HAFTA 7

TAHMİN EDİCİLERDE ARANAN ÖZELLİKLER

7.5. Tutarlılık (Consistency)

Tahmin edici örneklemin bir fonksiyonu olup, aynı zamanda örneklem hacmine de bağlıdır.

Tanım 7.5.1 $X_1, X_2, \ldots, X_n$ olasılık veya olasılık yoğunluk fonksiyonu $f(x;\theta)$ olan kitleden bir örneklem, $T_n = T_n(\underset{\sim}{X})$ de θ nın herhangi bir tahmin edicisi olsun. $n \to \infty$ iken $T_n \xrightarrow{P} \theta$ ise, T_n tahmin edicilerin dizisi θ için *tutarlıdır* denir$\otimes$

Örnek 7.5.1 $N(\mu,1)$ dağılımından $n-$ birimlik bir örneklem $X_1, X_2, \ldots, X_n$ olsun. Zayıf büyük sayılar yasasına göre, örneklem ortalaması kitle ortalamasına olasılıkta yakınsar ($n \to \infty$ iken $\bar{X}_n \xrightarrow{P} \mu$). Yani, $\bar{X}_n$ örneklem ortalaması μ için tutarlıdır$\oplus$

Yukarıdaki örnekte örneklem ortalamasının tutarlılığı için zayıf büyük sayılar yasası kullanıldı. Ayrıca, dağılım normal olduğundan olasılıkta yakınsaklık için $P(|\bar{X}_n - \mu| > \varepsilon)$ olasılığı da hesaplanabilirdi. Tahmin edici örneklem ortalaması ise, merkezi limit teoreminden olasılıkta yakınsama kolay elde edilebilir. Ancak, karşımıza her zaman örneklem ortalaması çıkmayabilir.

Tahmin edicilerin tutarlılık özelliği Chebyshev eşitsizliğinden elde edilebilir. $X_1, X_2, \ldots, X_n$ parametresi θ olan kitleden bir örneklem, T_n de θ nın tahmin edicilerinin bir dizisi olsun. Bu durumda yine Chebyshev eşitsizliğine göre her $\varepsilon > 0$ sayısı için,

$$P(|T_n - \theta| > \varepsilon) \leq \frac{E(T_n - \theta)^2}{\varepsilon^2}$$

yazılır. Buradan $n \to \infty$ iken $E(T_n - \theta)^2 \to 0$ ise $T_n \xrightarrow{P} \theta$ dir. Yani, T_n tahmin edicilerinin dizisinin θ için tutarlı olması için $E(T_n - \theta)^2 \to 0$ olması yeterlidir. Diğer taraftan,

$$E_\theta(T_n - \theta)^2 = E_\theta(T_n - E_\theta(T_n) + E_\theta(T_n) - \theta)^2 = Var_\theta(T_n) + (E_\theta(T_n) - \theta)^2$$

olduğundan T_n tahmin edicilerinin dizisinin θ için tutarlı olması için gerek ve yeter koşul $Var_\theta(T_n)$ ve $(E_\theta(T_n) - \theta)^2$ terimlerinin her ikisinin de $n \to \infty$ iken sıfıra gitmesidir. Bu sonuç aşağıdaki teoremde özetlenmiştir.

Teorem 7.5.1 T_n tahmin edicilerin bir dizisi olsun. Eğer

$$\lim_{n\to\infty} Var_\theta(T_n) = 0 \quad \text{ve} \quad \lim_{n\to\infty} (E_\theta(T_n) - \theta)^2 = 0$$

koşulları sağlanıyorsa, T_n tahmin edicilerin dizisi θ için tutarlıdır ◊

Örnek 7.5.2 a) $X_1, X_2, \ldots, X_n$ olasılık yoğunluk fonksiyonu $f(x;\theta)$ olan kitleden bir örneklem olsun. X lerin olasılık yoğunluk fonksiyonu

$$f(x;\theta) = \begin{cases} e^{-(x-\theta)} & , \quad x > \theta \\ 0 & , \quad d.y \end{cases}$$

ise $T_n = X_{(1)}$ sıra istatistiği θ için yeterlidir. T_n nin olasılık yoğunluk fonksiyonu,

$$f_{T_n}(t;\theta) = \begin{cases} n\, e^{-n\,(t-\theta)} & , \quad t > \theta \\ 0 & , \quad d.y \end{cases}$$

olup $E(T_n) = \theta + 1/n$ ve $Var_\theta(T_n) = 1/n^2$ dir. Buradan, $\lim\limits_{n\to\infty} Var_\theta(T_n) = 0$ olduğu açıktır. Bununla birlikte,

$$\lim_{n\to\infty} (E_\theta(T_n) - \theta)^2 = \lim_{n\to\infty} [(\theta + 1/n) - \theta]^2 = \lim_{n\to\infty} (1/n^2) = 0$$

olduğundan T_n tahmin edicilerin dizisi θ için tutarlıdır.

b) $X_1, X_2, \ldots, X_n$ parametresi θ olan düzgün dağılımdan bir örneklem olsun. Yani, $X \sim U(0,\theta)$ olmak üzere, $T_n = X_{(n)}$ sıra istatistiği θ için yeterli olup, T_n nin olasılık yoğunluk fonksiyonu $0 < t < \theta$ için $f_{T_n}(t;\theta) = n\,\theta^{-n}\, t^{n-1}$ dir. Buradan, T_n nin beklenen değeri ile varyansının

$$E_\theta(T_n) = \frac{n}{n+1}\theta \quad \text{ve} \quad Var_\theta(T_n) = \frac{n\,\theta^2}{(n+1)^2(n+2)}$$

olduğunu biliyoruz (Örnek (6.4.2a)). Ayrıca,

$$\lim_{n\to\infty} Var_\theta(T_n) = \lim_{n\to\infty}\left(\frac{n\,\theta^2}{(n+1)^2(n+2)}\right) = 0$$

$$\lim_{n\to\infty} \left(E_\theta(T_n) - \theta\right)^2 = \lim_{n\to\infty}\left(\frac{n}{n+1}\theta - \theta\right)^2 = \lim_{n\to\infty}\left(\frac{\theta^2}{(n+1)^2}\right) = 0$$

dır. Yani, T_n tahmin edicilerin dizisi θ için tutarlıdır ⊕

7.6. Etkinlik (Efficiency)

Tahmin edicilerde aranan özelliklerden biri de etkinliktir. Bu özellik, tahmin edicileri karşılaştırmak için de kullanılır. Bir parametre için bir çok tahmin edici önerilebilir. Elbette, tahmin edicilerin diğer özellikleri ile beraber etkin olması da beklenir. Varsa, en etkin tahmin edici kullanılmak istenir.

