HAFTA 3

TAHMİN PROBLEMİNE GİRİŞ

Buraya kadar, rasgele değişkenlerin dağılım fonksiyonları ile olasılık fonksiyonlarının yanında bazı moment özellikleri ele alındı. Rasgele değişkenlerin dönüşümleri, bu dönüşümlerin momentlerine ilişkin bazı eşitsizlikler ve uygulamada çok karşılaşılan bazı özel dağılımlar incelendi.

Hemen hemen bütün bilim dallarında olduğu gibi, istatistiğin de esas amaçlardan biri üzerinde çalışılan kitle hakkında bilgi sahibi olmaktır. Başka bir ifade ile kitlenin bilinmeyenleri (*parametreler*) hakkında bilgi sahibi olmaktır. Yani, kitlenin parametreleri hakkında tahminlerde bulunmaktır. Kitle parametresi θ ile gösterilecektir. Kitlenin $\theta_1, \theta_2 \ldots, \theta_k$ gibi k tane parametresi varsa, bu da $\underset{\sim}{\theta}$ ile gösterilecektir. Bundan sonraki bölümlerde, parametrelere ilişkin tahmin yöntemleri üzerinde durulacaktır. Bu tahminler, gerçekleşmesi tamamen raslantıya bağlı olan deneylerin sonucuna bağlıdır. Tek bir deney ile istatistiki sonuç çıkarım yapmak anlamlı değildir. Dolayısı ile deneyler tekrar edilir. Bir deney aynı koşullarda tekrarlandığında, aynı sonuç gözlenmeyebilir.

6.1. Örneklem ve Özellikleri

Kitle parametrelerini tahmin etmek için deneyler tekrarlanır. Her tekrarda rasgele değişkenin değeri gözlenir. Ancak, deneylerin belli kurallara göre yapılması gerekir. Deneyler birbirinden bağımsız olarak yapılabildiği gibi, bağımlı da olabilir. Her deney için rasgele değişkenin dağılımı da farklı olabilir. Burada, aksi söylenmedikçe, denemelerin birbirinden bağımsız olduğu varsayılacaktır. Yani, rasgele değişkenler birbirinden bağımsız aynı dağılıma sahiptir.

Tanım 6.1.1 Birbirinden bağımsız aynı dağılıma sahip $X_1, X_2 \ldots, X_n$ rasgele değişkenlerine bir *örneklem* denir $\otimes$

$X_1, X_2, \ldots, X_n$ olasılık veya olasılık yoğunluk fonksiyonu $f(x;\theta)$ olan kitleden bir örneklem olsun. Bu durumda örneklemin ortak olasılık yoğunluk fonksiyonu,

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n}(x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

şeklinde yazılabilir. $X_1, X_2 \ldots, X_n$ beklenen değeri θ olan üstel dağılımdan bir örneklem ise $P(X_1 > 2, X_2 > 2 \ldots, X_n > 2)$ olasılığı,

$$P(X_1 > 2, X_2 > 2 \ldots, X_n > 2) = \prod_{i=1}^{n} P(X_i > 2) = [P(X_i > 2)]^n = [e^{-2/\theta}]^n = e^{-2n/\theta}$$

şeklinde hesaplanır.

Örneklem, genellikle sonsuz elemanlı bir kitleden alınır. Ancak, örnek uzayın sonlu elemanlı olduğu durumlarda örnekleme kurallarına dikkat edilmelidir. Sonsuz elemanlı bir kitleden örneklem alındığında, X_2 rasgele değişkeninin değeri X_1 rasgele değişkeninin alacağı değerden etkilenmez. Örnek uzay sonlu elemanlı ise denemeler birbirine bağımlı olabilir. Örneğin, içinde belli sayıda sarı ve mavi topların bulunduğu bir kavanozdaki sarı topların oranı hakkında bilgi sahibi olmak istiyorsak kavanozdan çekilen topun tekrar yerine konması (iadeli çekiliş) ile belli sayıda top çekilir. Buna göre, kavanozun içindeki sarı topların oranı aynıdır. Çekilen top yerine konulmazsa ikinci çekilen topun sarı gelmesi olasılığı birinci çekilen topun sarı veya mavi olmasına bağlıdır.

Tanım 6.1.2 Örneklemin herhangi bir fonksiyonuna *tahmin edici* (veya *istatistik*) denir$\otimes$

$X_1, X_2 \ldots, X_n$ bir örneklem olsun. Örneklemin $T = T(X_1,\ldots,X_n)$ gibi herhangi bir fonksiyonu da (yani tahmin edici) bir rasgele değişkendir. Bu rasgele değişken sadece örneklemin bir fonksiyonu olup parametreye bağlı değildir. Tahmin edicinin aldığı değere bir *tahmin* denir.

