10. HAFTA

Çok Değişkenli Normal Dağılımdan Örneklem ve En Çok Olabilirlik Tahmin Yöntemi (MLE)

Bu bölümde, çok değişkenli normal bir kitleden alınan $\underline{X}_1, \underline{X}_2, ..., \underline{X}_n$ rasgele örneklemlerden elde edilen örneklem ortalama vektörü $\overline{\underline{X}}$ ve örneklem varyans-kovaryans matrisi S istatistiklerin örneklem dağılımları verilecektir.

Çok Değişkenli Normal Olabilirlik

$px1$ tipindeki $\underline{X}_1, \underline{X}_2, ..., \underline{X}_n$ rasgele vektörleri, ortalama vektörü $\underline{\mu}$ ve varyans-kovaryans matrisi Σ olan çok değişkenli normal bir kitleden rasgele bir örneklem olsun. $\underline{X}_1, \underline{X}_2, ..., \underline{X}_n$'ler karşılıklı bağımsız ve her birinin dağılımı $N_p(\underline{\mu}, \Sigma)$ olduğundan, bu rasgele vektörlerin ortak olasılık yoğunluk fonksiyonu, marjinal normal yoğunluk fonksiyonlarının çarpımına eşittir. Böylece ortak olasılık yoğunluk fonksiyonu

$$
\begin{aligned}
f_{\underline{X}_1, \underline{X}_2, ..., \underline{X}_n}(\underline{x}_1, \underline{x}_2, ..., \underline{x}_n) &= \prod_{j=1}^{n} f_{\underline{X}_j}(\underline{x}_j) \\
&= \prod_{j=1}^{n} \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\underline{x}_j - \underline{\mu})' \Sigma^{-1} (\underline{x}_j - \underline{\mu})} \\
&= \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} e^{-\frac{1}{2}\sum_{j=1}^{n}(\underline{x}_j - \underline{\mu})' \Sigma^{-1} (\underline{x}_j - \underline{\mu})}
\end{aligned}
$$

olarak elde edilir.

Ortak olasılık yoğunluk fonksiyonunu maksimum yapan parametre değerlerine en çok olabilirlik tahmin edicileri denir. En çok olabilirlik tahmin edicilerin bulunması için kullanılan tekniğe de en çok olabilirlik tahmin yöntemi adı verilir.

Sonuç: A , kxk tipinde simetrik bir matris ve $\underline{x}$, $kx1$ tipinde bir vektör olsun.

a) $\underline{x}' A \underline{x} = tr(\underline{x}' A \underline{x}) = tr(A \underline{x}\underline{x}')$

b) $tr(A) = \sum_{i=1}^{k} \lambda_i$, λ_i'ler A'nın özdeğerleridir.

İspat:

a) $\underline{x}'_{1xk} A_{kxk} \underline{x}_{kx1}$ ifadesinin değeri bir sabittir ve sabit bir değerin tr'i kendisidir. Böylece

$\underline{x}'A\underline{x} = tr(\underline{x}'A\underline{x})$ dır. Ayrıca B_{mxk} ve C_{kxm} matrisleri için $tr(BC) = tr(CB)$ olduğundan $tr(\underline{x}'A\underline{x}) = tr(A\underline{x}\underline{x}')$ dır ve sonuçta $\underline{x}'A\underline{x} = tr(\underline{x}'A\underline{x}) = tr(A\underline{x}\underline{x}')$ olur.

b) λ_i'ler A'nın özdeğerleri ve $\underline{e}_i$'ler bu özdeğerlere karşılık gelen birim özvektörler olsun. Buradan,

$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix}$ diagonel bir matris ve $P_{kxk} = \begin{bmatrix} \underline{e}_1 & \underline{e}_2 & \cdots & \underline{e}_k \end{bmatrix}$,

$PP' = P'P = I$ olan ortogonal bir matris olmak üzere, Spektral ayrışımdan $A = P'\Lambda P$ dır. Böylece,

$$\begin{aligned} tr(A) &= tr(P'\Lambda P) \\ &= tr(\Lambda PP') \\ &= tr(\Lambda) \\ &= \lambda_1 + \lambda_2 + \cdots + \lambda_k \\ &= \sum_{i=1}^{k} \lambda_i \end{aligned}$$

elde edilir.

