수리통계 II HW5

2020-15709 나승찬

1.15. $pdf_X(x) = \frac{1}{\theta} I_{[0,\theta]}(x)$, $L(\theta) = \frac{1}{\theta^n}$ $(\theta \ge X_{(n)})$

$l(\theta) = \log L(\theta) = -n\log\theta$ $(\theta \ge X_{(n)})$

$\dot{l}(\theta) = -\frac{n}{\theta} < 0$이므로 $l(\theta)$는 θ의 감소함수이고,

$\hat{\theta}^{MLE} = \hat{\theta} = X_{(n)}$이다. 이제 순서통계량의 분포에 따라 $pdf_{X_{(n)}}(x_{(n)}) = \frac{n!}{(n-1)!}\left(\frac{1}{\theta}x_{(n)}\right)^{n-1}\frac{1}{\theta}I_{[0,\theta]}(x_{(n)})$

$= \frac{n}{\theta^n}x_{(n)}^{n-1}I_{[0,\theta]}(x_{(n)})$이 $X_{(n)}$의 확률밀도함수이다.

(a) $H_0: \theta = \theta_0$ vs. $H_1: \theta \ne \theta_0$

영가설공간 MLE는 $\theta_0 \ge X_{(n)}$인 경우 $\hat{\theta}_0 = \theta_0$이다. 그러나 $\theta_0 < X_{(n)}$이면 $L(\theta)$가 정의되지 않는다. 이는 Unif$[0, \theta_0]$에서의 표본이 θ_0보다 큰 값을 가지는 사건이므로 영가설공간에서 절대 발생하지 않는 사건이다. 이 경우 H_0을 기각하고, $\theta_0 \ge X_{(n)}$인 경우를 살펴보자.

$2(l(\hat{\theta}) - l(\theta_0)) = -2n\log\frac{X_{(n)}}{\theta_0}I(\theta_0 \ge X_{(n)}) \ge$ const.

$\Leftrightarrow \log\frac{X_{(n)}}{\theta_0} \le$ const. $\left(\frac{X_{(n)}}{\theta_0} \le 1\right) \Leftrightarrow \frac{X_{(n)}}{\theta_0} \le C$ (기각역의 형태)

이제 유의수준 α에 대응하는 C의 값을 구하면

$P_{\theta=\theta_0}\left(\frac{X_{(n)}}{\theta_0} \le C\right) + P_{\theta=\theta_0}(X_{(n)} > \theta_0) = \alpha$이고 $P_{\theta=\theta_0}(X_{(n)} > \theta_0) = 0$.

$P_{\theta=\theta_0}(X_{(n)} \le C\theta_0) = \int_0^{C\theta_0}\frac{n}{\theta_0^n}x^{n-1}dx = C^n$이므로 $C = \sqrt[n]{\alpha}$.

$\therefore$ 기각역은 $X_{(n)} \le \theta_0 \cdot \sqrt[n]{\alpha}$ 또는 $X_{(n)} > \theta_0$

(b) $H_0: \theta \ge \theta_0$ vs. $H_1: \theta \ne \theta_0$인 복합가설이므로 영가설공간 MLE $\underset{\theta \ge \theta_0}{\arg\max}(-n\log\theta) = \hat{\theta}_0 = \max\{\theta_0, X_{(n)}\}$이다.

$$2(l(\hat{\theta}) - l(\hat{\theta}_0)) = -2n\log\frac{X_{(n)}}{\hat{\theta}_0} = \begin{cases} -2n\log\frac{X_{(n)}}{\theta_0} & X_{(n)} \le \theta_0 \\ 0 & X_{(n)} > \theta_0 \end{cases}$$

즉 $2(l(\hat{\theta}) - l(\hat{\theta}_0)) \ge$ const.인 기각역의 꼴은 (a)에서 계산한 것처럼 $X_{(n)} \le C\theta_0$이고, $X_{(n)} > \theta_0$에서 이 값을 기각역으로 할 수 없다($0 >$ const.가 될 수 없다.). 즉 유의수준 α인 $C = \sqrt[n]{\alpha}$이고, 유의수준 α 기각역은 $X_{(n)} \le \sqrt[n]{\alpha} \cdot \theta_0$이다.

