L16: Conditional distributions

1. Operators $E(\cdot)$ and $\mathrm{Cov}(\cdot,\cdot)$

   (1) Definitions

   For random matrix $Z=(Z_{st})_{m\times n}$ define expectation matrix $E(Z) \overset{def}{=\!=} (E(Z_{st}))_{m\times n}$.
   For random vectors $X=(X_1,..,X_m)^T\in R^m$ and $Y=(Y_1,...,Y_n)^T\in R^n$ define the covariance matrix of X and Y $\mathrm{Cov}(X,Y) \overset{def}{=\!=} (\mathrm{cov}(X_s,Y_t))_{m\times n}$. The variance-covariance matrix of X, $\mathrm{Cov}(X,X)$, often has simplified notation $\mathrm{Cov}(X)\in R^{p\times p}$.

   (2) Properties for $E(\cdot)$

   (i) $E(\alpha X+\beta Y)=\alpha E(X)+\beta E(Y)$
   (ii) For non-random D, $E(D)=D$
   (iii) $E(AXB)=AE(X)B$

   Proof: For (iii), we show $[E(AXB)]_{st}=[AE(X)B]_{st}$.

$$\begin{aligned}[E(AXB)]_{st} &= E(AXB)_{st}=E\left[(a_{s1},..,a_{sp})X\begin{pmatrix}b_{1t}\\ \vdots\\ b_{qt}\end{pmatrix}\right]\\ &= =E[\textstyle\sum_{i=1}^p\sum_{j=1}^q X_{ij}a_{si}b_{jt}]=\sum_{i=1}^p\sum_{j=1}^q E(X_{ij})a_{si}b_{jt}\\ &= (a_{s1},..,a_{sp})E(X)\begin{pmatrix}b_{1t}\\ \vdots\\ b_{qt}\end{pmatrix}=[AE(X)B]_{st}\end{aligned}$$

   (3) Properties for $\mathrm{Cov}(\cdot,\cdot)$

   (i) $\mathrm{Cov}(X,Y)=E\{[X-E(X)][Y-E(Y)]^T\}$
   (ii) $\mathrm{Cov}(X,Y)=E(XY^T)-E(X)E(Y^T)$
   (iii) $\mathrm{Cov}(AX+b,BY+d)=A\mathrm{Cov}(X,Y)B^T$

   Proof With $X\in R^m$ and $Y\in R^n$, $\mathrm{Cov}(X,Y)$ and $E[X-E(X)][Y-E(Y)]^T$ are in $R^{m\times n}$. For (i) we show $[\mathrm{Cov}(X,Y)]_{st}=\{E[X-E(X)][Y-E(Y)]^T\}_{st}$.

$$\begin{aligned}[\mathrm{Cov}(X,Y)]_{st} &= \mathrm{cov}(X_s,Y_t)=E[X_s-E(X_s)][Y_t-E(Y_t)]\\ &= E\{[X-E(X)][Y-E(Y)]^T\}_{st}=\{E[X-E(X)][Y-E(Y)]^T\}_{st}.\end{aligned}$$

   Ex1: $E(AXB+CYD+H)=AE(X)B+CE(Y)D+H$
   $\mathrm{Cov}(A_1X+B_1Y+C_1,\ A_2U+B_2V+C_2)$
   $=A_1\mathrm{Cov}(X,U)A_2^T+A_1\mathrm{Cov}(X,V)B_2^T+B_1\mathrm{Cov}(Y,U)A_2^T+B_1\mathrm{Cov}(Y,V)B_2^T.$

   Ex2: For $X\in R^{n\times n}$, $\mathrm{tr}[E(X)]=E[\mathrm{tr}(X)]$. But $|E(X)|\neq E(|X|)$.

2. Marginal distribution of normal vectors

   (1) A 1-1 transformation

   $y=h(x)\in R^p \Longleftrightarrow x=h^{-1}(y)\in R^p$ is a 1-1 mapping with $\frac{\partial y_1,..,y_p}{\partial x_1,..,x_p}\in R^{p\times p}$.

   If $X\in R^p$ has pdf $f(x)$, then $Y=h(X)$ has pdf $f_Y(y)=\left|\frac{f(x)}{\mathrm{abs}\left|\frac{\partial y_1,..,y_p}{\partial x_1,..,x_p}\right|}\right|_{x=h^{-1}(y)}$.