Örnek 6.1.1 $X_1, X_2 \ldots, X_n$, $n-$birimlik bir örneklem olsun. Buna göre,

$$T_1(\underset{\sim}{X}) = T_1(X_1,\ldots,X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n, \text{ örneklem ortalaması}$$

$$T_2(\underset{\sim}{X}) = T_2(X_1,\ldots,X_n) = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 = S_n^2, \text{ örneklem varyansı}$$

$$T_3(\underset{\sim}{X}) = T_3(X_1,\ldots,X_n) = \max_{1\le i\le n}\{X_i\} = X_{(n)}, \quad T_4(\underset{\sim}{X}) = T_4(X_1,\ldots,X_n) = \min_{1\le i\le n}\{X_i\} = X_{(1)}$$

birer tahmin edicidir. Ayrıca, örneklem değerleri $n = 5$ için,

$$X_1(w) = x_1 = 5,\ X_2(w) = x_2 = 3,\ X_3(w) = x_3 = 2,\ X_4(w) = x_4 = 1,\ X_5(w) = x_5 = 4$$

olarak gözlenmiş ise yukarıdaki tahmin ediciler için,

$$\bar{X}_n(w) = \bar{x}_n = \frac{5+3+2+1+4}{5} = 3$$

$$S_n^2(w) = s_n^2 = \frac{1}{4}\sum_{i=1}^{5}\left(x_i - \bar{x}_n\right)^2 = \frac{1}{4}\left[(5-3)^2 + ((3-3)^2 + (2-3)^2 + (1-3)^2 + (4-3)^2\right] = 2.5$$

$$X_{(1)}(w) = x_{(1)} = \min_{1\le i\le 5} x_i = 1\ , \quad X_{(5)}(w) = x_{(5)} = \max_{1\le i\le 5} x_i = 5$$

birer tahmindir$\oplus$

Teorem 6.1.1 $x_1, x_2 \ldots, x_n \in \mathbb{R}$ olmak üzere,

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i\,, \qquad s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2$$

olsun. Buna göre,

a) $(n-1)s_n^2 = \sum_{i=1}^{n} x_i^2 - n\,\bar{x}_n^2$ b) $\min_{a\in\mathbb{R}} \sum_{i=1}^{n}(x_i-a)^2 = \sum_{i=1}^{n}(x_i-\bar{x}_n)^2$

dir.

İspat a) s_n^2 açık olarak yazıldığında

$$(n-1)\,s_n^2 = \sum_{i=1}^{n}\left(x_i-\bar{x}_n\right)^2 = \sum_{i=1}^{n}\left[\,x_i^2 - 2\,\bar{x}_n x_i + \bar{x}_n^2\,\right] = \sum_{i=1}^{n} x_i^2 - 2\,\bar{x}_n\ \sum_{i=1}^{n} x_i +\ n\ \bar{x}_n^2$$

$$= \sum_{i=1}^{n} x_i^2 - 2\ n\ \bar{x}_n\left(\frac{1}{n}\ \sum_{i=1}^{n} x_i\right) +\ n\ \bar{x}_n^2 = \sum_{i=1}^{n} x_i^2 - 2n\ \bar{x}_n^2 + n\ \bar{x}_n^2 = \sum_{i=1}^{n} x_i^2 - n\ \bar{x}_n^2$$

şeklinde aranan sonuç elde edilir.

b) $f(a) = \sum_{i=1}^{n}(x_i-a)^2$ fonksiyonunu tanımlayalım ve fonksiyonu minimum yapan a değerini bulalım. $f(a)$ fonksiyonu a ya göre türevlenebilir olup, fonksiyonun minimumunu bulmak için birinci türev sıfıra eşitlenir. Birinci türevin sıfır olduğu yerde, fonksiyon minimum ya da maksimum değerini alır. Bu noktadaki fonksiyon değerinin minimum olması için fonksiyonun o noktadaki ikinci türevinin pozitif olmalıdır. Buna göre,

$$\frac{d\ f(a)}{dx} = -2\sum_{i=1}^{n}\left(x_i - a\right) = -2\sum_{i=1}^{n} x_i + 2\,n\,a = 0$$

eşitliğinden $a=\bar{x}_n$ bulunur. Yani, $f(a)$ fonksiyonu $a=\bar{x}_n$ noktasında minimum ya da maksimuma sahiptir. Fonksiyonun bu noktadaki ikinci türevi,

$$\left.\frac{d^2 f(a)}{d\,x^2}\right|_{a=\bar{x}_n} =\ 2n > 0$$

olduğundan fonksiyon, $a=\bar{x}_n$ noktasında minimuma sahiptir. Yani,

$$\min_{a\in\mathbb{R}} \sum_{i=1}^{n}\left(x_i-a\right)^2 = \sum_{i=1}^{n}\left(x_i-\bar{x}_n\right)^2$$

dir ◊

Teorem 6.1.2 $X_1, X_2 \ldots, X_n$ beklenen değeri μ varyansı σ^2 olan kitleden bir örneklem olsun. Buna göre,

a) $E(\bar{X}_n) = \mu$ b) $Var(\bar{X}_n) = \sigma^2 / n$ c) $E(S_n^2) = \sigma^2$

dir.

