Lecture 04

$*$ $$X = \begin{bmatrix} X_1 \\ \vdots \\ X_d \end{bmatrix}, \quad \mathbb{E}X = \begin{bmatrix} \mathbb{E}X_1 \\ \vdots \\ \mathbb{E}X_d \end{bmatrix}, \quad Z = \begin{bmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} \\ \vdots & & & \vdots \\ Z_{m1} & Z_{m2} & \cdots & Z_{mn} \end{bmatrix} \in \mathbb{R}^{m\times n}, \quad \mathbb{E}Z = \begin{bmatrix} \mathbb{E}Z_{11} & \mathbb{E}Z_{12} & \cdots & \mathbb{E}Z_{1n} \\ \vdots & & & \vdots \\ \mathbb{E}Z_{m1} & \mathbb{E}Z_{m2} & \cdots & \mathbb{E}_{mn} \end{bmatrix}$$

$$\mathrm{Var}(X) = \mathbb{E}\Big[\underbrace{(X-\mathbb{E}(X))}_{d\times 1}\underbrace{(X-\mathbb{E}(X))}_{d\times 1}^T\Big] \in \mathbb{R}^{d\times d} = \mathbb{E}(XX^T) - \mathbb{E}(X)\mathbb{E}(X)^T$$

(i,j) th element

$$= \begin{bmatrix} \mathrm{Var}(X_1) & & \\ & \ddots & \mathrm{Cov}(X_i, X_j) \\ & & \mathrm{Var}(X_d) \end{bmatrix}$$

$$\mathbb{E}(X_iX_j) - \mathbb{E}(X_i)\cdot\mathbb{E}(X_j) = \mathrm{Cov}(X_i, X_j)$$

Particularly, when $d=1$, $\mathrm{Var}(X) = \mathbb{E}[X-\mathbb{E}(X)]^2 = \mathbb{E}(X^2) - \mathbb{E}^2(X)$

$*$ Assume $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$, $\underset{k\times 1}{Y} = a + \underset{k\times d}{B}\underset{d\times 1}{X}$, $\mathbb{E}(Y) = a + B\cdot\mathbb{E}(X)$

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix}$$

$*$ random matrix : $\mathbb{E}(A\underset{m\times n}{Z}B) = A\cdot\mathbb{E}(Z)\cdot B$

Suppose $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix}$.

claim : $\mathbb{E}(BX) = B\cdot\mathbb{E}(X)$

Since $BX = \begin{pmatrix} b_{11}X_1 + b_{12}X_2 \\ b_{21}X_1 + b_{22}X_2 \\ b_{31}X_1 + b_{32}X_2 \end{pmatrix}$, So $\mathbb{E}(BX) = \begin{pmatrix} b_{11}\mathbb{E}(X_1) + b_{12}\mathbb{E}(X_2) \\ b_{21}\mathbb{E}(X_1) + b_{22}\mathbb{E}(X_2) \\ b_{31}\mathbb{E}(X_1) + b_{32}\mathbb{E}(X_2) \end{pmatrix} = B\cdot\mathbb{E}(X)$

$*$ $\mathrm{Var}(\underset{k\times d}{A}\underset{d\times 1}{X}) = A\,\mathrm{Var}(X)A^T$, $\mathrm{Var}(\underset{1\times d}{a^T}\underset{d\times 1}{X}) = a^T\,\mathrm{Var}(X)\,a \geq 0$ for every $a\in\mathbb{R}^d$

$\Rightarrow$ $\mathrm{Var}(X)$ is positive semi-definite

($Y = a^TX$ is a scalar)

$$\left[\begin{array}{l} M \text{ is PSD if } a^TMa \geq 0 \ \ \forall a\in\mathbb{R}^d \\ M \text{ is PD if } M \text{ is PSD and} \\ a^TMa = 0 \text{ only when } a = 0. \end{array}\right].$$

If M is PSD, $M = BB^T$
If M is PD, $M = BB^T$ with $\mathrm{rank}(B) = d$.

* $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X,Y) + \mathrm{Cov}(Y,X)$

$= \left(\mathrm{Cov}(x_i + y_i, x_j + y_j) \right) = \left(\mathrm{Cov}(x_i, x_j) + \mathrm{Cov}(x_i, y_j) + \mathrm{Cov}(y_i, x_j) + \mathrm{Cov}(y_i, y_j) \right)$

where $\mathrm{Cov}(X,Y) = \mathbb{E}(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))^T = \left(\mathrm{Cov}(x_i, y_j) \right)$

* 1-dim Normal distribution

$X \sim N(\mu, \sigma^2)$ if $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$, $x \in \mathbb{R}$.

$\mu \in \mathbb{R}$, $\sigma > 0$.

(" $X \sim N(0,0)$ " to mean that $X = 0$ w/ prob. 1)

If $X \sim N(\mu, \sigma^2)$, then $a + bX \sim N(a + b\mu, b^2\sigma^2)$

If $X \sim N(\mu, \sigma^2)$, $Y \sim N(\tilde{\mu}, \tilde{\sigma}^2)$, $X \perp\!\!\!\perp Y$, then $X + Y \sim N(\mu + \tilde{\mu}, \sigma^2 + \tilde{\sigma}^2)$

$X - Y \sim N(\mu - \tilde{\mu}, \sigma^2 + \tilde{\sigma}^2)$

$aX + bY \sim N(a\mu + b\tilde{\mu}, a^2\sigma^2 + b^2\tilde{\sigma}^2)$

* MGF.

