```
Math 582    HW 11, 2022  Due Tuesday, April 21.
Exam 3: Tuesday April 26, Final: Thursday, May 5: 12:30-2:30
```

This homework has 2 pages, 4 problems.

1) Let $\boldsymbol{\Sigma}_i$ be the nonsingular population covariance matrix of the ith 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 $\sqrt{n_i}(T_i - \boldsymbol{\mu}_i) \xrightarrow{D} N_m(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, and $\sqrt{n}(T_i - \boldsymbol{\mu}_i) \xrightarrow{D} N_m\left(\boldsymbol{0}, \frac{\boldsymbol{\Sigma}_i}{\pi_i}\right)$ for $i = 1, 2$. Assume $T_1 \perp\!\!\!\perp T_2$.

Then

$$\sqrt{n}\begin{bmatrix} T_1 - \boldsymbol{\mu}_1 \\ T_2 - \boldsymbol{\mu}_2 \end{bmatrix} \xrightarrow{D} \boldsymbol{u}. \tag{1}$$

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

Now

$$\sqrt{n}[T_1 - T_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] \xrightarrow{D} \boldsymbol{w}.$$

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

2) The *sample median absolute deviation* is $MAD(Y_i) = MAD(n) = MED(|Y_i - MED(n)|,\ i = 1, \ldots, n)$: 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 $Y_1, ..., Y_n$ 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 $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. What is $\boldsymbol{\Sigma}$?

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

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

$$\boldsymbol{\Sigma}_{\boldsymbol{w}^*} = E[(\boldsymbol{w}^* - E(\boldsymbol{w}^*))(\boldsymbol{w}^* - E(\boldsymbol{w}^*))^T] = E[\boldsymbol{w}^*\boldsymbol{w}^{*T}] - E(\boldsymbol{w}^*)[E(\boldsymbol{w}^*)]^T =$$

$$\frac{1}{n}\sum_{i=1}^n (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y})[(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y})]^T - [\frac{1}{n}\sum_{i=1}^n (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y})][\frac{1}{n}\sum_{i=1}^n (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y})]^T$$

$$= \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{w}} = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{w}_i - \overline{\boldsymbol{w}})(\boldsymbol{w}_i - \overline{\boldsymbol{w}})^T = \frac{1}{n}\sum_{i=1}^n \boldsymbol{w}_i\boldsymbol{w}_i^T - \left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{w}_i\right]\left[\frac{1}{n}\sum_{i=1}^n \boldsymbol{w}_i\right]^T$$

$$= \frac{1}{n}\sum_{i=1}^n \boldsymbol{w}_i\boldsymbol{w}_i^T - \overline{\boldsymbol{w}}[\overline{\boldsymbol{w}}]^T = \frac{1}{n}\sum_{i=1}^n \boldsymbol{z}_i\boldsymbol{z}_i^T - \overline{\boldsymbol{z}}[\overline{\boldsymbol{z}}]^T = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{v}_i\boldsymbol{v}_i^T - \overline{\boldsymbol{v}}[\overline{\boldsymbol{v}}]^T = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}}$$

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

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