2. ÖRNEKLEME DAĞILIMLARI

2.1. İSTATİSTİKTE KULLANILAN BAZI EŞİTSİZLİKLER

Olasılık ve istatistikte eşitsizlikler önemli bir yer tutar. Bazen, olasılıkların ve rasgele değişkenin momentlerinin hesaplanması zor olabilir. Böyle durumlarda, olasılık ve momentler için bir alt veya üst sınır verilir.

2.1.1.Markov Eşitsizliği

X bir rasgele değişken g de $g:\mathbb{R}\rightarrow\mathbb{R}^{+}$ tanımlı negatif değerler almayan bir fonksiyon olmak üzere, $c>0$ için

$$P\big(g(X)>c\big) \leq \frac{E\big(g(X)\big)}{c}$$

dir.

İspat: X sürekli olsun (kesikli durumda integral yerine toplam gelir). X in olasılık yoğunluk fonksiyonu $f(x)$ olmak üzere, $g(X)$ in beklenen değeri

$$\begin{aligned} E\big(g(X)\big) &= \int_{-\infty}^{\infty} g(x)\,f(x)dx = \int_{\{g(x)>c\}} g(x)\,f(x)dx + \int_{\{g(x)\leq c\}} g(x)\,f(x)dx \\ &\geq \int_{\{g(x)>c\}} g(x)\,f(x)dx \geq c \int_{\{g(x)>c\}} f(x)dx = c\,P(g(X)>c) \end{aligned}$$

şeklinde bir eşitsizlik elde edilir. Burada, $g(x)$ negatif olmayan değerler aldığından,

$$g(x)>c \Rightarrow \int_{\{g(x)>c\}} g(x)\,f(x)dx \geq c \int_{\{g(x)>c\}} f(x)dx$$

dir. Buradan da, $c>0$ için $E(g(X)) \geq c\,P(g(X)>c)$ şeklinde aranan eşitsizlik ispat edilmiş olur.

2.1.2.Chebyshev Eşitsizliği

X rasgele değişkeninin beklenen değeri μ, varyansı sonlu σ^2 olsun. $c > 0$ olmak üzere,

$$P(|X-\mu|> c) \leq \frac{Var(X)}{c^2}$$

dir.

İspat: $Y=(X-\mu)^2$ olarak alınsın. Bu durumda Y negatif değerler almayan bir rasgele değişkendir ve $E(Y)=E(X-\mu)^2=Var(X)$ olaur. Buna göre Markov eşitsizliğinden,

$$P(|X-\mu|> c) = P(Y > c^2) \leq \frac{E(Y)}{c^2} = \frac{Var(X)}{c^2}$$

$c = k\sigma$ alınırsa $c^2 = k^2\sigma^2$ olur. Buradan,

$$P(|X-\mu|> k\sigma) \leq 1/k^2$$

$$P(|X-\mu|< k\sigma) \geq 1-1/k^2$$

elde edilir. Chebyshev eşitsizliği genellikle olasılıklar için bir alt veya üst sınır belirlemek için kullanılır.

Chebyshev eşitsizliği bir rasgele değişkenin kendi ortalaması komşuluğunda bulunması olasılığı için,

$$P\left(\mu_X - k\sigma_X < X < \mu_X + k\sigma_X\right) \geq 1 - \frac{1}{k^2}$$

gibi bir sınır değer belirlemektedir.

Örneğin, varyansı var olan her hangi bir olasılık dağılımında, Chebyshev eşitsizliğine göre

$$P(\mu_X - 2\sigma_X < X < \mu_X + 2\sigma_X) \geq 1 - \frac{1}{2^2} = 0.75$$

dır.

Örneğin, $X \sim N(\mu_X, \sigma_X^2)$ olduğunda,

$$P(\mu_X - 2\sigma_X < X < \mu_X + 2\sigma_X) = 0.9544$$

dır.