   Proof:

$$\begin{aligned}P(Y\in D) &= P(h(X)\in D)=P(X\in h^{-1}(D))\\ &= \iint_{h^{-1}(D)} f(x)dx_1,...,dx_p\\ &\overset{x=h^{-1}(y)}{=\!=\!=} \iint_D f(x)|_{x=h^{-1}(y)}\,\mathrm{abs}\left|\frac{\partial x_1,..,x_p}{\partial y_1,..,y_p}\right| dy_1,..,dy_p\\ &= \iint_D \left.\frac{f(x)}{\mathrm{abs}\left|\frac{\partial y_1,..,y_p}{\partial x_1,..,x_p}\right|}\right|_{x=h^{-1}(y)} dy_1,..,dy_p.\end{aligned}$$

So $f_Y(y) = \left| \frac{f(x)}{\text{abs}\left|\frac{\partial y_1,..,y_p}{\partial x_1,..,x_p}\right|} \right|_{x=h^{-1}(y)}$ is the pdf for Y.

(2) $X \in N(\mu, \Sigma)$. With non-singular $A \in R^{p\times p}$, $Y = AX + b \sim N(A\mu + b, A\Sigma A^T)$.

Proof $y = Ax + b \Longleftrightarrow x = A^{-1}(y - b)$ with $\frac{\partial y_1,..,y_p}{\partial x_1,..,x_p} = A$.

$x \sim N(\mu, \Sigma) \Longleftrightarrow f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)]$.

$$\begin{aligned} \text{So } f_Y(y) &= \left.\frac{f(x)}{|A|}\right|_{x=A^{-1}(y-b)} \\ &= \frac{1}{(2\pi)^{p/2}|A\Sigma A^T|^{1/2}} \exp\left[-\tfrac{1}{2}(y - A\mu - b)^T(A\Sigma A^T)^{-1}(y - A\mu - b)\right]. \end{aligned}$$

Thus $Y \sim N(A\mu + b, A\Sigma A^T)$.

(3) $X \sim N(\mu, \Sigma)$. With full row rank A, $Y = AX + b \sim N(A\mu + b, A\Sigma A^T)$.

Pf: Expand A to non-singular $\begin{pmatrix} A \\ B \end{pmatrix}$. Then

$\begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix} X + \begin{pmatrix} b \\ 0 \end{pmatrix} \sim N\left(\begin{pmatrix} A \\ B \end{pmatrix}\mu + \begin{pmatrix} b \\ 0 \end{pmatrix}, \begin{pmatrix} A \\ B \end{pmatrix}\Sigma(A^T, B^T)\right)$. So

$$\begin{aligned} f_Y(y) &= \textstyle\iint_z f_{Y,Z}(y, z)\, dz_1, ..., dz_{p-q} \\ &= \cdots = \frac{1}{(2\pi)^{p/2}|A\Sigma A^T|^{1/2}} \exp\left[-\tfrac{1}{2}(y - A\mu - b)^T(A\Sigma A^T)^{-1}(y - A\mu - b)\right]. \end{aligned}$$

Thus $Y \sim N(A\mu + b, A\Sigma A^T)$.

Ex3: Marginal distribution of normal X is normal. For example

With $X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix}\right)$, $\begin{pmatrix} X_1 \\ X_3 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{13} \\ \sigma_{31} & \sigma_3^2 \end{pmatrix}\right)$

since $\begin{pmatrix} X_1 \\ X_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} X$.

Ex4: $X \sim (\mu, \Sigma) \Longrightarrow E(X^TAX) = \mu^TA\mu + \text{tr}(A\Sigma)$.

$$\begin{aligned} E(X^TAX) &= E[\text{tr}(X^TAX)] = E[\text{tr}(AXX^T)] = \text{tr}[E(AXX^T)] = \text{tr}[AE(XX^T)] \\ &= \text{tr}[A(\mu\mu^T + \Sigma)] = \text{tr}(A\mu\mu^T) + \text{tr}(A\Sigma) = \mu^TA\mu + \text{tr}(A\Sigma). \end{aligned}$$

3. Conditional distributions

Let $f_X(x)$, $f_Y(y)$ and $f(x, y)$ be pdfs for $X \in R^p$, $Y \in R^q$ and $\begin{pmatrix} X \\ Y \end{pmatrix} \in R^{p+q}$.

(1) Conditional pdf $f_{Y|x}(y)$

$f_{Y|x}(y) = \frac{f(x,y)}{f_X(x)} \geq 0$ and $\iint_{R^q} f_{Y|x}(y)dy_1, ..dy_q = \frac{f_X(x)}{f_X(x)} = 1$. So $f_{Y|x}(y)$ gives a class of pdfs for Y indexed by x, the conditional pdfs of Y given $X = x$.