İspat a) Beklenen değer operatörünün lineer olmasından dolayı aranan sonuç

$$E\left(\bar{X}_n\right)=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E\left(X_i\right)=\frac{1}{n}\sum_{i=1}^{n}\mu=\mu$$

şeklinde elde edilir.

b) Benzer şekilde,

$$Var\left(\bar{X}_n\right)=Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}Var\left(X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}\sigma^2=\frac{\sigma^2}{n}$$

dir.

c) S_n^2 Teorem (6.1.1a) daki gibi yazıldığında, $E(S_n^2)$ değeri

$$E(S_n^2)=E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2-n\bar{X}_n^2\right)$$

$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}E(X_i^2)-nE(\bar{X}_n^2)\right]=\frac{1}{n-1}\left[\sum_{i=1}^{n}(\mu^2+\sigma^2)-n(\mu^2+\sigma^2/n)\right]$$

$$=\frac{1}{n-1}\left[n(\mu^2+\sigma^2)-(n\mu^2-\sigma^2)\right]=\frac{(n-1)\sigma^2}{(n-1)}=\sigma^2$$

şeklinde bulunur ◊

6.2. Rasgele Değişken Dizilerinde Yakınsamalar

Rasgele değişken dizilerinde yakınsamalara geçmeden önce, reel sayı dizilerindeki yakınsama kavramını hatırlayalım. Elemanları reel sayılar olan bir dizi a_n olsun. Yani, her $n \in \mathbb{N}$ için $a_n \in \mathbb{R}$ dir.

Tanım 6.2.1 a) a_n elemanları reel sayılar olan bir dizi ve $a \in \mathbb{R}$ olsun. Her $\varepsilon > 0$ için bir $n_0(\varepsilon)$ sayısı var ve $n > n_0(\varepsilon)$ için $|a_n - a| < \varepsilon$ oluyorsa, a_n dizisi a *noktasına yakınsıyor* denir ve $a_n \to a$ (veya $\lim\limits_{n\to\infty} a_n = a$) ile gösterilir.

b) $a_n \to 0$ ise, $a_n = o(1)$ ve $(a_n / f_n) \to 0$ ise $a_n = o(f_n)$ dir.

c) Sonlu bir M sayısı için M ye bağlı bir n_M sayısı varsa ve her $n > n_M$ için $|a_n| < M$ oluyorsa a_n dizisi *sınırlıdır* denir ve $a_n = O(1)$ ile gösterilir. Benzer şekilde, $(a_n / g_n) = O(1)$ ise $a_n = O(g_n)$ dir ⊗

Örnek 6.2.1 $a_n = (3n+2)/n^2$ dizisi için, $n \to \infty$ iken $a_n \to 0$ olup $a_n = o(1)$ dir. Ayrıca,

$$\frac{a_n}{(1/n)} = n\,a_n = \frac{3n+2}{n} = 3 + \frac{2}{n}$$

olup her $n \in \mathbb{N}$ için $\left|\frac{a_n}{(1/n)}\right| = \left|3 + \frac{2}{n}\right| < 6$ yazılabilir. Yani, $a_n / (1/n) = n\,a_n = O(1)$ veya $a_n = O(1/n)$ dir$\oplus$

Reel sayı dizilerindeki yakınsama ve sınırlılık kavramlarının ayrıntılarına fazla girmeden, yukarıdaki kavramları rasgele değişken dizileri için yazalım. Burada, rasgele değişken dizilerinde birkaç yakınsama türü kısaca incelenecektir. Bunlardan, olasılıkta yakınsama, tahmin edicilerin tutarlılık özelliğinin gösterilmesi açısından önemlidir. Tahmin edicilerin asimptotik dağılımlarını bulmak için de dağılımda yakınsama kavramı önemlidir.

Tanım 6.2.2 (*Olasılıkta Yakınsama*): $n \in \mathbb{N}$ için X_n rasgele değişkenlerin herhangi bir dizisi olsun. Her $\varepsilon > 0$ için $\exists \delta > 0$ ve $\exists\, n_0(\varepsilon)$ öyle ki $n > n_0(\varepsilon)$ için X_n dizisi

$$P(\{w : |\, X_n(w) - X(w)\,| > \delta\}) < \varepsilon$$

özelliğini sağlanıyorsa, X_n rasgele değişkenlerinin dizisi X rasgele değişkenine *olasılıkta yakınsıyor* denir ve $X_n \xrightarrow{P} X$ şeklinde gösterilir$\otimes$

X_n rasgele değişken dizisinin X rasgele değişkenine olasılıkta yakınsaması genellikle,

$$\lim_{n\to\infty} P(\{w : |\, X_n(w) - X(w)\,| > \varepsilon\}) = 0 \quad \text{(veya kısaca } \lim_{n\to\infty} P(|\, X_n - X\,| > \varepsilon) = 0\text{)}$$

şeklinde ifade edilir. Ayrıca, reel sayı dizilerinde olduğu gibi, $X_n \xrightarrow{P} 0$ ise $X_n = o_P(1)$ ve f_n bir reel sayı dizisi olmak üzere $X_n / f_n = o_P(1)$ ise $X_n = o_P(f_n)$ yazılır.

