Lecture 4 Review:

Three definitions of normal random vectors:

1. Normal probability density function (p.d.f.),
2. Moment generating function,
3. Relationship with independent univariate normals.

5.1 Orthogonal Transformations of MVN Vectors

Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I})$, and let $\mathbf{T}_{n\times n}$ be an orthogonal matrix. Then $\mathbf{TY} \sim N_n(\mathbf{T}\boldsymbol{\mu}, \sigma^2\mathbf{I})$.

Proof: By 4.3.1,

$$\mathbf{TY} \sim N_n(\mathbf{T}\boldsymbol{\mu}, \sigma^2\mathbf{TT}') = N_n(\mathbf{T}\boldsymbol{\mu}, \sigma^2\mathbf{I}).$$

Why is this interesting/important?

Mutually independent normal random variables with common variance remain mutually independent with common variance under orthogonal transformations.

1. Orthogonal matrices correspond to rotations and reflections about the origin. In particular, they preserve vector length:

$$||\mathbf{Ty}||^2 = (\mathbf{Ty})'(\mathbf{Ty}) = \mathbf{y}'\mathbf{T}'\mathbf{Ty} = \mathbf{y}'\mathbf{y} = ||\mathbf{y}||^2.$$

2. Orthogonal matrices are transformations about $\mathbf{0}$. A similar result holds for transformations about $\boldsymbol{\mu}$. For $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}_n)$, let

$$\mathbf{w} = \mathbf{T}(\mathbf{Y} - \boldsymbol{\mu}) + \boldsymbol{\mu}$$

   for some orthogonal transformation T. Then

$$\mathbf{w} \sim N_n(\boldsymbol{\mu}, \sigma^2\mathbf{I}_n).$$

   $\mathbf{Y}$ is "spherically symmetric" about $\boldsymbol{\mu}$.

3. The common variance MVN distribution is the only MVN distribution with this invariance property (mutually independent univariate normal variables with the same variance remain mutually independent normal variables under orthogonal transformations). Other multivariate distributions and MVN distributions with unequal variances do not have this invariance.

Definition: For any positive integer d, χ^2_d is the distribution of $\sum_{i=1}^d Z_i^2$, where $Z_1, \ldots, Z_d$ are independent and identically distributed $N(0, 1)$ random variables.

Example (Independence of sample mean and variance):

Let $Y_1, \ldots, Y_n$ be independent $N(\mu, \sigma^2)$ r.v.'s. Then $\bar{Y}$ and $s^2 = \frac{1}{n-1}\sum_i (Y_i - \bar{Y})^2$ are independent and $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$.

Proof: Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$. We have $\mathbf{Y} \sim N_n(\mu\mathbf{1}, \sigma^2\mathbf{I})$. Let $\mathbf{T}$ be an orthogonal matrix with first row equal to $\mathbf{1}'/\sqrt{n}$. (*How do we know such a* $\mathbf{T}$ *exists?*) Let $\mathbf{X} = \mathbf{TY}$. Then $\mathbf{X} \sim N_n(\mu\mathbf{T1}, \sigma^2\mathbf{I})$ by 5.1. Now,

$$X_1 = \sum_{i=1}^{n} \frac{1}{\sqrt{n}} Y_i = \sqrt{n}\bar{Y}. \qquad (1)$$

Because $\mathbf{T}$ preserves length:

$$\sum_{i=1}^{n} X_i^2 = \mathbf{X}'\mathbf{X} = \mathbf{Y}'\mathbf{Y} = \sum_{i=1}^{n} Y_i^2. \qquad (2)$$

Therefore,

$$\begin{aligned}\sum_{i=2}^{n} X_i^2 &= [\textstyle\sum_{i=1}^{n} X_i^2] - X_1^2 \\ &= [\textstyle\sum_{i=1}^{n} Y_i^2] - X_1^2 \quad \text{by (2)} \\ &= [\textstyle\sum_{i=1}^{n} Y_i^2] - n\bar{Y}^2 \quad \text{by (1)} \\ &= (n-1)s^2.\end{aligned}$$

Because X_1 and $(X_2, \ldots, X_n)$ are independent, $\bar{Y}$ (a function of X_1) and s^2 (a function $X_2, \ldots, X_n$) are independent (`Seber, 1977, Theorem 1.9`).