Örnek: X rasgele değişkeninin beklenen değerinin $\mu = 10$ ve standart sapmasının $\sigma = 5$ olduğu biliniyor. Buna gore aşağıdaki olasılıklar için bir alt ve üst sınır belirleyiniz.

a) $P(X \leq 0 \text{ veya } X \geq 20) = ?$

b) $P(-5 < X < 25) = ?$

a) $P(|X - 10| > k\sigma) \leq 1/k^2$

$\sigma = 5$

$k = 2$ olduğunda $P(X \leq 0 \text{ veya } X \geq 20)$ durumu sağlanır. Buna göre,

$P(|X - 10| > 2 \times 5) \leq 1/4$

şeklinde bir üst sınır belirlenir.

b) Chebychev eşitsizliğinden $P(|X - \mu| < k\sigma) \geq 1 - 1/k^2$ olduğundan,

$P(|X - 10| < 3 \times 5) \geq 1 - 1/3^2$ olup

$P(-5 < X < 25) \geq 1 - 1/9 \cong 0.89$

biçiminde bir alt sınır elde edilir.

2.2. ZAYIF BÜYÜK SAYILAR KANUNU

$X_1, X_2, \ldots, X_n, \ldots$ bağımsız ve aynı μ ortalamalı $\sigma^2 < \infty$ varyanslı rasgele değişkenler ise $\bar{X} \xrightarrow{P} \mu$ dır.

Yani, $X_1, X_2, \ldots, X_n$ bağımsız rasgele değişkenler, $E(X_i) = \mu$, $Var(X_i) = \sigma^2$, $i = 1, 2, \ldots, n$ olmak üzere,

$$\lim_{n \to \infty} P\left(| \bar{X}_n - \mu | < \varepsilon \right) = 1$$

dir ($\bar{X}_n$ olasılıkta μ 'ye yakınsar). Bu Zayıf Büyük Sayılar Kanunu olarak bilinir. Bunun böyle olduğu gösterilsin.

$X_1, X_2, \ldots, X_n$ ‘lerin ortalaması,

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

olmak üzere, $E\left(\bar{X}_n \right) = \mu$ ve $Var(\bar{X}_n) = \frac{\sigma^2}{n}$ olup, Chebyshev eşitsizliğinden,

$$P\left(| \bar{X}_n - \mu_{\bar{X}_n} | < k\sigma_{\bar{X}_n} \right) = P\left(| \bar{X}_n - \mu | < k \frac{\sigma}{\sqrt{n}} \right) \geq 1 - \frac{1}{k^2}$$

dir. $k = \frac{\varepsilon \sqrt{n}}{\sigma}$ olarak alınırsa $\varepsilon = k \frac{\sigma}{\sqrt{n}} > 0$ için

$$P\left(| \bar{X}_n - \mu | < \varepsilon \right) \geq 1 - \frac{1}{\left(\frac{\varepsilon \sqrt{n}}{\sigma} \right)^2}$$

biçiminde elde edilir. Eşitsizliğin her iki tarafının $n \to \infty$ için limiti alınırsa,

$$\lim_{n\to\infty} P\left(|\bar{X}_n-\mu|<\varepsilon\right) \geq \lim_{n\to\infty}\left(1-\frac{1}{\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)^2}\right)$$

olup olasılık birden büyük olamayacağından,

$$\lim_{n\to\infty} P\left(|\bar{X}_n-\mu|<\varepsilon\right)=1$$

elde edilir.

Örnek. $X_1, X_2..., X_n,...$ düzgün bir paranın ard arda atılışında gelen tura sayısını (0 ya da 1 tane) göstermek üzere $X_i \sim b(1, p=\frac{1}{2})$, $i=1,2,3,...$ dağılımına sahip olur. Zayıf Büyük Sayılar Kanunu'na göre

$$\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{n \text{ denemedeki başarı sayısı}}{n} \xrightarrow[n\to\infty]{P} p$$

dir. İstenildiği kadar küçük $\varepsilon > 0$ değeri için,

$$\lim_{n\to\infty} P\left(|\bar{X}_n - \frac{1}{2}|<\varepsilon\right)=1$$

dır.Yani, düzgün bir para atıldıkça gelen tura sayısının ortalaması 1/2 değerine yakınsar. Zayıf Büyük Sayılar Kanunu'nun bu özel hali Bernoulli Büyük Sayılar Kanunu olarak bilinir.

