HAFTA 13

5.1.6. Hipergeometrik Dağılım

Bir kavanozun içinde iki farklı renkte (sarı ve lacivert diyelim) toplam N tane top bulunduğunu düşünelim. Bunlardan a tanesi sarı geri kalan $N-a$ tanesi lacivert olsun. Kavanozdan aynı anda n tane top çekelim (bir çeşit iadesiz çekiliş yapıyoruz). X bu n tane top içindeki sarı topların sayısı olsun. O zaman, x tane sarı top gelmesi olasılığı (dağılımın olasılık fonksiyonu),

$$P(X=x)=\frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}\ ,\ x=0,1,2,...,n$$

şeklinde olur. Böyle bir deneme için, X rasgele değişkeninin olasılık fonksiyonu bu şekilde verilmiş ise, X e Hipergeometrik dağılıma sahiptir denir ve $X \sim HYP(n,a,N)$ ile gösterilir. Şimdi bu fonksiyonun bir olasılık fonksiyonu olduğunu gösterelim. Bunun için,

$$\sum_{x=0}^{n} P(X=x)=\binom{N}{n}^{-1}\sum_{x=0}^{n}\binom{a}{x}\binom{N-a}{n-x}=1 \quad \text{veya} \quad \sum_{x=0}^{n}\binom{a}{x}\binom{N-a}{n-x}=\binom{N}{n}$$

olduğunun gösterilmesi gerekir. Bunun için tümevarım yöntemini kullanabiliriz. Önce iddianın $n=1$ için doğru olduğunu gösterelim. $n=1$ ise eşitliğin sağ tarafı N dir. Eşitliğin sol tarafı ise,

$$\sum_{x=0}^{1}\binom{a}{x}\binom{N-a}{n-x}=\binom{a}{0}\binom{N-a}{1-0}+\binom{a}{1}\binom{N-a}{1-1}=(N-a)+a=N$$

olup iddia $n=1$ için doğrudur. Şimdi, iddianın n için doğru olduğunu varsayalım ve $n+1$ için doğru olduğunu gösterelim. İşlemlerde, $m=n+1$ denirse $n=m-1$ olup, m.nci terimin eklenip çıkartılması ile,

$$\begin{aligned}\sum_{x=0}^{n+1}\binom{a}{x}\binom{N-a}{n+1-x}&=\sum_{x=0}^{n}\binom{a}{x}\binom{N-a}{n+1-x}+\binom{a}{n+1}\binom{N-a}{0}=\sum_{x=0}^{n}\binom{a}{x}\binom{N-a}{n+1-x}+\binom{a}{n+1}\\&=\sum_{x=0}^{m-1}\binom{a}{x}\binom{N-a}{m-x}+\binom{a}{n+1}=\sum_{x=0}^{m}\binom{a}{x}\binom{N-a}{m-x}-\binom{a}{m}\binom{N-a}{m-m}+\binom{a}{n+1}\\&=\binom{N}{m}-\binom{a}{m}+\binom{a}{n+1}=\binom{N}{m}=\binom{N}{n+1}\end{aligned}$$

sonucu elde edilir. Yani, iddia her n için doğrudur. Başka bir deyişle, verilen fonksiyon bir olasılık fonksiyonudur. Hipergeometrik dağılımın beklenen değer ve varyansının hesabı biraz karmaşıktır. X in beklenen değeri,

$$E(X)=\sum_{x=0}^{n} x\,P(X=x)=\binom{N}{n}^{-1}\sum_{x=0}^{n} x\binom{a}{x}\binom{N-a}{n-x}=\binom{N}{n}^{-1}\sum_{x=1}^{n} x\,\frac{a!}{x!(a-x)!}\binom{N-a}{n-x}$$

$$=\binom{N}{n}^{-1}\sum_{x=1}^{n}\frac{a!}{(x-1)!(a-x)!}\binom{N-a}{n-x}=\binom{N}{n}^{-1}\sum_{x=0}^{n-1}\frac{a(a-1)!}{y!((a-1)-y)!}\binom{N-a}{(n-1)-y}$$

$$=a\binom{N}{n}^{-1}\sum_{x=0}^{n-1}\binom{a-1}{y}\binom{N-a}{(n-1)-y}=a\binom{N}{n}^{-1}\binom{N-1}{n-1}$$

$$=a\,\frac{(N-1)!}{(n-1)!(N-n)!}\,\frac{n!(N-n)!}{N!}=\frac{na}{N}$$

olarak hesaplanmıştır. Benzer şekilde (biraz daha karmaşık olmakla birlikte) rasgele değişkenin varyansı da hesaplanabilir. Hipergeometrik dağılımın beklenen değer ve varyansı sırası ile

$$E(X)=\frac{na}{N}, \qquad Var(X)=\frac{N-n}{N-1}\,\frac{na}{N}\left(1-\frac{a}{N}\right)$$

dir.