Tanım 6.2.3 (*Olasılıkta Sınırlılık*) Her $n \in \mathbb{N}$ için X_n rasgele değişkenlerin herhangi bir dizisi olsun. Her $\varepsilon > 0$ sayısı için öyle M_ε ve N_ε sayıları varsa ve $n > N_\varepsilon$ için,

$$P(\{w : |\, X_n(w)\,| > M_\varepsilon\}) < \varepsilon$$

özelliğini sağlıyorsa, X_n rasgele değişkenlerinin dizisi olasılıkta sınırlıdır denir ve $X_n = O_P(1)$ ile gösterilir$\otimes$

Reel sayı dizilerinde olduğu gibi, f_n bir reel sayı dizisi olmak üzere $X_n / f_n = O_P(1)$ ise $X_n = O_P(f_n)$ dir. Olasılıkta sınırlılık için genellikle $P(\{w : |\, X_n(w)\,| > M_\varepsilon) < \varepsilon$ yerine $P(|\, X_n\,| > M_\varepsilon) < \varepsilon$ veya $\lim_{n\to\infty} P(|\, X_n\,| > M_\varepsilon) = 0$ gösterimleri kullanılır.

Örnek 6.2.2 X_n , $n = 1, 2, 3, ...$ beklenen değeri μ , varyansı σ^2 olan bağımsız aynı dağılımlı rasgele değişkenler olsun. Örneklem ortalaması kitle ortalamasına olasılıkta yakınsar. Yani, $n \to \infty$ iken $\bar{X}_n \xrightarrow{P} \mu$ dir. Burada, $X_1, X_2, \ldots, X_n$ rasgele değişkenleri bağımsız olmasına rağmen $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_n$ rasgele değişkenleri bağımsız değildir ($n = 1$ için $\bar{X}_n = X_1$, $n = 2$ ve $\bar{X}_n = (X_1 + X_2)/2$ olup $\bar{X}_1$ ile $\bar{X}_2$ bağımsız değildir) dir. Olasılıkta yakınsama Chebyshev eşitsizliğinden

$$0 \le P(|\bar{X}_n - \mu| > \varepsilon) \le \frac{E(\bar{X}_n - \mu)^2}{\varepsilon^2} = \frac{Var(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0 \ , \ n \to \infty$$

şeklinde elde edilir. Yani, $n \to \infty$ iken $\bar{X}_n \xrightarrow{P} \mu$ dir. Buradan $n \to \infty$ iken $\bar{X}_n - \mu \xrightarrow{P} 0$ olup, $\bar{X}_n - \mu = o_P(1)$ de yazılabilir. Diğer taraftan,

$$\frac{(\bar{X}_n - \mu)}{1/\sqrt{n}} = \sqrt{n}\,(\bar{X}_n - \mu) = O_P(1)$$

olduğundan, $\bar{X}_n = \mu + O_P(1/\sqrt{n})$ şeklinde de yazılabilir. Bu sonuç, olasılık ve istatistikte çok kullanılan *zayıf büyük sayılar kanunudur (WLLN)*. Zayıf büyük sayılar kanunu açık olarak " $X_1, X_2, \ldots, X_n \ldots$ beklenen değeri μ , varyansı σ^2 olan bağımsız rasgele değişkenler olmak üzere, örneklem ortalaması kitle ortalamasına olasılıkta yakınsar" şeklinde ifade edilir. Benzer şekilde, örneklem varyansı da kitle varyansına olasılıkta yakınsar ($n \to \infty$ iken $S_n^2 \xrightarrow{P} \sigma^2$). Bunu göstermek için Chebyshev eşitsizliği

$$P(|\bar{X}_n - \mu| > \varepsilon) \le \frac{E(S_n^2 - \sigma^2)^2}{\varepsilon^2} = \frac{Var(S_n^2)}{\varepsilon^2}$$

şeklinde yazılır. Olasılıkta yakınsamanın sağlanabilmesi (yani, $S_n^2 \xrightarrow{P} \sigma^2$) için $n \to \infty$ iken $Var(S_n^2) \to 0$ olmalıdır. Yani, $Var(S_n^2)$ değerinin hesaplanması gerekir. S_n^2 in varyansı bazı dağılımlar için kolay hesaplanabilmesine rağmen, genellikle zor ve karmaşıktır $\oplus$

Tanım 6.2.4 (*Hemen hemen her yerde yakınsama*) X_n rasgele değişkenlerin bir dizisi olsun. X_n rasgele değişkenlerinin dizisi her $\varepsilon > 0$ için,

$$P\left(\lim_{n \to \infty} \left\{ w : \left| X_n(w) - X(w) \right| > \varepsilon \right\} \right) = 0$$

özelliğini sağlıyorsa, X_n hemen hemen her yerde X rasgele değişkenine yakınsar denir ve $X_n \xrightarrow{hhhy} X$ ile gösterilir⊗

