556: MATHEMATICAL STATISTICS I

EXAMPLES CLASS NOTES

1. **Multivariate Normal Calculations:** In computing the sampling distributions for the sample mean and sample variance statistics, we used properties of the multivariate Normal distribution. Specifically we used results concerning linear transforms of the Normal random vectors.

   Recall that the multivariate normal distribution arises as a location-scale transform of a vector of iid standard Normal components: let $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top$ be a vector if independent rvs with $Z_i \sim Normal(0,1)$. Consider the transform

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{V}\mathbf{Z}$$

   where $\boldsymbol{\mu}$ is $n \times 1$ and $\mathbf{V}$ is $n \times n$ and non-singular. Then $\mathbf{X} \sim Normal_n(\boldsymbol{\mu}, \Sigma)$, with $\Sigma = \mathbf{V}\mathbf{V}^\top$. To see this we may use the multivariate transformation theorem: we have that

$$f_{\mathbf{Z}}(\mathbf{z}) = \prod_{i=1}^{n} \left(\frac{1}{2\pi}\right)^{1/2} \exp\left\{-\frac{1}{2}z_i^2\right\} = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}\mathbf{z}^\top\mathbf{z}\right\}$$

   and hence by the transformation theorem

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{Z}}\left(\mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)|J|$$

   where $|J|$ is the absolute value of the determinant of the transformation. For this linear transformation, basic linear algebra results allow us to conclude that

$$|J| = |\mathbf{V}|^{-1}$$

   that is, the reciprocal of the determinant of $\mathbf{V}$. Hence if $\Sigma = \mathbf{V}\mathbf{V}^\top$, we have that

$$f_{\mathbf{X}}(\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}$$

   as the $\mathbf{z}^\top\mathbf{z}$ term becomes

$$\begin{aligned}\{\mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\mu})\}^\top\{\mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\mu})\} &= (\mathbf{x} - \boldsymbol{\mu})^\top\{\mathbf{V}^{-1}\}^\top\{\mathbf{V}^{-1}\}(\mathbf{x} - \boldsymbol{\mu}) \\ &= (\mathbf{x} - \boldsymbol{\mu})^\top\{\mathbf{V}\mathbf{V}^\top\}^{-1}(\mathbf{x} - \boldsymbol{\mu})\end{aligned}$$

   and

$$|\Sigma| = |\mathbf{V}\mathbf{V}^\top| = |\mathbf{V}||\mathbf{V}^\top| = |\mathbf{V}||\mathbf{V}| = |\mathbf{V}|^2.$$

   The mgf of the multivariate normal is easily computed. If $\mathbf{t} = (t_1, \ldots, t_n)^\top$ is a vector of reals, we have by independence that

$$M_{\mathbf{Z}}(\mathbf{t}) = \mathbb{E}_{\mathbf{Z}}\left[\exp\{\mathbf{t}^\top\mathbf{Z}\}\right] = \prod_{i=1}^{n} \mathbb{E}_{Z_i}\left[\exp\{t_i Z_i\}\right] = \prod_{i=1}^{n} M_{Z_i}(t_i)$$

   and from the formula sheet we therefore have that

$$M_{\mathbf{Z}}(\mathbf{t}) = \prod_{i=1}^{n} \exp\left\{\frac{t_i^2}{2}\right\} = \exp\left\{\frac{\mathbf{t}^\top\mathbf{t}}{2}\right\}.$$

From this result, we compute that

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}_{\mathbf{X}}\left[\exp\{\mathbf{t}^\top \mathbf{X}\}\right] &= \mathbb{E}_{\mathbf{Z}}\left[\exp\{\mathbf{t}^\top(\boldsymbol{\mu} + \mathbf{V}\mathbf{Z})\}\right] \\ &= \exp\{\mathbf{t}^\top \boldsymbol{\mu}\}\mathbb{E}_{\mathbf{Z}}\left[\mathbf{t}^\top(\mathbf{V}\mathbf{Z})\}\right] \\ &= \exp\{\mathbf{t}^\top \boldsymbol{\mu}\} M_{\mathbf{Z}}(\mathbf{V}^\top \mathbf{Z}) \\ &= \exp\left\{\mathbf{t}^\top \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^\top(\mathbf{V}\mathbf{V}^\top)\mathbf{t}\right\} \\ &= \exp\left\{\mathbf{t}^\top \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^\top \Sigma \mathbf{t}\right\}. \end{aligned}$$