$X \sim HYP(n,a,N)$ olsun. Bu durumda $N\to\infty$ ve $a\to\infty$ için p pozitif sabit bir sayı olmak üzere $(a/N)\to p$ ise her bir $x=0,1,2,\ldots,n$ için

$$\lim_{N\to\infty}\binom{N}{n}^{-1}\binom{a}{x}\binom{N-a}{n-x}=\binom{n}{x}p^{x}(1-p)^{n-x} \ , \ x=0,1,2,3,\ldots,n$$

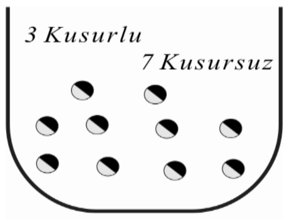
dir (Bain ve Engelhardt, 1992, sayfa 97). Yani, Hipergeometrik dağılım $N\to\infty$ iken Binom dağılımına yaklaşır.

Bazen sonlu elemanlı bir kitleden iadesiz çekiliş yapılmak istenebilir. Bir deneyin sadece iki sonucu varsa (sarı ve lacivert topların bulunduğu bir kavanozdan aynı anda ya da iadesiz çekiliş yapılması), çekilen örnek sayısı n sabit ve denemeler bağımlı (deneylerde birinin sonucu diğerini etkiliyorsa) ise böyle durumlarda Hipergeometrik dağılım kullanılır.

Örnek 5.1.3 Bir kavanozda 3 kusurlu, 7 kusursuz parça vardır. Bu kavanozdan aynı anda 3 parça rasgele seçilsin. Çekilen kusurlu parçaların olasılık fonksiyonunu bulalım. X çekilen kusurlu parçaların sayısı olup, $N=10$, $a=3$, $n=3$ dür.

X in değer kümesi $D_X=\{0,1,2,3\}$ olup, olasılıklar

$$P(X=x)=\frac{\binom{3}{x}\binom{7}{3-x}}{\binom{10}{3}} \ , \ x=0,1,2,3$$



Kavanoz

formülü ile hesaplanır.

X kusurlu parçaların sayısı olmak üzere olasılıklar,

$X=x$	0	1	2	3
$P(X=x)$	$35/120$	$63/120$	$21/120$	$1/120$

şeklinde tablo halide yazılabilir $\oplus$

5.1.7. Kesikli Düzgün Dağılım

Değer kümesi $D_X=\{x_1,x_2,...,x_N\}$ olan X rasgele değişkeninin olasılık fonksiyonu,

$$P(X=x)=\frac{1}{N}\ ,\ x=x_1,x_2,...,x_N$$

şeklinde ise X kesikli düzgün dağılıma sahiptir denir. Dağılımın beklenen değer ve varyansı,

$$E(X)=\frac{N+1}{2}\quad ,\quad Var(X)=\frac{N^2-1}{12}$$

dir.

5.1.8. Poisson Dağılımı

Bir çok deney sürekli bir zaman aralığında yapılır. Böyle bir ortamda gözlenen sonuçlar kesikli olabilir. Birim zaman aralıkları (dakika, saat, gün, ay, yıl gibi) veya birim uzunluk (alan veya hacim gibi) sürekli ortamlardır. Örneğin, bir mağazaya belli bir saat dilimi içinde gelen müşterilerin sayısı, böyle bir deneye örnektir. Bu tür deneylere Poisson deneyleri denir. X sürekli ortamdaki kesikli sonuçların sayısını göstermek üzere, X in (Poisson dağılımının) olasılık fonksiyonu, $\lambda>0$ için,

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

şeklindedir. X rasgele değişkeni bu olasılık fonksiyonuna sahipse, X Poisson dağılımına sahiptir denir ve $X\sim Poisson(\lambda)$ ile gösterilir. e^{λ} fonksiyonunun sıfır noktası komşuluğundaki Taylor serisi açılımı

$$\sum_{x=0}^{\infty}P(X=x)=\sum_{x=0}^{\infty}\frac{e^{-\lambda}\lambda^x}{x!}=e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!}=e^{-\lambda}e^{\lambda}=1$$

