# ST5215 Advanced Statistical Theory, Lecture 8

HUANG Dongming

National University of Singapore

4 Sep 2020

# Overview

Last time

- Minimal sufficiency

Today

- Completeness
- Basu's theorem

# Recap: Minimal sufficiency

We want to find sufficient statistics that provide the greatest reduction

### Definition

Let $T$ be a sufficient statistic for $P \in \mathcal{P}$. $T$ is called a minimal sufficient statistic if and only if, for any other statistic $S$ sufficient for $P \in \mathcal{P}$, there is a measurable function $\psi$ such that $T = \psi(S)$ $\mathcal{P}$-a.s.

- Minimal sufficient statistics are **unique**: they are equivalent (i.e., there is a one-to-one measurable mappings between them)
- Minimal sufficient statistics **exist** under weak conditions, e.g., $\mathcal{P}$ is on $\mathcal{R}^n$ and dominated by a $\sigma$-finite measure
- Tools to check whether a sufficient statistic $T$ is minimal
    1. Find a countable sub-family $\{f_i\}$ and show $T$ is equivalent to the vector of density ratios $f_i/f_\infty$ (or $f_i/f_0$)
    2. Check if this holds: the density ratio $f_\theta(y)/f_\theta(x)$ as a function of $\theta$ is constant $\Rightarrow T(x) = T(y)$

# Ancillary statistics

- A minimal sufficient statistic might not be "simplest sufficient statistic"
    - May still contain redundant information
    - e.g. if $\overline{X}$ is minimal sufficient, then so is $(\overline{X}, \exp(\overline{X}))$
    - Need a notion that describes whether a statistic can be further reduced
- A statistic $V(X)$ is said to be *ancillary* if its distribution does not depend on the population $P$; $V(X)$ is said to be *first-order ancillary* if $E_P[V(X)]$ does not depend on $P$
    - e.g.: trivial ancillary statistic: $V(X) = c$
- If $V(X)$ is a nontrivial ancillary statistic, then $\sigma(V(X))$ is a nontrivial $\sigma$-field that does not contain any information about $P$
- Similarly, for a statistic $S(X)$, if $V(S(X))$ is ancillary, then $\sigma(S(X))$ contains a nontrivial $\sigma$-field $\sigma(V(S(X)))$ that does not contain any information about $P$
    - The "data" $S(X)$ may be further reduced
- A sufficient statistic $T$ is good in reducing data **if no non-constant function of $T$ is ancillary or even first-order ancillary**

# Completeness

## Definition

- A statistic $T(X)$ is said to be *complete* for $P \in \mathcal{P}$ if and only if, for any Borel function $f$, $E_P f(T) = 0$ for all $P \in \mathcal{P}$ implies that $f(T) = 0$ $\mathcal{P}$-a.s.
- $T$ is said to be *boundedly complete* if and only if the previous statement holds for any bounded Borel functions $f$.

- Intuition: A complete statistic contains "completely" useful information about $P$; no redundancy
- Clearly, a complete statistic is boundedly complete
- If $T$ is complete and $S = \psi(T)$ for a measurable function $\psi$, then $S$ is also complete
  - Similar result holds for bounded completeness
- A complete sufficient statistic is effective in reducing the data, so we expect that a complete sufficient statistic is always minimal.

# Completeness + Sufficiency $\Rightarrow$ Minimal Sufficiency

**Proposition**

*Suppose $X$ is a sample from unknown $P \in \mathcal{P}$, and suppose a minimal sufficient statistic exists.*
*If a statistic $U$ is sufficient and boundedly complete, then $U$ is minimal sufficient.*

