

Math 582 HW 11, 2024 Due Wednesday, Nov. 20.  
 Exam 3: Wednesday Dec. 4, Final: TBAday, Dec. : time TBA

**This homework has 2 pages, 4 problems.**

1) Let  $\Sigma_i$  be the nonsingular population covariance matrix of the  $i$ th treatment group or population. To simplify the large sample theory, assume  $n_i = \pi_i n$  where  $0 < \pi_i < 1$  and  $\sum_{i=1}^2 \pi_i = 1$ . Let  $T_i$  be a multivariate location estimator such that  $\sqrt{n_i}(T_i - \mu_i) \xrightarrow{D} N_m(\mathbf{0}, \Sigma_i)$ , and  $\sqrt{n}(T_i - \mu_i) \xrightarrow{D} N_m\left(\mathbf{0}, \frac{\Sigma_i}{\pi_i}\right)$  for  $i = 1, 2$ . Assume  $T_1 \perp T_2$ .

Then

$$\sqrt{n} \begin{bmatrix} T_1 - \mu_1 \\ T_2 - \mu_2 \end{bmatrix} \xrightarrow{D} \mathbf{u}. \quad (1)$$

You found the distribution of  $\mathbf{u}$  in homework 10 problem 5.

Now

$$\sqrt{n}[T_1 - T_2 - (\mu_1 - \mu_2)] \xrightarrow{D} \mathbf{w}.$$

Find the distribution of  $\mathbf{w}$ . Hint: multiply both sides of (1) by  $\mathbf{A} = [\mathbf{I}_m \quad -\mathbf{I}_m]$  and find the distribution of  $\mathbf{w} = \mathbf{A}\mathbf{u}$ .

2) The *sample median absolute deviation* is  $MAD(Y_i) = MAD(n) = MED(|Y_i - MED(n)|)$ ,  $i = 1, \dots, n$ : find the sample median and go out the distance  $MAD(n)$  that covers at least half of the cases. Then  $MAD(n)$  estimates the population median absolute deviation  $MAD(Y)$ : find the population median and go out the distance  $MAD(Y)$  that covers at least half of the mass. For  $Y_1, \dots, Y_n$  iid  $N(\mu, \sigma^2)$ , a  $MAD(n) \xrightarrow{P} \sigma$  where  $a \approx 1.483$ .

a) If  $X$  and  $Y$  are random variables, show that

$$Cov(X, Y) = [V(X + Y) - V(X - Y)]/4.$$

b) Suppose  $(X_i, Y_i)^T$  are iid from a bivariate normal distribution. Suggest a consistent estimator of  $Cov(X, Y)$  that is a function of  $MAD(X_i + Y_i)$  and  $MAD(X_i - Y_i)$ .

Hint:  $W_i = X_i + Y_i \sim N(E(X) + E(Y), V(X + Y))$  and

$Z_i = X_i - Y_i \sim N(E(X) - E(Y), V(X - Y))$ .

3) GLMs (and MA and ARMA time series models) are fit by maximum likelihood.

Thus  $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} N_p(\mathbf{0}, \Sigma)$ . What is  $\Sigma$ ?

4) Suppose that  $\mathbf{u}_i = (\mathbf{x}_i^T, Y_i)^T$  are iid for  $i = 1, \dots, n$ . Let  $\boldsymbol{\mu}_\mathbf{x} = E(\mathbf{x})$  and  $\mu_Y = E(Y)$ . Let  $\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\Sigma}}_{\mathbf{x}Y}$  and  $\boldsymbol{\eta} = \boldsymbol{\Sigma}_{\mathbf{x}Y} = Cov(\mathbf{x}, Y)$ . Then  $\sqrt{n}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{x}Y} - \boldsymbol{\Sigma}_{\mathbf{x}Y}) = \sqrt{n}(\tilde{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{D} N_p(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{w}})$  where  $\boldsymbol{\Sigma}_{\mathbf{w}} = Cov(\mathbf{w})$  and  $\mathbf{w}_i = (\mathbf{x}_i - \boldsymbol{\mu}_\mathbf{x})(Y_i - \mu_Y)$ .

The nonparametric bootstrap samples the  $\mathbf{u}_i = (\mathbf{x}_i^T, Y_i)^T$  with replacement. This bootstrap model has the  $\mathbf{u}_i^* = (\mathbf{x}_i^{*T}, Y_i^*)^T$  iid with respect to the bootstrap distribution. Then  $E(\mathbf{x}_i^*) = \bar{\mathbf{x}}$ ,  $E(Y_i^*) = \bar{Y}$ ,  $\mathbf{w}_i^* = (\mathbf{x}_i^* - \bar{\mathbf{x}})(Y_i^* - \bar{Y})$ . Fix  $n$ . Then  $\sqrt{m}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{x}Y}^* - \tilde{\boldsymbol{\Sigma}}_{\mathbf{x}Y}) = \sqrt{m}(\tilde{\boldsymbol{\eta}}^* - \tilde{\boldsymbol{\eta}}) \xrightarrow{D} N_p(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{w}^*})$ . Since the empirical distribution is used,

$$\begin{aligned} \boldsymbol{\Sigma}_{\mathbf{w}^*} &= E[(\mathbf{w}^* - E(\mathbf{w}^*))(\mathbf{w}^* - E(\mathbf{w}^*))^T] = E[\mathbf{w}^* \mathbf{w}^{*T}] - E(\mathbf{w}^*)[E(\mathbf{w}^*)]^T = \\ &= \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(Y_i - \bar{Y})[(\mathbf{x}_i - \bar{\mathbf{x}})(Y_i - \bar{Y})]^T - \left[ \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(Y_i - \bar{Y}) \right] \left[ \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(Y_i - \bar{Y}) \right]^T \\ &= \tilde{\boldsymbol{\Sigma}}_{\mathbf{w}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{w}_i - \bar{\mathbf{w}})(\mathbf{w}_i - \bar{\mathbf{w}})^T = \frac{1}{n} \sum_{i=1}^n \mathbf{w}_i \mathbf{w}_i^T - \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{w}_i \right] \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{w}_i \right]^T \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{w}_i \mathbf{w}_i^T - \bar{\mathbf{w}}[\bar{\mathbf{w}}]^T = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T - \bar{\mathbf{z}}[\bar{\mathbf{z}}]^T = \tilde{\boldsymbol{\Sigma}}_{\mathbf{z}} = \frac{1}{n} \sum_{i=1}^n \mathbf{v}_i \mathbf{v}_i^T - \bar{\mathbf{v}}[\bar{\mathbf{v}}]^T = \tilde{\boldsymbol{\Sigma}}_{\mathbf{v}} \end{aligned}$$

where  $\mathbf{z}_i = (\mathbf{x}_i - \bar{\mathbf{x}})Y_i$  and  $\mathbf{v}_i = (\mathbf{x}_i - \bar{\mathbf{x}})(Y_i - \bar{Y})$ .

Use the bootstrap proof technique from Exam 3 review 107) to find the limiting distribution of  $\sqrt{n}(\tilde{\boldsymbol{\Sigma}}_{\mathbf{x}Y}^* - \tilde{\boldsymbol{\Sigma}}_{\mathbf{x}Y}) = \sqrt{n}(\tilde{\boldsymbol{\eta}}^* - \tilde{\boldsymbol{\eta}})$ .