1.12 (1) $S_i^2 := \frac{1}{n_i}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2$, $\bar{X}_i := \frac{1}{n_i}\sum_{j=1}^{n_i}X_{ij}$ $(i=1,2)$ 로 두면 $(X_{11}, \cdots, X_{1n_1}, X_{21}, \cdots X_{2n_2})$로부터의 로그가능도는

$l(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = -\frac{n_1}{2}\log(2\pi\sigma_1^2) - \frac{n_2}{2}\log(2\pi\sigma_2^2) - \frac{n_1}{2\sigma_1^2}(S_1^2 + (\bar{X}_1 - \mu_1)^2) - \frac{n_2}{2\sigma_2^2}(S_2^2 + (\bar{X}_2 - \mu_2)^2)$

즉 $\hat{\mu}_i = \bar{X}_i$, $\hat{\sigma}_i^2 = S_i^2$ from $l_{\mu_i} = \frac{2n_i}{2\sigma_i^2}(\bar{X}_i - \mu_i)$, $l_{\sigma_i^2} = -\frac{n_i}{2\sigma_i^2} + \frac{n_i}{2\sigma_i^4}(S_i^2 + (\bar{X}_i - \hat{\mu}_i)^2)$이고 고정된 σ_i^2에 대해 l은 μ의 증가함수)

(2) $H_0: \sigma_1^2 \le \sigma_2^2$에서 MLE를 구하면 앞과 같은 방식으로 $\hat{\mu}_i = \bar{X}_i$, $\underset{\sigma_1^2 \le \sigma_2^2}{\arg\min}\left(n_1\log(2\pi\sigma_1^2) + n_2\log(2\pi\sigma_2^2) + \frac{n_1S_1^2}{\sigma_1^2} + \frac{n_2S_2^2}{\sigma_2^2}\right)$을 찾자.

Θ_0 MLE는 (i) 전역 MLE S_1^2, S_2^2에 대해 $S_1^2 \le S_2^2$이 성립하면 영공간 MLE = 전역 MLE이고 $\hat{\sigma}_{i,0}^2 = S_i^2$이고 (ii) $S_1^2 > S_2^2$인 경우 위의 식은 $\sigma_1^2 = \sigma_2^2$일 때 최소이므로 공통분산추정량 $\hat{\sigma}_0^2$을 구하면 $\hat{\sigma}_0^2 = \frac{n_1S_1^2 + n_2S_2^2}{n_1 + n_2} = \hat{\sigma}_1^2 = \hat{\sigma}_2^2$이다. 즉 영공간 MLE는

$$(\hat{\mu}_{1,0}, \hat{\mu}_{2,0}, \hat{\sigma}_{1,0}^2, \hat{\sigma}_{2,0}^2)^T = \begin{cases} (\bar{X}_1, \bar{X}_2, S_1^2, S_2^2)^T & (S_1^2 \le S_2^2) \\ \left(\bar{X}_1, \bar{X}_2, \frac{n_1S_1^2 + n_2S_2^2}{n_1+n_2}, \frac{n_1S_1^2 + n_2S_2^2}{n_1+n_2}\right)^T & (S_1^2 > S_2^2) \end{cases}$$

(3) $2(l(\hat{\theta}) - l(\hat{\theta}_0)) = \{-n_1\log(2\pi S_1^2) - n_2\log(2\pi S_2^2) + n_1\log(2\pi\hat{\sigma}_{1,0}^2) + n_2\log(2\pi\hat{\sigma}_{2,0}^2)\}I\{S_1^2 > S_2^2\}$이고, $S_1^2 \le S_2^2$이면 $l(\hat{\theta}) = l(\hat{\theta}_0)$이어서 H_0을 채택하게 되므로 $S_1^2 > S_2^2$인 경우만 생각하면 위의 식은 $n_1\log\frac{\hat{\sigma}_{1,0}^2}{S_1^2} + n_2\log\frac{\hat{\sigma}_{2,0}^2}{S_2^2}$이다.

식을 정리하면 $n_1\log\left(\frac{n_1S_1^2 + n_2S_2^2}{(n_1+n_2)S_1^2}\right) + n_2\log\left(\frac{n_1S_1^2 + n_2S_2^2}{(n_1+n_2)S_2^2}\right)$

$= n_1\log\left(\frac{n_1}{n_1+n_2}\right)^{-1} + (n_1+n_2)\log\left(\frac{n_1}{n_1+n_2} \times \frac{S_1^2}{S_2^2} + \frac{n_2}{n_1+n_2}\right) \ge C$

$\Leftrightarrow \frac{S_1^2}{S_2^2} \ge C$, $S_1^2 \ge S_2^2$인 경우(기각역의 형태)

(4) 검정력함수 $\gamma(\theta) = \gamma(\sigma_1^2, \sigma_2^2) = P_\theta(S_1^2/S_2^2 \ge C)$

$= P\left(\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \ge \frac{C\sigma_1^2}{\sigma_2^2}\right)$이고, $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$이므로

$\gamma\left(\frac{\sigma_1^2}{\sigma_2^2}\right) = P\left(\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \ge C \cdot \frac{\sigma_1^2}{\sigma_2^2}\right)$은 $\frac{\sigma_1^2}{\sigma_2^2}$의 증가함수이다.

