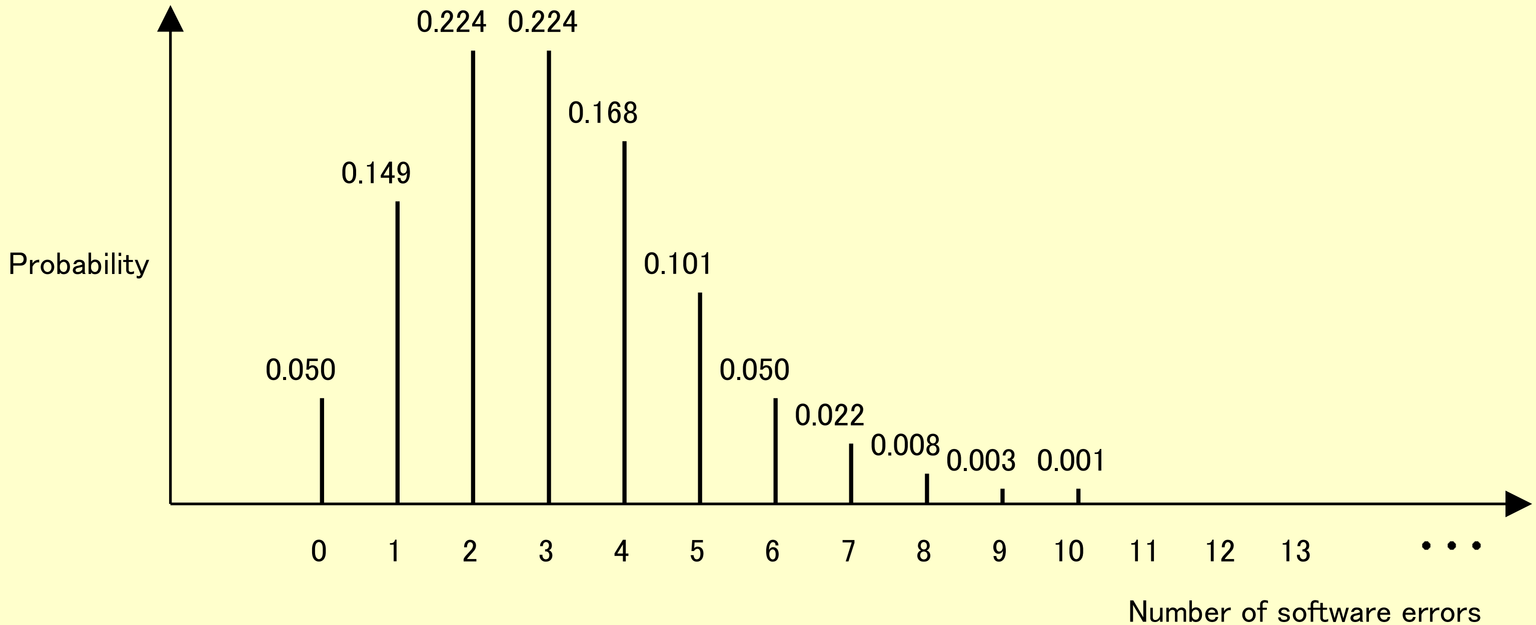
$$P(X \geq 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$$

$$= 1 - \frac{e^{-3} \times 3^0}{0!} - \frac{e^{-3} \times 3^1}{1!} - \frac{e^{-3} \times 3^2}{2!}$$

$$= 1 - e^{-3}\left(\frac{1}{1} + \frac{3}{1} + \frac{9}{2}\right)$$

$$= 1 - 0.423 = 0.577$$

Örnek: Yazılım Hataları (3/3)



x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	• • •
$P(X \leq x)$	0.050	0.149	0.224	0.224	0.168	0.101	0.050	0.022	0.008	0.003	0.001	0.000	0.000	0.000	

3.5 Çokterimli Dağılım
3.5.1 Çokterimli Dağılımın Tanımı (1/2)

- Her biri $p_1,\ldots, p_k$ ($p_1+\cdots+p_k = 1$) sabit olasılıklarıyla k sonuç içeren n tane bağımsız deneme olduğunu düşünün.
- Her bir sonucun meydana gelme sayısını gösteren $\mathbf{X_1,\ldots, X_k}$ rasgele değişkeni, bir çoklu dağılıma sahip olduğu söylenir.
- Birleşik olasılık fonksiyonu;

$$P\left(X_1 = x_1,\ldots, X_k = x_k\right) = \frac{n!}{x_1!\cdots x_k!} \times p_1^{x_1} \times \cdots \times p_k^{x_k}$$

x_i negatif olmayan tam sayı değerlerini alır ve $x_1 + \cdots + x_k = n.$

3.5.1 Çokterimli Dağılımın Tanımı (2/2)

- $X_1,\ldots, X_k$ rasgele değişkenlerinin beklenen değer ve varyansı;

$$E\left(X_i\right)=np_i \qquad Var\left(X_i\right)=np_i(1-p_i)$$

ama X ler bağımsız değil. (*neden?)*

Örnek: Makine Bozulmaları (1/4)

- Makine arızalarının elektriksel arızalara, mekanik arızalara ve operatörün yanlış kullanımına bağlı olduğunu ve bu nedenlerin sırasıyla 0.2, 0.5 ve 0.3 olasılıkları ile meydana geldiğini varsayalım.
- Mühendis, önümüzdeki on bozulmanın nedenlerini tahmin etmekle ilgileniyor.

- X_1: elektriksel sebeplerden dolayı arıza sayısı.
- X_2: Mekanik sebeplerden kaynaklanan arıza sayısı.
- X_3: operatörün yanlış kullanımından kaynaklanan arıza sayısı.

Örnek: Makine Bozulmaları (2/4)

- $X_1+X_2+X_3=10$
- Arıza nedenleri birbirinden bağımsız olarak kabul edilebilir ise, o zaman olasılık fonksiyonu,

$$P\left(X_1=x_1, X_2=x_2, X_3=x_3\right)=\frac{10!}{x_1!x_2!x_3!}\times 0.2^{x_1}\times 0.5^{x_2}\times 0.3^{x_3}$$

- Üç elektrik arızası, beş mekanik arıza ve iki hatalı arıza oluşması olasılığı,

$$P\left(X_1=3, X_2=5, X_3=2\right)=\frac{10!}{3!5!2!}\times 0.2^{3}\times 0.5^{5}\times 0.3^{2}=0.057$$

Örnek: Makine Bozulmaları (3/4)

- Elektriksel sebeplerden kaynaklanan beklenen arıza sayısı,

$$E\left(X_1\right)=np_1=10\times 0.2=2$$

- Mekanik sebeplerden kaynaklanan beklenen arıza sayısı,

$$E\left(X_2\right)=np_2=10\times 0.5=5$$

- Operatörün yanlış kullanımından kaynaklanan beklenen arıza sayısı,

$$E\left(X_3\right)=np_3=10\times 0.3=3$$

Örnek: Makine Bozulmaları (4/4)

- Eğer mühendis, en fazla iki elektrik kesintisi olasılığı ile ilgileniyorsa, bu $X_1 \sim B(10, 0.2)$ ile hesaplanabilir. Böylece,

$$P(X_1 \le 2) = P(X_1 = 0) + P(X_1 = 1) + P(X_1 = 2)$$

$$= \binom{10}{0} \times 0.2^0 \times 0.8^{10} + \binom{10}{1} \times 0.2^1 \times 0.8^9 + \binom{10}{2} \times 0.2^2 \times 0.8^8$$

$$= 0.107 + 0.268 + 0.302 = 0.677$$