(2) Conditional pdf $f_{X|y}(x)$

Similarly, $f_{X|y}(x) = \frac{f(x,y)}{f_Y(y)}$ gives a class of pdfs of X indexed by y, the conditional pdfs of X given $Y = y$.

Ex4: $\begin{pmatrix} X \\ Y \end{pmatrix}$ has pdf $f(x, y) = 2$ on $0 < x < 1$ and $0 < y < 1 - x$. Find the conditional pdf for X given $Y = y$.

(i) $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_0^{1-y} 2\, dx = 2(1 - y)$ on $0 < y < 1$

(ii) $f_{X|Y=y}(x) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{1-y}$ on domain $0 < x < 1 - y$ where $0 < y < 1$.

So $X|y$ has uniform distribution over the interval $(0, 1 - y)$.

L17 Parameters of conditional distributions and independence

1. Parameters of conditional distributions

   (1) $E(Y|x)$ and $\mathrm{Cov}(Y,\, Y|x)$
   At $X = x \in R^p$ for $Y \in R^q$, $f_{Y|x}(y) = \frac{f(x,y)}{f_X(x)}$. Then

$$E(Y|x) = (E(Y_j|x))_{q\times 1} = \left(\iint_{R^q} y_j f_{Y|x}(y)dy_1, ..., dy_q\right)_{q\times 1}$$

   is a vector-valued function of x. The variance-covariance matrix

$$\mathrm{Cov}(Y,\, Y|x) = (\mathrm{cov}(Y_i,\, Y_j|x))_{q\times q} = E(YY^T|x) - E(Y|x)E(Y^T|x)$$

   is a matrix-valued function of x.

   (2) The mean of $E(Y|X)$ and the mean of $\mathrm{Cov}(Y,\, Y|X)$
   $E(Y|X)$ and $\mathrm{Cov}(Y,\, Y|X)$ are vector-valued and matrix-valued functions of $X \in R^p$.

   (i) $E[E(Y|X)] = E(Y)$
   For $E[E(Y|X)] \in R^q$ and $E(Y) \in R^q$ we show $\{E[E(Y|X)]\}_j = [E(Y)]_j$.

$$\begin{array}{rcl} \{E[E(Y|X)]\}_j & = & E[E(Y_j|X)] = \iint_{R^p} E(Y_j|x)f_X(x)dx_1, ..., dx_p \\ & = & \iint_{R^p} \left[\iint_{R^q} y_j f_{Y|x}(y)dy_1, ..., dy_q\right] f_X(x)dx_1, ..., dx_p \\ & = & \iint_{R^{p+q}} y_j f(x,\, y)dx_1, ..., dy_q = E(Y_j) = [E(Y)]_j. \end{array}$$

   (ii) $E[\mathrm{Cov}(Y,\, Y|X)] = E(YY^T) - E[E(Y|X)E(Y^T|X)]$
   By (1) and (i) of (2),

$$\begin{array}{rcl} E[\mathrm{Cov}(Y,\, Y|X)] & = & E[E(YY^T|X) - E(Y|X)E(Y^T|X)] \\ & = & E[E(YY^T|X)] - E[E(Y|X)E(Y^T|X)] \\ & = & E(YY^T) - E[E(Y|X)E(Y^T|X)]. \end{array}$$

   (3) Variance-covariance matrices for $E(Y|X)$ and Y

   (i) $\mathrm{Cov}(E(Y|X),\, E(Y|X)) = E[E(Y|X)E(Y^T|X)] - E(Y)E(Y^T)$

$$\begin{array}{rcl} \mathrm{Cov}(E(Y|X),\, E(Y|X)) & = & E[E(Y|X)E(Y^T|X)] - E[E(Y|X)]E[E(Y^T|X)] \\ & = & E[E(Y|X)E(Y^T|X)] - E(Y)E(Y^T). \end{array} .$$

   (ii) $\mathrm{Cov}(Y,\, Y) = E[\mathrm{Cov}(Y,\, Y|X)] + \mathrm{Cov}(E(Y|X),\, E(Y|X))$.

$$\begin{array}{l} E[\mathrm{Cov}(Y,\, Y|X)] + \mathrm{Cov}(E(Y|X),\, E(Y|X)) \\ = E(YY^T) - E[E(Y|X)E(Y^T|X)] + E[E(Y|X)E(Y^T|X) - E(Y)E(Y^T) \\ = E(YY^T) - E(Y)E(Y^T) = \mathrm{Cov}(Y,\, Y). \end{array}$$

   Comments: $E(Y|X)$ is a function of X; It has the mean $E(Y)$. Its variance-covariance matrix $\mathrm{Cov}(E(Y|X),\, E(Y|X)) = \mathrm{Cov}(Y,\, Y) - E[\mathrm{Cov}(Y,\, Y|X)] \leq \mathrm{Cov}(Y,\, Y)$.