Now if Σ has a block diagonal structure

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{bmatrix}$$

where Σ_{11} is $n_1 \times n_1$ and Σ_{22} is $n_2 \times n_2$, then

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22}^{-1} \end{bmatrix}$$

and $|\Sigma| = |\Sigma_{11}||\Sigma_{22}|$. Hence if we consider the partition $\mathbf{X} = (\mathbf{X}_1^\top, \mathbf{X}_2^\top)^\top$ where $\mathbf{X}_1$ is an $n_1 \times 1$ vector, then

$$(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = (\mathbf{x}_1 - \boldsymbol{\mu}_1)^\top \Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) + (\mathbf{x}_2 - \boldsymbol{\mu}_2)^\top \Sigma_{22}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_2)$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the relevant components of $\boldsymbol{\mu}$. Therefore in this block diagonal case, we deduce that

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}_1}(\mathbf{x}_1) f_{\mathbf{X}_2}(\mathbf{x}_2)$$

where

$$\begin{aligned} f_{\mathbf{X}_1}(\mathbf{x}_1) &= \left(\frac{1}{2\pi}\right)^{n_1/2} \frac{1}{|\Sigma_{11}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\mu}_1)^\top \Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)\right\} \\ f_{\mathbf{X}_2}(\mathbf{x}_2) &= \left(\frac{1}{2\pi}\right)^{n_2/2} \frac{1}{|\Sigma_{22}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}_2 - \boldsymbol{\mu}_2)^\top \Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)\right\} \end{aligned}$$

and hence $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent. Similarly for the mgf, we have that

$$\mathbf{t}^\top \boldsymbol{\mu} + \frac{1}{2}\mathbf{t}^\top \Sigma \mathbf{t} = \left(\mathbf{t}_1^\top \boldsymbol{\mu}_1 + \frac{1}{2}\mathbf{t}_1^\top \Sigma_{11}\mathbf{t}_1\right) + \left(\mathbf{t}_2^\top \boldsymbol{\mu}_2 + \frac{1}{2}\mathbf{t}_2^\top \Sigma_{22}\mathbf{t}_2\right)$$

and we conclude independence in the same way.

These results confirm that for the Normal case, the zero blocks in the variance-covariance matrix Σ indicate independence of the components $\mathbf{X}_1$ and $\mathbf{X}_2$. In general for two variables, having a zero covariance does not imply that the variables are independent, although the converse is true.

2. **Order statistics:** If $X_1, \ldots, X_n$ is a random sample, we have in the **continuous** case that the marginal cdf of $Y_j = X_{(j)}$ is

$$F_{Y_j}(x) = \sum_{k=j}^{n} \binom{n}{k} \{F_X(x)\}^k \{1 - F_X(x)\}^{n-k}$$

and the marginal pdf is

$$f_{Y_j}(x) = \frac{n!}{(j-1)!(n-j)!} \{F_X(x)\}^{j-1} \{1 - F_X(x)\}^{n-j} f_X(x)$$

To see this in the continuous case, if the jth order statistic is at x, then we have

(i) a single observation at x, which contributes $f_X(x)$;
(ii) $j-1$ observations which have values less than x, which contributes $\{F_X(x)\}^{j-1}$;
(iii) $n-j$ observations which have values greater than x, which contributes $\{1 - F_X(x)\}^{n-j}$;

Thus the required mass/density is proportional to

$$\{F_X(x)\}^{j-1} f_X(x) \{1 - F_X(x)\}^{n-j}.$$

The normalizing constant is the number of ways of labelling the original x values to obtain this configuration of order statistics: this is

$$n \times \binom{n-1}{j-1} = \frac{n!}{(j-1)!(n-j)!}$$

we may choose the value in step (i) in n ways, and then the $j-1$ data in step (ii) in $\binom{n-1}{j-1}$ ways.