Tanım 7.6.1 *(Etkinlik)* T_n ve T_n^* herhangi bir θ parametresi için iki tahmin edici olsun. Her θ için

$$Var_\theta(T_n) \leq Var_\theta(T_n^*)$$

oluyorsa T_n tahmin edicisi T_n^* tahmin edicisine göre *daha etkindir* denir$\otimes$

Örnek 7.6.1 a) $X \sim U(0,\theta)$ dağılımından bir örneklem $X_1, X_2, \ldots, X_n$ olsun. $X_{(n)}$ sıra istatistiği θ için yeterli olup bu istatistiğin beklenen değeri ile varyansı

$$E_\theta(X_{(n)}) = \frac{n}{n+1}\theta \quad \text{ve} \quad Var_\theta(X_{(n)}) = \frac{n\,\theta^2}{(n+1)^2(n+2)}$$

dir (Örnek (7.5.2b). Buradan, $T_1 = \frac{n+1}{n} X_{(n)}$ denirse,

$$E_\theta(T_1) = \theta \quad \text{ve} \quad Var_\theta(T_1) = \frac{\theta^2}{n(n+2)}$$

bulunur.

Ayrıca, $T_2 = 2\bar{X}_n$ tahmin edicisinin beklenen değeri ile varyansı da $E_\theta(T_2) = \theta$ ve $Var_\theta\left(T_2\right) = \theta^2/(3n)$ olarak hesaplanmıştır (her iki tahmin edici de tutarlıdır). Buradan,

$$Var_\theta(T_1) = [\theta^2/(n(n+2))] \leq [\theta^2/(3n)] = Var_\theta(T_2)$$

olduğundan yeterli istatistiğin fonksiyonu olan T_1 tahmin edicisi T_2 ye göre daha etkindir.

b) $X_1, X_2, \ldots, X_n$ parametresi λ olan Poisson dağılımından bir örneklem olsun. Poisson dağılımının beklenen değeri ve varyansı aynı olup $E_\lambda(\bar{X}_n) = E_\lambda(S_n^2) = \lambda$ dır. Örneklem ortalaması ile örneklem varyansının beklenen değerleri aynıdır. Bu tahmin edicilerin etkinliği için varyanslarına ihtiyaç vardır. Önce, $E(S_n^2 \,|\, \bar{X}_n) = \bar{X}_n$ olduğunu gösterelim. Koşullu beklenen değerin tanımından,

$$E\left(S_n^2 \,|\, \bar{X}_n\right) = \frac{1}{n-1} E\left[\left(\sum_{i=1}^{n} X_i^2 - n\,\bar{X}_n^2\right) | \bar{X}_n\right] = \frac{n}{n-1} E\left[\left(X_1^2 - \bar{X}_n^2\right) | \bar{X}_n\right]$$

$$=\frac{n}{n-1}\left\{\left[E\left(X_1^2 \mid \bar{X}_n\right)-\bar{X}_n^2\right]\right\}$$

olup $E(X_1^2 \mid \bar{X}_n)$ koşullu beklenen değerinin hesaplanması gerekir. Bunun için, $Y=n\bar{X}_n$ diyelim. $Y \sim Poisson(n\lambda)$ ve $X_2+X_3+\ldots+X_{n-1} \sim Poisson((n-1)\lambda)$ olduğu dikkate alındığında koşullu beklenen değer için, $P(X_1=x_1 \mid \bar{X}_n=\bar{x})$ koşullu olasılığı,

$$P(X_1=x_1 \mid \bar{X}_n=\bar{x})=\frac{P(X_1=x_1,\ \bar{X}_n=\bar{x})}{P(\bar{X}_n=\bar{x})}=\frac{P(X_1=x_1,\ Y=n\bar{x})}{P(Y=n\bar{x})}=\frac{P(X_1=x_1,\ Y=y)}{P(Y=y)}$$

$$=\frac{P\left(X_1=x_1,\ \sum_{i=2}^{n} X_i=y-x_1\right)}{P\left(Y=y\right)}=\frac{P\left(X_1=x_1\right)P\left(\sum_{i=2}^{n} X_i=y-x_1\right)}{P\left(Y=y\right)}=\binom{y}{x_1}\left(\frac{1}{n}\right)^{x_1}\left(1-\frac{1}{n}\right)^{y-x_1}$$

olarak hesaplanmıştır. Yani, $y=n\bar{x}$ olmak üzere $(X_1 \mid \bar{X}_n=\bar{x}) \sim Binom(y,1/n)$ olup koşullu olasılık fonksiyonu

$$P\left(X_1=x_1 \mid \bar{X}_n=\bar{x}\right)=\binom{y}{x_1}\left(\frac{1}{n}\right)^{x_1}\left(1-\frac{1}{n}\right)^{y-x_1},\ x_1=0,1,2,\ldots,y$$

