HAFTA 5

TAHMİN EDİCİLERDE ARANAN ÖZELLİKLER

7.9. Tahmin Edicileri Bulma Yöntemleri

Buraya kadar, tahmin edicilerin bazı özellikleri incelendi. Tahmin edicilerin bu özellikleri yanında başka özellikler de aranabilir. Yani, tahmin edicilerin özellikleri bunlarla sınırlı değildir. Eğer varsa, en iyi tahmin edicinin bulunması istenebilir. “En iyi tahmin edici” göreceli bir kavram olup bunun iyi tanımlanması gerekir. Bundan önceki kısımlarda, tahmin edicicilerin nasıl bulunacağı hakkında hiçbir şey söylenmedi. Yeterli tahmin edicilerin faktörizasyon teoremi yardımı ile bulunabileceğini biliyoruz. Ayrıca, yeterli tahmin ediciler tek olmayıp başka yeterli tahmin ediciler de vardır. Bu kısımda, tahmin edicilerin bulunma yöntemlerinden birkaç tanesi incelenecektir. Bunlar literatürde çok kullanılan yöntemlerdir. Momentler yöntemi, en çok olabilirlik yöntemi, Bayes yöntemi ve en küçük kareler yöntemi bu kısımda incelenecek yöntemlerdir. En küçük kareler yöntemi, iki değişken arasındaki ikişki ile ilgili olup çok geniş uygulama alanına sahiptir. Bu kısımda, en küçük kareler yöntemi üzerinde fazla durulmayacak, ileride ayrı bir bölüm olarak ele alınacaktır.

7.9.1. Momentler Yöntemi

Parametrelerin momentler tahmin edicilerinin bulunabilmesi için kitle momentlerinin var olması gerekir. Örneğin, Cauchy dağılımının parametresi için momentler tahmin edicisi bulunamaz. Momentler tahmin edicileri, kitle momentleri hesaplanarak örneklem momentlerine eşitlenmesi ile bulunur.

$X_1, X_2, \ldots, X_n$ olasılık veya olasılık yoğunluk fonksiyonu $f(x;\underset{\sim}{\theta})$ olan kitleden bir örneklem olsun. $\underset{\sim}{\theta} = (\theta_1, \theta_2, ..., \theta_k)'$ olmak üzere, kitle momentleri bu parametrelere bağlıdır. Yani, $s = 1, 2, 3, ..., k$ için $E_{\underset{\sim}{\theta}}(X^s) = g_s(\underset{\sim}{\theta})$ kitle momentleri ve

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \,,\ \ m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \,,\ \ m_3 = \frac{1}{n}\sum_{i=1}^{n} X_i^3 \,,\ \ldots,\ m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

örneklem momentleri hesaplanır. $\hat{\underset{\sim}{\theta}}_M = (\hat{\theta}_{1,M}, \hat{\theta}_{2,M}, ..., \hat{\theta}_{k,M})'$ olmak üzere,

$$g_s(\hat{\theta}_{1,M}, \hat{\theta}_{2,M}, ..., \hat{\theta}_{k,M}) = m_s \ ,\ s = 1, 2, 3, ..., k$$

eşitliklerinden elde edilecek $\hat{\theta}_{s,M}$ çözümü θ_s parametresinin *momentler tahmin edicisidir*. Burada, parametre sayısı kadar kitle momenti ile örneklem momenti hesaplanır. Kitlenin bir tane parametresi varsa, kitlenin beklenen değeri örneklem ortalamasına eşitlenir.

Örnek 7.9.1.1 a) $N(\mu,\sigma^2)$ dağılımından bir örneklem $X_1, X_2, \ldots, X_n$ olsun. Kitlenin iki tane parametresi olup dağılımın ilk iki momenti,

$$E_{\mu,\sigma^2}(X) = g_1(\mu,\sigma^2) = \mu \quad \text{ve} \quad E_{\mu,\sigma^2}(X^2) = g_1(\mu,\sigma^2) = \mu^2 + \sigma^2$$

dir. Buradan,

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n \quad \text{ve} \quad m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

şeklindeki m_1 ve m_2 örneklem momentleri kullanılarak,

$$\hat{\mu}_M = \bar{X}_n \quad \text{ve} \quad \hat{\mu}_M^2 + \hat{\sigma}_M^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

eşitliklerinden μ ve σ^2 nin momentler tahmin edicileri sırası ile

$$\hat{\mu}_M = \bar{X}_n \quad , \quad \hat{\sigma}_M^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

olarak bulunur. Momentler tahmin edicileri yanlı olabilir. Örneğin,

$$E\left(\hat{\sigma}_M^2\right) = E\left(\frac{n-1}{n} S_n^2\right) = \left(\frac{n-1}{n}\right)\sigma^2 \neq \sigma^2$$

olup yanlılığı $Bias_{\sigma^2}(\hat{\sigma}_M^2) = E(\hat{\sigma}_M^2) - \sigma^2 = -\sigma^2 / n$ dir.

b) $X_1, X_2, \ldots, X_n$ olasılık yoğunluk fonksiyonu,

$$f(x;\theta) = \begin{cases} \theta\, x^{\theta-1} & , \quad 0 < x < 1 \\ 0 & , \quad d.y. \end{cases}$$

olan dağılımdan bir örneklem olsun. Bu kitlenin bir tane parametresi vardır. Dağılımın beklenen değeri,

$$E_\theta(X) = \int_0^1 x\, f(x;\theta)\, dx = \int_0^1 x\,\theta\, x^{\theta-1}\, dx = \theta\int_0^1 x^\theta dx = \frac{\theta}{\theta+1}$$

olup $\hat{\theta}_M / (\hat{\theta}_M + 1) = \bar{X}_n$ eşitliğinden θ nın momentler tahmin edicisi $\hat{\theta}_M = \bar{X}_n / (1 - \bar{X}_n)$ olarak bulunur. Bu tahmin edicinin beklenen değerini ve varyansını hesaplamak kolay değildir (bu tahmin edici aslında karşıt orana (odds ratio) karşılık gelmektedir). Bu değerlere ihtiyaç duyulduğunda asimptotik sonuçlar kullanılabilir.

