Mathematical Statistics2 Tutoring8

Seung Bong Jung

Seoul National University

December 1, 2021

Properties of Normal Distribution

- Suppose that $X \sim N(\mu, \sigma^2 I)$, and let A and B be real symmetric and idempotent matrices. If $AB = 0$, then, $X^t AX$ and $X^t BX$ are independent.
- Suppose that $Z \sim N(0, I)$, and let A and B be real symmetric matrices. Then, $Z^t AZ$ and $Z^t BZ$ are independent if and only if $AB = 0$.

Fisher-Cochran Theorem

Assume that $Z \sim N(0, I)$. Then, the following three conditions are equivalent.

- Z^tAZ and Z^tBZ are independent $\chi^2(r_1)$ and $\chi^2(r_2)$ random variables, respecitvely.
- $Z^t(A+B)Z \sim \chi^2(r)$ and $r = \text{rank}(A) + \text{rank}(B)$.
- $Z^t(A+B)Z \sim \chi^2(r)$, $z^tAz \sim \chi^2(r_1)$ and B is nonnegative definite.

Noncentral Chi-Square Distributions

Let Y be a random variable. The distribution of Y is called chi-square distribution with degree of freedom k and noncentrality parameter δ, denoted by $\chi^2(k;\delta)$, if one of the following equivalent conditions is met.

- $Y \overset{d}{\sim} \sum_{i=1}^{k} Z_i^2$, where Z_i are i.i.d. $N(\mu_i, 1)$ and $\delta = \sum_{i=1}^{k} \mu_i^2$.
- $\text{mgf}_Y(t) = (1-2t)^{-k/2} \exp(\delta t/(1-2t))$, $t < 1/2$;
- $\text{cgf}_Y(t) = \sum_{r=1}^{\infty} (t^r/r!) 2^{r-1} (r-1)!(n + r\delta);$, $t < 1/2$;
- $Y \overset{d}{=} V + Z^2$, where $V \sim \chi^2(k-1)$ and $Z \sim N(\theta, 1)$ are independent and $\theta^2 = \delta$;
- $\text{pdf}_Y(y) = \sum_{j=0}^{\infty} \text{pdf}_{\mathcal{P}(\delta/2)}(j) \cdot \text{pdf}_{\chi^2(k+2j)}(y)$, where $\mathcal{P}(\delta/2)$ denotes the Poisson$(\delta/2)$ distribution.

Noncentral t and F Distributions

- Noncentral t distribution: The distribution of Y is called noncentral t distribution with degree of freedom k and noncentrality parameter δ, denoted by $t(k;\delta)$, if $Y \overset{d}{=} Z/\sqrt{V/k}$, where $Z \sim N(\delta, 1)$ and $V \sim \chi^2(k)$ are independent.
- Noncentral F distribution: The distribution of Y is called noncentral F distribution with degrees of freedom k_1 and k_2 and noncentrality parameter $\delta > 0$, denoted by $F(k_1, k_2; \delta)$, if $Y \overset{d}{=} V_1/k_1/(V_2/k_2)$, where $V_1 \sim \chi^2(k_1;\delta)$ and $V_2 \sim \chi^2(k_2)$ are independent.

Distribution of Quadratic Forms: $N(\mu, I)$ Case

Suppose that Z is $d-$variate $N(\mu, I)$ random vector. Let A be a real symmetric matrix. Then, $Z^t AZ \sim \chi^2(r; \delta)$ if and only if $A^2 = A$, $r =$ trace(A) and $\delta = \mu^t A\mu$.

LRT in One-Way Classification Model

Assume that we observe independent $X_{ij} \sim N(\mu_i, \sigma^2)$ for $j = 1, \ldots, n_i$; $i = 1, \ldots, b$. Let $n = n_1 + \cdots + n_b$. We wish to test

$$H_0 : \mu_1 = \cdots = \mu_b \text{ versus } H_1 : \text{ not } H_0.$$

- MLE under the full model:

$$\hat{\mu}_i = \bar{X}_{i\cdot}, \hat{\sigma}^2 = SSW/n,$$

  where $SSW = \sum_{i,j}(X_{ij} - \bar{X}_{i\cdot})^2$.