şeklindedir. Buradan, koşullu beklenen değer ile koşullu varyans

$$E\left(X_1 \mid \bar{X}_n\right)=\frac{Y}{n}=\bar{X}_n \quad \text{ve} \quad Var\left(X_1 \mid \bar{X}_n\right)=\left(\frac{1}{n}\right)\left(1-\frac{1}{n}\right)Y=\frac{n-1}{n}\bar{X}_n$$

olup

$$E(X_1^2 \mid \bar{X}_n)=\frac{n-1}{n}\bar{X}_n+\bar{X}_n^2$$

olduğu açıktır. Buradan da,

$$E\left(S_n^2 \mid \bar{X}_n\right)=\frac{n}{n-1}\left\{\left[E\left(X_1^2 \mid \bar{X}_n\right)-\bar{X}_n^2\right]\right\}=\frac{n}{n-1}\left\{\frac{n-1}{n}\bar{X}_n+\bar{X}_n^2-\bar{X}_n^2\right\}=\bar{X}_n$$

elde edilir. Dolayısı ile,

$$Var_\lambda\left(S_n^2\right)=Var_\lambda\left(E\left(S_n^2 \mid \bar{X}_n\right)\right)+E_\lambda\left(Var\left(S_n^2 \mid \bar{X}_n\right)\right)\geq Var_\lambda\left(E\left(S_n^2 \mid \bar{X}_n\right)\right)=Var_\lambda\left(\bar{X}_n\right)$$

olduğundan $\bar{X}_n$ tahmin edicisi S_n^2 tahmin edicisine göre daha etkindir $\oplus$

$X_1, X_2, \ldots, X_n$ olasılık veya olasılık yoğunluk fonksiyonu $f(x;\theta)$ olan kitleden bir örneklem, T_n de bir tahmin edici olmak üzere, $E_\theta(T_n)=\tau(\theta)$ olsun. Ayrıca, $\tau(\theta)$ nın θ ya göre türevlenebilir olduğunu kabul edelim.

Tanım 7.6.2 *(Asimptotik Etkinlik)* $E_\theta(T_n)=\tau(\theta)$ olacak şekilde herhangi bir tahmin edici T_n olsun. Eğer

$$I_n(\theta)=\frac{(\tau'(\theta))^2}{-\,n\,E_\theta\left(\dfrac{\partial^2}{\partial\theta^2}\ln\left(f(X;\theta)\right)\right)} \quad \text{olmak üzere} \quad \lim_{n\to\infty}\frac{Var_\theta\left(T_n\right)}{I_n(\theta)}=1$$

oluyorsa T_n tahmin edicisine *asimptotik etkindir* denir $\otimes$

Örnek 7.6.2 $Poisson(\lambda)$ dağılımından $n-$birimlik bir örneklem $X_1,X_2,\ldots,X_n$ olsun. $T_n=\bar{X}_n$ olmak üzere, $E_\lambda(\bar{X}_n)=\lambda$ ve $Var_\lambda(\bar{X}_n)=\lambda/n$ olup $\tau(\lambda)=\lambda$ için $\tau'(\lambda)=1$ dir. Diğer taraftan $\ln(f(X,\lambda))=\ln(e^{-\lambda}\lambda^X/X!)=-\lambda+X\ln(\lambda)-\ln(X!)$ olup türevler,

$$\frac{\partial}{\partial\lambda}\ln\left(f\left(X;\lambda\right)\right)=-1+\frac{X}{\lambda} \qquad \text{ve} \qquad \frac{\partial^2}{\partial\lambda^2}\ln\left(f\left(X;\lambda\right)\right)=-\frac{X}{\lambda^2}$$

şeklindedir. Buradan, $I_n(\theta)$ nın paydasındaki değer,

$$-\,n\,E_\lambda\left(\frac{\partial^2}{\partial\lambda^2}\ln\left(f(X;\lambda)\right)\right)=-n\,E_\lambda\left(-X/\lambda^2\right)=n/\lambda \quad \text{ve} \quad I_n(\lambda)=\frac{1}{n/\lambda}=\lambda/n$$

olup,

$$\lim_{n\to\infty}\left(\frac{Var_\lambda\left(\bar{X}_n\right)}{I_n(\lambda)}\right)=\lim_{n\to\infty}\left(\frac{\lambda/n}{\lambda/n}\right)=1$$

dir. Yani, $\bar{X}_n$ tahmin edicisi λ için asimptotik etkindir $\oplus$

7.7. Yansızlık (Unbiasedness)

Tahmin edicilerin yansızlığı, tahmin teorisinde en çok kullanılan özelliklerden biridir. Bir parametre için birden çok yansız tahmin edici bulunabildiği gibi, bazen yansız tahmin ediciler olmayabilir.

Tanım 7.7.1 $X_1,X_2,\ldots,X_n$ parametresi θ olan kitleden bir örneklem, T de θ nın herhangi bir tahmin edicisi olsun. T nin beklenen değeri mevcut ve $E_\theta(T)=\tau(\theta)$ ise T ye $\tau(\theta)$ için *yansız* bir tahmin edici denir $\otimes$

Herhangi bir parametre için yansız tahmin edici bulunamayabilir. Örneğin, Cauchy(θ) dağılımının beklenen değeri yoktur. Dolayısı ile, θ için yansız tahmin edici bulunamaz.

Örnek 7.7.1 a) $X \sim U(0,\theta)$ dağılımından bir örneklem $X_1, X_2, \ldots, X_n$ olsun. $T_1 = ((n+1)/n)X_{(n)}$ ve $T_2 = 2\overline{X}_n$ tahmin edicilerinin beklenen değerlerinin $E_\theta(T_1) = \theta$ ve $E_\theta(T_2) = \theta$ olduğunu biliyoruz (Örnek (7.6.1a)). Yani, T_1 ve T_2 nin her ikisi de θ için yansızdır. Ayrıca, $Var_\theta(T_1) \le Var_\theta(T_2)$ olduğu da aynı örnekte gösterildi. θ için yine sıra istatistiklerine bağlı olarak başka yansız tahmin ediciler de bulunabilir. Örneğin, $T = X_{(1)}$ denirse, T nin olasılık yoğunluk fonksiyonu,

$$f_T(t;\theta) = \begin{cases} \dfrac{n}{\theta}\left(1-\dfrac{t}{\theta}\right)^{n-1} & , \quad 0 < t < \theta \\ 0 & , \quad d.y. \end{cases}$$

olup, $E_\theta(T) = E_\theta(X_{(1)}) = \theta/(n+1)$ dir. Buradan da $T_3 = (n+1)X_{(1)}$ tahmin edicisi de θ için yansızdır. Ancak, n. sıra istatistiğine bağlı T_1 yansız tahmin edicisinin varyansı T nin varyansından küçüktür.