$\varphi_X(t) := \mathbb{E}(e^{tX})$ $[\varphi_X(0) = 1]$

If $\varphi_X(t) = \varphi_Y(t)$ (both finite and equal) for every $\underbrace{-\varepsilon < t < \varepsilon}_{\text{for some } \varepsilon > 0}$, then X and Y have the same distn.

Denoted by $X \stackrel{d}{=} Y$.

* $X \sim t_\nu \Rightarrow \varphi_X(t)$ is infinite $\forall t > 0$.

× [MGF completely determines the distribution.] w/ caveats

✓ [Characteristic function $\phi_X(t) = \mathbb{E}[e^{itX}]$ completely determines the distn]

* MGF of Gaussian :

$X \sim N(\mu, \sigma^2)$, $\varphi_X(t) = \mathbb{E}(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} dx$

$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \exp\left\{ -\frac{(x-\mu)^2 - 2t\sigma^2 x}{2\sigma^2} \right\} dx$

$$(x-\mu)^2 - 2t\sigma^2 x = x^2 - 2\mu x + \mu^2 - 2t\sigma^2 x$$

Completing squares $= x^2 - 2(\mu + t\sigma^2)\, x + \mu^2$

$$= (x - \mu - t\sigma^2)^2 - (\mu + t\sigma^2)^2 + \mu^2$$

$$= (x - \mu - t\sigma^2)^2 - 2t\mu\sigma^2 - t^2\sigma^4$$

So

$$\psi_X(t) = \exp\left\{ t\mu + \frac{1}{2} t^2 \sigma^2 \right\}$$

For example, $\psi_X(t) = e^{2t + t^2}$, then $X \sim N(2, 2)$

* Use MGF to prove some conclusions:

① If $X \sim N(\mu, \sigma^2)$, $Y = a + bX$,

$$\psi_{a+bX}(t) = \mathbb{E}(e^{t(a+bX)}) = e^{at} \cdot \psi_X(bt)$$

So $\psi_Y(t) = e^{at} \cdot e^{\mu b t + \frac{1}{2} b^2 t^2 \sigma^2}$

$$= e^{(a+b\mu)t + \frac{1}{2}\sigma^2 (b^2 t^2)}$$

$$Y \sim N(a + b\mu, b^2 t^2)$$

② If $X \perp\!\!\!\perp Y$,

$$\psi_{X+Y}(t) = \mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(e^{tX} \cdot e^{tY}) = \mathbb{E}(e^{tX}) \cdot \mathbb{E}(e^{tY}) = \psi_X(t) \cdot \psi_Y(t).$$

So if $X \sim N(\mu, \sigma^2)$, $Y \sim N(\tilde{\mu}, \tilde{\sigma}^2)$

then $\psi_{X+Y}(t) = \psi_X(t) \cdot \psi_Y(t) = \exp\{\mu t + \frac{1}{2}\sigma^2 t^2\} \cdot \exp\{\tilde{\mu} t + \frac{1}{2}\tilde{\sigma}^2 t^2\}$

$$= \exp\{(\mu + \tilde{\mu})t + \frac{1}{2}(\sigma^2 + \tilde{\sigma}^2) t^2\}$$

$\Rightarrow X + Y \sim N(\mu + \tilde{\mu}, \sigma^2 + \tilde{\sigma}^2)$.

* Multivariate Gaussian

$$f(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right\}, \quad x \in \mathbb{R}^d, \ \mu \in \mathbb{R}^d, \ \Sigma \in \mathbb{R}^{d \times d}$$

If $X \sim N(0,1)$, $Y \sim N(0,1)$ $\Rightarrow$ Is $\begin{pmatrix} X \\ Y \end{pmatrix}$ a Normal?

→ If $X \perp\!\!\!\perp Y$, yes

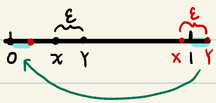
→ If $X \not\!\perp\!\!\!\perp Y$, not necessary!

00:29:00

* If X, Y have the same distn.
say $X, Y \sim U(0,1)$.
$P(X<Y)$ What's the max possible value?

Consider:

$$Y = (X+\varepsilon) \bmod 1 = \begin{cases} X+\varepsilon, & \text{if } X+\varepsilon \le 1 \\ X+\varepsilon-1, & \text{if } X+\varepsilon > 1 \end{cases}$$



* First check that $Y \sim U(0,1)$. Consider $\mathbb{P}(Y \le y)$ for $y \in [0,1]$.
When $y \in [0,\varepsilon)$, $\mathbb{P}(Y \le y) = \mathbb{P}(X+\varepsilon-1 \le y,\ X+\varepsilon > 1) = \mathbb{P}(1-\varepsilon < X < y+1-\varepsilon) = y$.
When $y \in [\varepsilon, 1]$, $\mathbb{P}(Y \le y) = \mathbb{P}(X+\varepsilon-1 \le y,\ X+\varepsilon > 1) + \mathbb{P}(X+\varepsilon \le y,\ X+\varepsilon \le 1)$

$$= \mathbb{P}(1-\varepsilon < X < y+1-\varepsilon) + \mathbb{P}(X+\varepsilon \le y)$$
$$= \mathbb{P}(1-\varepsilon < X < 1) + \mathbb{P}(X \le y-\varepsilon)$$
$$= \varepsilon + y - \varepsilon = y.$$

* Next, compute $P(X<Y) = P(X+\varepsilon \le 1) = 1-\varepsilon \xrightarrow{\varepsilon \to 0} 1$
However, when $\varepsilon = 0$, $X = Y$, so $P(X<Y) = 0$.
This indicates that $P(X<Y) < 1$ but cannot reach to 1.