Örnek 6.2.3 X_n , $n = 1, 2, 3, ...$ beklenen değeri μ, varyansı σ^2 olan bağımsız aynı dağılımlı rasgele değişkenler olsun. Örneklem ortalaması hemen hemen her yerde kitle ortalamasına yakınsar. Yani, $n \to \infty$ iken $\bar{X}_n \xrightarrow{hhhy} \mu$ dir. Bu yakınsama, literatürde *güçlü büyük sayılar kanunu* olarak bilinir⊕

Tanım 6.2.5 (*Momentlerde yakınsama*) X_n rasgele değişkenlerin bir dizisi, X de herhangi bir rasgele değişken olsun. $n \to \infty$ iken $E(X_n - X)^r \to 0$ ise X_n rasgele değişkenlerinin dizisi X rasgele değişkenine r. momentte yakınsar denir ve $X_n \xrightarrow{r} X$ ile gösterilir⊗

X_n rasgele değişkenlerinin dizisi X rasgele değişkenine momentlerde yakınsıyor ise, olasılıkta da yakınsar. Ancak, bunun tersi her zaman doğru değildir.

Örnek 6.2.4 Chebyshev eşitsizliğinden $X_n \xrightarrow{r} X$ ise $X_n \xrightarrow{P} X$ olduğu görülür. Her $\varepsilon > 0$ için Chebyshev eşitsizliği

$$0 \leq P(|X_n - X| > \varepsilon) \leq \frac{E(X_n - X)^r}{\varepsilon^r}$$

şeklinde olup $X_n \xrightarrow{r} X$ ise $E(X_n - X)^r \to 0$ dir. Yani $n \to \infty$ iken $P(|X_n - X| > \varepsilon) \to 0$ dir. X_n dizisi momentlerde X rasgele değişkenine yakınsıyorsa olasılıkta da yakınsar. Ancak bunun tersi doğru değildir. Bunun için X_n rasgele değişkenlerini,

$$P(X_n = n^3) = 1/n^2 \quad \text{ve} \quad P(X_n = 0) = 1 - (1/n^2)$$

olacak şekilde Bernoulli resgele değişkenlerinin bir dizisi olarak seçelim. Yani,

$$X_n = \begin{cases} n^3 & , \quad n^{-2} \text{ olasılıkla} \\ 0 & , \quad 1 - n^{-2} \text{ olasılıkla} \end{cases}$$

olarak verilmiş olsun. Buradan, her $\varepsilon > 0$ için $n \to \infty$ iken $P(|X_n| > \varepsilon) = n^{-2} \to 0$ olduğundan $X_n \xrightarrow{P} 0$ dır. Yani, olasılıkta yakınsama gerçekleşir. Olasılıkta yakınsama gerçekleşmiş olmasına rağmen, $n \to \infty$ iken $r \geq 1$ için

$$E(X_n - 0)^r = E(X_n)^r = 0^r P(X = 0) + n^{3r} P(X = n^3) = n^{3r} n^{-2} = n^{3r-2} \to \infty$$

dir. Yani, momentlerde yakınsama gerçekleşmez⊕

Tanım 6.2.6 (*Dağılımda yakınsama*) X_n dağılım fonksiyonları F_n olan rasgele değişkenlerin bir dizisi olsun. Dağılım fonksiyonu F olan X rasgele değişkeni için, F nin sürekli olduğu yerlerde $n \to \infty$ iken $F_n(x) \to F(x)$ ise X_n rasgele değişkenlerinin dizisi *dağılımda* X *rasgele değişkenine yakınsıyor* denir ve $X_n \xrightarrow{\mathcal{D}} X$ ile gösterilir$\otimes$

Dağılımda yakınsamaya en iyi örnek merkezi limit teoremi (MLT) ve uygulamalarıdır. Bu uygulamalara geçmeden önce, bilerek veya bilmeyerek bilimin her alanında uygulamalarına rastlanan merkezi limit teoremini ifade ve ispat edelim. Literatürde, teoremin değişik türlerine rastlanabilir. Teoremdeki koşulların sağlanmadığı hallerde yeni koşullar ekleyerek merkezi limit teoremi sağlatılmaya çalışılır. Aşağıda bu teoremin Casella ve Berger (2002) de verilen hali dikkate alınmıştır.

Teorem 6.2.1 (*Merkezi Limit Teoremi*) $X_1, X_2 \ldots, X_n, \ldots$ beklenen değeri μ varyansı σ^2 olan bağımsız aynı dağılıma sahip rasgele değişkenler olsun. Rasgele değişkenlerin moment çıkaran fonksiyonu sıfır noktası komşuluğunda varsa $n \to \infty$ iken,

$$\frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma} \xrightarrow{\mathcal{D}} N(0,1)$$

dir.