$\sup_{\sigma_1^2 \le \sigma_2^2}\gamma(\theta) = \alpha$인 C는 $\sup_{\frac{\sigma_1^2}{\sigma_2^2} \le 1}\gamma\left(\frac{\sigma_1^2}{\sigma_2^2}\right) = \alpha$를 찾으면 되므로 $\sigma_1^2 = \sigma_2^2$이면 된다. 즉 $P(S_1^2/S_2^2 \ge C) = \alpha$, $S_1^2/S_2^2 \sim F(n_1-1, n_2-1)$이도록 C를 정하면

기각역: $\frac{S_1^2}{S_2^2} \ge F_\alpha(n_1-1, n_2-1)$

1.13 (1) $L(\mu,\sigma)=\prod_{i=1}^{n}\left(\frac{1}{\sigma}e^{-\frac{x_i-\mu}{\sigma}}\right) \quad (\mu \le x_{(1)})$

$l(\mu,\sigma)=\log L(\mu,\sigma)=-n\left(\log\sigma+\frac{\bar{x}-\mu}{\sigma}\right)$
$(-\infty<\mu\le x_{(1)},\ 0<\sigma<\infty)$

전역 MLE를 구하면 $l_\mu=\frac{n}{\sigma}>0$이므로 $l(\mu,\sigma)$는 각 σ에서 μ의 증가함수이므로 $\hat{\mu}=X_{(1)}$이다.
한편 $-l(\sigma):=\log\sigma+\frac{\bar{x}-x_{(1)}}{\sigma}$에서 $\dot{l}(\sigma)=0$인 σ는 $\bar{x}-x_{(1)}$이므로 $\hat{\sigma}=\bar{X}-X_{(1)}$이다. ($\because -\dot{l}(\sigma)=\frac{1}{\sigma}-\frac{\bar{x}-x_{(1)}}{\sigma^2}$, $(-\ddot{l})(\sigma)=-\frac{1}{\sigma^2}+\frac{2(\bar{x}-x_{(1)})}{\sigma^3}$)

(2) 이제 $H_0: \mu=\mu_0$에서의 영공간 MLE를 구하면
(i) $X_{(1)}<\mu_0$이면 X_i가 Θ_0 분포로부터 추출된 가능성은 0이고, (ii) $X_{(1)}\ge\mu_0$이면 이런 문제는 없다. 우선 Θ_0의 $\hat{\mu}_0=\mu_0$이고, $\hat{\sigma}_0$을 구하면 $\dot{l}(\sigma)=-\frac{n}{\sigma}+\frac{\bar{x}-\mu}{\sigma^2}$이므로 $\hat{\sigma}_0=\bar{x}-\hat{\mu}_0=\bar{x}-\mu_0$이다.

(3) 한편, 표준지수분포의 순서통계량에 대해 다음의 대의적 표현이 성립한다. (예4.3.3) 「$(X_{(r)})_{1\le r\le n}\overset{d}{=}\left(\frac{1}{n}Z_1+\cdots+\frac{1}{n-r+1}Z_r\right)_{1\le r\le n}$, $Z_r\overset{iid}{\sim}\mathrm{Exp}(1)$, $X_{(r)}$은 $\mathrm{Exp}(1)$로부터의 n개 표본에 기초한 순서통계량 of rth」
또한, 7.13~7.14 문제에 주어진 $X\sim\mathrm{Exp}(\mu,\sigma)$에 대해 $(X-\mu)/\sigma\sim\mathrm{Exp}(1)$이다. ($y=\frac{x-\mu}{\sigma}$로 두면 이는 one-to-one이고 $pdf_Y(y)=\sigma\cdot\frac{1}{\sigma}e^{-\frac{\sigma y+\mu-\mu}{\sigma}}I_{(\mu,\infty)}(\sigma y+\mu)=e^{-y}I_{(0,\infty)}(y)$, $Y\sim\mathrm{Exp}(1)$)
또한 (예4.3.3)으로부터 $(X-\mu)/\sigma\sim\mathrm{Exp}(1)$임을 적용하면 $n(X_{(1)}-\mu)/\sigma,\ (n-r+1)(X_{(r)}-X_{(r-1)})/\sigma\ (r=2,\cdots,n)\overset{iid}{\sim}\mathrm{Exp}(1)$이다.

(4) $2(l(\hat{\theta})-l(\hat{\theta}_0))=-2n\left(\log\hat{\sigma}+\frac{\bar{x}-\hat{\mu}}{\hat{\sigma}}\right)+2n\left(\log\hat{\sigma}_0+\frac{\bar{x}-\hat{\mu}_0}{\hat{\sigma}_0}\right)$
$=2n\log\frac{\hat{\sigma}_0}{\hat{\sigma}}=2n\log\frac{\bar{x}-\mu_0}{\bar{x}-x_{(1)}}$이고, (i)에서 밝혔듯 이 값은 $X_{(1)}\ge\mu_0$에서만 정의되어 $\bar{x}-\mu_0>\bar{x}-x_{(1)}$이다. 이 범위에서
$2n\log\frac{\bar{x}-\mu_0}{\bar{x}-\hat{\mu}}\ge c\Leftrightarrow\frac{\bar{x}-\mu_0}{\bar{x}-\hat{\mu}}=1+\frac{\hat{\mu}-\mu_0}{\bar{x}-\hat{\mu}}\ge c\Leftrightarrow\frac{\hat{\mu}-\mu_0}{\hat{\sigma}}\ge c$