We first assume $U$ is one-dimensional

- Suppose $T$ is minimal sufficient. We have $T = \varphi(U)$
- We want to show that $U = \psi(T)$ for some $\psi$
- Let $\phi : \mathcal{R} \mapsto \mathcal{R}$ be a bounded 1-1 continuous function (such as $\phi(t) = \arctan(t), 1/(1 + e^{-t})$, *etc.*)
- Note that $E_P\left[\phi(U) \mid T\right]$ does not depend on $P$ and can be written as $\eta(T)$. Easy to see that $\eta$ is also bounded
- Since $E_P\left[\phi(U) - \eta(\phi(U))\right] = E_P\left[\phi(U) - \eta(T)\right] = E_P\left[\phi(U) - E_P\left[\phi(U) \mid T\right]\right] = 0$ for all $P \in \mathcal{P}$ and $U$ is boundedly complete, we have $\phi(U) - \eta(T) = 0$ $\mathcal{P}$-a.s.
- Therefore $U = \phi^{-1}(\eta(T))$ $\mathcal{P}$-a.s.

If $U$ is $p$-dimensional, we use the same argument with two modifications:

- Use the notation $\phi(U)$ to denote the vector$(\phi(U_1), \phi(U_2), \ldots, \phi(U_p))$, and $\eta(T)$ to denote the vector $(E_P[\phi(U_1) \mid T], \ldots, E_P[\phi(U_p) \mid T])$
- For each $i = 1, \ldots, p$, it holds that $E_P\left[(\phi(U))_i - (\eta(\varphi(U)))_i\right] = 0$. By the completeness of $U$, we conclude that $\phi(U) = \eta(T)$ a.s.

**Remark.** A minimal sufficient statistic is **not** necessarily complete
Example:

- $P \in \{\text{Unif}(\theta, \theta + 1) : \theta \in \mathcal{R}\}$
- $T(X) = (X_{(1)}, X_{(n)})$ is minimal sufficient
- But $T_2 - T_1 = X_{(n)} - X_{(1)}$ does not depend on $P$ so $T$ is not complete

# Complete Sufficient Statistics in Exponential Families

## Proposition (A)

*If $\mathcal{P}$ is a natural exponential family of full rank with p.d.f.'s given by*

$$f_\eta(x) = exp\Big\{\eta^\tau T(x) - \zeta(\eta)\Big\}h(x), \tag{1}$$

*then $T(X)$ is complete and sufficient for $\eta \in \Xi$.*

**Remark.** This result provides another way to show $T$ is minimal sufficient.
**Proof:** We have shown that $T$ is sufficient. Now show $T$ is complete.

The argument here is standard.
Suppose $f$ is a function such that $E_\eta[f(T)] = 0$ for all $\eta \in \Xi$, i.e.,

$$\int f(t)\exp\{\eta^\tau t - \zeta(\eta)\}d\lambda = 0 \quad \text{for all } \eta \in \Xi, \tag{2}$$

where $\lambda$ is a measure on $(\mathcal{R}^p, \mathcal{B}^p)$ (Theorem 2.1 in the textbook)

# Exponential Families (Cont.)

- Let $\eta_0$ be an interior point of $\Xi$
- Then we can find a neighborhood of $\eta_0$,
  $N(\eta_0) = \{\eta \in \mathcal{R}^p : \|\eta - \eta_0\| < \epsilon\}$ for some $\epsilon > 0$, such that

$$\int f_+(t) e^{\eta^\tau t} d\lambda = \int f_-(t) e^{\eta^\tau t} d\lambda, \qquad \forall \eta \in N(\eta_0). \tag{3}$$

- In particular,

$$\int f_+(t) e^{\eta_0^\tau t} d\lambda = \int f_-(t) e^{\eta_0^\tau t} d\lambda = c. \tag{4}$$

- If $c = 0$, then $f = 0$ a.e. $\lambda$
- If $c > 0$, then $c^{-1} f_+(t) e^{\eta_0^\tau t}$ and $c^{-1} f_-(t) e^{\eta_0^\tau t}$ are p.d.f.'s w.r.t. $\lambda$
  - Eq. (3) implies that their m.g.f.'s are the same in a neighborhood of 0
  - By the uniqueness of the m.g.f., we conclude the two p.d.f.'s are the same $\lambda$-a.e., which implies $f = f_+ - f_- = 0$ $\lambda$-a.e.
- Hence $T$ is complete.