Çok değişkenli ortak olasılık yoğunluk fonksiyonundaki üstel ifade yukarıda verilen sonucun (a) seçeneğinden,

$$\begin{aligned} (\underline{x}_j - \underline{\mu})' \Sigma^{-1} (\underline{x}_j - \underline{\mu}) &= tr\left[(\underline{x}_j - \underline{\mu})' \Sigma^{-1} (\underline{x}_j - \underline{\mu}) \right] \\ &= tr\left[\Sigma^{-1} (\underline{x}_j - \underline{\mu})(\underline{x}_j - \underline{\mu})' \right] \end{aligned}$$

dir ve her iki tarafın toplamı alınır ise

$$\begin{aligned} \sum_{j=1}^{n} (\underline{x}_j - \underline{\mu})' \Sigma^{-1} (\underline{x}_j - \underline{\mu}) &= \sum_{j=1}^{n} tr\left[\Sigma^{-1} (\underline{x}_j - \underline{\mu})(\underline{x}_j - \underline{\mu})' \right] \\ &= tr\left[\Sigma^{-1} \left(\sum_{j=1}^{n} (\underline{x}_j - \underline{\mu})(\underline{x}_j - \underline{\mu})' \right) \right] \end{aligned}$$

elde edilir. $\sum_{j=1}^{n} (\underline{x}_j - \underline{\mu})(\underline{x}_j - \underline{\mu})'$ ifadesindeki $(\underline{x}_j - \underline{\mu})$'ye $\overline{\underline{x}} = \frac{1}{n}\sum_{j=1}^{n} \underline{x}_j$ bir eklenip, bir çıkarılır ise

$$\sum_{j=1}^{n}(\underline{x}_j-\underline{\bar{x}}+\underline{\bar{x}}-\underline{\mu})(\underline{x}_j-\underline{\bar{x}}+\underline{\bar{x}}-\underline{\mu})'=\sum_{j=1}^{n}(\underline{x}_j-\underline{\bar{x}})(\underline{x}_j-\underline{\bar{x}})'+\sum_{j=1}^{n}(\underline{\bar{x}}-\underline{\mu})(\underline{\bar{x}}-\underline{\mu})'$$
$$=\sum_{j=1}^{n}(\underline{x}_j-\underline{\bar{x}})(\underline{x}_j-\underline{\bar{x}})'+n(\underline{\bar{x}}-\underline{\mu})(\underline{\bar{x}}-\underline{\mu})'$$

sonucu elde edilir. Burada, $\sum_{j=1}^{n}(\underline{x}_j-\underline{\bar{x}})(\underline{\bar{x}}-\underline{\mu})'$ ve $\sum_{j=1}^{n}(\underline{\bar{x}}-\underline{\mu})(\underline{x}_j-\underline{\bar{x}})'$ çapraz çarpım ifadelerine ilişkin matrisler sıfırdır.

Bu sonuçlardan, çok değişkenli normal kitleden alınan rasgele örnekleme ilişkin ortak olasılık yoğunluk fonksiyonu

$$f_{\underline{X}_1,\underline{X}_2,\cdots,\underline{X}_n}(\underline{x}_1,\underline{x}_2,\cdots,\underline{x}_n)=\frac{1}{(2\pi)^{np/2}\left|\Sigma\right|^{n/2}}e^{-\frac{1}{2}tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\underline{x}_j-\underline{\bar{x}})(\underline{x}_j-\underline{\bar{x}})'+n(\underline{\bar{x}}-\underline{\mu})(\underline{\bar{x}}-\underline{\mu})'\right)\right]}$$

biçimindedir.