$E_\theta(X) = \mu = \theta/(\theta+1)$ ve $E_\theta(X^2) = \theta/(\theta+2)$

olup kitle varyansı

$$Var_\theta(X) = \sigma^2 = \frac{\theta}{\theta+2} - \left(\frac{\theta}{\theta+1}\right)^2 = \frac{\theta(\theta^2+\theta+1)}{(\theta+1)^2(\theta+2)}$$

dır. Buradan, $g(x) = x/(1-x)$ fonksiyonu için

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

şeklinde birinci dereceden Taylor serisi açılımından asimptotik dağılım

$$g(\bar{X}_n) \sim AN(g(\mu)\ ,\ [g'(\mu)]^2\sigma^2/n)$$

şeklinde yazılır. Burada, $\mu = \theta/(\theta+1)$ için, $g(\mu(\theta)) = \theta$ ve $g'(\mu(\theta)) = (\theta+1)^2$ dir. Dolayısı ile,

$$\sigma^2\left[g'(\mu)\right]^2 = \frac{\theta(\theta^2+\theta+1)}{(\theta+1)^2(\theta+2)}\left[(\theta+1)^2\right]^2 = \frac{\theta(\theta^2+\theta+1)(\theta+1)^2}{(\theta+2)}$$

olduğundan asimptotik dağılım $n \to \infty$ iken

$$g\left(\bar{X}_n\right) = \frac{\bar{X}_n}{1-\bar{X}_n} \sim AN\left(\theta\ ,\frac{\theta(\theta^2+\theta+1)(\theta+1)^2}{n(\theta+2)}\right)$$

olarak bulunur. Burada, $E_\theta(g(\bar{X}_n)) \cong \theta$ olup $E_\theta(g(\bar{X}_n)) \neq \theta$ dir.

c) $X_1, X_2, \ldots, X_n$ parametreleri n ve p olan Binom dağılımından bir örneklem olsun. $X_i \sim Binom(n, p)$ dağılımının ilk iki momenti,

$$E_{n,p}(X) = n\,p \text{ ve } E_{n,p}\left(X^2\right) = Var_{n,p}(X) + \left[E_{n,p}(X)\right]^2 = n\,p(1-p) + n^2p^2$$

olup $\tilde{n}_M$ ve $\tilde{p}_M$ parametrelerin momentler tahmin edicilerini göstermek üzere,

$$\tilde{n}_M\ \tilde{p}_M = \bar{X}_n \text{ ve } \tilde{n}\,\tilde{p}(1-\tilde{p}) + \tilde{n}^2\tilde{p}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

denklemlerinin çözümünden n ve p parametrelerinin momentler tahmin edicileri sırası ile,

$$\tilde{n}_M = \frac{\bar{X}_n^{\,2}}{\bar{X}_n - \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2} \quad , \quad \tilde{p}_M = \frac{\bar{X}_n}{\tilde{n}_M}$$

olarak bulunur $\oplus$

7.9.2. En Çok Olabilirlik Yöntemi

Olabilirlik yönteminin temel prensibi "Örneklem değerlerine bakarak, örneklem değerlerini elde etme olasılıklarının (veya olasılık yoğunluklarının) en yüksek olduğu değerlere karşılık gelen örneklem değerinin bilinmeyen parametre için bir tahmin olarak seçimidir."

En çok olabilirlik yöntemi, tahmin edicileri bulmak için kullanılan bir yöntemdir. En çok olabilirlik tahmin edicileri aşağıda tanımı verilen olabilirlik fonksiyonunu maksimum yapan değerlerdir.

Tanım 7.9.2.1 *(Olabilirlik fonksiyonu)* $X_1, X_2, \ldots, X_n$ olasılık veya olasılık yoğunluk fonksiyonu $f(x;\theta)$ olan kitleden bir örneklem olsun. θ nın olabilirlik fonksiyonu (likelihood function) $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)'$ olmak üzere,

$$L(\theta | \underset{\sim}{X} = \underset{\sim}{x}) = f(\underset{\sim}{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

dir $\otimes$

Olabilirlik fonksiyonunu maksimum yapan değer (örneklemin bir fonksiyonu) θ nın *en çok olabilirlik tahmin edicisidir*. Yani, θ nın en çok olabilirlik tahmin edicisi

$$\hat{\theta}_n(\underset{\sim}{X}) = \max_{\theta\in\Theta} L(\theta | \underset{\sim}{X})$$

dir. Olabilirlik fonksiyonunun maksimizasyonu yerine genellikle fonksiyonun logaritması (log-likelihood, $\ell(\theta) = \log(L(\theta | \underset{\sim}{X} = \underset{\sim}{x}))$) maksimize edilir. θ nın en çok olabilirlik tahmin edicisi olabilirlik fonksiyonunu veya log-olabilirlik fonksiyonunu maksimum yapan değerdir. Yani,

$$\hat{\theta}_n\left(\underset{\sim}{X}\right) = \max_{\theta\in\Theta} \ell\left(\theta\right) = \max_{\theta\in\Theta} L\left(\theta | \underset{\sim}{X}\right)$$

dir. Ancak, bazı durumlarda log-olabilirlik fonksiyonu tanımlı olmayabilir. Olabilirlik fonksiyonu parametrenin bir fonksiyonu olup, fonksiyonun maksimumu bulunmayabilir. Aşağıdaki örnekte, bazı dağılımların parametrelerine ilişkin en çok olabilirlik tahmin edicileri elde edilmiştir.