b) $Poisson(\lambda)$ dağılımından alınan $n-$birimlik bir örneklem $X_1, X_2, \ldots, X_n$ olmak üzere $\tau(\lambda) = P_\lambda(X = 0) = e^{-\lambda}$ parametresi için yansız bir tahmin edici,

$$W_n = \begin{cases} 1 & , \quad X_1 = 0 \\ 0 & , \quad d.d. \end{cases}$$

olarak alınabilir. Burada, W , 0 ve 1 değerlerini alan Bernoulli rasgele değişkeni olup

$$E_\lambda(W) = 0\,P_\lambda(X_1 \neq 0) + 1\,P_\lambda(X_1 = 0) = P_\lambda(X_1 = 0) = e^{-\lambda}$$

dır.

c) Yine $Poisson(\lambda)$ dağılımından bir örneklem $X_1, X_2, \ldots, X_n$ olsun. Bu defa $\tau(\lambda) = \lambda^2$ için yansız bir tahmin edici bulmak isteyelim. Bunun için

$$T = X_1 + X_2 + \ldots + X_n \sim Poisson(n\lambda)$$

olduğunu biliyoruz. Poisson dağılımının beklenen değer ve varyansı aynı olduğundan,

$$E_\lambda(T^2) = Var_\lambda(T) + (E_\lambda(T))^2 = n\lambda + n^2\lambda^2$$

yazılır. Buradan, $U = (T^2 - T)/n^2$ denirse, $E_\lambda(U) = \lambda^2$ olur. Yani U , λ^2 için yansızdır.

d) Şimdi, $X_1, X_2, \ldots, X_n$ parametresi p olan Bernoulli dağılımından bir örneklem olsun ve p^2 için yansız bir tahmin edici bulmaya çalışalım.

$T = X_1 + X_2 + \ldots + X_n \sim Binom(n, p)$ olup, $E_p(T) = n\,p$ ve $Var_p(T) = n\,p(1-p)$ dir. Buna göre, T nin ikinci momenti

$$E_p(T^2) = Var_p(T) + (E_p(T))^2 = np(1-p) + n^2p^2 = np - np^2 + n^2p^2$$

olarak hesaplanmıştır. Buradan da,

$$E_p(T^2 - T) = (np - np^2 + n^2p^2) - np = p^2(n^2 - n) = n(n-1)p^2$$

olduğundan $U = (T^2 - T)/n(n-1)$ istatistiği p^2 için yansızdır⊕

Tanım 7.7.2 *(Yan)* $X_1, X_2, \ldots, X_n$ parametresi θ olan kitleden bir örneklem, T de θ nın herhangi bir tahmin edicisi olsun. $E_\theta(T)$ nin mevcut olması halinde $E_\theta(T) - \theta$ değerine T nin *yanı* (veya *yanlılığı*) denir ve $Bias_\theta(T)$ ile gösterilir. Yani, $Bias_\theta(T) = E_\theta(T) - \theta$ dır⊗

Tahmin edicilerin bu özelliklerinin yanında, genellikle küçük varyanslı tahmin ediciler tercih edilir. Bazen yanlı bir tahmin edicinin varyansı, yansız tahmin edicinin varyansından küçük olabilir. Yanlı tahmin edicinin varyansı küçük ve yanlılığı herhangi bir şekilde azaltılabilirse, bu yanlı tahmin edici tercih edilebilir. Yanlı tahmin edicilerin yanlılığı değişik yöntemlerler ile azaltılabilir. *Jackknife metodu* bunlardan biridir.

Örnek 7.7.2 $X_1, X_2, \ldots, X_n$ parametresi p olan Bernoulli dağılımından bir örneklem olsun. p^2 parametresini tahmin etmek isteyelim. Jackknife metodu tahmin edicilerin yanlılığını azaltmak için kullanılan bir yöntemdir. Bu tahmin ediciler aşağıdaki gibi hesaplanır (1-adım). θ nın tahmin edicileri $T_n(X_1, X_2, \ldots, X_n)$ şeklinde verilmiş olsun. $i = 1, 2, ..., n$ için

$$T_n^{(i)} = T_n(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, ..., X_n)$$

tahmin edicilerini tanımlayalım (örneklemden X_i çıkartılıyor). Buradan, θ nın Jackknife tahmin edicisi,

$$JK(T_n) = nT_n - \frac{n-1}{n}\sum_{i=1}^{n} T_n^{(i)}$$

şeklinde hesaplanır. Genellikle, $Bias_\theta(JK(T_n)) \leq Bias_\theta(T_n)$ dir.

Şimdi, p nin en çok olabilirlik tahmin edicisi (ileride bahsedilecektir) $\bar{X}_n$ olmak üzere, p^2 nin en çok olabilirlik tahmin edicisi $\bar{X}_n^2$ dir (Teorem 7.9.2.1). Bu en çok olabilirlik tahmin edicisine bağlı Jackknife tahmin edicisini bulalım (başka tahmin ediciler kullanarak da Jackknife tahmin edicileri yazılabilir). Buna göre, $T_n = \bar{X}_n^2$ için $T_n^{(i)}$

$$T_n^{(i)} = \frac{1}{(n-1)^2}\left(\sum_{j=1,\ j\neq i}^{n} X_j\right)^2 = \frac{1}{(n-1)^2}\left[\sum_{j=1}^{n} X_j - X_i\right]^2 = \frac{1}{(n-1)^2}\left[n\bar{X}_n - X_i\right]^2$$

olarak yazılır. Ayrıca, X_i ler Bernoulli dağılımına sahip rasgele değişkenler olup sadece 0 ve 1 değerlerini aldığından $\sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} X_i$ dir. Buradan $\sum_{i=1}^{n} T_n^{(i)}$ tahmin edicisi

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

$$= \frac{1}{(n-1)^2} \left[n^3 \bar{X}_n^2 - 2n^2 \bar{X}_n^2 + \sum_{i=1}^{n} X_i^2 \right] = \frac{1}{(n-1)^2} \left[n^3 \bar{X}_n^2 - 2n^2 \bar{X}_n^2 + \sum_{i=1}^{n} X_i \right]$$

