Hipergeometrik Dağılım

Tanım : N birimden oluşan bir kitleden a tanesi istenen özelliğe sahip, $N - a$ tanesinde ise istenen özellik yoktur. Bu kitleden n genişliğinde iadesiz olarak bir örneklem çekildiğinde istenen özellikteki birimlerin sayısı X tesadüfi değişkeni ile gösterildiğinde X'e hipergeometrik tesadüfi değişken denir ve $X \sim HGeometrik(n, a, N)$ ile gösterilir. X tesadüfi değişkeninin olasılık fonksiyonu aşağıdaki gibidir:

$$P(X = x) = \left\{ \begin{array}{cc} \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}, & x = 0,1, \dots, n \\ 0, & d.d \end{array} \right\}$$

Bu dağılım özellikle örnekleme kuramı ve kalite denetiminde sıkça kullanılır.

Beklenen Değer ve Varyans

$$E(X) = \sum_{D_x} x\mathrm{P(X=x)}$$

$$E(X) = \sum_{x=0}^{n} x \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$$

$$= \frac{na}{N}$$

$$\mathrm{E(X^2)} = \sum_{D_x} x^2\mathrm{P(X=x)}$$

$$\mathrm{E(X^2)} = \sum_{x=0}^{n} x^2 \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$$

$$= \frac{na}{N(N-1)}[(a-1)(n-1) + (N-1)]$$

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

$$= \frac{N-n}{N(N-1)} n \frac{a}{N}\left(1 - \frac{a}{N}\right)$$

Moment çıkaran fonksiyonun çıkarılışı karışık olduğundan burada verilmeyecektir.

Örnek : 150 kişilik bir sınıftaki öğrencilerin %20 si Karadeniz bölgesindendir. Bu sınıftan tesadüfi olarak 6 öğrenci seçildiğinde bu öğrencilerden,

a) İkisinin Karadeniz bölgesinden olması olasılığı nedir?
b) En az üçünün Karadeniz bölgesinden olması olasılığı nedir?
c) En çok dördünün Karadeniz bölgesinden olmaması olasılığı nedir?
d) Seçilen bu 6 öğrenciden kaçının Karadeniz bölgesinden olması beklenilir.
e) Karadeniz bölgesinden olmayanların sayısının varyansını hesaplayınız.

Çözüm. X ve Y sırasıyla Karadeniz bölgesinden olanların ve olmayanların sayısını göstersin. $N = 150,\ a = 30\ ve\ n = 6$ olmak üzere,

a) $P(X = 2) = \frac{\binom{30}{2}\binom{120}{4}}{\binom{150}{6}}$

$=0,067$

b) $P(X \geq 3) = 1 - P(X < 3)$

$$= 1 - [P(X = 0) + P(X = 1) + P(X = 2)$$

$$= 1 - \sum_{x=0}^{2} \frac{\binom{3U}{x}\binom{120}{6-x}}{\binom{150}{6}} = 0,0946$$

c) $\bar{a} = 120$ olmak üzere Y tesadüfi değişkeninin olasılık fonksiyonu,

$$P(Y = y) = \begin{cases} \dfrac{\binom{\bar{a}}{y}\binom{N-\bar{a}}{n-y}}{\binom{N}{n}}, & y = 0,1,\dots,n \\ 0, & d.d \end{cases}$$

$$P(Y \leq 4) = 1 - P(Y > 4)$$

$$= 1 - [P(Y = 5) + P(Y = 6)]$$

$$=1 - \left[\frac{\binom{120}{5}\binom{30}{1}+\binom{120}{6}\binom{30}{0}}{\binom{150}{6}}\right]$$

$$=0,3446$$

d) $E(X) = \frac{6.30}{150} = \frac{6}{5}$

e) $Var(Y) = \frac{N-1}{N(N-1)} n \frac{\bar{a}}{N}(1 - \frac{\bar{a}}{N}) = 0,93$

Örnek : Bir firma bir üreticiden 500 kasa kiraz alacaktır. Bunun için tesadüfi olarak seçtiği 100 kasa kirazı kontrol etmek istiyor ve en çok iki kasa kirazın bozuk olabileceğini hoş görüyor. Bu bağlamda %1, %2, %3, %4, %5, %8, %10 bozuk oranı için kirazların (parti mallarının) kabul edilme olasılıklarını hesaplayınız.