2. Independence

(1) Independence

Let $f_X(x)$, $f_Y(y)$ and $f(x,y)$ be pdfs of $X \in R^p$, $Y \in R^q$ and $\begin{pmatrix} X \\ Y \end{pmatrix} \in R^{p+q}$

$$X \text{ and } Y \text{ are independent } \overset{def}{\Longleftrightarrow} f(x,\, y) = f_X(x)\, f_Y(y).$$

(2) Property I

If X and Y are independent, then $f_{Y|x}(y) = f_Y(y)$ and $f_{X|y}(x) = f_X(x)$.
Hence $P(Y \in A|X = x) = P(Y \in A)$ and $P(X \in B|Y = y) = P(X \in B)$

Proof: $f(x,y) = f_x(x)\, f_Y(y) \Longrightarrow f_{Y|x}(y) = \frac{f(x,y)}{f_X(x)} = f_Y(y)$. So
$P(Y \in A|X = x) = \iint_A f_{Y|x}(y)dy_1, ..., dy_q = \iint_A f_Y(y)dy_1, .., dy_q = P(Y \in A)$.

Comment: Knowing or not knowing X does not affect the pdf and probability for Y.

(3) Property II

If X and Y are independent, then $E(XY^T) = E(X)E(Y^T)$ and hence $\mathrm{Cov}(X,\, Y) = 0$.
In such a case it is said that X and Y are uncorrelated.

Proof: For $E(XY^T)$ and $[E(X)E(Y^T)]$ in $R^{p\times q}$ show $[E(XY^T)]_{ij} = [E(X)E(Y^T)]_{ij}$.

$$\begin{aligned}
[E(XY^T)]_{ij} &= E(X_iY_j) = \iint_{R^{p+q}} x_iy_jf(x,y)dx_1, ..., dy_q \\
&= \iint_{R^{p+q}} x_iy_jf_X(x)f_Y(y)dx_1, ..., dy_q \\
&= \iint_{R^p} x_if_X(x)dx_1, .., dx_p \iint_{R^q} y_jf_Y(y)dy_1, ..., dy_q \\
&= E(X_i)E(Y_j) = [E(X)E(Y^T)]_{ij}.
\end{aligned}$$

Comment: $E(XY^T) = E(X)E(Y^T) \Longleftrightarrow \mathrm{Cov}(X,\, Y) = 0$.
So if X and Y are independent, then X and Y are uncorrelated.

(4) Suppose $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$. Then

$$X \text{ and } Y \text{ are independent} \Longleftrightarrow X \text{ and } Y \text{ are uncorrelated.}$$

Proof: If X and Y are uncorrelated, then $\Sigma_{12} = \mathrm{Cov}(X,\, Y) = 0$ and $\Sigma_{21} = \Sigma_{12}^T = 0$.
So $f(x,\, y) = f_X(x)\, f_Y(y)$. Hence X and Y are independent.

Ex: If $\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$, then $Y|(X = x) \sim N(\mu_y - \Sigma_{21}\Sigma_{11}^{-1}(x - \mu_1),\, \Sigma_{22.1})$
and $X|(Y = y) \sim N(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(y - \mu_1),\, \Sigma_{11.2})$ where $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$
and $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

Proof Let $A = \begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}$. Then $A\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22.1} \end{pmatrix}\right)$.
So $\left(Y - \Sigma_{21}\Sigma_{11}^{-1}X\right)|(X = x) \sim Y - \Sigma_{21}\Sigma_{11}^{-1}x \sim N(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1,\, \Sigma_{22.1})$.
Consequently $Y|(X = x) \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x - \mu_1),\, \Sigma_{22.1})$.

Comment: $E(Y|X) = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X - \mu_1) \sim N(\mu_2,\, \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$.
So $E[E(Y|X)] = \mu_2 = E(Y)$ and
$\mathrm{Cov}(E(Y|X),\, E(Y|X)) = \Sigma_{22} - \Sigma_{22.1} \le \Sigma_{22} = \mathrm{Cov}(Y,\, Y)$.