Örnek 7.9.2.1 a) $N(\mu,\sigma^2)$ dağılımından bir örneklem $X_1, X_2, \ldots, X_n$ olsun. Bu durumda olabilirlik fonksiyonu,

$$L(\mu,\sigma^2 | \underset{\sim}{X} = \underset{\sim}{x}) = f(\underset{\sim}{x} | \mu,\sigma^2) = \prod_{i=1}^{n} f(x_i | \mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

$$= \left(\frac{1}{2\pi}\right)^{n/2} \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

şeklindedir. Olabilirlik fonksiyonunun logaritması (log-olabilirlik fonksiyonunu) da

$$\ell(\mu,\sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

dir. Fonksiyonun birinci türevlerini sıfır yapan nokta veya noktalarda fonksiyon maksimum ya da minimuma sahiptir. Bu nokta veya noktaların maksimum olduğunu görmek için de ikinci türevlere bakılır. Buna göre birinci türevler;

$$\frac{\partial\ell(\mu,\sigma^2)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \quad , \quad \frac{\partial\ell(\mu,\sigma^2)}{\partial\sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

olup türevlerin sıfıra eşitlenmesi ile çözümler

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \quad , \quad -\frac{n}{\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

eşitliklerinden $\hat{\mu}_n = \bar{X}_n$ ve $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ şeklinde bulunmuştur. Yani, olabilirlik fonksiyonu bu noktalarda maksimum ya da minimum değerini alır. Bu noktalardaki ikinci türevler,

$$\left.\frac{\partial^2\ell(\mu,\sigma^2)}{\partial\mu^2}\right|_{\mu=\hat{\mu}_n,\sigma^2=\hat{\sigma}_n^2} = -\left.\frac{n}{\sigma^2}\right|_{\mu=\hat{\mu}_n,\sigma^2=\hat{\sigma}_n^2} < 0$$

ve

$$\left.\frac{\partial^2\ell(\mu,\sigma^2)}{\partial(\sigma^2)^2}\right|_{\mu=\hat{\mu}_n,\sigma^2=\hat{\sigma}_n^2} = \left.\frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2\right|_{\mu=\hat{\mu}_n,\sigma^2=\hat{\sigma}_n^2} < 0$$

şeklinde olduğundan μ ve σ^2 parametrelerinin en çok olabilirlik tahmin edicileri,

$$\hat{\mu}_n = \bar{X}_n \quad , \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

dir.

b) Şimdi, $X_1, X_2, \ldots, X_n$ Örnek (7.9.1.1b) de verilen kitleden bir örneklem olsun. X lerin olasılık yoğunluk fonksiyonu $0 < x < 1$ için $f(x;\theta) = \theta x^{\theta-1}$ olup θ nın olabilirlik fonksiyonu,

$$L\left(\theta \mid \underset{\sim}{X} = \underset{\sim}{x}\right) = \prod_{i=1}^{n} f\left(x_i;\theta\right) = \prod_{i=1}^{n}\theta x_i^{\theta-1} = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta} \prod_{i=1}^{n}\frac{1}{x_i}$$

ve log-olabilirlik fonksiyonu da,

$$\ell\left(\theta\right) = \ln\left(L\left(\theta \mid \underset{\sim}{X} = \underset{\sim}{x}\right)\right) = n\ln\left(\theta\right) + \theta\sum_{i=1}^{n}\ln\left(x_i\right) - \sum_{i=1}^{n}\ln(x_i)$$

şeklindedir. Birinci türevin sıfıra eşitlenmesi ile

$$\frac{\partial \ell(\theta)}{\partial \ell \theta}=\frac{n}{\theta}+\sum_{i=1}^{n}\ln(x_i)=0$$

eşitliğinden çözüm

$$\hat{\theta}=-\frac{n}{\sum_{i=1}^{n}\ln(X_i)}$$

olarak bulunur. Bu çözümün ikinci türevde yerine konulması ile

$$\left.\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}}=-\left.\frac{n}{\theta^2}\right|_{\theta=\hat{\theta}}<0$$

olduğu görülür. Buradan θ nın en çok olabilirlik tahmin edicisi,

$$\hat{\theta}=-\frac{n}{\sum_{i=1}^{n}\ln(X_i)}$$

dir. Bu tahmin ediciyi, $Y_i=-\ln(X_i)$ olmak üzere $\hat{\theta}_n=1/\overline{Y}_n$ şeklinde yazabiliriz. Şimdi $\hat{\theta}_n$ nın bazı özelliklerini inceleyelim. Önce, $Y=-\ln(X)$ denirse Y nin değer kümesi $D_Y=\mathbb{R}^+$ olup olasılık yoğunluk fonksiyonunun,

