STAT 350: Lecture 8

Reading: Chapter 5, chapter 15.

Distribution Theory
Linear and Quadratic Functions of Normals

So far we have defined:

- The standard normal density
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad -\infty < x < \infty$$
- $Z \sim N(0,1)$
- If $Z \sim N(0,1)$ then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$.
- If $Z_1, \ldots, Z_n$ are independent $N(0,1)$ then
$$Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \sim MVN_n(0, I)$$
- $$\mu_X \equiv \mathrm{E}\left(\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \right) = \begin{bmatrix} E(X_1) \\ \vdots \\ \mathrm{E}(X_n) \end{bmatrix}$$
- $$\mathrm{E} \begin{bmatrix} M_{11} & \cdots & M_{1c} \\ \vdots & \cdots & \vdots \\ M_{r1} & \cdots & M_{rc} \end{bmatrix} = \begin{bmatrix} \mathrm{E}(M_{11}) & \cdots & \mathrm{E}(M_{1c}) \\ \vdots & \cdots & \vdots \\ \mathrm{E}(M_{r1}) & \cdots & \mathrm{E}(M_{rc}) \end{bmatrix}$$
- $$\mathrm{Var}(X) = \mathrm{E}\left[(X - \mu_X)(X - \mu_X)^T\right]$$
- ijth entry of $\mathrm{Var}(X)$ is $\mathrm{Cov}(X_i, X_j)$.
- iith entry is $\mathrm{Cov}(X_i, X_i) = \mathrm{Var}(X_i)$.

Now suppose $Z \sim MVN_n(o, I)$. Then

$$E(Z) = \begin{bmatrix} E(Z_1) \\ \vdots \\ E(Z_n) \end{bmatrix} = \mathbf{0}_n$$

The $\mathbf{0}_n$ matches the 0 in $MVN_n(0, I)$.

Next we compute the variance of Z. Note that $E\left[(Z-0)(Z-0)^T\right]$ has ijth entry

$$E(Z_i Z_j) = \begin{cases} 0 & i \neq j \quad \text{(independence)} \\ E(Z_i^2) = 1 & i = j \end{cases}$$

So

$$\mathrm{Var}(Z) = I_{n \times n}$$

the $n \times n$ identity matrix.

Now suppose that

$$X = AZ + \mu$$

where

- A is an $m \times n$ matrix of constants.
- μ is a vector in R^m
- $Z \sim MVN_n(0, I)$

Then we say that X has a $MVN_m(\mu, AA^T)$ distribution.

Now $E(X) = E(AX + \mu)$ has ith component

$$\begin{aligned} E[(AZ)_i] + \mu_i &= E(\sum_j A_{ij} Z_j) + \mu_i \\ &= \sum_j A_{ij} E(Z_j) + \mu_i \\ &= \mu_i \end{aligned}$$

Moreover,

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}(\big[(X-\mu)(X-\mu)^T\big] \\
&= \mathrm{E}[(AZ)(AZ)^T] \\
&= \mathrm{E}[AZZ^TA^T] \\
&= A\mathrm{E}[ZZ^T]A^T \\
&= AIA^T \\
&= AA^T
\end{aligned}$$

The last three lines need some justification. The point is that matrix multiplication by A, whose entries are constants, is just like multiplication by a constant — you can pull the constant outside of the expected value sign. Here is the justification. The ijth entry in $\mathrm{E}[AZZ^TA^T]$ is

$$\begin{aligned}
\mathrm{E}(\sum_k \sum_\ell A_{ik}Z_kZ_\ell A_{j\ell}) &= \sum_{k,\ell} A_{ik}\mathrm{E}(Z_kZ_\ell)A_{j\ell} \\
&= \sum_{k,\ell} A_{ik}I_{k\ell}A_{j\ell} \\
&= \sum_{k,\ell} A_{ik}I_{k\ell}A^T_{\ell j} \\
&= (AIA^T)_{ij} \\
&= (AA^T)_{ij}
\end{aligned}$$

So, $\mathrm{Var}(X) = AA^T$.
Thus $X \sim MVN(\mu, \Sigma)$ means that

- $\mathrm{E}(X) = \mu$
- $\mathrm{Var}(X) = \Sigma$
- X is "normal".

Things to notice along the way

1. $\mathrm{E}(AX + b) = A\mathrm{E}(X) + b$ when $A_{m\times n}$, $X_{n\times 1}$ $b_{m\times 1}$ and A and b are constant.