2.3 MERKEZI LIMIT TEOREMI

$X_1, X_2,..., X_n,...$ bağımsız ve aynı dağılıma (aynı beklenen değer μ, varyans $\sigma^2<\infty$) sahip rasgele değişkenler olmak üzere,

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - E(\sum_{i=1}^{n} X_i)}{\sqrt{Var(\sum_{i=1}^{n} X_i)}} \le t \right) = \int_{-\infty}^{t} \frac{1}{\sqrt{2n}} e^{-z^2/2} dz$$

dır. Burada, $Z \sim N(0,1)$ dir.

Büyük *n* ler için,

$$P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le t \right) \approx \int_{-\infty}^{t} \frac{1}{\sqrt{2n}} e^{-z^2/2} dz$$

yani, $Z \sim N(0,1)$ olmak üzere

$$P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le t \right) \approx P(Z \le t)$$

dır. Bu şu demektir, yaklaşık olarak yapılabilen olasılık hesaplarında, büyük *n* ler için, $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ 'nin dağılım fonksiyonu yerine standart normal dağılımın dağılım fonksiyonu alınabilir. Hattâ $\bar{X}_n$ nin dağılım fonksiyonu yerine $N\left(\mu, \frac{\sigma^2}{n}\right)$ normal dağılımın dağılım fonksiyonu alınabilir.

Örnek: Düzgün bir tavla zarının 100 kez atılışında gelen nokta sayılarının toplamının [340,360] aralığında olması olasılığı nedir?

$P(340 \le \sum_{i=1}^{100} X_i \le 360) = ?$

X_i	1	2	3	4	5	6
$P(X = x_i)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

$E(X_i) = 7/2 = 3.5$

$Var(X_i) = 35/12$

olmak üzere, kesikli dağılımlarda Merkezi Limit Teoremini kullanabilmek için süreklilik düzeltmesi denen aşağıdaki işlem yapılmaktadır.

$$P(340 \le \sum_{i=1}^{100} X_i \le 360) = P(339.5 \le \sum_{i=1}^{100} X_i \le 360.5)$$

$$= P\left(\frac{339.5 - E(\sum_{i=1}^{100} X_i)}{\sqrt{Var(\sum_{i=1}^{100} X_i)}} \le \frac{\sum_{i=1}^{n} X_i - E(\sum_{i=1}^{n} X_i)}{\sqrt{Var(\sum_{i=1}^{n} X_i)}} \le \frac{360.5 - E(\sum_{i=1}^{100} X_i)}{\sqrt{Var(\sum_{i=1}^{100} X_i)}}\right)$$

$$= P\left(\frac{339.5 - 100 \times 3.5}{\sqrt{100 \times \frac{35}{12}}} \le Z \le \frac{360.5 - 100 \times 3.5}{\sqrt{100 \times \frac{35}{12}}}\right)$$

$$= P(-0.6148 \le Z \le 0.6148) = 2P(0 < Z < 0.6148)$$

$$= 2(0.7317 - 0.5) = 0.4634$$

biçiminde bulunur. Düzgün bir tavla zarının n=100 atılışında gelen nokta sayısı toplamının 350 olması olasılığı nedir sorusu sorulduğunda süreklilik düzeltmesine göre,

$$P(\sum_{i=1}^{100} X_i = 350) = P(350 - 0.5 \le \sum_{i=1}^{100} X_i \le 350 + 0.5)$$

$$\approx P\left(\frac{349.5 - 100 \times 3.5}{\sqrt{100 \times \frac{35}{12}}} \le Z \le \frac{350.5 - 100 \times 3.5}{\sqrt{100 \times \frac{35}{12}}}\right)$$

$$= P\left(-0.05 \times \sqrt{\frac{12}{35}} \le Z \le 0.05 \times \sqrt{\frac{12}{35}}\right)$$

$$= P(-0.0293 \le Z \le 0.0293)$$

$$= 0.0234$$

değeri bulunur.