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

olduğu gösterilebilir. Buna göre $Y \sim \ddot{U}stel(1/\theta)$ olup,

$$W=-\sum_{i=1}^{n}\ln(X_i)=\sum_{i=1}^{n}Y_i \sim Gamma(n,1/\theta)$$

dır. Ayrıca W nun olasılık yoğunluk fonksiyonu,

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

olup $\hat{\theta}_n$ nın beklenen değeri,

$$E_\theta(\hat{\theta}_n)=E_\theta(n/W)=n\int_0^\infty \frac{1}{w}\frac{\theta^n w^{n-1}e^{-w\theta}}{\Gamma(n)}dw=\frac{n\theta^n}{\Gamma(n)}\int_0^\infty w^{n-2}e^{-w\theta}dw$$

$$=\frac{n\theta^n}{\Gamma(n)}\frac{\Gamma(n-1)}{\theta^{n-1}}=\frac{n}{n-1}\theta$$

olarak bulunmuştur. Bu tahmin edici yansız değildir. Ancak, $\lim_{n\to\infty} E_\theta(\hat{\theta}_n) = \theta$ olduğundan en çok olabilirlik tahmin edicisi asimptotik yansızdır. Benzer şekilde, $\hat{\theta}_n$ nin ikinci momenti,

$$E_\theta(\hat{\theta}_n^2) = E_\theta(n^2 / W^2) = n^2 \int_0^\infty \frac{1}{w^2} \frac{\theta^n w^{n-1} e^{-w\theta}}{\Gamma(n)} dw = \frac{n^2\,\theta^n}{\Gamma(n)} \int_0^\infty w^{n-2} e^{-w\theta} dw = \frac{n^2\theta^2}{(n-1)(n-2)}$$

olarak hesaplanmıştır. Buradan, en çok olabilirlik tahmin edicisinin varyansı da,

$$Var_\theta(\hat{\theta}_n) = E_\theta(\hat{\theta}_n^2) - \left[E_\theta\left(\hat{\theta}_n\right)\right]^2 = \frac{n^2\theta^2}{(n-1)(n-2)} - \left(\frac{n\,\theta}{n-1}\right)^2 = \frac{n^2\theta^2}{(n-1)^2(n-2)}$$

olur. Buna göre, $n\to\infty$ iken $E_\theta(\hat{\theta}_n) \to \theta$ ve $Var_\theta(\hat{\theta}_n) \to 0$ olduğundan $\hat{\theta}_n$ en çok olabilirlik tahmin edicisi tutarlıdır.

c) $X_1, X_2, \ldots, X_n$ beklenen değeri θ olan Üstel dağılımdan bir örneklem olsun. θ nın olabilirlik fonksiyonu,

$$L(\theta | \underset{\sim}{X} = \underset{\sim}{x}) = \prod_{i=1}^n f(x_i;\theta) = \prod_{i=1}^n \frac{1}{\theta} e^{-x_i/\theta} = \frac{1}{\theta^n} \exp\left(-\frac{1}{\theta}\sum_{i=1}^n x_i\right)$$

ve log-olabilirlik fonksiyonu da,

$$\ell(\theta) = \ln(L(\theta | \underset{\sim}{X} = \underset{\sim}{x})) = -n\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^n x_i$$

şeklindedir. log-olabilirlik fonksiyonunun birinci türevinin sıfıra eşitlenmesi ile,

$$\frac{\partial \ell(\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n x_i = 0$$

elde edilir. Buradan da çözüm $\hat{\theta}_n = \bar{X}_n$ olur. Ayrıca,

$$\left.\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right|_{\theta=\hat{\theta}} = \frac{n}{\bar{X}_n^{\,2}} - \frac{2n\bar{X}_n^{\,2}}{\bar{X}_n^{\,3}} = -\frac{n}{\bar{X}_n^{\,2}} < 0$$

olduğundan θ nın en çok olabilirlik tahmin edicisi $\hat{\theta}_n = \bar{X}_n$ dir.

d) Bernoulli dağılımının parametresinin en çok olabilirlik tahmin edicisini bulalım. Olabilirlik fonksiyonu,

$$L\left(p | \underset{\sim}{X} = \underset{\sim}{x}\right) = \prod_{i=1}^n f\left(x_i;p\right) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

den log-olabilirlik fonksiyonu,

$$\ell(p)=\ln\left(L\left(p|\underset{\sim}{X}=\underset{\sim}{x}\right)\right)=\ln(p)\sum_{i=1}^{n}x_i+\ln(1-p)\left(n-\sum_{i=1}^{n}x_i\right)$$

olarak yazılır. Buradan, birinci türevin sıfıra eşitlenmesi ile

$$\frac{\partial\ell(p)}{\partial p}=\frac{\sum_{i=1}^{n}x_i}{p}-\frac{n-\sum_{i=1}^{n}x_i}{1-p}=0$$

eşitliğinden çözüm $\hat{p}_n=\bar{X}_n$ olarak elde edilir. $p=\bar{X}_n$ noktasında ikinci türev negatif olup p nin en çok olabilirlik tahmin edicisi $\hat{p}_n=\bar{X}_n$ dir$\oplus$

Bazen parametrelerin yerine, parametrelerin fonksiyonunun en çok olabilirlik tahmin edicisinin değerine ihtiyaç duyulur. Örneğin, Poisson dağılımında λ yerine $P_\lambda(X=0)=e^{-\lambda}$ şeklinde $P_\lambda(X=0)$ olasılığının en çok olabilirlik tahmini bulunmak istenebilir. Dağılımın esas parametresinin en çok olabilirlik tahmin edicisi bulunduktan sonra parametrenin fonksiyonunun en çok olabilirlik tahmin edicisi bulunabilir (Hogg, McKean ve Craig (2005), sayfa 316). Aşağıdaki teorem bunu ifade etmektedir. Fonksiyon üzerinde herhangi bir kısıtlama söz konusu olmayıp, fonksiyonun bire bir olması halinde teoremin ispatı açıktır.