şeklinde yazılabilir. Buradan da, p^2 nin Jackknife tahmin edicisi,

$$JK(T_n) = nT_n - \frac{n-1}{n} \sum_{i=1}^{n} T_n^{(i)} = n\left(\bar{X}_n^2\right) - \frac{n-1}{n} \left\{ \frac{1}{(n-1)^2} \left[\left(n^3 \bar{X}_n^2 - 2n^2 \bar{X}_n^2 + \sum_{i=1}^{n} X_i \right) \right] \right\}$$

$$= n\left(\bar{X}_n^2\right) - \frac{1}{n(n-1)} \left[n^3 \bar{X}_n^2 - 2n\bar{X}_n^2 + n\bar{X}_n \right] = n\left(\bar{X}_n^2\right) - \frac{n^2\left(\bar{X}_n^2\right)}{n-1} + \frac{2n\bar{X}_n^2}{n-1} - \frac{\bar{X}_n}{n-1}$$

$$= \bar{X}_n^2 \left[n - \frac{n^2}{n-1} + \frac{2n}{n-1} \right] - \frac{\bar{X}_n}{n-1} = \bar{X}_n^2 \left[\frac{n^2 - n - n^2 + 2n}{n-1} \right] - \frac{\bar{X}_n}{n-1} = \frac{n}{n-1} \bar{X}_n^2 - \frac{\bar{X}_n}{n-1}$$

$$= \frac{\left(\sum_{i=1}^{n} X_i \right)^2}{n(n-1)} - \frac{\sum_{i=1}^{n} X_i}{n(n-1)} = \frac{\left(\sum_{i=1}^{n} X_i \right)^2 - \sum_{i=1}^{n} X_i}{n(n-1)}$$

olarak elde edilir. Bu tahmin edici Örnek (7.7.1d) deki yansız tahmin edici ile aynıdır. Jackknife tahmin edicilerinin yanlılığı genellikle daha küçüktür. Burada, bir önceki örnekte verilen yansız tahmin edici ile Jackknife tahmin edicisi aynıdır. Bu her zaman böyle olmayabilir $\oplus$

7.8. Asimptotik Normallik

Tahmin edicilerin dağılımları istatistiki sonuç çıkarım açısından çok önemlidir. İstatistiki sonuç çıkarım için de normallik çok önemlidir. Tahmin edicilerin dağılımları normal olmasa bile, merkezi limit teoremine ile limit durumunda normallik elde edilebilir. Tahmin edicilerin dağılımları MLT de olduğu gibi bazen kolay bulunabilmesine rağmen, asimptotik dağılımı bulmak bazen zordur. Böyle durumlarda limit dağılımları elde edilmeye çalışılır.

$X_1, X_2, \ldots, X_n$ beklenen değeri μ, varyansı σ^2 olan kitleden bir örneklem ise merkezi limit teoremine göre, $n \to \infty$ iken $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{D} N(0,1)$ olup $\bar{X}_n$ asimptotik normaldir diyebiliriz. Bunu, $\bar{X}_n \sim AN(\mu, \sigma^2/n)$ olarak ifade edelim. Tahmin ediciler bazen $\bar{X}_n$ nin bir fonksiyonu olduğu gibi, bazen karmaşık olabilir. Örneğin, S_n^2 nin dağılımı, örneklem normal ise serbestlik derecesi $n-1$ olan ki-karedir. Ancak, örneklem normal değilse S_n^2 nin dağılımını bulmak kolay değildir. Bu durumda, limit dağılımından faydalanılır. Bunun için de Taylor serisi yaklaşımları kullanılır.

$g(x)$ fonksiyonu r. dereceden türevlenebilir (ve sıfırdan farklı) ise, herhangi bir a sabiti için $g(x)$ fonksiyonunun a noktası komşuluğundaki Taylor serisi açılımı,

$$T_r(x)=\sum_{i=0}^{r} g^{(i)}(a)\frac{(x-a)^i}{i!}+R(x;\, a)$$

olarak verilir. Burada, $g^{(i)}(a)=\left.\frac{d^i}{d\,x^i}g(x)\right|_{x=a}$ dir. Eğer $g^{(i)}(a)=\left.\frac{d^i}{d\,x^i}g(x)\right|_{x=a}$ varsa,

$$\lim_{x\to a}\frac{g(x)-T_r(x)}{(x-a)^r}=0$$

dır (Casella ve Berger, 2002, sayfa 241).

Benzer şekilde, $X_1, X_2, \ldots, X_k$ rasgele değişkenleri $E(X_i)=\theta_i\,, i=1,2,3,...,k$ olmak üzere, $\underset{\sim}{X}=(X_1, X_2, \ldots, X_k)'$ ve $\underset{\sim}{\theta}=(\theta_1,\theta_2,...,\theta_k)'$ için $g(\underset{\sim}{x})$ fonksiyonu θ_i noktalarında x_i lere göre türevlenebilir olsun. Buradan da,

$$g_i'(\underset{\sim}{a})=\left.\frac{d}{dx_i}g(\underset{\sim}{x})\right|_{x_1=\theta_1,\, x_2=\theta_2,\ldots,x_k=\theta_k}$$

varsa, $g(\underset{\sim}{x})$ fonksiyonunun $\underset{\sim}{\theta}$ noktasındaki birinci dereceden Taylor serisi açılımı

$$g(\underset{\sim}{x})=g(\underset{\sim}{\theta})+\sum_{i=1}^{k} g_i{}'(\underset{\sim}{\theta})(x_i-\theta_i)\;+K(\underset{\sim}{x},\underset{\sim}{\theta})$$