İspat: Standart normal dağılımın moment çıkaran fonksiyonu, $M_Z(t) = e^{t^2/2}$ olup, teoremin ispatı için $Z_n = \sqrt{n}\,(\bar{X}_n - \mu)/\sigma$ olmak üzere Z_n nin moment çıkaran fonksiyonunun $n \to \infty$ iken $M_Z(t)$ ye yakınsadığını göstermek yeterlidir. $i = 1, 2, 3, \ldots$ için X_i lerin beklenen değeri μ varyansı σ^2 ise $Y_i = (X_i - \mu)/\sigma$ rasgele değişkenleri de beklenen değeri sıfır varyansı 1 olan bağımsız aynı dağılımlı rasgele değişkenlerdir. Buradan,

$$Z_n = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$$

olup Z_n nin moment çıkaran fonksiyonu

$$M_{Z_n}(t) = M_{(\bar{X}_n - \mu)/\sigma}(t) = M_{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i}(t) = M_{\sum_{i=1}^{n} Y_i}(t/\sqrt{n}) = \left[M_Y(t/\sqrt{n})\right]^n$$

şeklinde yazılabilir. $M_Y(t/\sqrt{n})$ nin sıfır noktası komşuluğundaki Taylor serisi açılımı,

$$M_Y(t/\sqrt{n}) = \sum_{k=0}^{\infty} M_Y^{(k)}(0) \frac{(t/\sqrt{n})^k}{k!}$$

olup $M_Y^{(k)}(0)$, $M_Y(t)$ fonksiyonunun k. türevinin sıfır noktasındaki değeridir. Moment çıkaran fonksiyonunun özelliklerinden

$$M_Y^{(0)}(0) = 1,\ M_Y^{(1)}(0) = E(Y) = 0 \text{ ve } M_Y^{(2)}(0) = E(Y^2) = Var(Y) = 1$$

olup $M_y(t/\sqrt{n})$ fonksiyonunun Taylor serisi açılımını

$$M_Y(t/\sqrt{n}) = 1 + 0 + \frac{(t/\sqrt{n})^2}{2!} + \sum_{k=3}^{\infty} M_Y^{(k)}(0)\frac{(t/\sqrt{n})^k}{k!} = 1 + \frac{t^2}{2n} + R_n(t/\sqrt{n})$$

şeklinde de yazabiliriz. Diğer taraftan sabit her t ($t \neq 0$) için

$$\lim_{n\to\infty} \frac{R_n(t/\sqrt{n})}{(1/\sqrt{n})^2} = \lim_{n\to\infty}\left[n R_n(t/\sqrt{n})\right] = 0$$

dır. $R_n(t/\sqrt{n})$ içindeki terimlerin hepsi pozitif ve en büyük terim $k=3$ deki değer olup,

$$k=3 \text{ için } n\,(t/\sqrt{n})^3/3! \to 0 \text{ ve } k=4 \text{ için } n\,(t/\sqrt{n})^4/4! \to 0$$

dır. Bütün terimler pozitif olduğundan, $n \to \infty$ iken $n\,R_n(t/\sqrt{n}) \to 0$ dır. Ayrıca, elemanları reel sayılar olan bir a_n dizisi için $a_n \to a$ ise $(1 + a_n/n)^n \to e^a$ olduğunu da biliyoruz. Buna göre $n \to \infty$ iken,

$$a_n = \frac{t^2}{2} + n R_n\left(t/\sqrt{n}\right) \to \frac{t^2}{2}$$

olup,

$$\lim_{n\to\infty}\left[M_Y\left(t/\sqrt{n}\right)\right]^n = \lim_{n\to\infty}[1 + \frac{1}{n}(\frac{t^2}{2} + n R_n(t/\sqrt{n}))]^n = e^{t^2/2}$$

şeklinde aranan yakınsama elde edilir. Bu da teoremin ispatı için yeterlidir ◊

$X_1, X_2 \ldots, X_n, \ldots$ beklenen değeri μ varyansı σ^2 olan bağımsız rasgele değişkenler olsun. Bazen,

$$\frac{\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)}{\sqrt{Var\left(\sum_{i=1}^{n} X_i\right)}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{n\left(\bar{X}_n - \mu\right)}{\sqrt{n}\,\sigma} = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}$$

olduğundan merkezi limit teoremi,

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \xrightarrow{\mathcal{D}} N(0,1) \quad \text{yerine} \quad \frac{\sum_{i=1}^{n} X_i - E\left(\sum_{i=1}^{n} X_i\right)}{\sqrt{Var\left(\sum_{i=1}^{n} X_i\right)}} \xrightarrow{\mathcal{D}} N(0,1)$$

şeklinde ifade edilir. Şimdi, MLT nin uygulamalarına ilişkin birkaç örnek verelim.