(5) 유의수준 α에 대응하는 기각역의 c를 구하자.
기각역의 형태: $X_{(1)}\ge\mu_0$에서 $\frac{\hat{\mu}-\mu_0}{\hat{\sigma}}\ge c$, $X_{(1)}<\mu_0$
$\sup_{\sigma}\left[P_{\mu_0,\sigma}\left(\frac{\hat{\mu}-\mu_0}{\hat{\sigma}}\ge c\right)+P_{\mu_0,\sigma}(X_{(1)}<\mu_0)\right]=\alpha$인 c 찾자.
$P_{\mu_0}(X_{(1)}<\mu_0)=0$이고, (3)으로부터 $2n(X_{(1)}-\mu)/\sigma\sim\mathrm{Gamma}(1,2)\overset{d}{=}\chi^2(2)$, $2\sum_{r=2}^{n}(n-r+1)(X_{(r)}-X_{(r-1)})/\sigma=\frac{2n\hat{\sigma}}{\sigma}\sim\mathrm{Gamma}(n-1,2)\overset{d}{=}\chi^2(2n-2)$

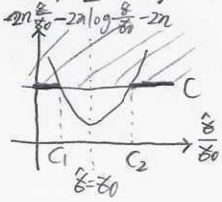
$\hat{\sigma}^{UN}:=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-X_{(1)})$에 대해 (H₀이 사실일 때) 분산의 ~~불편추정량~~ $\hat{\sigma}^{UN}$, 모분산 σ는 $\frac{(2n-2)\hat{\sigma}^{UN}}{\sigma}\sim\chi^2(2n-2)$이므로

$\frac{\hat{\mu}-\mu_0}{\hat{\sigma}}=\frac{X_{(1)}-\mu_0}{\frac{n}{n-1}\hat{\sigma}^{UN}}=\frac{n(X_{(1)}-\mu_0)/2}{\hat{\sigma}^{UN}/(2n-2)}=\frac{\left(\frac{n(X_{(1)}-\mu_0)}{\sigma}\right)/2}{(\hat{\sigma}^{UN}/\sigma)/(2n-2)}$
$\overset{d}{=}\frac{\chi^2(2)/2}{\chi^2(2n-2)/(2n-2)}\overset{d}{=}F(2,2n-2)$이므로 유의수준 α에 대응하는 $P\left(\frac{\hat{\mu}-\mu_0}{\hat{\sigma}}\ge c\right)=\alpha$는 F분포 따르는 검정통계량에서 결정, 기각역: $\begin{cases}\frac{\hat{\mu}-\mu_0}{\hat{\sigma}}\ge F_\alpha(2,2n-2)\\ X_{(1)}<\mu_0\end{cases}$

1.14. $H_0: \sigma=\sigma_0$으로 두고, (1)과 (3), (5)의 결과를 활용한다.

(6) Θ_0에서의 MLE는 $\sigma=\sigma_0$이고, 이 경우에도 $l(\mu)$는 μ의 증가함수이므로 $\hat{\mu}_0=X_{(1)}$이다. 이 경우에도 $\sigma_0>0$이다.

(7) $l(\hat{\theta})-l(\hat{\theta}_0)=-2n\left(\log\hat{\sigma}+\frac{\bar{x}-\hat{\mu}}{\hat{\sigma}}\right)+2n\left(\log\sigma_0+\frac{\bar{x}-\hat{\mu}_0}{\sigma_0}\right)$
$=2n\log\frac{\sigma_0}{\hat{\sigma}}+\frac{2n\hat{\sigma}}{\sigma_0}-2n$이므로 기각역의 형태는
$\begin{cases}\frac{\hat{\sigma}}{\sigma_0}\le C_1 \text{ or } \frac{\hat{\sigma}}{\sigma_0}\ge C_2\\ C_1-\log C_1=C_2-\log C_2\end{cases}$ $(C_1, C_2>0)$



(8) $\frac{2n\hat{\sigma}}{\sigma}\sim\chi^2(2n-2)$이므로
유의수준 α 기각역은
$P\left(\frac{\hat{\sigma}}{\sigma_0}\le C_1\right)+P(\hat{\sigma}/\sigma_0\ge C_2)=\alpha$
$\Leftrightarrow\begin{cases}P(C_1\sigma_0\le\hat{\sigma}\le C_2\sigma_0)=1-\alpha\\ C_1-\log C_1=C_2-\log C_2\end{cases}$를 만족시킨 C_1, C_2에 대해 $\hat{\sigma}\le\sigma_0C_1$ 또는 $\hat{\sigma}\ge\sigma_0C_2$로 한다.
(단, $P(C_1\sigma_0\le\hat{\sigma}\le C_2\sigma_0)=\int_{C_1\sigma_0}^{C_2\sigma_0}\frac{1}{\Gamma(2n-2)2^{2n-2}}x^{2n-3}e^{-\frac{x}{2}}dx$)

7.16.