şeklinde yazılır. Burada, $K(\underset{\sim}{x},\underset{\sim}{\theta})$ kalan terimi göstermektedir. Buradan, $g(\underset{\sim}{x})$ fonksiyonu için,

$$g(\underset{\sim}{x})\;\approx\; g(\underset{\sim}{\theta})+\sum_{i=1}^{k} g_i{}'(\underset{\sim}{\theta})(x_i-\theta_i)$$

yaklaşımı elde edilir. Buna göre, $g(\underset{\sim}{X})$ rasgele değişkeninin beklenen değer ve varyansı için

$$E_{\underset{\sim}{\theta}}(g(\underset{\sim}{X}))\;\approx\; g(\underset{\sim}{\theta})+\sum_{i=1}^{k} g_i{}'(\underset{\sim}{\theta})E_{\underset{\sim}{\theta}}(X_i-\theta_i)$$

$$Var_{\underset{\sim}{\theta}}(g(\underset{\sim}{X}))\;\approx\;\sum_{i=1}^{k}[g_i{}'(\underset{\sim}{\theta})]^2 Var_{\underset{\sim}{\theta}}(X_i)+2\sum_{i>j} g_i(\underset{\sim}{\theta})g_j(\underset{\sim}{\theta})Cov(X_i, X_j)$$

yaklaşımları kullanılabilir. $E(X_i)=\theta_i, i=1,2,3,...,k$ olduğundan $E_{\underset{\sim}{\theta}}(g(\underset{\sim}{X})) \approx g(\underset{\sim}{\theta})$ dir. $\bar{X}_n \sim AN(\mu,\sigma^2/n)$ ise $g(x)$ fonksiyonunun μ noktası komşuluğundaki birinci dereceden Taylor serisi açılımından $g(\bar{X}_n) \sim AN(g(\mu),[g'(\mu)]^2\sigma^2/n)$ yazılır.

$X_1, X_2, \ldots, X_n$ beklenen değeri μ olan kitleden bir örneklem ve $E_\mu(X)=\mu \neq 0$ olsun. μ için bir tahmin edici $\bar{X}_n$ ise $g(\mu)$ parametresini tahmin etmek için

$$g(\bar{X}_n) \cong g(\mu)+g'(\mu)(\bar{X}_n-\mu)$$

yaklaşımı kullanılır. Burada $g(\bar{X}_n)$ nin beklenen değer ve varyansı için de,

$$E_\mu(g(\bar{X}_n)) \cong g(\mu) \text{ ve } Var_\mu(g(\bar{X}_n)) \cong [g'(\mu)]^2 Var_\mu(\bar{X}_n)$$

yaklaşımları kullanılabilir. T_n, θ parametresi için tahmin edcicilerin bir dizisi ise $n \to \infty$ iken

$$\sqrt{n}(T_n-\mu) \xrightarrow{D} N(0,\sigma^2)$$

olsun. Belirli bir g fonksiyonu için $g'(\theta)$ var ve sıfırdan farklı ise $n \to \infty$ iken

$$\sqrt{n}(g(T_n)-g(\mu)) \xrightarrow{D} N(0,\sigma^2[g'(\theta)]^2)$$

dir. Eğer, $g'(\theta)=0$ ise ikinci dereceden Taylor serisi açılımı kullanılır ve $g(x)$ fonksiyonunun ikinci dereceden Taylor serisi açılımı

$$g(T_n)=g(\theta)+g'(\theta)(T_n-\theta)+\frac{g''(\theta)}{2}(T_n-\theta)^2=g(\theta)+\frac{g''(\theta)}{2}(T_n-\theta)^2$$

olup $n \to \infty$ iken $\sqrt{n}(T_n-\mu) \xrightarrow{D} N(0,\sigma^2)$ olduğundan

$$\frac{n(T_n-\theta)^2}{\sigma^2} \xrightarrow{\mathcal{D}} \chi_1^2$$

yazılır. Buradan da, $n \to \infty$ iken

$$n\left(g(T_n)-g(\theta)\right)^2/\sigma^2 \xrightarrow{\mathcal{D}} (0.5)\ g''(\theta)\chi_1^2$$

asimptotik yaklaşımı elde edilmiş olur.

Örnek 7.8.1 $Poisson(\lambda)$ dağılımından bir örneklem $X_1, X_2, \ldots, X_n$ olsun. $\bar{X}_n$ örneklem ortalaması λ nın bir tahmin edici ise $e^{-\bar{X}_n}$ de $e^{-\lambda}$ nın bir tahmin edicisi olarak alınabilir. Merkezi limit teoreminden $\bar{X}_n \sim AN(\lambda,\lambda/n)$ olduğunu biliyoruz ($E_\lambda(X)=Var_\lambda(X)=\lambda$). $g(x)=e^{-x}$ denirse $g'(x)=-e^{-x}$ olup, $e^{-\bar{X}_n}$ in asimptotik dağılımı

$$g(\bar{X}_n)=e^{-\bar{X}_n}\sim AN\,(g(\lambda),[g'(\lambda)]^2\ \sigma^2/n)=AN\left(e^{-\lambda}\,,\,e^{-2\lambda}\lambda/n\right)$$

olur. Benzer şekilde, $g(\lambda)=1/\lambda$ nın tahmini için $1/\bar{X}_n$ önerilebilir. Bu durumda $g(x)=1/x$ alındığında, $g'(x)=-1/x^2$ olup $1/\bar{X}_n$ in asimptotik dağılımı

$$(\bar{X}_n)^{-1}=g(\bar{X}_n)\sim AN\,(g(\lambda),[g'(\lambda)]^2\ \sigma^2/n)=AN(\lambda^{-1},\lambda^{-4}\lambda/n)=AN(\lambda^{-1},\lambda^{-3}/n)$$

şeklinde yazılır. $(\bar{X}_n)^{-1}$ tahmin edicisinin varyansı için bir tahmin de,

$$Var_\lambda\left(\bar{X}_n^{-1}\right)=\left(\bar{X}_n^{-4}\right)Var_\lambda\left(\bar{X}_n\right)=\left(\bar{X}_n^{-4}\right)\frac{\bar{X}_n}{n}=\frac{1}{n\bar{X}_n^{3}}$$

nin değeri alınabilir$\oplus$

Tahmin edicilerin asimptotik normalliği ile ilgili bazı özellikler aşağıdaki örnek üzerinde açıklanmaya çalışılmıştır. Burada önemli olan, verilen fonksiyonun birinci dereceden Taylor serisine açılımıdır. Neredeyse, bütün asimptotik özellikler (normallik ile ilgili) fonksiyonun Taylor serisi açılımına bağlıdır.