Örnek 6.2.5 a) Bir para 100 defa atıldığında en az 60 defa tura gelmesi olasılığını hesaplayalım. Bu olasılık Binom dağılımından hesaplanabilir. $X \sim Binom(100,1/2)$ olduğundan aranan olasılık $P(X \geq 60)$ dır. Bu olasılık da,

$$P(X \geq 60) = \sum_{x=60}^{100} P(X = x) = \sum_{x=60}^{100} \binom{100}{x} \left(\frac{1}{2}\right)^{x} \left(\frac{1}{2}\right)^{100-x} = \left(\frac{1}{2}\right)^{100} \sum_{x=60}^{100} \binom{100}{x} = 0.02844$$

olarak bilgisayar desteği ile hesaplanmıştır. Aynı olasılık MLT teoremi ile yaklaşık olarak hesaplanabilir. Paranın her bir atılışında (örneğin i. atılış) gelen turaların sayısı X_i olsun. Buradan, $X_i \sim Bern(1/2)$ olup $X = X_1 + X_2 + \ldots + X_{100}$ paranın 100 defa atılmasında toplam turaların sayısı olur. Denemeler birbirinden bağımsız olup $X \sim Binom(100,1/2)$ ve $E(X) = 100(1/2) = 50$, $Var(X) = 100(1/2)(1/2) = 25$ dir. Bu örnek için 100 denemenin yeterince büyük olduğu göz önüne alındığında, MLT ne göre $P(X \geq 60)$ olasılığı

$$P(X \geq 60) = P\left(\sum_{i=1}^{100} X_i \geq 60\right) = P\left(\frac{\sum_{i=1}^{100} X_i - E\left[\sum_{i=1}^{100} X_i\right]}{\sqrt{Var\left[\sum_{i=1}^{100} X_i\right]}} \geq \frac{60-50}{5}\right) \cong P\left(Z \geq 2\right) = 0.0228$$

şeklinde yaklaşık olarak hesaplanmıştır. Aradaki fark çok küçüktür. n örneklem hacmi büyüdükçe aradaki fark küçülür.

b) Bir elektronik ürünün dayanma süresi (yıl olarak) X olsun. X in olasılık yoğunluk fonksiyonu,

$$f(x) = \begin{cases} 1/x^2 & , \quad x > 1 \\ 0 & , \quad d.y. \end{cases}$$

olarak verilmiş ise rasgele seçilen 100 üründen en az 25 tanesinin 5 yıldan fazla dayanması olasılığını hesaplamak isteyelim. Bunun için önce,

$$Y_i = \begin{cases} 1 & , \quad X_i > 5 \\ 0 & , \quad d.y. \end{cases}$$

rasgele değişkenlerini tanımlayalım. Seçilen 100 üründen 5 yıldan fazla dayananların toplam sayısı Y_i lerin toplamıdır. Ayrıca, her bir ürünün 5 yıldan fazla dayanma olasılığı,

$$p = P(Y_i = 1) = P(X_i > 5) = \int_{x=5}^{\infty} \frac{1}{x^2} dx = -\frac{1}{x}\Big|_{x=5}^{\infty} = \frac{1}{5}$$

olup, $Y_1, Y_2, \ldots, Y_{100}$ başarı olasılığı $1/5$ olan bağımsız Bernoulli dağılımına sahip rasgele değişkenlerdir. Yani, $Y_i \sim Bern(1/5)$ olup, $Y = Y_1 + Y_2 + \ldots + Y_{100} \sim Binom(100, 1/5)$ dir. Yani aranan olasılık, $P(Y \geq 25)$ olarak yazılabilir. Ayrıca,

$$E(Y) = n\,p = 100(1/5) = 20 \;\text{ ve }\; Var(Y) = n\,p\,q = 100(1/5)(4/5) = 16$$

olup aranan olasılık MLT'ne göre yaklaşık olarak

$$P(Y \geq 25) = P\left(\sum_{i=1}^{100} Y_i \geq 25\right) = P\left(\frac{\sum_{i=1}^{100} Y_i - E\left[\sum_{i=1}^{100} Y_i\right]}{\sqrt{Var\left[\sum_{i=1}^{100} Y_i\right]}} \geq \frac{25-20}{4}\right) \cong P\left(Z \geq 1.25\right) = 0.1056$$

şeklinde bulunmuştur.

c) Aşağıdaki limit

$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}$$

MLT yardımı ile hesaplanabilir. Şimdi bu limit değerini hesaplayalım.

$X_1, X_2, \ldots, X_n, \ldots$ beklenen değeri 1 olan Poisson dağılımlı bağımsız rasgele değişkenler olsun. $X = X_1 + X_2 + \ldots + X_n \sim Poisson(n)$ olup,

$$P(X \leq n) = \sum_{k=0}^{n} P(X = k) = \sum_{k=0}^{n} e^{-n} \frac{n^k}{k!} = e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!}$$

dir. $E(X) = n$ ve $Var(X) = n$ olduğundan aranan limit

$$\lim_{n\to\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} = \lim_{n\to\infty} P(X \leq n) = \lim_{n\to\infty} P\left(\sum_{i=1}^{n} X_i \leq n\right)$$

$$= \lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - E\left[\sum_{i=1}^{n} X_i\right]}{\sqrt{Var\left[\sum_{i=1}^{n} X_i\right]}} \leq \frac{n-n}{\sqrt{n}}\right) = P(Z \leq 0) = \frac{1}{2}$$

olarak bulunmuş olur.