(a) $L(p_1, \cdots, p_k) = \prod_{i=1}^{k} p_i^{x_i}(1-p_i)^{n_i - x_i}$

$$l(p_1, \cdots, p_k) = \log L(p_1, \cdots, p_k) = \sum_{i=1}^{k}\left(x_i \log p_i + (n_i - x_i)\log(1-p_i)\right)$$

where $x_i = \sum_{j=1}^{n_i} x_{ij} = X_{i\cdot}$

$$\frac{\partial l}{\partial p_i} = \frac{x_i}{p_i} - \frac{n_i - x_i}{1-p_i}, \quad \frac{\partial^2 l}{\partial p_i \partial p_j} = \begin{cases} 0 & (i \neq j) \\ -\frac{x_i}{p_i^2} - \frac{(n-x_i)}{(1-p_i)^2} & (i=j) \end{cases}$$

이므로 $\hat{p} = (\hat{p}_1, \cdots, \hat{p}_k)$ (전역 MLE)는 가능도방정식의 근인 $\hat{p}_i = \frac{x_i}{n_i}$ 로 주어진다.

(영공간 MLE) p_0, $\hat{p}_0$에 대해 $\sum_{i=1}^{k}\left(x_i \log p_0 + (n_i - x_i)\log(1-p_0)\right) = l(p_0)$

$\dot{l}(p_0) = \frac{\Sigma(x_i)}{p_0} - \frac{\Sigma(n_i - x_i)}{1-p_0}$, $\ddot{l}(p_0) = -\frac{\Sigma x_i}{p_0^2} - \frac{\Sigma(n_i - x_i)}{(1-p_0)^2} < 0$이므로 $\hat{p}_{i0} = \frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{k} n_i}$

$$2\left(l(\hat{p}_1, \cdots, \hat{p}_k) - l(\hat{p}_{10}, \cdots, \hat{p}_{k0})\right) = 2\sum_{i=1}^{k}\left\{x_i \log\frac{\hat{p}_i}{\hat{p}_0} + (n_i - x_i)\log\frac{1-\hat{p}_i}{1-\hat{p}_0}\right\}$$

이때 $1-\hat{p}_i = \hat{q}_i$, $1-\hat{p}_0 = \hat{q}_0$으로 두면 $x_i = n_i\hat{p}_i$, $n_i - x_i = n_i - n_i\hat{p}_i = n_i\hat{q}_i$이므로 위 식은 $2\sum_{i=1}^{k} n_i\left\{\hat{p}_i \log\frac{\hat{p}_i}{\hat{p}_{i0}} + \hat{q}_i \log\frac{\hat{q}_i}{\hat{q}_{i0}}\right\}$

(b) $x\log x = (x-1) + \frac{1}{2}(x-1)^2 + o(x^2)$이므로 $(x \approx 1)$

$\frac{\hat{p}_i}{\hat{p}_{i0}} \xrightarrow{n_i \to \infty} 1$, $\frac{\hat{q}_i}{\hat{q}_{i0}} \xrightarrow{n_i \to \infty} 1$을 위의 매클로린 급수에 대입한다.

$$2\sum_{i=1}^{k} n_i\left\{\hat{p}_{i0}\left(\frac{\hat{p}_i}{\hat{p}_{i0}} - 1\right) + \frac{\hat{p}_0}{2}\left(\frac{\hat{p}_i}{\hat{p}_{i0}} - 1\right)^2 + \hat{q}_0\left(\frac{\hat{q}_i}{\hat{q}_{i0}} - 1\right) + \frac{\hat{q}_0}{2}\left(\frac{\hat{q}_i}{\hat{q}_{i0}} - 1\right)^2\right\}$$

$$= \sum_{i=1}^{k} n_i\left\{2(\hat{p}_i + \hat{q}_i) - 2(\hat{p}_{i0} + \hat{q}_{i0}) + \frac{(\hat{p}_i - \hat{p}_{i0})^2}{\hat{p}_{i0}} + \frac{(\hat{q}_i - \hat{q}_{i0})^2}{\hat{q}_{i,0}}\right\}$$

$$= \sum_{i=1}^{k}\left\{\frac{(n_i\hat{p}_i - n_i\hat{p}_{i0})^2}{n_i\hat{p}_{i0}} + \frac{(n_i\hat{p}_i - n_i\hat{p}_{i0})^2}{n_i\hat{q}_{i0}}\right\} \quad (n_i\hat{p}_i = X_{i\cdot} = x_i)$$

$$= \sum_{i=1}^{k}\frac{(x_i - n_i\hat{p}_{i0})^2}{n_i}\left(\frac{\hat{p}_{i0} + \hat{q}_{i0}}{\hat{p}_{i0}\hat{q}_{i0}}\right) = \sum_{i=1}^{k}\left(\frac{x_i - n_i\hat{p}_{i0}}{\sqrt{n_i\hat{p}_{i0}\hat{q}_{i,0}}}\right)^2$$

$$\approx 2\left(l(\hat{p}) - l(\hat{p}_0)\right)$$

(c)

