Math 3236 Statistical Theory

3/30/23

$X_i \quad i = 1 \ldots N$

$Y_i \quad i = 1 \ldots M$

X_i are Normal

$\mu_X \quad \sigma_X$

Y_i are Normal

$\mu_Y \quad \sigma_Y$

How big is $\mu_X - \mu_Y$?

$\sigma_X = \sigma_Y$ unknown

σ_X and σ_Y Known.

$$\bar{X}_N = \frac{1}{N} \sum_{i=1}^{N} X_i \qquad N\left(\mu_X , \frac{\sigma_X^2}{N}\right)$$

$$\bar{Y}_M = \frac{1}{M} \sum_{i=1}^{M} Y_i \qquad N\left(\mu_Y , \frac{\sigma_Y^2}{M}\right)$$

$$Z = \frac{\sqrt{NM}\left(\bar{X}_N - \bar{Y}_M - \mu_X + \mu_Y\right)}{\bar{\sigma}}$$

$$\bar{\sigma} = (N+M)\left(\frac{\sigma_X^2}{N} + \frac{\sigma_Y^2}{M}\right)$$

Z is $N(0,1)$

coeff. γ conf. int.

$$\bar{X}_N - \bar{Y}_M - \frac{\bar{\sigma}}{\sqrt{N+M}} z_{\frac{1-\gamma}{2}} \leq \mu_X - \mu_Y \leq$$

$$\bar{X}_N - \bar{Y}_M + \frac{\bar{\sigma}}{\sqrt{N+M}} z_{\frac{1-\gamma}{2}}$$

$$\mu_X - \mu_Y \geq \bar{X}_N - \bar{Y}_M - \frac{\sigma}{\sqrt{N+M}} z_{1-\gamma}$$

coeff γ lower bound on

$$\mu_X - \mu_Y$$

σ

$\sigma_X = \sigma_Y = \sigma$ unknown

Find coeff γ upper limit for σ^2

$$S_X^2 = \frac{1}{\sigma^2} \sum_{i=1}^{N} (X_i - \bar{X}_N)^2 \qquad \chi^2_{N-1}$$

$$S_Y^2 = \frac{1}{\sigma^2} \sum_{i=1}^{M} (Y_i - \bar{Y}_N)^2 \qquad \chi^2_{M-1}$$

$S_X^2 + S_Y^2$ is χ^2_{N+M-2} d.o.f.

$$\sigma^2 \le \frac{1}{\chi^2_{\gamma, N+M-2}} \left(\sum_{i=1}^{N} (X_i - \overline{X}_N)^2 + \sum_{i=1}^{M} (Y_i - \overline{Y}_M)^2 \right)$$

γ conf. interval on

$$\mu_X - \mu_Y$$

$$Z = \frac{\sqrt{NM}\,(\overline{X}_N - \overline{Y}_M - \mu_X + \mu_Y)}{\sigma \sqrt{N+M}}$$

$$U = \frac{1}{\sigma^2} \left(\sum_{i=1}^{N} (X_i - \overline{X}_N)^2 + \sum_{j=1}^{M} (Y_j - \overline{Y}_M)^2 \right)$$

$$\overline{T} = \frac{Z}{\sqrt{\frac{U}{M+N-2}}} \qquad \tau_{N+M-2} \text{ d.o.f.}$$

$$\Sigma^2 = \frac{1}{N+M-2}\left(\sum_{i=1}^{N}(X_i - \overline{X}_N)^2 + \sum_{j=1}^{M}(Y_j - \overline{Y}_M)^2\right)$$

$$\overline{X}_N - \overline{Y}_M - T_{1-\frac{\gamma}{2}, N+M-2}\, \Sigma \sqrt{\frac{1}{N} + \frac{1}{M}} \leq$$

$$\mu_X - \mu_Y \leq$$

$$\overline{X}_N - \overline{Y}_M - T_{1-\frac{\gamma}{2}, N+M-2}\, \Sigma \sqrt{\frac{1}{N} + \frac{1}{M}}$$

$X_i \qquad i = 1 \quad \ldots \quad N \qquad F$ c.d.f.