Teorem 7.9.2.1 θ nın en çok olabilirlik tahmin edicisi $\hat{\theta}_n$ ise $\tau(\hat{\theta}_n)$ de $\tau(\theta)$ nın en çok olabilirlik tahmin edicisidir$\Diamond$

Parametrelerin en çok olabilirlik tahmin edicilerinin, olabilirlik fonksiyonunun türevlenebilir olması halinde nasıl bulunacağını yukarıda gördük. Olabilirlik fonksiyonu bazen türevlenemeyebilir.

Örnek 7.9.2.2 a) $X_1,X_2,\ldots,X_n$ beklenen değeri μ, varyansı σ^2 olan normal dağılımdan bir örneklem ise, μ nün en çok olabilirlik tahmin edicisi $\bar{X}_n$ olup μ^2 nin en çok olabilirlik tahmin edicisi $\bar{X}_n{}^2$ dir (Teorem (7.9.2.1)).

b) $X_1,X_2,\ldots,X_n$ olasılık yoğunluk fonksiyonu

$$f(x;\theta)=\frac{1}{2}e^{-|x-\theta|}\quad,\ x\in\mathbb{R}$$

olan kitleden bir örneklem olsun. Bu dağılım Laplace dağılımı veya çift üstel (double exponential) dağılım olarak da bilinir. Buna göre log-olabilirlik fonksiyonu,

$$\ell(\theta)=-n\ln(2)-\sum_{i=1}^{n}\left|X_i-\theta\right|$$

olup fonksiyonun birinci türevi

$$\frac{\partial \ell(\theta)}{\partial \theta} = \sum_{i=1}^{n} \operatorname{sgn}(x_i - \theta)$$

dir.

Burada, $\operatorname{sgn}(x)$ işaret fonksiyonu olup değeri x in durumuna göre, -1, 0 veya 1 olabilir. Ayrıca, $x \neq 0$ için $\partial(|x|)/\partial x = \operatorname{sgn}(x)$ dir. Birinci türevin sıfıra eşitlenmesi ile, θ nın en çok olabilirlik tahmin edicisi $\hat{\theta}_n = Medyan\{X_1, X_2, \ldots, X_n\} = Q_2$ olur (medyan verileri tam ortadan ikiye ayırır).

c) $X_1, X_2, \ldots, X_n$ parametresi θ olan düzgün dağılımdan bir örneklem olsun. Yani, X lerin olasılık yoğunluk fonksiyonu,

$$f(x;\theta) = \begin{cases} 1/\theta & , \quad 0 < x \leq \theta \\ 0 & , \quad d.y. \end{cases}$$

olup θ nın olabilirlik fonksiyonu, $L(\theta | \underset{\sim}{X} = \underset{\sim}{x}) = \theta^{-n} I_{\{0 < x_{(n)} \leq \theta\}}(\underset{\sim}{x})$ şeklinde yazılabilir.

Burada, olabilirlik fonksiyonu θ ya göre türevlenemez. Fonksiyon, $\theta \geq x_{(n)}$ için θ nın azalan bir fonksiyonudur (diğer yerlerde sıfırdır). Böylece, θ nın en küçük değerinde olabilirlik fonksiyonu maksimum değere ulaşır. Dolayısı ile, θ nın en çok olabilirlik tahmin edicisi, $\hat{\theta}_n = X_{(n)}$ dır$\oplus$

Tahmin edici, örneklem içindeki bilgiyi özetleyen bir rasgele değişkendir. Bu özet bilgi ile, kitle parametreleri hakkında bazı istatistiki sonuç çıkarımlar yapılır. En çok olabilirlik yöntemine göre, bu veri indirgemedeki temel prensip, gözlenen herhangi iki değer için (bunlar $\underset{\sim}{x}$ ve $\underset{\sim}{y}$ olsun) olabilirlik fonksiyonları $L(\theta | \underset{\sim}{x})$ ve $L(\theta | \underset{\sim}{y})$ olmak üzere,

$$L(\theta | \underset{\sim}{x}) / L(\theta | \underset{\sim}{y}) = c(\underset{\sim}{x}, \underset{\sim}{y})$$

özelliğine sahip ise $\underset{\sim}{x}$ veya $\underset{\sim}{y}$ üzerinden yapılacak sonuç çıkarımlar aynıdır.

7.9.3. Bayes Tahmin Edicileri

Kitle parametreleri genellikle rasgele olmayan sabitlerdir. Bayes yönteminde, kitle parametreleri de rasgele değişken olarak göz önüne alınır. Bu parametreler, alabileceği değerlere ilişkin inancın gücünü yansıtan önsel dağılımlara uyan rasgele değişkenlerdir. Bayes yönteminde, θ nın bir dağılımına (*önsel dağılım* ya da prior distribution) ihtiyaç duyulur. Önsel sezgilerin örneklemden çıkarılan bilgi ile karşılaştırılması yapılır. θ nın önsel dağılımın yanında, örneklem bilgisini de yansıtan bir sonsal dağılım kullanılır.