olup verilen fonksiyon bir olasılık fonksiyonudur. İkinci ve üçüncü bölümlerde dağılımın ismi verilmeden bu olasılık fonksiyonuna sahip bir rasgele değişkenin momentleri elde edildi. Ayrıca, dağılımın üretici fonksiyonları ve özellikleri ikinci ve üçüncü bölümlerde incelendi. Bunları tekrar hatırlayalım. Dağılımın ilk iki momenti ile varyansı Örnek (2.5.2) de $E(X)=\lambda$, $E(X^2)=\lambda+\lambda^2$ ve $Var(X)=\lambda$ olarak elde edilmişti. Dağılımın moment çıkaran fonksiyonu da $M_X(t)=e^{\lambda(e^t-1)}$ dir (Örnek (2.5.3a)). Örnek (2.6.2) de dağılımın kümülant üreten fonksiyonu, $K_X(t)=\lambda(e^t-1)$ olarak bulunmuş ve bütün kümülantların aynı olduğu ($K_n=\lambda$) görülmüştü. Dağılımın çarpımsal moment üreten fonksiyonu $N_X(t)=e^{\lambda(t-1)}$ olup çarpımsal momentlerin bütün n ler için,

$$\left.\frac{d^n N_X(t)}{dt^n}\right|_{t=1}=\left.\frac{d^n}{dt^n}\left(e^{\lambda(t-1)}\right)\right|_{t=1}=\lambda^n$$

şeklinde olduğunu biliyoruz ((Örnek (2.6.3)). Yani,

$$E(X)=\lambda,\ \ E\big(X(X-1)\big)=\lambda^2,\ \ E\big(X(X-1)(X-2)\big)=\lambda^3,$$
$$E\big(X(X-1)(X-2)(X-3)\big)=\lambda^4,\ ...\big(X(X-1)(X-2)...(X-n+1)\big)=\lambda^n$$

dir.

Binom dağılımı ile Poisson dağılımı arasında bir benzerlik kurulabilir. Örneğin, belirli bir saat dilimi içinde bir mağazaya giren müşterilerin sayısı Poisson rasgele değişkenidir. Oysa, mağazaya giren (maksimum müşteri sayısı sabit) her müşteriyi başarı olarak değerlendirirsek, mağazaya giren müşterilerin sayısı Binom dağılımına sahip olur.

$X \sim Binom(n,p)$ ise dağılımın olasılık fonksiyonunun,

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

olduğunu biliyoruz. Bu dağılımın beklenen değeri $\mu=n\,p$ ($p=\mu/n$) dir. Buradan, $P(X=x)$ olasılığı için

$$P(X=x)=P_n(x)=\binom{n}{x}p^x q^{n-x}=\frac{n!}{x!(n-x)!}\left(\frac{\mu}{n}\right)^x\left(1-\frac{\mu}{n}\right)^{n-x}$$
$$=\frac{n\ (n-1)\ (n-2)\ (n-3)\ \ldots(n-x+1)}{x!\ n^x}\mu^x\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}$$
$$=\frac{1\ \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\left(1-\frac{3}{n}\right)\ldots\left(1-\frac{x-1}{n}\right)}{x!}\mu^x\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}$$

eşitliği elde edilir. Ayrıca, n yeterince büyük ($\mu=n\,p$ sabit kalacak şekilde p sabit) ise,

$$\lim_{n\to\infty}\left[1\ \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\left(1-\frac{3}{n}\right)\ldots\left(1-\frac{x-1}{n}\right)\right]=1$$

ve

$$\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{-x}=1\ ,\ \lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{n}=e^{-\mu}$$

dir. Yani, $X \sim Binom(n,p)$ ve $Y \sim Poisson(\mu)$ olmak üzere, n yeterince büyük ($\mu = n\,p$ sabit kalacak şekilde p sabit) ise

$$P_n(x)=\binom{n}{x}p^x q^{n-x}\ \cong\ \frac{e^{-\mu}\mu^x}{x!}=P(Y=x)$$

yaklaşımı elde edilir. Yani, Binom dağılımına Poisson dağılımı olarak yaklaşılabilir. Bunun tersi de olabilir.

$P(X=1)=\binom{n}{1}\left(\frac{\mu}{n}\right)^1\left(1-\frac{\mu}{n}\right)^{n-1}$							$P(Y=1)=\frac{e^{-\mu}\mu^1}{1!}$
	$n=10$	$n=20$	$n=30$	$n=40$	$n=50$	$n=100$	
$\mu=1$	0.3874	0.3774	0.3741	0.3726	0.3716	0.3697	*0.3679*
$\mu=2$	0.2684	0.2702	0.2705	0.2706	0.2706	0.2707	*0.2707*
$\mu=3$	0.1211	0.1368	0.1413	0.1434	0.1447	0.1471	*0.1494*
$\mu=4$	0.0403	0.0576	0.0631	0.0657	0.0672	0.0703	*0.0733*
$\mu=5$	0.0098	0.0211	0.0253	0.0274	0.0286	0.0312	*0.0337*

Yukarıda, değişik μ değerleri için $Y \sim Poisson(\mu)$ ve $X \sim Binom(n,p)$ olmak üzere, $p=\mu/n$ alınarak $P(X=1)$ ve $P(Y=1)$ olasılıkları hesaplanarak (değerler dördüncü basamaktan sonra yuvarlatılmıştır) tablo halinde verilmiştir. Tablodaki değerlerden de görüldüğü gibi, n değeri büyüdükçe, $P(X=1)$ Binom olasılıkları $P(Y=1)$ olasılıklarına yaklaşmaktadır.