(1) H_0하 $\hat{p}_i$의 극한분포

$x_{ij} \overset{iid}{\sim} \text{Bernoulli}(p_i)$이므로 $E(x_{ij}) = p_i$, $Var(x_{ij}) = p_i q_i$

CLT로부터 $n_i \to \infty$이면 $\sqrt{n_i}\left(\frac{\sum_{j=1}^{n_i} x_{ij}}{n_i} - p_i\right) \xrightarrow{d} N(0, p_i q_i)$이고,

식을 정리하면 $\sqrt{\frac{n_i}{p_i q_i}}\left(\hat{p}_i - p_i\right) \xrightarrow{d} N(0, 1^2)$이다.

$Z_i := \sqrt{\frac{n_i}{p_i q_i}}(\hat{p}_i - p_i)$로 두면 H_0하에서 $Z_i = \sqrt{\frac{n_i}{p_0 q_0}}(\hat{p}_i - p_0)$

이고, $p_1 = \cdots = p_k = p_0$이므로 이 값도 $N(0, 1^2)$으로 분포수렴한다.

즉 $Z := (Z_1, \cdots, Z_k)^T \xrightarrow{d} N(0, I_k)$이다.

(2) 검정통계량의 극한분포

$n_i \to \infty$일 때 $\hat{p}_{i,0} \to p_0$이므로

$$\sum_{i=1}^{k}\left(\frac{x_i - n_i\hat{p}_{i0}}{\sqrt{n_i\hat{p}_{i0}\hat{q}_{i0}}}\right)^2 = \sum_{i=1}^{k}\frac{(\hat{p}_i - \hat{p}_{i,0})^2}{\frac{\hat{p}_{i,0}\hat{q}_{i,0}}{n_i}}$$

$$\longrightarrow \sum_{i=1}^{k}\frac{(\hat{p}_i - \hat{p}_{i,0})^2}{p_0 q_0 / n_i}$$

$$= \sum_{i=1}^{k}\frac{(\hat{p}_i - p_0)^2}{p_0 q_0/n_i} + \left(\sum_{i=1}^{k}\frac{(\hat{p}_{i,0} - p_0)^2}{p_0 q_0/n_i}\right) - 2\sum_{i=1}^{k}\frac{(\hat{p}_i - p_0)(\hat{p}_{i,0} - p_0)}{p_0 q_0/n_i}$$ 영

$$= \sum_{i=1}^{k}\left(\frac{\hat{p}_i - p_0}{\sqrt{p_0 q_0/n_i}}\right)^2 - \left(\frac{\hat{p}_{i0} - p_0}{\sqrt{p_0 q_0 / \sum_{i=1}^{k} n_i}}\right)^2$$

앞에서 정의한 H_0하 $Z = (Z_1, \cdots Z_k)^T$에 대해

$Z_i = \sqrt{\frac{n_i}{p_0 q_0}}(\hat{p}_i - p_0)$, $W := (\sqrt{n_1}, \cdots, \sqrt{n_k})^T$로 두면 위 식의 값은 $Z^T Z - \frac{(W^T Z)^2}{W^T W}$

$= Z^T Z - Z^T W (W^T W)^{-1} W^T Z$의 이차형식이므로

$= Z^T (I - W(W^T W)^{-1} W^T) Z \sim \chi^2(k-1)$.

20.

(1) 가능도, 로그가능도

(a) $l(\mu_1,\cdots,\mu_k, \sigma_1,\cdots,\sigma_k) = \log L(\mu_1,\cdots,\mu_k, \sigma_1,\cdots,\sigma_k)$

$= -\sum_{i=1}^{k} \frac{n_i}{2}\left\{\log(2\pi\sigma_i^2) + \frac{1}{\sigma_i^2}\{S_i^2 + (\bar{X}_i - \mu_i)\}\right\}$

Where $\bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ij}$, $S_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)^2$

(전역 MLE)

$\frac{\partial l}{\partial \mu_i} = \frac{n_i}{\sigma_i^2}(\bar{X}_i - \mu_i)$이고 μ_i의 감소함수이므로 $\hat{\mu}_i = \bar{X}_i$

$\frac{\partial l}{\partial \sigma_i^2} = -\frac{n_i}{2\sigma_i^2} + \frac{n_i S_i^2}{2\sigma_i^4} = 0$ iff $\sigma_i^2 = S_i^2$이고 σ_i^2의 감소함수이므로 $\hat{\sigma}_i^2 = S_i^2$