- MLE under H_0:

$$\hat{\mu}_i \equiv \bar{X}_{\cdot\cdot}, \hat{\sigma}^2 = \sum_{i,j}(X_{ij} - \bar{X}_{\cdot\cdot})^2/n = (SSW + SSB)/n,$$

  where $SSB = \sum_{i,j}(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$.

LRT in One-Way Classification Model

- LRT statistic:

$$2\left(\ell(\hat{\theta}^{\Theta}) - \ell(\hat{\theta}^{\Theta_0})\right) = n\log(1 + SSB/SSW).$$

- Null distribution of LRT statistic: Under H_0,

$$\frac{SSB/(b-1)}{SSW/(n-b)} \sim F(b-1, n-b).$$

- The level α LRT rejects H_0 if

$$\frac{SSB/(b-1)}{SSW/(n-b)} > F_{\alpha}(b-1, n-b).$$

LRT in One-Way Classification Model

Reparametrization: Let $\mu_i = \mu + \alpha_i$ in the one-way classification model, where $\mu = \sum_{i=1}^{b} n_i \mu_i / n$, $\alpha_i = \mu_i - \mu$.
The model is then expressed as

$$X_{ij} = \mu + \alpha_i + e_{ij},\ e_{ij} \text{ are i.i.d. } N(0, \sigma^2)$$

with α_i satisfying $\sum_{i=1}^{b} n_i \alpha_i = 0$. The hypothesis $H_0 : \mu_1 = \cdots = \mu_b$ is then equivalent to $H_0 : \alpha_1 = \cdots = \alpha_b = 0$.

Confidential Region and Simultaneous Confidence Interval in One-Way Classification Model

Suppose C is $k \times r$ $(k > r)$ real matrix with full-rank and denote the column space of C by $\mathcal{C}(C)$. Let $D = \mathrm{Var}_{\mu,\sigma^2}(\hat{\mu})/\sigma^2 = \mathrm{diag}(1/n_i)$. Then the followings hold in One-Way Classification Model.

$$(1) \quad P_{\mu,\sigma^2}\{(C^t\mu - C^t\hat{\mu})^t(C^tDC)^{-1}(C^t\mu - C^t\hat{\mu}) \leq p\hat{\sigma}^2 F_\alpha(r, n-k)\} = 1 - \alpha$$

$$(2) \quad P_{\mu,\sigma^2}\left\{|c^t\mu - c^t\hat{\mu}| \leq \sqrt{\sum_{i=1}^k c_i\hat{\sigma}^2/n_i}\sqrt{rF_\alpha(r, n-k)}, {}^\forall c \in \mathcal{C}(C)\right\} = 1 - \alpha$$

Simultaneous Confidence Interval for the Contrast in One-Way Classification Model

Let $\alpha_i = \mu_i - \bar{\mu}$ and $\alpha^* = \alpha/m$ for $m = k(k-1)/2$.

$$(1) \quad P_{\mu,\sigma^2}\{|c^t\alpha - c^t\hat{\alpha}| \leq \sqrt{\sum_{i=1}^{k} c_i^2\hat{\sigma}^2/n_i}\sqrt{(k-1)F_\alpha(k-1,n-k)},$$
$$^\forall c : c_1 + \cdots + c_k = 0\} = 1-\alpha$$

$$(2) \quad P_{\mu,\sigma^2}\{|(\alpha_i - \alpha_j) - (\hat{\alpha}_i - \hat{\alpha}_j)| \leq \sqrt{(n_i^{-1} + n_j^{-1})\hat{\sigma}^2}$$
$$\sqrt{(k-1)F_\alpha(k-1,n-k)},^\forall i \neq j\} \geq 1-\alpha$$

$$(3) \quad P_{\mu,\sigma^2}\{|(\alpha_i - \alpha_j) - (\hat{\alpha}_i - \hat{\alpha}_j)| \leq \sqrt{(n_i^{-1} + n_j^{-1})\hat{\sigma}^2}t_{\alpha^*/2}(n-k)$$
$$,^\forall i \neq j\} \geq 1-\alpha$$

Power of LRT in One-Way Classification Model

Under H_1, $Z_{ij} \equiv X_{ij}/\sigma \sim N((\mu + \alpha_i)/sigma, 1)$.