$X_1, X_2, \ldots, X_n$ örneklemi verildiğinde önce önsel dağılım belirlenir. Daha sonra verilen örneklemden sonsal dağılım bulunur. Sonsal dağılımın beklenen değeri θ parametresinin *Bayes tahmin edicisidir*. Bayes tahmin edicilerini bulmak için aşağıdaki yol izlenir.

Önce önsel dağılım belirlenir (bu $\pi(\theta)$ olsun). θ verildiğinde, örneklemin olasılık veya olasılık yoğunluk fonksiyonu yazılır. Bu olasılık fonksiyonu $f(x \mid \theta)$ olsun. X ler ile θ nın ortak olasılık yoğunluk fonksiyonu da $f(\underset{\sim}{x};\theta) = f(\underset{\sim}{x} \mid \theta)\, \pi(\theta)$ eşitliğinden elde edilir. X ler verildiğinde θ nın *koşullu dağılımı* (*posterior* veya *sonsal dağılım,* $\pi(\theta \mid \underset{\sim}{X} = \underset{\sim}{x})$) bulunarak, θ nın Bayes tahmin edicisi bu sonsal dağılımın beklenen değeridir. Yani, θ nın Bayes tahmin edicisi

$$\hat{\theta}_B = E\left(\theta \mid X_1, X_2, \ldots, X_n\right)$$

dir.

Örnek 7.9.3.1 a) p verildiğinde $X_1, X_2, \ldots, X_n$ Bernoulli dağılımından bir örneklem olsun. p de parametreleri α ve β olan Beta dağılımına sahip bir rasgele değişken olsun. $p \sim Beta(\alpha, \beta)$, p verildiğinde $X \sim Bern(p)$ ve $Y = \sum_{i=1}^{n} X_i \sim Binom(n, p)$ dir. Burada, Y yeterli olduğundan, parametre hakkında örneklem içindeki tüm bilgi Y tarafından özetlenir. Önsel dağılımın olasılık yoğunluk fonksiyonu $0 < p < 1$ için

$$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, p^{\alpha-1}(1-p)^{\beta-1}$$

olup Y ile p nın ortak olasılık fonksiyonu,

$$f(y;p) = f\left(y \mid p\right)\pi(p) = \left[\binom{n}{y} p^{y}(1-p)^{n-y}\right]\left[\frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}\, p^{\alpha-1}(1-p)^{\beta-1}\right]$$

$$= \binom{n}{y} \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}\, p^{\alpha+y-1}(1-p)^{n-y+\beta-1} \quad , \; y = 0,1,2,\ldots,n \text{ ve } 0 < p < 1$$

dir. Y nin marjinal olasılık fonksiyonu $y = 0,1,2,3,\ldots,n$ için

$$f_Y(y) = \int_{p=0}^{1} f(y;p)\,dp = \binom{n}{y} \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)} \int_{0}^{1} p^{\alpha+y-1}(1-p)^{n-y+\beta-1} dp$$

$$= \binom{n}{y} \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)} \frac{\Gamma\left(\alpha+y\right)\Gamma\left(n-y+\beta\right)}{\Gamma\left(\alpha+\beta+n\right)}$$

şeklinde olup sonsal (posterior) dağılım (sadeleştirmelerden sonra) $0 < p < 1$ için

$$\pi(p \mid y) = \frac{f(y;p)}{f_Y(y)} = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+y)\Gamma(n-y+\beta)} p^{\alpha+y-1}(1-p)^{n-y+\beta-1}$$

olarak bulunmuştur (Casella ve Berger, 2002, sayfa 324). Yani, sonsal dağılım da parametreleri $y+n$ ve $n-y+\beta$ olan Betadır. Buradan p nin Bayes tahmin edicisi (sonsal dağılımın beklenen değeri)

$$\hat{p}_B = E\left(p \mid \underset{\sim}{X}\right) = \frac{y+\alpha}{n+\alpha+\beta} = \frac{\sum_{i=1}^{n} x_i}{n+\alpha+\beta} + \frac{\alpha}{n+\alpha+\beta}$$

$$= \frac{\alpha+\beta}{n+\alpha+\beta}\frac{\alpha+\beta}{n+\alpha+\beta} + \frac{n}{n+\alpha+\beta}\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\alpha+\beta}{n+\alpha+\beta}\mu + \frac{n}{n+\alpha+\beta}\bar{X}_n = c_1\mu + c_2\bar{X}_n$$

şeklinde elde edilmiştir. Burada, μ önsel dağılımın beklenen değeridir. Önsel dağılım parametreleri α ve β olan Beta dağılımı seçildiğinde sonsal dağılım da farklı parametreler ile yine Beta dağılımı olarak bulundu. Önsel dağılım ile sonsal dağılım farklı olabilir. Ancak, örnekte de görüldüğü gibi, Bayes tahmin edicisi önsel dağılımın beklenen değeri ile örneklem ortalamasının lineer birleşimidir. Bu genellikle doğrudur.