(영가설 MLE)

$\hat{\mu}_{i,0}$을 먼저 구하면 $\underset{\Theta_0}{\operatorname{argmin}}\left\{\sum_{i=1}^{k} n_i\left\{\log(2\pi\sigma_i^2) + \frac{S_i^2 + (\bar{X}_i - \mu_i)^2}{\sigma_i^2}\right\}\right\}$의 값은 $\hat{\mu}_i$에 대해서는 (for fixed σ_i^2) $\bar{X}_i$으로 결정되므로 $\hat{\mu}_{i,0} = \bar{X}_i$이고, $\sigma_{i,0}$을 구하기 위해 $\sigma_0 = \sigma_{i,0}$을 대입하면 $\underset{\Theta_0}{\operatorname{argmin}}\left\{\sum_{i=1}^{k} n_i \log(2\pi\sigma_{i,0}^2) + \frac{n_i S_i^2}{\sigma_{i,0}^2}\right\}$을 구하기 위해 미분하자.

$\frac{\partial l_0}{\partial \sigma_{i,0}^2} = \sum_{i=1}^{k}\frac{n_i}{\sigma_{i,0}^2} - \sum_{i=1}^{k}\frac{n_i S_i^2}{\sigma_{i,0}^4}$이므로 $\sigma_0^2 = \sigma_{i,0}^2 = \frac{\sum_{i=1}^{k} n_i S_i^2}{\sum_{i=1}^{k} n_i}$

이므로 영가설공간 MLE $\hat{\sigma}_0^2 = \hat{\sigma}_{i,0}^2 = \frac{\sum_{i=1}^{k} n_i S_i^2}{\sum_{i=1}^{k} n_i}$ 이다.

(검정통계량)

$2\{l(\hat{\mu}_1,\cdots,\hat{\mu}_k,\hat{\sigma}_1^2,\cdots,\hat{\sigma}_k^2) - l(\hat{\mu}_{1,0},\cdots,\hat{\mu}_{k,0},\hat{\sigma}_{1,0}^2,\cdots,\hat{\sigma}_{k,0}^2)\}$

$= \sum_{i=1}^{k}\left\{-n_i\log(2\pi\hat{\sigma}_i^2) - \frac{n_i S_i^2}{\hat{\sigma}_i^2} + n_i\log(2\pi\hat{\sigma}_0^2) + \frac{n_i S_i^2}{\hat{\sigma}_0^2}\right\}$

$= -\sum_{i=1}^{k} n_i \log\frac{\hat{\sigma}_i^2}{\hat{\sigma}_0^2} + \sum_{i=1}^{k} n_i S_i^2\left(\frac{1}{\hat{\sigma}_0^2} - \frac{1}{\hat{\sigma}_i^2}\right)$

$= -\sum_{i=1}^{k} n_i \log\frac{\hat{\sigma}_i^2}{\hat{\sigma}_0^2}$ ($\because$ (*))

(*) $\sum_{i=1}^{k}\left(\frac{1}{\hat{\sigma}_0^2} - \frac{1}{\hat{\sigma}_i^2}\right) n_i S_i^2 = \sum_{i=1}^{k}\left\{\frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} n_i S_i^2} - \frac{1}{S_i^2}\right\} n_i S_i^2$

$= \left(\sum_{i=1}^{k} n_i\right) \times \frac{\sum_{i=1}^{k} n_i S_i^2}{\sum_{i=1}^{k} n_i S_i^2} - \sum_{i=1}^{k}\frac{n_i S_i^2}{S_i^2}$

$= \sum_{i=1}^{k}(n_i - n_i) = 0$

(b) 근사

$n_i \to \infty$, $\frac{n_i}{\sum_{i=1}^{k} n_i} \to \gamma_i$일 때 $\frac{\hat{\sigma}_i^2}{\hat{\sigma}_0^2} = \frac{\hat{\sigma}_i^2/\sigma_0^2}{\hat{\sigma}_0^2/\sigma_0^2} \to 1$ (H_0일 때)

이때 $t=0$ 근방에서 $\log(1+t) \approx t - \frac{t^2}{2}$에 $\frac{\hat{\sigma}_i^2}{\hat{\sigma}_0^2} - 1$을 대입하면

$-\sum_{i=1}^{k} n_i \log\frac{\hat{\sigma}_i^2}{\hat{\sigma}_0^2} \approx -\sum_{i=1}^{k}\left\{\left(\frac{\hat{\sigma}_i^2}{\hat{\sigma}_0^2} - 1\right) - \frac{1}{2}\left(\frac{\hat{\sigma}_i^2}{\hat{\sigma}_0^2} - 1\right)^2\right\} n_i = \frac{1}{2}\sum_{i=1}^{k}\left(\frac{\hat{\sigma}_i^2}{\hat{\sigma}_0^2} - 1\right)^2 n_i$