This heuristic argument can verified using direct computation. Recall that in the continuous case the joint pdf of order statistics $Y_1, \ldots, Y_n$ with $Y_j = X_{(j)}$ is

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = n! f_X(y_1) \ldots f_X(y_n) = n! \prod_{i=1}^{n} f_X(y_j) \qquad y_1 < \ldots < y_n$$

as there are $n!$ configurations of the xs that yield identical order statistics, and the result follows by the Theorem of Total Probability. We obtain the marginal pdf of Y_j by integrating out the other $n-1$ variables: we do this in the order $y_1, y_2, \ldots, y_{j-1}$, then $y_n, y_{n-1}, \ldots, y_{j+1}$, and remember that there is a constraint on the support of the pdf

$$y_1 < y_2 < \cdots < y_{j-1} < y_j < y_{j+1} < \cdots < y_{n-1} < y_n$$

- Integrate out y_1:

$$n! \prod_{i=2}^{n} f_X(y_j) \int_{-\infty}^{y_2} f_X(y_1)\, dy_1 = n! \prod_{i=2}^{n} f_X(y_j) F_X(y_2)$$

- Integrate out y_2:

$$n! \prod_{i=3}^{n} f_X(y_j) \int_{-\infty}^{y_3} f_X(y_2) F_X(y_2)\, dy_2 = \frac{n!}{2} \prod_{i=3}^{n} f_X(y_j) \{F_X(y_3)\}^2$$

  using the general calculus result that

$$\int_a^b \frac{dg(t)}{dt} g(t)\, dt = \left[\frac{1}{2}\{g(t)\}^2\right]_a^b = \frac{1}{2}\left(\{g(b)\}^2 - \{g(a)\}^2\right)$$

- Integrate out y_3:

$$\frac{n!}{2}\prod_{i=4}^{n} f_X(y_j)\int_{-\infty}^{y_4} f_X(y_3)\{F_X(y_3)\}^2\,dy_3 = \frac{n!}{2.3}\prod_{i=4}^{n} f_X(y_j)\{F_X(y_3)\}^3$$

Repeating this to finally integrate out up to $j-1$ leaves

$$\frac{n!}{2.3.\ldots.(j-1)}\prod_{i=j}^{n} f_X(y_j)\{F_X(y_j)\}^{j-1}.$$

Now we begin integrating from y_n downwards:

- Integrate out y_n:

$$\frac{n!}{(j-1)!}\prod_{i=j}^{n-1} f_X(y_j)\{F_X(y_j)\}^{j-1}\int_{y_{n-1}}^{\infty} f_X(y_n)\,dy_n$$

$$= \frac{n!}{(j-1)!}\prod_{i=j}^{n-1} f_X(y_j)\{F_X(y_j)\}^{j-1}\{1-F_X(y_{n-1})\}$$

- Integrate out y_{n-1}:

$$\frac{n!}{(j-1)!}\prod_{i=j}^{n-2} f_X(y_j)\{F_X(y_j)\}^{j-1}\int_{y_{n-2}}^{\infty} f_X(y_n)\{1-F_X(y_{n-1})\,dy_{n-1}$$

$$= \frac{n!}{(j-1)!.2}\prod_{i=j}^{n-2} f_X(y_j)\{F_X(y_j)\}^{j-1}\{1-F_X(y_{n-2})\}^2$$

and so on until we have integrated out y_{j+1} to obtain

$$f_{Y_j}(y_j) = \frac{n!}{(j-1)!(n-j)!}\left\{F_X(y_j)\right\}^{j-1}\left\{1-F_X(y_j)\right\}^{n-j} f_X(y_j).$$