- Distribution of SSW: $SSW/\sigma^2 \sim \chi^2(n-b)$.
- Distribution of SSB: $SSB/\sigma^2 \sim \chi^2(b-1;\delta)$.
- Distribution of test statistic: Under H_1, the test statistic

$$F \equiv \frac{SSB/(b-1)}{SSW/(n-b)} \sim F(b-1, n-b; \delta).$$

- Power function $\gamma(\delta) \equiv P(F \geq F_\alpha(b-1, n-b))$. The power function $\gamma(\delta)$ is non-decreasing in δ.

LRT in Two-Way Classification Model Without Interaction

Assume that we observe X_{ij}, $1 \le i \le a$, $1 \le j \le b$ such that

$$X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij},$$

where ϵ_{ij} are i.i.d. $N(0, \sigma^2)$, and α_i and β_j are unknown constants that add up to zero, i.e., $\sum_{i=1}^{a} \alpha_i = 0 = \sum_{j=1}^{b} \beta_j$. We wish to test

$$H_0 : \alpha_1 = \cdots = \alpha_a = 0 \text{ versus } H_1 : \text{ not } H_0.$$

- Log-likelihood:

$$\ell(\theta) = -\frac{1}{2\sigma^2} \sum_{i,j} (x_{ij} - \mu - \alpha_i - \beta_j)^2 - \frac{ab}{2} \log(2\pi\sigma^2).$$

LRT in Two-Way Classification Model Without Interaction

- MLE under the full model:

$$\hat{\mu} = \bar{X}_{..}, \hat{\alpha}_i = \bar{X}_{i.} - \bar{X}_{..}, \hat{\beta}_j = \bar{X}_{.j} - \bar{X}_{..},$$
$$\hat{\sigma}^2 = \sum_{i,j} (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2/ab \overset{\text{let}}{=} \mathsf{SS}_{AB}/ab.$$

- MLE under H_0:

$$\begin{aligned}
\hat{\mu} &= \bar{X}_{..}, \hat{\alpha}_i = 0, \hat{\beta}_j = \bar{X}_{.j} - \bar{X}_{..}, \\
\hat{\sigma}^2 &= \sum_{i,j} (X_{ij} - \bar{X}_{.j})^2/ab \\
&= \sum_{i,j} (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2/ab + \sum_{i,j} (\bar{X}_{i.} - \bar{X}_{..})^2/ab \\
&\overset{\text{let}}{=} (\mathsf{SS}_{AB} + \mathsf{SS}_A)/ab.
\end{aligned}$$

LRT in Two-Way Classification Model Without Interaction

- LRT statistic:

$$2\left(\ell(\hat{\theta}^{\Theta}) - \ell(\hat{\theta}^{\Theta_0})\right) = ab \cdot \log\left(1 + \frac{\mathsf{SS}_A}{\mathsf{SS}_{AB}}\right).$$

- Null distribution of LRT statistic: Under H_0

$$\frac{\mathsf{SS}_A/(a-1)}{\mathsf{SS}_{AB}/(a-1)(b-1)} \sim F(a-1,(a-1)(b-1)).$$

- The level α LRT rejects H_0 if

$$\frac{\mathsf{SS}_A/(a-1)}{\mathsf{SS}_{AB}/(a-1)(b-1)} > F_\alpha(a-1,(a-1)(b-1)).$$

LRT in Two-Way Classification Model With Interaction

Assume that we observe X_{ijk}, $1 \le i \le a$, $1 \le j \le b$, $1 \le k \le r$ such that

$$X_{ijjk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk},$$

where ϵ_{ijk} are i.i.d. $N(0, \sigma^2)$, and α_i, β_j and γ_{ij} are unknown constants with $\sum_{i=1}^a \alpha_i = 0 = \sum_{j=1}^b \beta_j$ and $\sum_{i=1}^a \gamma_{ij} = 0 = \sum_{j=1}^b \gamma_{ij}$. We wish to test

$$H_0 : \gamma_{ij} \overset{i,j}{\equiv} 0 \text{ versus } H_1 : \text{ not } H_0.$$

- Log-likelihood:

$$\ell(\theta) = -\frac{1}{2\sigma^2} \sum_{i,j,k} (x_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2 - \frac{abr}{2} \log(2\pi\sigma^2).$$