Bu yoğunluk fonksiyonunda $\underline{x}_1,\underline{x}_2,\cdots,\underline{x}_n$ gözlem değerleri yazılır ise olabilirlik fonksiyonu elde edilmiş olur ve $L(\underline{\mu},\Sigma)$ ile gösterilir. Ortak olasılık yoğunluk fonksiyonu $\underline{x}_j$ 'lerin fonksiyonu iken, olabilirlik fonksiyonu bilinmeyen $\underline{\mu}$ ve Σ kitle parametrelerinin bir fonksiyonudur. Böylece $\underline{x}_j$ vektörü birimlere ilişkin gözlem değerleri ise olabilirlik fonksiyonu

$$L(\underline{\mu},\Sigma)=\frac{1}{(2\pi)^{np/2}\left|\Sigma\right|^{n/2}}e^{-\frac{1}{2}tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\underline{x}_j-\underline{\bar{x}})(\underline{x}_j-\underline{\bar{x}})'+n(\underline{\bar{x}}-\underline{\mu})(\underline{\bar{x}}-\underline{\mu})'\right)\right]}$$

dir. Ayrıca

$$tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\underline{x}_j-\underline{\bar{x}})(\underline{x}_j-\underline{\bar{x}})'+n(\underline{\bar{x}}-\underline{\mu})(\underline{\bar{x}}-\underline{\mu})'\right)\right]=tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\underline{x}_j-\underline{\bar{x}})(\underline{x}_j-\underline{\bar{x}})'\right)\right]+ntr\left[\Sigma^{-1}(\underline{\bar{x}}-\underline{\mu})(\underline{\bar{x}}-\underline{\mu})'\right]$$
$$=tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\underline{x}_j-\underline{\bar{x}})(\underline{x}_j-\underline{\bar{x}})'\right)\right]+n(\underline{\bar{x}}-\underline{\mu})'\,\Sigma^{-1}(\underline{\bar{x}}-\underline{\mu})$$

olmak üzere

$$L(\underline{\mu},\Sigma)=\frac{1}{(2\pi)^{np/2}\left|\Sigma\right|^{n/2}}e^{-\frac{1}{2}tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\underline{x}_j-\underline{\bar{x}})(\underline{x}_j-\underline{\bar{x}})'\right)\right]+n(\underline{\bar{x}}-\underline{\mu})'\,\Sigma^{-1}(\underline{\bar{x}}-\underline{\mu})}$$

biçiminde elde edilir.

$\underline{\mu}$ ve Σ kitle parametrelerinin En Çok Olabilirlik Tahmin Edicileri

Bu bölümde $\underline{\mu}$ ve Σ kitle parametrelerinin en çok olabilirlik tahmin edicileri elde edilecek.

Sonuç 9: B, pxp tipinde simetrik ve pozitif tanımlı bir matris ve $b>0$ olan bir sabit için

$$\frac{1}{|\Sigma|^{b}} e^{-\frac{1}{2}tr(\Sigma^{-1}B)} \leq \frac{1}{|B|^{b}}(2b)^{pb} e^{-bp}$$

dir, bu eşitlik bütün pozitif tanımlı Σ_{pxp} matrisi için geçerlidir. $\Sigma = \frac{1}{2b}B$ için eşitlik sözkonusudur.

Sonuç: $\underline{X}_1, \underline{X}_2, \cdots, \underline{X}_n$, ortalama vektörü $\underline{\mu}$ ve varyans-kovaryans matrisi Σ olan p boyutlu normal dağılımdan rasgele bir örneklem olsun. Buradan

$$\hat{\underline{\mu}} = \overline{\underline{X}} = \frac{1}{n}\sum_{j=1}^{n} \underline{X}_{ij} \text{ ve } \hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(\underline{X}_{ij} - \overline{\underline{X}})(\underline{X}_{ij} - \overline{\underline{X}})' = \frac{(n-1)}{n} S = S_n$$

$\underline{\mu}$ ve Σ kitle parametrelerinin en çok olabilirlik tahmin edicileridir. Bunların gözlem değerleri $\overline{\underline{\mathrm{x}}} = \frac{1}{n}\sum_{j=1}^{n} \underline{\mathrm{x}}_{ij}$ ve $\hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(\underline{\mathrm{x}}_{ij} - \overline{\underline{\mathrm{x}}})(\underline{\mathrm{x}}_{ij} - \overline{\underline{\mathrm{x}}})'$ 'ye $\underline{\mu}$ ve Σ kitle parametrelerinin en çok olabilirlik tahmin değerleri denir.