# Example: normal families

Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables having the $N(\mu, \sigma^2)$ distribution, $\mu \in \mathcal{R}$, $\sigma > 0$ are **unknown**

- It is easy to check that the joint p.d.f. is

$$(2\pi)^{-n/2} \exp\left\{\eta_1 T_1 + \eta_2 T_2 - n\zeta(n)\right\}, \tag{5}$$

  where
  - $T_1 = \sum_{i=1}^n X_i$
  - $T_2 = -\sum_{i=1}^n X_i^2$
  - $\eta = (\eta_1, \eta_2) = (\mu/\sigma^2, 1/(2\sigma^2))$
- This is a natural exponential family of full rank: $\Xi = \mathcal{R} \times (0, \infty)$ is an open set of $\mathcal{R}^2$
- So $T(X) = (T_1(X), T_2(X))$ is complete and sufficient for $\eta$
- There is a one-to-one correspondence between $\eta$ and $\theta = (\mu, \sigma^2)$
  - $T$ is also complete and sufficient for $\theta$
- There is a one-to-one correspondence between $(\overline{X}, S^2)$ and $(T_1, T_2)$
  - $(\overline{X}, S^2)$ is complete and sufficient for $\theta$

# Example: Uniform Family with Varying Right Ends

Let $X_1, \ldots, X_n \sim P_\theta = U(0, \theta)$ be i.i.d. for $\theta > 0$. We will show that the largest order statistic, $X_{(n)}$, is complete and sufficient for $\theta$

- The sufficiency follows from Factorization theorem: the joint p.d.f. is $\theta^{-n} I_{(0,\theta)}(x_{(n)})$
- The CDF of $X_{(n)}$ is

$$
\begin{aligned}
&= P_\theta(X_{(n)} \le x) = P_\theta(X_1 \le x, \ldots, X_n \le x) \\
&= \prod_{i=1}^{n} P_\theta(X_i \le x) = \frac{x^n}{\theta^n} I_{(0,\theta)}(x).
\end{aligned}
$$

- The p.d.f. of $X_{(n)}$ is then

$$
\frac{n x^{n-1}}{\theta^n} I_{(0,\theta)}(x).
$$

# Example (Cont.)

- Let $g$ be a Borel function on $[0, \infty)$ s.t. $E_\theta[g(X_{(n)})] = 0$ for all $\theta > 0$. Then
$$\int_0^\theta g(x)x^{n-1}\,\mathrm{d}x = 0$$
for all $\theta > 0$
- **Differentiate** the above w.r.t. $\theta$
$$g(\theta)\theta^{n-1} = 0$$
- Thus, $g(\theta) = 0$ for all $\theta > 0$
- By definition, $X_{(n)}$ is complete for $\theta$

# Basu's Theorem

The relationship between an ancillary statistic and a complete and sufficient statistic is characterized in the following result

**Theorem (Basu)**

*Let $V$ and $T$ be two statistics of $X$ from a population $P \in \mathcal{P}$. If $V$ is ancillary and $T$ is boundedly complete and sufficient for $P \in \mathcal{P}$, then $V$ and $T$ are independent w.r.t. any $P \in \mathcal{P}$.*

- Intuition: $\sigma(V)$ does not contain information about $P$, while $\sigma(T)$ carries non-redundant and sufficient information about $P$. This suggests that $V$ and $T$ are independent
- Basu's theorem is useful in proving the independence of two statistics.