Bayes yönteminde önsel dağılımın seçimi önemlidir. Bu seçim için bazı olasılık kuralları dikkate alınmalıdır. Örneğin bu örnekte, önsel dağılım olarak başka bir dağılım da ele alınabilirdi. Ancak, aranan Bayes tahmin edicisi Binom dağılımının başarı olasılığıdır. Dolayısı ile, öncel dağılım olarak tanım kümesi $(0,1)$ aralığı olan bir dağılım seçilmelidir.

b) θ beklenen değeri μ olan üstel dağılıma sahip bir rasgele değişken (önsel dağılım üstel) ve θ verildiğinde $X_1, X_2, \ldots, X_n$ de beklenen değeri θ olan Poisson dağılımından bir örneklem olsun. θ nın Bayes tahmin edicisini bulalım. Koşullu dağılım Poisson olduğundan ortak olasılık fonksiyonu,

$$P(\underset{\sim}{X} = \underset{\sim}{x} \mid \theta) = \prod_{i=1}^{n} P(X_i = x_i \mid \theta) = e^{-n\theta}\, \theta^{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} 1/x_i\,!$$

şeklinde olup faktörizasyon teoreminden $T = \sum_{i=1}^{n} X_i$, θ için yeterlidir. Ayrıca, θ verildiğinde, T nin koşullu dağılımı beklenen değeri $n\theta$ olan Poissondur. Buradan T ve θ nın ortak olasılık fonksiyonu,

$$f_{T,\theta}(t,\theta) = f_T(t\,|\,\theta)\,\pi(\theta) = \frac{1}{\mu} e^{-\theta/\mu} e^{-n\theta} \theta^t \prod_{i=1}^{n} 1/x_i! = \frac{1}{\mu} e^{-(n+1/\mu)\theta} \theta^t \prod_{i=1}^{n} 1/x_i!$$

olup T nin marjinal olasılık fonksiyonu da $t = 0,1,2,3,...$ için

$$f_T(t) = \frac{1}{\mu}\prod_{i=1}^{n} 1/x_i! \left[\int_{\theta=0}^{\infty} e^{-(n+1/\mu)\theta}\theta^t\,d\theta\right] = \left[\frac{1}{\mu}\prod_{i=1}^{n} 1/x_i!\right]\left[\frac{\mu^{t+1}\Gamma(t+1)}{(n\mu+1)^{t+1}}\right]$$

şeklindedir. Sonsal dağılım da koşullu olasılık yoğunluk fonksiyonunun tanımından $\theta > 0$ için

$$\pi(\theta|T=t) = \frac{f(t,\theta)}{f_T(t)} = \frac{(n\mu+1)^{t+1}}{\mu^{t+1}\Gamma(t+1)}\theta^t e^{-(n+1/\mu)\theta}$$

olarak bulunmuştur. Dolayısı ile, $T = t$ verildiğinde θ nın beklenen değeri

$$E(\theta|T=t) = \frac{(n\mu+1)^{t+1}}{\mu^{t+1}\Gamma(t+1)}\int_{\theta=0}^{\infty}\theta^t e^{-(n+1/\mu)\theta}d\theta = \frac{(n\mu+1)^{t+1}}{\mu^{t+1}\Gamma(t+1)}\frac{\mu^{t+2}\Gamma(t+2)}{(n\mu+1)^{t+2}} = \frac{\mu(t+1)}{n\mu+1}$$

dir. Koşullu beklenen değerde, t yerine T yazıldığında, θ nın Bayes tahmin edicisi

$$\hat{\theta}_B = E(\theta|T) = \frac{\mu(T+1)}{n\mu+1} = \frac{1}{n\mu+1}\mu + \frac{n}{n\mu+1}\bar{X}_n = c_1\mu + c_2\bar{X}_n$$

şeklinde düzenlenebilir. Yine, Bayes tahmin edicisi önsel dağılımın beklenen değeri ile örneklem ortalamasının lineer birleşimidir.

c) θ verildiğinde $X_1, X_2, \ldots, X_n$ beklenen değeri θ, varyansı σ^2 olan normal dağılımdan bir örneklem olsun. θ da beklenen değeri μ, varyansı τ^2 olan normal dağılıma sahip bir rasgele değişken olsun (önsel dağılım da normal). Buna göre, θ nın Bayes tahmin edicisini bulalım (μ, σ^2 ve τ^2 biliniyor). Faktörizasyon teoreminden, $\bar{X}_n$ örneklem ortalaması θ için yeterlidir. θ verildiğinde $\bar{X}_n$ nin koşullu dağılımı da normaldir ($\bar{X}_n \sim N(\theta, \sigma^2/n)$). Buradan, $\bar{X}_n$ ve θ nın ortak olasılık yoğunluk fonksiyon, $\bar{x} \in \mathbb{R}$ ve $\theta \in \mathbb{R}$ için,

$$\begin{aligned} f(\bar{x};\theta) = f(\bar{x}\,|\,\theta)\,\pi(\theta) &= \frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2\right)\frac{1}{\sqrt{2\pi\tau^2}}\exp\left(-\frac{1}{2\tau^2}(\theta-\mu)^2\right) \\ &= \frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}}\frac{1}{\sqrt{2\pi\tau^2}}\exp\left(-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2\right) \\ &= \frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}}\frac{1}{\sqrt{2\pi\tau^2}}\exp\left(-\frac{1}{2\kappa^2}(\theta-\delta(x))^2 - \frac{1}{\tau^2+\sigma^2/n}(\bar{x}-\mu)^2\right) \end{aligned}$$

