ST 661 Note

Frances Lin

1/25/2022

Outline

- mean, covariance
- mgf
- normal distribution & MVN

Defintion?

Let $Z = (z_1, ...z_n)^T$, then

$$E(z) = \begin{bmatrix} E(z_1) \\ E(z_2) \\ \vdots \\ E(z_n) \end{bmatrix}$$

and

$$cov(z) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \ddots & & \\ \vdots & & \ddots & \\ \sigma_{n1} & \cdots & \cdots & \sigma_{nn} \end{bmatrix}$$

Suppose

$$E(Y) = \mu,$$

then

$$cov(Y) = \Sigma = E((y - \mu)(y - \mu)^T)$$

can be shown to be

$$= E(yy^T) - \mu\mu^T.$$

$$\sim var(\mu) = E((\mu - E(\mu))^2) = E(\mu^2) - (E(\mu))^2$$

in the univariate case.

Similarly,

$$\rho = (\rho_{ij})_{mxn} = \begin{bmatrix} 1 & & \rho_{ij} \\ & \ddots & \\ & & 1 \end{bmatrix},$$

ρ_{ij} : correlation between y_i, y_j, and $\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$, where σ_i : sd of $y_i = \sqrt(\sigma_{ij})$.

Note.

1. $\Sigma = cov(y)$ is psd. HW. Use $a^T \Sigma a = cov(a^T y) \geq 0$.
2. $\Sigma = D_\sigma P D_\sigma$, where

$$D_\sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} = (diag(\Sigma))^{\frac{1}{2}}.$$

One way to decompose Σ

$$P = {D_\sigma}^{-1} \Sigma {D_\sigma}^{-1}.$$

Partitioning

Suppose x is dim p, y is is dim q so that z is dim $p+q$, let

$$z = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ \vdots \\ y \\ \vdots \end{bmatrix},$$

then

$$\mu = E(z) = \begin{bmatrix} E(x) \\ E(y) \end{bmatrix},$$

and

$$\Sigma = \begin{bmatrix} (\Sigma_{xx})_{pxp} & (\Sigma_{xy})_{pxq} \\ (\Sigma_{yx})_{qxp} & (\Sigma_{yy})_{qxq} \end{bmatrix},$$

where

$cov(x) = (\Sigma_{xx})_{pxp}$

Σ_{xy} : is the cross covariance matrix.

Σ_{xy} is sometimes denoted by $cov(x, y)$. It is not to be confused by

$$cov(\begin{bmatrix} x \\ y \end{bmatrix})_{(p+q)x(p+q)}.$$

This forms a vector z.

$$\Sigma yx = cov(y, x) = \Sigma xy^T$$

$$\Sigma yx = \begin{bmatrix} (\sigma_{x_1 y_1}) & \cdots & (\Sigma_{x_1 y_q}) \\ \vdots & \ddots & \\ (\sigma_{x_p y_1}) & \cdots & (\Sigma_{x_p y_q}) \end{bmatrix}$$
$$= E(x - E(x))(y - E(y))^T.$$

We need a vertical vector and a horizontal vector to make a matrix.

Once we have

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix},$$

let

$$z = a^T ya = a_1 y_1 + ... + a_n y_n$$

a scalar, then

$$E(z) = E(a^T ya) = E(a_1 y_1 + ... + a_n y_n) = a_1 E(y_1) + ...a_n E(y_n) = a^T E(y)$$

$$=> E(a^T y) = a^T E(y)$$

Next, we want to generalize this to a more general form, but now let

$$z_{px1} = A_{pxn} Y_{nx1}$$

and

$$E(z) = AE(y)$$

Thm (Linearity of expectation)

X : matrix

a, b : constant vector

i) $E(Ay) = AE(y)$

ii) $E(a^T xb) = a^T E(x)b$

iii) $E(AxB) = AE(x)B$

Whatever they are, so long they are constant. (a, b, A, B)

iv) $E(Ay + b) = AE(y) + b$

b is the intercept

Let $z = a^T y$, $cov(y) = \Sigma$, and $E(y) = \mu$, then

$$var(a^T y) = a^T \Sigma a$$
$$= E((a^T y - E(a^T y))^2) = E((a^T y - a^T \mu)^2) = E((a^T (y - \mu))^2)$$

using $(AB)^T = B^T A^T$

$$= E((a^T (y - \mu))((y - \mu)^T a)) = a^T E((y - \mu)(y - \mu)^T) a = a^T \Sigma a$$

using definition of Σ.