Remember that rows 2 through n of $\mathbf{T}$ are orthogonal to row 1, and in particular orthogonal to $\mathbf{1}$. This implies that $E[\mathbf{X}] = \mu \mathbf{T} \mathbf{1} = (\sqrt{n}\mu, 0, \ldots, 0)'$. Altogether, we have that $X_2, \ldots, X_n$ are independent $N(0, \sigma^2)$ random variables. Therefore,

$$(n-1)s^2/\sigma^2 = \sum_{i=2}^{n} X_i^2/\sigma^2 = \sum_{i=2}^{n} (X_i/\sigma)^2 \sim \chi^2_{n-1}.$$

5.2 Partitioned MVN distributions:

Let $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be partitioned as

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix},$$

where $\mathbf{Y}_1$ is $p \times 1$ and $\mathbf{Y}_2$ is $q \times 1$, $(p+q=n)$. Then the mean and covariance matrix are correspondingly partitioned as

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} = \begin{pmatrix} \operatorname{cov}(\mathbf{Y}_1) & \operatorname{cov}(\mathbf{Y}_1, \mathbf{Y}_2) \\ \operatorname{cov}(\mathbf{Y}_2, \mathbf{Y}_1) & \operatorname{cov}(\mathbf{Y}_2) \end{pmatrix}.$$

Some Facts and Results:

1. Marginal distributions: $\mathbf{Y}_1 \sim N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$, $\mathbf{Y}_2 \sim N_q(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$.
   Proof: Use 4.3.1 with $\mathbf{Y}_1 = \mathbf{C}_1\mathbf{Y} = (\mathbf{I}_{p\times p}, \mathbf{0}_{p\times q})\mathbf{Y}$ and $\mathbf{Y}_2 = \mathbf{C}_2\mathbf{Y} = (\mathbf{0}_{q\times p}, \mathbf{I}_{q\times q})\mathbf{Y}$.

2. Uncorrelated implies independent for normal random vectors:
   $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent if and only if $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}'_{21} = \mathbf{0}$.

   Proof:

   Let $\mathbf{t} = (\mathbf{t}_1, \mathbf{t}_2)'$. Then the m.g.f. of $\mathbf{Y}$ can be written

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= \exp(\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}) \\ &= \exp\{\boldsymbol{\mu}_1'\mathbf{t}_1 + \boldsymbol{\mu}_2'\mathbf{t}_2 + \frac{1}{2}(\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1 + \mathbf{t}_1'\boldsymbol{\Sigma}_{12}\mathbf{t}_2 + \\ &\qquad \mathbf{t}_2'\boldsymbol{\Sigma}_{21}\mathbf{t}_1 + \mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2)\} \end{aligned}$$

   $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent iff this equals

$$\begin{aligned} M_{\mathbf{Y}_1}(\mathbf{t}_1)M_{\mathbf{Y}_2}(\mathbf{t}_2) &= \exp(\boldsymbol{\mu}_1'\mathbf{t}_1 + \frac{1}{2}\mathbf{t}_1'\boldsymbol{\Sigma}_{11}\mathbf{t}_1) \\ &\quad \times \exp(\boldsymbol{\mu}_2'\mathbf{t}_2 + \frac{1}{2}\mathbf{t}_2'\boldsymbol{\Sigma}_{22}\mathbf{t}_2), \end{aligned}$$

   for all $\mathbf{t}_1, \mathbf{t}_2$. This holds exactly when

$$\mathbf{t}_1'\boldsymbol{\Sigma}_{12}\mathbf{t}_2 + \mathbf{t}_2'\boldsymbol{\Sigma}_{21}\mathbf{t}_1 = 0,$$

   for all $\mathbf{t}_1, \mathbf{t}_2$, i.e, when $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}'_{21} = \mathbf{0}$.

3. Conditional distributions: If $\mathbf{\Sigma}$ is p.d. then the conditional distribution of $\mathbf{Y}_1$ given $\mathbf{Y}_2$ is

$$\mathbf{Y}_1|\mathbf{Y}_2 = \mathbf{y}_2 \sim N_p(\boldsymbol{\mu}_1+\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}(\mathbf{y}_2-\boldsymbol{\mu}_2), \mathbf{\Sigma}_{11}-\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21}).$$