2. $\mathrm{E}(AMB) = A\mathrm{E}(M)B$ whenever A, B and M are matrices whose dimensions make the multiplication possible and A and B are non-random constant matrices while M is a random matrix.

3. $\mathrm{Var}(AX + b) = A\mathrm{Var}(X)A^T$ where A and X are as in 1). The notation $\mathrm{Cov}(X)$ is sometimes used for $\mathrm{Var}(X)$. This matrix is called the variance-covariance matrix of X.

Application to Least Squares

The following do not use the normal assumption:

$$\begin{aligned}
\hat{\beta} &= (X^TX)^{-1}X^TY \\
&= (X^TX)^{-1}X^T(X\beta + \epsilon) \\
&= (X^TX)^{-1}(X^TX)\beta + (X^TX)^{-1}X^T\epsilon \\
&= \beta + (X^TX)^{-1}X^T\epsilon \\
\mathrm{E}(\hat{\beta}) &= \beta + (X^TX)^{-1}X^T\mathrm{E}(\epsilon) \\
&= \beta
\end{aligned}$$

So $\hat{\beta}$ is **unbiased**.

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}) &= \mathrm{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] \\
&= (X^TX)^{-1}X^T\mathrm{E}(\epsilon\epsilon^T)\left((X^TX)^{-1}X^T\right)^T \\
&= (X^TX)^{-1}X^T\sigma^2 I\left((X^TX)^{-1}X^T\right)^T \\
&= \sigma^2(X^TX)^{-1}
\end{aligned}$$

If, also, $\epsilon_i \sim N(0, \sigma^2)$ then $\epsilon_i/\sigma \sim N(0,1)$ and

$$\begin{aligned}
\hat{\beta} &= \beta + (X^TX)^{-1}X^T\epsilon \\
&= \beta + \sigma(X^TX)^{-1}X^T\left(\frac{\epsilon}{\sigma}\right) \\
&\sim MVN(\beta, \sigma^2\underbrace{(X^TX)^{-1}X^TX(X^TX)^{-1}}_{(X^TX)^{-1}})
\end{aligned}$$

The fitted vector is

$$\hat{\mu} = X\hat{\beta} \sim MVN(X\beta, \sigma^2 X(X^T X)^{-1} X^T)$$

The residual vector is

$$\begin{aligned}
\hat{\epsilon} &= Y - X\hat{\beta} \\
&= X\beta + \epsilon - X\beta - X(X^T X)^{-1} X^T \epsilon \\
&= (I - X(X^T X)^{-1} X^T)\epsilon \\
&\sim MVN(0, \sigma^2 M M^T)
\end{aligned}$$

where $M = I - X(X^T X)^{-1} X^T)$.
Notation:

$$H = X(X^T X)^{-1} X^T$$

Jargon: H is called the **hat** matrix. Notice that

$$\hat{\mu} = HY$$

Extras

I tried, as I was calculating things for the vectors $\hat{\beta}$, $\hat{\mu}$ and $\hat{\epsilon}$ to emphasize which things needed which assumptions.

So for instance we have following matrix identities which depend only on the model equation

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \beta + (X^T X)^{-1} X^T \epsilon$$

$$\hat{\mu} = X\hat{\beta} = \mu + H\epsilon$$

where $\mathbf{H}$ is the 'hat' matrix $X(X^T X)^{-1} X^T$,

$$\hat{\epsilon} = (I - H)\epsilon$$

If we add the assumption that $\mathrm{E}(\epsilon_i) = 0$ for each i then we get

$$\mathrm{E}(\hat{\beta}) = \beta$$

$$\mathrm{E}(\hat{\mu}) = \mu$$

and

$$\mathrm{E}(\hat{\epsilon}) = 0\,.$$

If we add the assumption that the errors are homoscedastic ($\mathrm{Var}(\epsilon_i) = \sigma^2$ for all i) and uncorrelated ($\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$ for all $i \neq j$) then we can compute variances and get

$$\mathrm{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$

$$\mathrm{Var}(\hat{\mu}) = \sigma^2 \mathbf{H}$$

and

$$\mathrm{Var}(\hat{\epsilon}) = \sigma^2 (\mathbf{I} - \mathbf{H}) .$$

NOTE: usually we assume that the ϵ_i are independent and identically distributed which guarantees the homoscedastic, uncorrelated assumption above.

Next we add the assumption that the errors ϵ_i are independent normal variables. Then we conclude that each of $\hat{\beta}$, $\hat{\mu}$ and $\hat{\epsilon}$ have Multivariate Normal distributions with the means and variances as just described.