- MLE under the full model:

$$
\begin{aligned}
\hat{\mu} &= \bar{X}_{...}, \hat{\alpha}_i = \bar{X}_{i..} - \bar{X}_{...}, \hat{\beta}_j = \bar{X}_{.j.} - \bar{X}_{...}, \\
\hat{\gamma}_{ij} &= \bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...}, \\
\hat{\sigma}^2 &= \sum_{i,j,k} (X_{ijk} - \bar{X}_{ij.})^2/(abr) \overset{\text{let}}{=} \mathsf{SSW}/(abr).
\end{aligned}
$$

- MLE under H_0:

$$
\begin{aligned}
\hat{\mu} &= \bar{X}_{...}, \hat{\alpha}_i = \bar{X}_{i..} - \bar{X}_{...}, \hat{\beta}_j = \bar{X}_{.j.} - \bar{X}_{...}, \\
\hat{\sigma}^2 &= \sum_{i,j,k} (X_{ijk} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...})^2/(abr) \\
&= \sum_{i,j,k} (X_{ijk} - \bar{X}_{ij.})^2/(abr) + \sum_{i,j,k} (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...})^2/(abr) \\
&\overset{\text{let}}{=} (\mathsf{SSW} + \mathsf{SS}_{AB})/(abr).
\end{aligned}
$$

LRT in Two-Way Classification Model With Interaction

- LRT statistic:

$$2\left(\ell(\hat{\theta}^{\Theta}) - \ell(\hat{\theta}^{\Theta_0})\right) = abr \cdot \log\left(1 + \frac{\mathsf{SS}_{AB}}{\mathsf{SSW}}\right).$$

- Null distribution of LRT statistic: Under H_0

$$\frac{\mathsf{SS}_{AB}/(a-1)(b-1)}{\mathsf{SSW}/ab(r-1)} \sim F((a-1)(b-1), ab(r-1)).$$

- The level α LRT rejects H_0 if

$$\frac{\mathsf{SS}_{AB}/(a-1)(b-1)}{\mathsf{SSW}/ab(r-1)} > F_\alpha((a-1)(b-1), ab(r-1)).$$

LRT in Two-Way Classification Model With Interaction

For the hypothesis

$$H_0 : \alpha \overset{i}{\equiv} 0 \text{ versus } H_1 : \text{ not } H_0,$$

you may also prove that the level α LRT rejects H_0 if

$$\frac{\mathsf{SS}_A/(a-1)}{\mathsf{SSW}/ab(r-1)} > F_\alpha(a-1, ab(r-1)),$$

where $\mathsf{SS}_A = \sum_{i,j,k}(\bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot\cdot\cdot})^2$.

ANOVA Table for Two-Way Classification Model

Source	SS	df	MS	F ratio
$A \times B$	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	$\text{MS}_{AB}/\text{MSE}$
A	SS_A	$a-1$	MS_A	MS_A/MSE
B	SS_B	$b-1$	MS_B	MS_B/MSE
Error	SSW	$ab(r-1)$	MSE	
Total	SST	$abr-1$		

Here, SS and MS (=SS/df) stand for the sum of squares and the mean of squares, respectively, and SST $= \sum_{i,j,k}(X_{ijk} - \bar{X}_{...})^2$.
Note: The so-called interaction effect $A \times B$ explains the "variation" due to possible non-additivity of the effects A and B.

Power of LRT for Interaction in Two-Way Classification Model

We consider the hypothesis

$$H_0 : \gamma_{ij} \overset{i,j}{\equiv} 0 \text{ versus } H_1 : \text{ not } H_0.$$

Under H_1, $Z_{ijk} \equiv X_{ijk}/\sigma \sim N((\mu + \alpha_i + \beta_j + \gamma_{ij})/\sigma, 1)$.

- Distribution of SSW: $\text{SSW}/\sigma^2 \sim \chi^2(ab(r-1))$.
- Distribution of SS_{AB}: $\text{SS}_{AB}/\sigma^2 \sim \chi^2((a-1)(b-1); \delta)$.
- Distribution of test statistic: Under H_1, the test statistic

$$F \equiv \frac{\text{SS}_{AB}/(a-1)(b-1)}{\text{SSW}/ab(r-1)} \sim F((a-1)(b-1), ab(r-1); \delta).$$

- Power function: $\gamma(\delta) \equiv P(F \geq F_\alpha((a-1)(b-1), ab(r-1)))$. The power function $\gamma(\delta)$ is non-decreasing in δ.