Çözüm. $X \sim HGeometrik\ (n, a, N)$ olmak üzere burada belli özelliğe sahip birim sayısı $a = Np$ alınır ve X tesadüfi değişkeni de bozuk kiraz kasası sayısını gösterir.

$$P(X = x) = \begin{Bmatrix} \dfrac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}, & x = 0,1,\dots,n \\ 0, & d.d \end{Bmatrix}$$

En çok 2 kasa bozuk kirazın hoş görülebileceği kabul edildiğinden istenen olasılık

$$\text{P(X≤2)} = \sum_{x=0}^{2} \frac{\binom{500\cdot\rho}{x}\binom{500.1}{100-x}}{\binom{5\quad 0\quad 0}{1\quad 0\quad 0}}$$

$a = Np \; ve \; N = 500$

$p = \%0 \; için \; a = 0$

$$P(X = 0) = 1, \; P(X = 1) = 0, P(X = 2) = 0$$

olduğundan

$$P(X \leq 2) = 1$$

bulunur.

$p = 0.01 \; için \; a = 5$

$$P(X \leq 2) = \sum_{x=0}^{2} \frac{\binom{500\cdot\rho}{x}\binom{500.q}{100-x}}{\binom{5\quad 0\quad 0}{1\quad 0\quad 0}}$$

$$= \frac{\binom{5}{0}\binom{4\quad 9\quad 5}{1\quad 0\quad 0}}{\binom{5\quad 0\quad 0}{1\quad 0\quad 0}} + \frac{\binom{5}{1}\binom{495}{99}}{\binom{500}{100}} + \frac{\binom{5}{2}\binom{495}{98}}{\binom{500}{100}}$$

$$= 0{,}3260 + 0{,}4116 + 0{,}2053$$

$$= 0{,}9429$$

$p = 0.02 \; için \; a = 10$

$$P(X \leq 2) = \sum_{x=0}^{2} \frac{\binom{500\cdot\rho}{x}\binom{500.q}{100-x}}{\binom{5\quad 0\quad 0}{1\quad 0\quad 0}}$$

$$= \frac{\binom{10}{0}\binom{490}{100}}{\binom{5\quad 0\quad 0}{1\quad 0\quad 0}} + \frac{\binom{10}{1}\binom{490}{99}}{\binom{500}{100}} + \frac{\binom{5}{2}\binom{490}{98}}{\binom{500}{100}}$$

=0,678

$p = 0.03\ için\ a = 15$

$$P(X\leq 2) = \sum_{x=0}^{2} \frac{\binom{15}{x}\binom{485}{100-x}}{\binom{500}{100}}$$

$$= 0{,}3830$$

$p = 0{,}04\ için\ a = 20\ \ P(X \leq 2) = 0{,}1985$

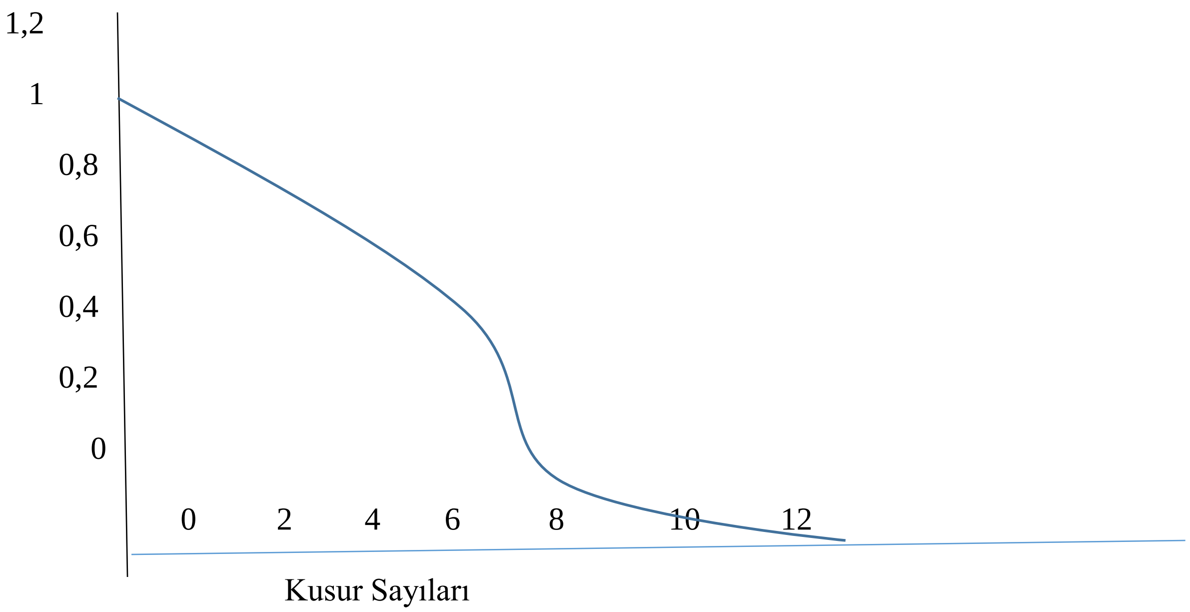
$p = 0{,}05\ için\ a = 25\ \ P(X \leq 2) = 0{,}0920$