The cdf is also readily computable by direct calculation using integration by parts:

$$\begin{aligned}
F_{Y_j}(y_j) &= \int_{-\infty}^{y_j} f_{Y_j}(t)\,dt = \frac{n!}{(j-1)!(n-j)!}\int_{-\infty}^{y_j} f_X(t)\{F_X(t)\}^{j-1}\left\{1-F_X(t)\right\}^{n-j}\,dt \\
&= \frac{n!}{(j-1)!(n-j)!}\left[\frac{1}{j}\{F_X(t)\}^{j}\left\{1-F_X(t)\right\}^{n-j}\right]_{-\infty}^{y_j} \\
&\qquad + \frac{n!}{(j-1)!(n-j)!}\int_{-\infty}^{y_j}\frac{n-j}{j} f_X(t)\{F_X(t)\}^{j}\left\{1-F_X(t)\right\}^{n-j-1}\,dt \\
&= \binom{n}{j}\{F_X(y_j)\}^{j}\left\{1-F_X(y_j)\right\}^{n-j} \\
&\qquad + \binom{n}{j}(n-j)\int_{-\infty}^{y_j} f_X(t)\{F_X(t)\}^{j}\left\{1-F_X(t)\right\}^{n-j-1}\,dt
\end{aligned}$$

Note that in the integrand the power on the second term is reduced to $n-j-1$. Therefore, using this calculation recursively to we obtain

$$\sum_{k=j}^{n} \binom{n}{k} \{F_X(y_j)\}^k \{1 - F_X(y_j)\}^{n-k}$$

Using either the heuristic approach, or direct computation, it is possible to construct the joint pdf for $Y_j = X_{(j)}$ and $Y_k = X_{(k)}$ for $j < k$ as

$$f_{Y_j,Y_k}(y_j, y_k) = n(n-1)\binom{n-2}{j-1}\binom{n-j-1}{n-k}$$
$$f_X(y_j)f_X(y_k)\{F_X(y_j)\}^{j-1}\{F_X(y_k) - F_X(y_j)\}^{k-j-1}\{1 - F_X(y_k)\}^{n-k}$$

for $y_j < y_k$ and zero otherwise. In the special case of $j = 1$ and $k = n$, we obtain

$$f_{Y_1,Y_n}(y_1, y_n) = n(n-1)f_X(y_1)f_X(y_n)\{F_X(y_n) - F_X(y_1)\}^{n-2} \qquad y_1 < y_n$$

In this case, we can also construct the joint cdf: we have that

$$\begin{aligned} F_{Y_1,Y_n}(y_1, y_n) &= P_{Y_1,Y_n}[Y_1 \leq y_1, Y_n \leq y_n] \\ &= P_{X_1,\ldots,X_n}\left[\bigcap_{i=1}^{n}(X_i \leq y_n)\right] - P_{X_1,\ldots,X_n}\left[\bigcap_{i=1}^{n}(y_1 < X_i \leq y_n)\right] \\ &= \{F_X(y_n)\}^n - \{F_X(y_n) - F_X(y_1)\}^n \end{aligned}$$

by independence, as we have the partition

$$(Y_n \leq y_n) = ((Y_1 \leq y_1) \cap (Y_n \leq y_n)) \cup ((Y_1 > y_1) \cap (Y_n \leq y_n))$$

where

- the event $(Y_n \leq y_n)$ corresponds to the event that all of $X_1, \ldots, X_n$ are less than or equal to the value y_n, so
$$(Y_n \leq y_n) = \bigcap_{i=1}^{n}(X_i \leq y_n)$$
- the event $((Y_1 > y_1) \cap (Y_n \leq y_n))$ is equivalent to all X_i lying between y_1 and y_n

$$(Y_1 > y_1) \cap (Y_n \leq y_n) = \bigcap_{i=1}^{n}(y_1 < X_i \leq y_n)$$

and the result follows by probability Axiom III.