Proof: see Seber & Lee, Example 2.9

EXERCISE: Linear Regression with Random $X's$

Suppose we have:

$$Y_i = \boldsymbol{\beta}'\mathbf{X}_i + \varepsilon_i, \qquad i = 1, 2, \ldots, n,$$

where

- Y_i is the response variable,
- $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of regression parameters,
- $\mathbf{X}_i$ is a $(p \times 1)$ vector of random variables,
- ε_i is a normal mean zero error term.
- Assume $\mathbf{X}_i \sim N_p(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{XX}})$, $Cov(\mathbf{X}_i, \varepsilon_i) = \mathbf{0}$,
- the joint distribution $(Y_i, \mathbf{X}_i)$ is normal

$$\begin{pmatrix} Y_i \\ \mathbf{X}_i \end{pmatrix} \sim N_{p+1}\left(\begin{bmatrix} \mu_Y \\ \boldsymbol{\mu}_{\mathbf{X}} \end{bmatrix}, \begin{bmatrix} \sigma_Y^2 & \boldsymbol{\sigma}_{Y\mathbf{X}} \\ \boldsymbol{\sigma}_{\mathbf{X}Y} & \boldsymbol{\Sigma}_{\mathbf{XX}} \end{bmatrix} \right).$$

1. Rewrite the joint distribution of $(Y_i, \mathbf{X}_i)$ using only $\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{XX}}, \boldsymbol{\beta}, \sigma_\varepsilon^2$.

$$\begin{pmatrix} Y_i \\ \mathbf{X}_i \end{pmatrix} \sim N_{p+1}\left(\begin{bmatrix} \boldsymbol{\beta}'\boldsymbol{\mu}_{\mathbf{X}} \\ \boldsymbol{\mu}_{\mathbf{X}} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} + \sigma_\varepsilon^2 & \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\mathbf{XX}} \\ \boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} & \boldsymbol{\Sigma}_{\mathbf{XX}} \end{bmatrix} \right).$$

2. Derive the conditional distribution of $Y_i|\mathbf{X}_i = \mathbf{x}_i$.

Using 5.2.3, $Y_i|\mathbf{X}_i = \mathbf{x}_i$ is normal with mean

$$\boldsymbol{\beta}'\boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}}) = \boldsymbol{\beta}'\mathbf{X}$$

and with variance

$$\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} + \sigma_\varepsilon^2 - \boldsymbol{\beta}'\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\Sigma}_{\mathbf{XX}}\boldsymbol{\beta} = \sigma_\varepsilon^2$$

Therefore $\hat{Y}(\mathbf{x}_i) = [Y_i|\mathbf{X}_i = \mathbf{x}_i] \sim N_1(\boldsymbol{\beta}'\mathbf{x}_i, \sigma_\varepsilon^2)$.

5.3 Linear Regression, Random $X's$, and the Multiple Correlation Coefficient

Suppose we have:

$$Y_i = \boldsymbol{\beta}'\mathbf{X}_i + \varepsilon_i, \qquad i = 1, 2, \ldots, n,$$

where

- Y_i is the response variable,
- $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of regression parameters,
- $\mathbf{X}_i$ is a $(p \times 1)$ vector of random variables,
- ε_i is a normal mean zero error term.
- Assume $\mathbf{X}_i \sim N_p(\boldsymbol{\mu}_{\mathbf{X}}, \boldsymbol{\Sigma}_{\mathbf{XX}})$, $Cov(\mathbf{X}_i, \varepsilon_i) = \mathbf{0}$,

You just derived the joint distribution $(Y_i, \mathbf{X}_i)$ and the conditional distribution of $Y_i|\mathbf{X}_i = \mathbf{x}_i$.

Definition: The *multiple correlation coefficient of* Y *with* $\mathbf{X}$ is defined as

$$\rho_{Y:\mathbf{X}} = \mathrm{corr}(Y, \hat{Y}(\mathbf{X})) = \sqrt{\frac{\boldsymbol{\sigma}_{Y\mathbf{X}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\sigma}_{\mathbf{X}Y}}{\sigma_Y^2}}$$

It can be shown that $\rho_{Y:\mathbf{X}}$ is the largest possible correlation between Y and any linear function of the components of $\mathbf{X}$, i.e., for any constant vectors $\mathbf{a}$ and $\mathbf{b}$,

$$\mathrm{corr}(Y, \mathbf{a}'\mathbf{X} + \mathbf{b}) \leq \rho_{Y:\mathbf{X}}.$$

Also,

$$\mathrm{var}(Y|\mathbf{X} = \mathbf{x}) = \sigma_Y^2 - \boldsymbol{\sigma}_{Y\mathbf{X}}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\boldsymbol{\sigma}_{\mathbf{X}Y} = \sigma_Y^2(1 - \rho_{Y:\mathbf{X}}^2).$$

Because of this equation, $\rho_{Y:\mathbf{X}}^2$ is sometimes called "the proportion of the variance of Y explained by $\mathbf{X}$."