şeklinde yazılabilir. Burada,

$$\delta(x)=\frac{\tau^2\bar{x}+\left(\sigma^2/n\right)\mu}{\tau^2+\sigma^2/n} \quad \text{ve } \kappa^2=\frac{\tau^2\sigma^2/n}{\tau^2+\sigma^2/n}$$

dir. $\bar{X}_n$ nin marjinal olasılık yoğunluk fonksiyonu da,

$$c=\frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}}\frac{1}{\sqrt{2\pi\tau^2}}=\sqrt{n}/\left(2\pi\sigma\tau\right)$$

olmak üzere,

$$f_{\bar{X}_n}(\bar{x})=\int_{-\infty}^{\infty} f(\bar{x};\theta)\,d\theta=c\exp\left(-\frac{1}{\tau^2+\sigma^2/n}(\bar{x}-\mu)^2\right)\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2\kappa^2}(\theta-\delta(x))^2\right)d\theta$$

$$=c\exp\left(-\frac{1}{\tau^2+\sigma^2/n}(\bar{x}-\mu)^2\right)\sqrt{2\pi\kappa^2}=\exp\left(-\frac{1}{\tau^2+\sigma^2/n}(\bar{x}-\mu)^2\right)\frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}}\frac{\sqrt{2\pi\kappa^2}}{\sqrt{2\pi\tau^2}}$$

olarak elde edilmiştir. Bu ifade biraz daha düzenlendiğinde, $\bar{X}_n$ nin marjinal olasılık yoğunluk fonksiyonu $\bar{x}\in\mathbb{R}$ için

$$f_{\bar{X}_n}(\bar{x})=\exp\left(-\frac{1}{\tau^2+\sigma^2/n}(\bar{x}-\mu)^2\right)\frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}}\frac{\sqrt{2\pi\kappa^2}}{\sqrt{2\pi\tau^2}}=\exp\left(-\frac{1}{\tau^2+\sigma^2/n}(\bar{x}-\mu)^2\right)\frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}}\frac{\kappa}{\tau}$$

$$=\exp\left(-\frac{1}{\tau^2+\sigma^2/n}(\bar{x}-\mu)^2\right)\frac{\sqrt{n}}{\sqrt{2\pi\sigma^2}}\frac{\sqrt{\dfrac{\tau^2\sigma^2/n}{\tau^2+\sigma^2/n}}}{\tau}=\frac{1}{\sqrt{\tau^2+\sigma^2/n}}\exp\left(-\frac{1}{\tau^2+\sigma^2/n}(\bar{x}-\mu)^2\right)$$

şeklinde yazılabilir. Yani, $\bar{X}_n$ nin marjinal dağılımı, $\bar{X}_n\sim N(\mu,\,(\tau^2+\sigma^2/n))$ dir. Ayrıca $\bar{X}_n=\bar{x}$ olarak verildiğinde θ nın koşullu olasılık yoğunluk fonksiyonu (sonsal dağılım) $\theta\in\mathbb{R}$ için,

$$\pi(\theta\,|\,\bar{x})=\frac{f(\bar{x};\theta)}{f_{\bar{X}_n}(\bar{x})}=\frac{\dfrac{\sqrt{n}}{\sqrt{2\pi\sigma^2}}\dfrac{1}{\sqrt{2\pi\tau^2}}\exp\left(-\dfrac{1}{2\kappa^2}(\theta-\delta(x))^2-\dfrac{1}{\tau^2+\sigma^2/n}(\bar{x}-\mu)^2\right)}{\dfrac{1}{\sqrt{\tau^2+\sigma^2/n}}\exp\left(-\dfrac{1}{\tau^2+\sigma^2/n}(\bar{x}-\mu)^2\right)}$$

$$=\frac{\sqrt{n}\sqrt{2\pi\left(\tau^2+\sigma^2/n\right)}}{\sqrt{2\pi\sigma^2}\sqrt{2\pi\tau^2}}\exp\left(-\frac{1}{2\kappa^2}(\theta-\delta(x))^2\right)$$

$$=\frac{\sqrt{\left(n\tau^2+\sigma^2\right)}}{\sqrt{2\pi\sigma^2\tau^2}}\exp\left(-\frac{1}{2\kappa^2}(\theta-\delta(x))^2\right)=\frac{1}{\sqrt{2\pi\kappa^2}}\exp\left(-\frac{1}{2\kappa^2}(\theta-\delta(x))^2\right)$$

olarak bulunmuştur. Yani, sonsal dağılım da normaldir ($\theta | \bar{X}_n = \bar{x} \sim N(\delta(x), \kappa^2)$). Dolayısı ile θ nın Bayes tahmin edicisi (sonsal dağılımın beklenen değeri),

$$\hat{\theta}_B = \delta\left(\bar{X}_n\right) = \frac{\tau^2 \bar{X}_n + (\sigma^2 / n)\mu}{\tau^2 + \sigma^2 / n} = \frac{\sigma^2 / n}{\tau^2 + \sigma^2 / n}\mu + \frac{\tau^2}{\tau^2 + \sigma^2 / n}\bar{X}_n = c_1\mu + c_2\bar{X}_n$$

şeklinde bulunmuştur $\oplus$

Bayes tahmin edicileri genel olarak yanlıdır. Önsel dağılımın seçimine göre değişir. Yansızlık tahmin edicilerde aranan önemli özelliklerden biri olmasına rağmen, bazen yanlı tahmin ediciler tercih edilebilir. Örneğin, en çok olabilirlik tahmin edicileri bazen yanlı (asimptotik yansız) olup daha küçük varyanslı olduğundan tercih edilebilir.