Örnek 7.8.2 a) $W_1,W_2,\ldots,W_n$ beklenen değeri μ varyansı σ^2 olan bağımsız rasgele değişkenler olsun. $\bar{W}_n^2-\mu^2$ ifadesi

$$\bar{W}_n^2-\mu^2=(\bar{W}_n-\mu)(\bar{W}_n+\mu)=O_P(1/\sqrt{n})\,O_P(1)=O_P(1/\sqrt{n})$$

şeklinde yazılabildiğinden $n\to\infty$ iken $\bar{W}_n\xrightarrow{P}\mu$ ise $\bar{W}_n^2\xrightarrow{P}\mu^2$ olduğu görülür. Ayrıca, $\sqrt{n}(\bar{W}_n^2-\mu^2)=O_P(1)$ olduğundan asimptotik dağılım, Z standart normal rasgele değişkeni göstermek üzere, $\sqrt{n}(\bar{W}_n^2-\mu^2)$ nin asimptotik dağılımı $n\to\infty$ iken

$$\sqrt{n}(\bar{W}_n^2-\mu^2)=\sqrt{n}(\bar{W}_n-\mu)(\bar{W}_n+\mu)\xrightarrow{D}\sigma Z\,2\mu\equiv N(0,4\sigma^4\,\mu^2)$$

şeklinde olur.

b) Şimdi, $\mu>0$ ve her $i=1,2,3,\ldots$ için $W_i>0$ olsun. Bu durumda, $n\to\infty$ iken

$$\log(\bar{W}_n)\xrightarrow{P}\log(\mu)\ \text{ ve }\ \sqrt{n}(\log(\bar{W}_n)-\log(\mu))\xrightarrow{D}N(0,\,\sigma^2/\mu^2)$$

dir. $g(x)=\log(x)$ fonksiyonunun μ noktası komşuluğundaki birinci dereceden Taylor serisi açılımı,

$$g(\bar{W}_n)=g(\mu)+g'(\mu)(\bar{W}_n-\mu)+\frac{1}{2!}g''(\mu)(\bar{W}_n-\mu)^2+\frac{1}{3!}g^{(3)}(\mu)(\bar{W}_n-\mu)^3+\ldots$$

olup bunu

$$g(\bar{W}_n) = g(\mu) + g'(\mu)(\bar{W}_n - \mu) + \frac{1}{2!} g''(\mu_n^*)(\bar{W}_n - \mu)^2$$

şeklinde yazabiliriz. Burada $n \to \infty$ iken,

$$g(\bar{W}_n) - g(\mu) = g'(\mu)(\bar{W}_n - \mu) + O_P(1/\sqrt{n})$$

olacağından $n \to \infty$ iken $\log(\bar{W}_n) \xrightarrow{P} \log(\mu)$ elde edilir. Ayrıca, $g(x) = \log(x)$ için $g'(x) = 1/x$ ve $g''(x) = -1/x^2$ olup $\mu_n^* \to 1/\mu^2$ dir. Türevler yerine yazıldığında,

$$g(\bar{W}_n) - g(\mu) = \frac{1}{\mu}(\bar{W}_n - \mu) - \frac{1}{2\mu_n^*}(\bar{W}_n - \mu)^2$$

ve

$$g(\bar{W}_n) - g(\mu) = \frac{1}{\mu}(\bar{W}_n - \mu) + O_P(1/n)$$

eşitlikleri yazılır. Buradan da,

$$\sqrt{n}(g(\bar{W}_n) - g(\mu)) = \mu^{-1}\sqrt{n}(\bar{W}_n - \mu) + O_P(1/\sqrt{n})$$

eşitliğinden asimptotik dağılım

$$\sqrt{n}(g(\bar{W}_n) - g(\mu)) \xrightarrow{D} \mu^{-1}\sigma Z \equiv N(0\ ,\ \sigma^2 / \mu^2)$$

şeklinde elde edilir.

c) $g(x) = 1/x$ alınırsa, $g(\mu) = 1/\mu$ parametrik fonksiyonu için $1/\bar{W}_n$ tahmin edicisi düşünülebilir. Böylece, $g(\bar{W}_n) = 1/\bar{W}_n$ fonksiyonunun μ noktası komşuluğundaki Taylor serisi açılımı,

$$g(\bar{W}_n) = 1/\bar{W}_n = \mu^{-1} - \mu^{-2}(\bar{W}_n - \mu) + 0.5\mu^{-3}(\bar{W}_n - \mu)^2 - + \ldots$$

olup birinci derece açılım

$$\sqrt{n}(\bar{W}_n^{-1} - \mu^{-1}) = -\mu^{-2}\sqrt{n}(\bar{W}_n - \mu) + O_P(1/\sqrt{n})$$

şeklindedir. Buradan asimptotik dağılım $n \to \infty$ iken,

$$\sqrt{n}(\bar{W}_n^{-1} - \mu^{-1}) \xrightarrow{D} -\mu^{-2}\sigma Z \equiv N(0\ ,\ \sigma^2 / \mu^4)$$

şeklinde elde edilir.

d) (X_i, Y_i) iki boyutlu bağımsız ortalama vektörü ile varyans-kovaryans matrisi

$$\underset{\sim}{\mu} = \left(\mu_x, \mu_y\right)' \text{ ve } \Sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix}$$

olan aynı dağılımlı rasgele vektörlerin bir dizisi olsun. $Z_n = \bar{X}_n \bar{Y}_n$ istatistiğinin asimptotik dağılımını bulmak isteyelim.