($\because \sum_{i=1}^{k}\left(\frac{\hat{\sigma}_i^2}{\hat{\sigma}_0^2} - 1\right) n_i = \sum_{i=1}^{k} n_i\left(\frac{1}{\hat{\sigma}_0^2} - \frac{1}{\hat{\sigma}_i^2}\right)$이고 이는 (*)과 같이 0)

(C) (1) S_i^2의 근사분포

$\frac{n_i}{2}\left(\frac{\hat{\sigma}_i^2}{\hat{\sigma}_0^2} - 1\right)^2$의 H_0하 근사분포를 구하기 위해

$Z_i := \sqrt{\frac{n_i}{2}}\left(\frac{S_i^2}{\sigma_0^2} - 1\right)$을 정의하자. 이때

$\frac{n_i S_i^2}{\sigma_i^2} = \frac{n_i S_i^2}{\sigma_0^2} \sim \chi^2(n_i - 1)$이므로

$\frac{S_i^2}{\sigma_0^2} \sim \frac{1}{n_i}\left\{\sum_{j=1}^{n_i}(Y_j) - Y_{n_i+1}\right\}$, $Y_j \overset{iid}{\sim} \chi^2(1)$, $1 \le j \le n_i + 1$

이고 $\frac{Y_{n_i+1}}{n_i} \to 0$ as $n_i \to \infty$, $\frac{1}{n_i}\sum_{j=1}^{n_i} Y_j \sim N\left(1, \frac{2}{n_i}\right)$이므로

$\sqrt{\frac{n_i}{2}}\left(\frac{S_i^2}{\sigma_0^2} - 1\right) \xrightarrow{d} N(0, 1^2)$이다.

$Z := (Z_1,\cdots,Z_k)^T$로 두면 $Z \sim N(0, I_k)$

(2) 검정통계량의 극한분포

$\sum_{i=1}^{k}\frac{n_i}{2}\left(\frac{\hat{\sigma}_i^2}{\hat{\sigma}_0^2} - 1\right)^2 \approx \sum_{i=1}^{k} n_i\left(\frac{\hat{\sigma}_i^2 - \hat{\sigma}_{i,0}^2}{\sqrt{2}\sigma_0^2}\right)^2$

$= \sum_{i=1}^{k}\left(\sqrt{\frac{n_i}{2}} \times \frac{\hat{\sigma}_i^2 - \sigma_0^2}{\sigma_0^2}\right)^2 - \left(\sqrt{\frac{\sum_{i=1}^{k} n_i}{2}} \times \frac{\hat{\sigma}_{i,0}^2 - \sigma_0^2}{\sigma_0^2}\right)^2$

이때 $w := (\sqrt{n_1},\cdots,\sqrt{n_k})$로 두면 이 값을

$= Z^T Z - Z^T(W(W^T W)^{-1} W^T)Z$의 이차형식으로 표현가능하므로 $\chi^2(k-1)$을 따른다.

7.19.

(a) 전역 MLE : 7.20 참고.

영가설공간 MLE : $\mu_1 = \cdots = \mu_k$ 이고

$$\frac{\partial}{\partial \mu_0}\sum_{i=1}^{k} n_i\left\{\log(2\pi\sigma_i^2) + \frac{S_i^2+(\bar{X}_i-\mu_0)^2}{\sigma_i^2}\right\} = \frac{-2}{\sigma_i^2}\sum_{i=1}^{k}(\bar{X}_i-\mu_0)$$

$$\hat{\mu}_0 = \frac{\sum_{i=1}^{k} n_i\bar{X}_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}}{n} = \bar{X}$$

$$\hat{\sigma}_0^2 \Rightarrow \frac{\partial}{\partial \sigma_0^2}\sum_{i=1}^{k} n_i\left\{\log(2\pi\sigma_i^2) + \frac{S_i^2+(\bar{X}_i-\bar{X})^2}{\sigma_i^2}\right\}$$

$$= \frac{n_i}{\sigma_i^2} - \sum_{i=1}^{k}\left\{\frac{S_i^2 n_i}{\sigma_i^4} + \frac{n_i(\bar{X}_i-\bar{X})^2}{\sigma_i^4}\right\} = 0 \text{이어야 하므로}$$

$$\hat{\sigma}_{0}^2 = \sum_{i=1}^{k}\left\{S_i^2 + (\bar{X}_i-\bar{X})^2\right\}$$

$$2\left\{l(\hat{\theta}) - l(\hat{\theta}^0)\right\} = -2\left\{\sum \log\left(\frac{\sigma_i^2}{\sigma_{i,0}^2}\right)n_i\right\} - 2\sum n_i + 2\sum n_i$$

$$= -2\sum_{i=1}^{k}\log\frac{\sigma_i^2}{\sigma_0^2}$$

(b) $\mu_0 = \cdots = \mu_k$로 두면 $\hat{\mu}_{i,0} \to \mu_0$, $\hat{\mu}_i \to \mu_0$ under H_0.