Örnek 5.1.4 $X \sim Poisson(\lambda)$ olsun. $E(X)=Var(X)=\lambda$ olduğunu biliyoruz.

a) X rasgele değişkeninin çarpıklık ve basıklık katsayılarını hesaplayalım. Üçüncü ve dördüncü merkezi momentler $E(X-\lambda)^3=\lambda$ ve $E(X-\lambda)^4=\lambda+3\lambda^2$ olup Poisson dağılımının çarpıklık ve basıklık katsayıları $\gamma=1/\sqrt{\lambda}$ ve $\eta=1/\lambda$ dir (Örnek (2.5.5).

b) $E(g(X))$ ve $g(-1)$ sonlu olacak şekilde herhangi bir g fonksiyonu için, $E(\lambda\,g(X))=E(X\,g(X-1))$ olduğunu gösterelim (Casella ve Berger, 2002, sayfa 126). $E(X\,g(X-1))$ ifadesi $y=x-1$ ve $x/x!=1/(x-1)!$ olduğu göz önüne alındığında,

$$E\big(X\,g(X-1)\big)=\sum_{x=0}^{\infty}xg(x-1)P(X=x)=\sum_{x=1}^{\infty}xg(x-1)\frac{e^{-\lambda}\lambda^x}{x!}$$

$$=\sum_{x=1}^{\infty}g(x-1)\frac{e^{-\lambda}\lambda^x}{(x-1)!}=\sum_{y=0}^{\infty}g(y)\frac{e^{-\lambda}\lambda^{y+1}}{y!}=\sum_{y=0}^{\infty}\lambda g(y)\frac{e^{-\lambda}\lambda^y}{y!}=E\big(\lambda g(X)\big)$$

şeklinde aranan eşitlik elde edilir.

c) Bu sonuç kullanılarak, Poisson dağılımının momentleri ardışık olarak hesaplanabilir. Önce, $E(X)=\lambda$ ve $Var(X)=\lambda$ olduğunu biliyoruz. Aslında, dağılımın ikinci momentinin $E(X^2)=\lambda+\lambda^2$ olduğunu da biliyoruz. $g(x)=x$ denirse ikinci moment

$$\begin{aligned} E(X^2) &= E(XX) = E(X(X-1+1)) = E\big(X(X-1)\big)+E(X) \\ &= E\big(\lambda g(X)\big)+E(X) = E\big(\lambda X\big)+E(X) = \lambda^2+\lambda \end{aligned}$$

olarak da hesaplanır. Benzer şekilde, $g(x)=x^2$ için dağılımın üçüncü momenti de

$$\begin{aligned} E(X^3) &= E(XX^2) = E(X(X-1+1)^2) = E(X(X-1)^2)+2E\big(X(X-1)\big)+E(X) \\ &= E\big(\lambda g(X)\big)+2E\big(X(X-1)\big)+E(X) = E(\lambda X^2)+2\lambda^2+\lambda \\ &= \lambda(\lambda^2+\lambda)+2\lambda^2+\lambda = \lambda^3+3\lambda^2+\lambda \end{aligned}$$

şeklinde hesaplanır $\oplus$

Bu kısma son vermeden önce, burada verilen dağılımlar arasındaki ilişkileri özetleyelim.

1) $X_1, X_2, \ldots, X_n$ bağımsız Bernoulli dağılımına sahip rasgele değişkenler ise, $X = X_1 + X_2 + \ldots + X_n \sim \text{Binom}(n,p)$ dir.

2) $X_i,\ i=1,2,...,k$ için bağımsız ve $X_i \sim Binom(n_i, p)$ ise $n = n_1+n_2+...+n_k$ için, $X = X_1+X_{2+}...+X_k \sim Binom(n, p)$ dir.

3) $i=1,2,...,k$ için bağımsız ve $X_i \sim Geo(p)$ ise $X = X_1+...X_k \sim NB(k, p)$ dir.

4) $X_i,\ i=1,2,...,k$ için bağımsız ve Poisson(λ_i) ise, $\lambda=\lambda_1+...+\lambda_k$ olmak üzere $X = X_1+...+X_k \sim Poisson(\lambda)$ dır.