$Y_1 = \min X_i$

$Y_2 = \min\{X_i \mid X_i > Y_1\}$

$Y_3 =$ order statistics

$\vdots$

$Y_N = \max X_i$

$$Y_k = \min_{\substack{I \subset \{1,\dots,N\} \\ |I| = k}} \max_{i \in I} X_i$$

$$\mathbb{P}(Y_k \leq y)$$

$\mathbb{P}($ k of the X_i are $\leq y$ while

$N-k$ are $> y$ $) =$

$$\binom{N}{k} F(y)^k (1 - F(y))^{N-k}$$

$$\mathbb{P}(Y_k \leq y) = \mathbb{P}(\text{at least } k \ X_i \leq y)$$

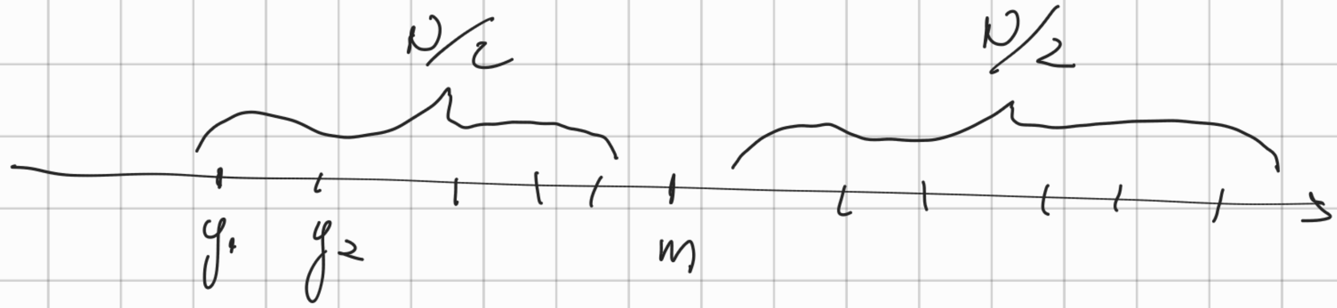
$$= \sum_{j=k}^{N} \binom{N}{j} F(y)^j (1 - F(y))^{N-j}$$

Median

$$F(m) = \frac{1}{2}$$

$$\left(f(m) = F'(m) > 0 \right)$$

$$\mathbb{P}\left(Y_i \le m \le Y_{N-i+1}\right)$$



$$\mathbb{P}\left(Y_i \le m \le Y_{N-i+1}\right) =$$

$$\mathbb{P}\left(m \le Y_{N-i+1}\right) - \mathbb{P}\left(m \le Y_i\right) =$$

$$= \mathrm{Bin}(N-i, N, 0.5) - \mathrm{Bin}(i-1, N, 0.5)$$

$$= 1 - 2\,\mathrm{Bin}(i-1, N, 0.5) =$$

$$= 2^{-N} \sum_{j=i}^{N-i} \binom{N}{j}$$

$N = 20$ 0.95 C.I. for m

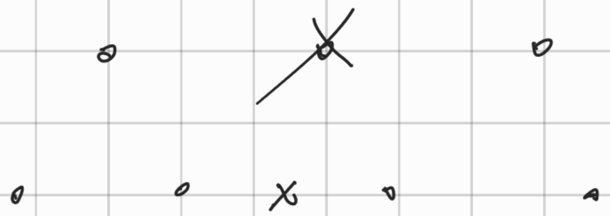
$$\mathbb{P}\left(Y_6 \le m \le Y_{15}\right) = 0.9586$$

$$Y_6 \le m \le Y_{15}$$

$$\mathrm{Bin}(i, N, 0.5) \simeq \Phi\left(\frac{2i+1-N}{\sqrt{N}}\right)$$

$$\Phi\left(\frac{2i+1-N}{\sqrt{N}}\right) \leq \frac{1-\gamma}{2}$$

$$\hat{m}(\underline{X}) = \begin{cases} Y_{\frac{N+1}{2}} & N \text{ odd} \\ \frac{1}{2}\left(Y_{\frac{N}{2}} + Y_{\frac{N}{2}+1}\right) & N \text{ even} \end{cases}$$



$N = 3$

$N = 4$

$\hat{m}(\underline{X})$ is a consistent estimator for m.