$p = 0{,}08\ için\ a = 40\ \ P(X \leq 2) = 0{,}0062$

$p = 0{,}10\ için\ a = 50\ \ P(X \leq 2) = 0{,}0001$

Verilen kusur oranları ve bulunan olasılıklar aşağıdadır:

Kusur % si	0	1	2	3	4	5	8	10
Kabul olasılığı	1	0,9429	0,6780	0,3830	0,1985	0,0920	0,062	0,0001



Şekil. İşlem Eğrisi Grafiği

Grafikten görüldüğü gibi parti mallarında (kirazlarda) hiç kusurlu mal yoksa parti malının kabul edilme olasılığı kesindir. %1 kusurlu mal varsa kabul edilme olasılığı 0,9429, %2 kusurlu mal varsa kabul edilme olasılığı 0,678, %3 ve daha fazlası için de alınacak malların kabul edilme olasılığı %50'nin altındadır. Yani 2 den çok kusurlu parti malları reddedilmektedir.

Çok Değişkenli Hipergeometrik Dağılım

Tanım . k farklı özelliğe sahip ve N biriminden oluşan bir kitle göz önüne alınıyor: N_1 tanesi 1. özelliğe, N_2 tanesi 2. özelliğe sahip ve bunun gibi N_k tanesi de k. özelliğe sahiptir.

$$N = \sum_{k=1}^{k} N_k$$

olur. Her bir alt kitleden sırasıyla $n_1, n_2, \dots \dots n_k$birimlik örneklem iadesiz olarak çekilmek istendiğinde

$$n = \sum_{i=1}^{k} n_i$$

ve X_i tesadüfi değişkenleri şöyle tanımlanıyor:

$$X_i = \left\{ \begin{array}{rl} n_i\,; & \text{Çekilen birimlerdem } i.\text{özelliğe sahip birim sayısı} \\ N_i - n_i\,; & \text{Çekilen birimlerden } i.\text{özelliğe sahip olmayan birim sayısı} \end{array} \right\}$$

Bu bağlamda $X_1, X_2, \dots, X_k$ tesadüfi değişkenlerinin ortak olasılık fonksiyonu aşağıdaki gibidir:

$$P(X_1 = n_1, X_2 = n_2, .., X_k = n_k = p(n_1, n_2, .., n_k) = \frac{\binom{N_1}{n_1}\binom{N_2}{n_2}..\binom{N_k}{n_k}}{\binom{N}{n}}$$

Beklenen Değer ve Varyans

X_i tesadüfi değişkeninin beklenen değer ve varyansı

$$E(X_i) = \frac{nn_i}{N}$$

$$Var(X_i) = \frac{N-n}{N(N-1)} n \frac{n_i}{N}\left(1 - \frac{n_i}{N}\right)$$

Örnek : Bir çocuk, içinde 10 mavi, 10 yeşil, 10 sarı ve 20 kırmızı boncuğun olduğu bir torbadan tesadüfi olarak 6 boncuk çekmektedir. Çektiği bu boncukların 1 mavi, 2 yeşil, 2 sarı ve 1 kırmızı olma olasılığı nedir?

Çözüm: $N_1 = 10$, $N_2 = 10, N_3 = 10\ ve\ N_4 = 20$ olmak üzere $X_1, X_2, X_3\ ve\ X_4$ tesadüfi değişkenleri de sırasıyla çocuğun çektiği mavi, yeşil, sarı ve kırmızı boncukların sayısını göstermek üzere istenen olasılık aşağıdadır.

$$P(X_1 = 1, X_2 = 2, X_3 = 2\ ve\ X_4 = 1) = \frac{\binom{10}{1}\binom{10}{2}\binom{10}{2}\binom{20}{1}}{\binom{50}{6}} = 0{,}0255$$

Düzgün(Uniform) Dağılım

Bir deneyin sonucunda meydana gelen sonlu sayıdaki farklı durumun eşit olasılığa sahip olduğunu düşünelim. Örneğin bir paranın atılması, bir zarın atılması veya n hacimli bir kitleden oluşturulan n hacimli örneklemlerden herhangi birinin seçilmesi deneyleri sonucu eşit olasılıklı durumlar oluşur.