İspat: Olabilirlik fonksiyonundaki *e* üzerindeki ifade ½ çarpanı gözönüne alınmadan

$$tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\underline{\mathrm{x}}_{j} - \overline{\underline{\mathrm{x}}})(\underline{\mathrm{x}}_{j} - \overline{\underline{\mathrm{x}}})'\right)\right] + n(\overline{\underline{\mathrm{x}}} - \underline{\mu})'\,\Sigma^{-1}(\overline{\underline{\mathrm{x}}} - \underline{\mu})$$

dır. Σ^{-1} pozitif tanımlı olduğundan ve $\underline{\mu}=\overline{\underline{\mathrm{x}}}$ olmadığı müddetçe $(\overline{\underline{\mathrm{x}}} - \underline{\mu})'\,\Sigma^{-1}(\overline{\underline{\mathrm{x}}} - \underline{\mu}) > 0$ dır. Böylece olabilirlik fonksiyonu $\hat{\underline{\mu}}=\overline{\underline{\mathrm{x}}}$ de $\underline{\mu}$'ye göre maksimum olur. Buradan olabilirlik fonksiyonu

$$L(\hat{\underline{\mu}}, \Sigma) = \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} e^{-\frac{1}{2}tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(\underline{\mathrm{x}}_{j} - \overline{\underline{\mathrm{x}}})(\underline{\mathrm{x}}_{j} - \overline{\underline{\mathrm{x}}})'\right)\right]}$$

ifadesi Σ üzerinden maksimize edilebilir. Bir önceki sonuçta $b = \frac{n}{2}$ ve $B = \sum_{j=1}^{n} (\underline{\mathrm{x}}_j - \underline{\bar{\mathrm{x}}})(\underline{\mathrm{x}}_j - \underline{\bar{\mathrm{x}}})'$ alınırsa, olabilirlik fonksiyonu $\hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n} (\underline{\mathrm{x}}_j - \underline{\bar{\mathrm{x}}})(\underline{\mathrm{x}}_j - \underline{\bar{\mathrm{x}}})'$ de maksimuma ulaşır.

$\underline{\bar{X}}$ en çok olabilirlik tahmin edicisi bir rasgele vektör ve $\hat{\Sigma}$ en çok olabilirlik tahmin edicisi bir rasgele matristir. Bu tahmin edicilerin, en çok olabilirlik tahminleri verilen veri kümesinden elde edilen değerlerdir. Bununla birlikte, olabilirlik fonksiyonunun maksimumu

$$L(\underline{\hat{\mu}}, \hat{\Sigma}) = \frac{1}{(2\pi)^{np/2}} e^{-\frac{np}{2}} \frac{1}{\left|\hat{\Sigma}\right|^{n/2}}$$

dir.

$\underline{\bar{X}}$ ve S ‘nin Örneklem Dağılımı

Tek değişkenli durumda $X_1, X_2, \cdots, X_n \sim N(\mu, \sigma^2)$ dağılımdan rasgele bir örneklem olsun.

Buradan, $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$ örneklem ortalaması ve $S^2 = \frac{1}{n-1}\sum_{j=1}^{n} (X_j - \bar{X})^2$ örneklem varyansı tahmin edicilerin, örneklem dağılımları sırasıyla $\bar{X} \sim N(\mu, \frac{1}{n}\sigma^2)$ ve $(n-1)S^2 = \sum_{j=1}^{n} (X_j - \bar{X})^2 \sim \sigma^2 \chi^2_{n-1}$ dir.

Çok değişkenli durumda $\underline{X}_1, \underline{X}_2, \cdots, \underline{X}_n \sim N_p(\underline{\mu}, \Sigma)$ dağılımdan rasgele bir örneklem olsun.

Buradan, $\underline{\bar{X}} = \frac{1}{n}\sum_{j=1}^{n} \underline{X}_j$ örneklem ortalama vektörü ve $S = \frac{1}{n-1}\sum_{j=1}^{n} (\underline{X}_j - \underline{\bar{X}})(\underline{X}_j - \underline{\bar{X}})'$ örneklem varyans-kovaryans matrisi tahmin edicilerin örneklem dağılımları sırasıyla $\underline{\bar{X}} \sim N_p(\underline{\mu}\ ,\ \frac{1}{n}\Sigma)$ ve $(n-1)S \sim W_{n-1}(\bullet|\Sigma)$ dir. Burada W Wishart dağılımını göstermektedir. Bununla birlikte $\underline{\bar{X}}$ ve S bağımsızdır.