Bunun için, $g(x,y) = x\,y$ denirse fonksiyonun her iki bileşenine göre türevleri vardır ve kısmi türevler,

$$\frac{\partial g(x,y)}{\partial x} = y\ ,\ \frac{\partial g(x,y)}{\partial y} = x\ ,\ \frac{\partial^2 g(x,y)}{\partial x^2} = \frac{\partial^2 g(x,y)}{\partial y^2} = 0 \quad \text{ve} \quad \frac{\partial^2 g(x,y)}{\partial x \partial y} = \frac{\partial^2 g(x,y)}{\partial y \partial x} = 1$$

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

$$\sqrt{n}\,(\bar{X}_n - \mu_x)/\sigma_x \xrightarrow{D} N(0,1) \quad \text{ve} \quad \sqrt{n}\,(\bar{Y}_n - \mu_y)/\sigma_y \xrightarrow{D} N(0,1)$$

dir. Diğer taraftan

$$(\bar{X}_n - \mu_x)(\bar{Y}_n - \mu_y) = O_P(1/\sqrt{n})O_P(1/\sqrt{n}) = O_P(1/n)$$

dir. Buradan, $g(x,y)$ fonksiyonunun $\underset{\sim}{\mu}$ noktası komşuluğundaki Taylor serisi açılımı,

$$g\left(\bar{X}_n, \bar{Y}_n\right) = g\left(\mu_x, \mu_y\right) + \left.\frac{\partial g}{\partial x}\right|_{x=\mu_x, y=\mu_y} \left(\bar{X}_n - \mu_x\right) + \left.\frac{\partial g}{\partial y}\right|_{x=\mu_x, y=\mu_y} \left(\bar{Y}_n - \mu_x\right)$$

$$+\frac{1}{2}\left(\bar{X}_n - \mu_x\,, \bar{Y}_n - \mu_x\right) \begin{bmatrix} \dfrac{\partial^2 g}{\partial x^2} & \dfrac{\partial^2 g}{\partial x \partial y} \\ \dfrac{\partial^2 g}{\partial y \partial x} & \dfrac{\partial^2 g}{\partial y^2} \end{bmatrix}_{x=\mu_x, y=\mu_y} \begin{bmatrix} \bar{X}_n - \mu_x \\ \bar{Y}_n - \mu_x \end{bmatrix} + kalan\ terim$$

şeklinde olup türevler yerine konulduğunda,

$$g\left(\bar{X}_n, \bar{Y}_n\right) = \bar{X}_n \bar{Y} = g\left(\mu_x, \mu_y\right) + \mu_y\left(\bar{X}_n - \mu_x\right) + \mu_x\left(\bar{Y}_n - \mu_x\right)$$

$$+\frac{1}{2}\left(\bar{X}_n - \mu_x\right)\left(\bar{Y}_n - \mu_x\right) + O_P\left(1/n\right) + \frac{1}{2}\left(\bar{Y}_n - \mu_x\right)\left(\bar{X}_n - \mu_x\right) + O_P\left(1/n\right)$$

eşitliği elde edilir. Buradan da asimptotik dağılım $n \to \infty$ iken,

$$\sqrt{n}\left(\bar{X}_n \bar{Y}_n - \mu_x \mu_y\right) = \mu_y \sqrt{n}\left(\bar{X}_n - \mu_x\right) + \mu_x \sqrt{n}\left(\bar{Y}_n - \mu_y\right)$$

$$+\frac{\sqrt{n}}{2}\left(\bar{X}_n - \mu_x\right)\left(\bar{Y}_n - \mu_y\right) + \frac{\sqrt{n}}{2}\left(\bar{Y}_n - \mu_y\right)\left(\bar{X}_n - \mu_x\right) + O_P\left(1/\sqrt{n}\right)$$

$$= \left(\mu_y,\ \mu_x\right) \begin{bmatrix} \sqrt{n}\left(\bar{X}_n - \mu_x\right) \\ \sqrt{n}\left(\bar{Y}_n - \mu_y\right) \end{bmatrix} + O_P\left(1/\sqrt{n}\right) \xrightarrow{D} N\left(0\ ,\ \sigma_z^2\right)$$

olarak elde edilir. Burada,

$$\sigma_z^2 = \left(\mu_y, \mu_x\right) \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix} \begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}$$

dir. Aslında bu sonuç, asimptotik dağılımın normal olduğunu (yani, $(\bar{X}_n - \mu_x, \bar{Y}_n - \mu_y)'$ nin iki boyutlu) olduğunu göstermeye yetmez.

Problemin çözümünü tamamlamak için, asimptotik dağılımın iki boyutlu normal olduğunun gösterilmesi gerekir. Merkezi limit teoremi tek değişkenli rasgele değişken dizileri için verildi. Bu sonuç, çok değişkenli rasgele değişken dizileri için genellenemeyebilir. Bu iki istatistiğin, herhangi bir lineer birleşiminin asimptotik normal olduğunun gösterilmesi gerekir (Serfling, 1980). Bunun için, $\underset{\sim}{\lambda} = (\lambda_1, \lambda_2)' \neq \underset{\sim}{0}$ ve $\underset{\sim}{\lambda}' \Sigma \underset{\sim}{\lambda} > 0$ olacak şekilde bir $\underset{\sim}{\lambda}$ için

$$\underset{\sim}{\lambda}' \begin{bmatrix} \sqrt{n}\left(\bar{X}_n - \mu_x\right) \\ \sqrt{n}\left(\bar{Y}_n - \mu_y\right) \end{bmatrix} = \left(\lambda_1, \lambda_2\right) \begin{bmatrix} \sqrt{n}\left(\bar{X}_n - \mu_x\right) \\ \sqrt{n}\left(\bar{Y}_n - \mu_y\right) \end{bmatrix} = \lambda_1 \sqrt{n}\left(\bar{X}_n - \mu_x\right) + \lambda_2 \sqrt{n}\left(\bar{Y}_n - \mu_y\right)$$

$$= \sqrt{n}\left[\lambda_1\left(\bar{X}_n - \mu_x\right) + \lambda_2\left(\bar{Y}_n - \mu_y\right)\right]$$

$$= \frac{\sqrt{n}}{n}\left[\sum_{i=1}^{n}\left(\lambda_1(X_i - \mu_x) + \lambda_2(Y_i - \mu_y)\right)\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i = \sqrt{n}\ \bar{Z}_n \xrightarrow{D} N(0, \sigma_z^2)$$

elde edilir. Burada, $Z_i = \lambda_1(X_i - \mu_x) + \lambda_2(Y_i - \mu_y)$ ler bağımsız beklenen değeri sıfır varyansı $\sigma_z^2 = \lambda_1^2 \sigma_{xx} + \lambda_2^2 \sigma_{yy} + 2\lambda_1 \lambda_2 \sigma_{yy} = \underset{\sim}{\lambda}' \Sigma \underset{\sim}{\lambda}$ dir$\oplus$