Tanım: X tesadüfi değişkenin alabileceği sonlu sayıda farklı $x_1, x_2, \dots, x_k$ değerlerini eşit olasılıklarla alıyor ise bu tesadüfi değişkene düzgün kesikli tesadüfi değişken denir ve olasılık fonksiyonu

$$P(X = x) = \left\{ \begin{array}{ll} \frac{1}{k}, & x = x_1, x_2, \dots, x_k \\ 0, & d.d \end{array} \right\}$$

dur. X tesadüfi değişkeninin dağılım fonksiyonu

$$F(x) = \left\{ \begin{array}{ll} 0, & x < 1 \\ \frac{x}{k}, & 1 \le x \le k \\ 1, & x > k \end{array} \right\}$$

Olur.

Beklenen Değer ve Varyans

$E(X) = \sum_{D_x} xP(X = x)$

$E(X) = \sum_{x=1}^{k} x\left(\frac{1}{k}\right)$

$= \frac{k+1}{2}$

$E(X^2) = \sum_{D_x} x^2 P(X=x)$

$E(X^2) = \sum_{x=1}^{k} x^2\left(\frac{1}{k}\right)$

$= \frac{(k+1)(2k+1)}{6}$

$Var(X) = E(X^2) - (E(X))^2$

$= \frac{(k-1)(k+1)}{12}$

Ayrıca $x_1 = a, x_2 = a + 1, \dots\dots, x_k = a + k - 1 = b$ olarak alınırsa

$$E(X) = \frac{a+b}{2}, \qquad Var(X) = \frac{(b-a+2)(b-a)}{12}$$

olur.

Moment Çıkaran Fonksiyon

$$M_X(t) = E(e^{tX})$$
$$=\sum_{x=1}^{k} e^{tx}\left(\frac{1}{k}\right) = \frac{1}{k}\sum_{x=1}^{k} e^{tx}$$
$$= \frac{1}{k}(e^{t} + e^{2t} + \cdots + e^{kt})$$
$$=\frac{1}{k}\left[\frac{e^{t}-e^{(k+1)t}}{1-e^{t}}\right]$$

Olasılık Çıkaran Fonksiyon

$$g_X(s) = E(s^X)$$
$$=\sum_{x=1}^{k} s^{x}\left(\frac{1}{k}\right) = \frac{1}{k}\sum_{x=1}^{k} s^{x}$$
$$= \frac{1}{k}(s + s^2 + \cdots + s^k)$$
$$=\frac{s}{k}\left[\frac{1-s^{k}}{1-s}\right]$$

Örnek : Kesikli bir X tesadüfi değişkeninin olasılık fonksiyonu aşağıdaki gibi veriliyor:

$$p(x) = \begin{cases} c, & x = 1, \ldots.,5 \\ 0, & d.d \end{cases}$$

Buna göre $Var(X)$'ı bulunuz.

Çözüm.

$$\sum_{x=1}^{5} p(x) = 1$$
$$\sum_{x=1}^{5} c = 1$$

olduğundan

$$c = \frac{1}{5}$$

olur.

Dağılım düzgün olduğundan

$$Var(X) = \frac{24}{12} = 2 \qquad dir.$$

Örnek : X tesadüfi değişkeni 10,11, ... ,20 tamsayıları üzerinde düzgün dağılıma sahip ise, bu tesadüfi değişkenin bir çift tamsayı değeri alması olasılığı nedir?

Çözüm.

$$p(x) = \begin{cases} c, & x = 10,11,\ldots.,20 \\ 0, & d.d \end{cases}$$

$$\sum_{x=10}^{20} p(x) = \sum_{x=10}^{20} c = 1$$

$$c = \frac{1}{11}$$

olarak elde edilir.

$$p(x) = \begin{cases} \frac{1}{11}, & x = 10,11,\ldots.,20 \\ 0, & d.d \end{cases}$$

$$P(X = 10) + P(X = 12) + P(X = 14) + P(X = 16) + P(X = 18) + P(X = 20) = \frac{6}{11}$$