d) Stirling formülü olarak bilinen $n! \cong e^{-n} n^{n+1/2} \sqrt{2\pi}$ yaklaşımı da MLT nin bir sonucudur. $X_1, X_2, \ldots, X_n, \ldots$ beklenen değeri 1 olan üstel dağılımlı bağımsız rasgele değişkenler olsun. Bu durumda, $X = X_1 + X_2 + \ldots + X_n \sim Gamma(n,1)$ olup, X in olasılık yoğunluk fonksiyonu,

$$f(x)=\begin{cases}\dfrac{1}{\Gamma(n)}x^{n-1}e^{-x} & , \quad x>0\\ & , \quad d.y.\end{cases}$$

şeklindedir. Merkezi limit teoremine göre $n\to\infty$ iken,

$$P\left(\frac{\sum_{i=1}^{n}X_i-E\left[\sum_{i=1}^{n}X_i\right]}{\sqrt{Var\left[\sum_{i=1}^{n}X_i\right]}}\le x\right)\to\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$$

olup bu ifade

$$P\left(\frac{X-E(X)}{\sqrt{Var(X)}}\le x\right)=P\left(\frac{X-n}{\sqrt{n}}\le x\right)=P\left(X\le n+x\sqrt{n}\right)\to\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$$

şeklinde yazılabilir. Ayrıca $X\sim Gamma(n,1)$ olduğundan, Gamma dağılımının olasılık yoğunluk fonksiyonu yerine yazılırsa $n\to\infty$ iken,

$$P\left(X\le n+x\sqrt{n}\right)=\int_{0}^{n+x\sqrt{n}}\frac{1}{\Gamma(n)}t^{n-1}e^{-t}dt\to\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$$

yaklaşımı elde edilir. Buradan, n yeterince büyük ise,

$$\int_{0}^{n+x\sqrt{n}}\frac{1}{\Gamma(n)}t^{n-1}e^{-t}dt\;\cong\;\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$$

olup her iki tarafın x e göre türevinden

$$\frac{\sqrt{n}}{\Gamma(n)}\left(n+x\sqrt{n}\right)^{n-1}e^{-\left(n+x\sqrt{n}\right)}\;\cong\;\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$

yaklaşımı elde edilir. Bu son ifade de $x=0$ yazılırsa,

$$\frac{\sqrt{n}}{\Gamma(n)}n^{n-1}e^{-n}\;\cong\;\frac{1}{\sqrt{2\pi}}\quad\text{olup}\quad\Gamma(n)\cong n^{n-1}e^{-n}\sqrt{n}\;\sqrt{2\pi}$$

elde edilir. Ayrıca $\Gamma(n)=(n-1)!$ olduğundan $(n-1)!\cong e^{-n}n^{n-1}\sqrt{2\pi}$ dir. Son olarak her iki taraf n ile çarpılırsa, sol taraf $n(n-1)!=n!$ olur ve aranan sonuç $n^{1/2}=\sqrt{n}$ yazılması ile, $n!\cong e^{-n}n^{n+1/2}\sqrt{2\pi}$ olarak bulunmuş olur $\oplus$

MLT aşağıda ifadesi verilen Slutsky teoremi (Casella ve Berger, 2002, sayfa 239-240) ile beraber, tahmin edicilerin asimptotik dağılımının bulunmasında kolaylıklar sağlar.

Teorem 6.2.2 (*Slutsky Teoremi*) X_n ve Y_n rasgele değişken dizileri, $a \in \mathbb{R}$ olmak üzere, $n \to \infty$ iken $X_n \xrightarrow{\mathcal{D}} X$ ve $Y_n \xrightarrow{P} a$ olsun. Bu durumda $n \to \infty$ iken,

a) $X_n \pm Y_n \xrightarrow{\mathcal{D}} X \pm a$ b) $X_n Y_n \xrightarrow{\mathcal{D}} X\, a$ c) $X_n / Y_n \xrightarrow{\mathcal{D}} X / a \quad , a \neq 0$

dir ◊

Örnek 6.2.6 $X_1, X_2, \ldots, X_n, ...$ beklenen değeri μ varyansı σ^2 olan bağımsız rasgele değişkenler olmak üzere MLT den $n \to \infty$ iken $\sqrt{n}\,(\bar{X}_n - \mu)/\sigma \xrightarrow{\mathcal{D}} N(0,1)$ dir. σ^2 bilinmiyorsa S_n^2 örneklem varyansı ile tahmin edilir. $\lim\limits_{n\to\infty} Var(S_n^2) = 0$ ise $n \to \infty$ iken $S_n^2 \xrightarrow{P} \sigma^2$ olasılıkta yakınsama sağlanır (Örnek (6.2.2)). Buradan da $\sigma / S_n \xrightarrow{P} 1$ olur. Slutsky teoremine göre $n \to \infty$ iken,

$$t_n^* = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{S_n} = \frac{\sigma}{S_n} \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \xrightarrow{\mathcal{D}} N(0,1)$$

elde edilir ⊕