Similarly, $cov(a^T y, b^T y) = a^T b$. Because $var(a^T y) = cov(a^T y, a^T y) = a^T \Sigma a$, the previous result is a special case of this result.

Thm

i) $cov(Ay) = A\Sigma A^T$

ii) $cov(Ay, By) = A\Sigma B^T$

iii) $cov(Ay + b) = A\Sigma A^T$

b is constant

Remark. In general, Σ is psd so $A\Sigma A^T$ is also psd. How about pd?

$cov(y) = \Sigma$ is pd.

?? $var(a^T y) \neq 0$. No trivial direction in there. (e.g. $var(a^T y) = 0$. Constant.)

This is the same as saying everything is random. A has to have full row rank for $A\Sigma A^T$ to be pd (i.e. Have to either lower the dim or maintain the dim & not introduce any redundancy).

mgf

Recall that for x

$$M_x(t) = E(e^{tx})$$

$\mathbb{R} \to \mathbb{R}$, uniquely determine the distribution of x. x is random. Even though mgf is deterministic, but it is equally complex.

Next let $y = (y_1, ...y_n)^T$, then

$$M_y(t_1, ...t_n) = M_y(t)$$

$\mathbb{R}^n \to ?$.

Let us take a look of the marginal mgf

$$M_{y_1}(t_1) = E(e^{t_1 y_1})$$

$$\vdots$$

$$M_{y_i}(t_i) = E(e^{t_i y_i}) \quad i = 1, ...n$$

$$M_y(t) = \begin{bmatrix} E(e^{t_1 y_1}) \\ E(e^{t_2 y_2}) \\ \vdots \\ E(e^{t_n y_n}) \end{bmatrix}$$

It works, but does it have the unique property? No. This proposed mgf only talks about the marginal behavior of the dist.

(Another) Def of mgf

Notice that

$$t^T y = t_1 y_1 + ... + t_n y_n$$

$$M_y(t) = E(e^{t^T y}) = E(e^{t_1 y_1 + ... + t_n y_n}),$$

where $(t_1 y_1 + ... + t_n y_n)$ is a scalar so $M_y(t) : \mathbb{R}^n \to \mathbb{R}$. Now. It does uniquely determine the dist of both joint & marginal dist.

e.g.

$$t = \begin{bmatrix} t_1 & 0 & \cdots & 0 \end{bmatrix}$$

$t_1 = t_1$, $t_2 = 0$,… $t_n = 0$

$$M_y(t) = E(e^{t_1 y_1}) = M_{y_1}(t_1).$$

e.g.

$$t = \begin{bmatrix} t_1 & t_2 & \cdots & 0 \end{bmatrix}$$

$t_1 = t_1$, $t_2 = t_2$,… $t_n = 0$

$$M_y(t) = E(e^{t_1 y_1 + t_2 y_2}) = M_{y_1, y_2}(t_1, t_2).$$

Joint mgf gives us not only the marginal but also the joint of any subvectors.

e.g.

$$t = \begin{bmatrix} 0 & \cdots & 0 & t_i & 0 & \cdots & t_j & 0 & \cdots & 0 \end{bmatrix}$$

$t_1 = 0$, $t_2 = 0$,… $t_i = t_i$, …

$$M_y(t) = M_{y_i, y_j}(t_i, t_j).$$

In multivariate, taking the derivative, we get

1st order is 1st moment

$$\frac{\partial}{\partial t_i} M_y(t)\Big|_{t=0} = E(y_i)$$

2nd order is 2nd moment

$$\frac{\partial^2}{\partial t_i t_j} M_y(t)\Big|_{t=0} = E(y_i y_j)$$

$i = j$ get *var*

$i \neq j$ get *cov*

3rd order is 3rd moment

$$\frac{\partial^3}{\partial t_i t_j t_k} M_y(t)\Big|_{t=0} = E(y_i y_j y_k)$$

Important results

page 6

To check

? Has comma $M_y(t) = M_{y_i, y_j}(t_i, t_j)$

? Has comma $\frac{\partial^3}{\partial t_i t_j t_k} M_y(t)\Big|_{t=0} = E(y_i y_j y_k)$