Proof:

- Let $A$ be an event on the range of $T$ and $B$ be an event on the range of $V$. Need to show

$$P(T \in A, V \in B) = P(T \in A)P(V \in B), \qquad \forall P \in \mathcal{P}$$

- Since $V$ is ancillary, $P(V \in B)$ is a constant $c_B$
- Since $T$ is sufficient, $E_P[I_B(V)|T]$ can be written as $h_B(T)$
- By tower property, for any $P \in \mathcal{P}$, it holds that $E_p[h_B(T) - c_B] = E_P\{E_P[I_B(V)|T] - P(V \in B)\} = 0$
- By the bounded completeness of $T$, $h_B(T) - c_B = 0$ $\mathcal{P}$-a.s.
- By the properties of conditional expectation, for all $P \in \mathcal{P}$,

$$\begin{aligned} P(T \in A, V \in B) &= E_P\{E_P[I_A(T)I_B(V)|T]\} \\ &= E_P\{I_A(T)E[I_B(V)|T]\} \\ &= E_P\{I_A(T)P(V \in B)\} \\ &= P(T \in A)P(V \in B), \end{aligned}$$

where the 3rd equation is due to $h_B(T) - c_B = 0$, $P$-a.s.

# Example: $\bar{X}$ and $S^2$ of a normal sample

Suppose $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ are i.i.d.
We want to show that the sample mean $\bar{X}$ and sample variance $S^2$ are independent

- This result does not rely on any model, but we can postulate a model to prove it using Basu's theorem
- Consider the family $\{N(\mu, \sigma^2) : \mu \in \mathcal{R}\}$ with $\sigma^2$ known
- Natural parameter: $\eta = \mu/\sigma^2$
- Easy to show it is an exponential family of full rank
- Then $\overline{X}$ is complete and sufficient for $\mu$ according to Proposition A
- Rewrite $S^2 = (n-1)^{-1} \sum_{i=1}^n (Z_i - \overline{Z})^2$, where $Z_i = X_i - \mu \sim N(0, \sigma^2)$ and $\bar{Z}$ is the sample mean of $Z_i$'s.
  $S^2$ is ancillary w.r.t. $\mu$, since its distribution does not depend on $\mu$
- By Basu's theorem, $\overline{X}$ and $S^2$ are independent w.r.t. $N(\mu, \sigma^2)$ for any $\mu \in \mathcal{R}$
- Since $\sigma^2$ is arbitrary, $\overline{X}$ and $S^2$ are independent w.r.t. $N(\mu, \sigma^2)$ for any $\mu \in \mathcal{R}$ and $\sigma^2 > 0$

# Example (Cont.)

Using the independence of $\bar{X}$ and $S^2$, we can show that $(n-1)S^2/\sigma^2$ has the chi-square distribution $\chi^2_{n-1}$

- Note that

$$n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2 + \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2,$$

  whose m.g.f. equals to the product of the m.g.f.'s of $n(\bar{X}-\mu)^2/\sigma^2$ and $(n-1)S^2/\sigma^2$
- $n(\bar{X}-\mu)^2/\sigma^2$ has the chi-square distribution $\chi^2_1$ with the m.g.f. $(1-2t)^{-1/2}, t < 1/2$
- $\sum_{i=1}^{n}(X_i-\mu)^2/\sigma^2$ has the chi-square distribution $\chi^2_n$ with the m.g.f. $(1-2t)^{-n/2}, t < 1/2$
- By the independence of $\bar{X}$ and $S^2$, the m.g.f. of $(n-1)S^2/\sigma^2$ is

$$(1-2t)^{-n/2}/(1-2t)^{-1/2} = (1-2t)^{-(n-1)/2}$$

  for $t < 1/2$
- This is the m.g.f. of the chi-square distribution $\chi^2_{n-1}$

# Tutorial

1. Ex 1.6.35 in JS
Let $\{a_n\}$ be a sequence of positive numbers satisfying $\sum_{n=1}^{\infty} a_n = 1$ and let $\{P_n\}$ be a sequence of probability measures on a common measurable space. Define $P = \sum_{n=1}^{\infty} a_n P_n$.
(a) Show that $P$ is a probability measure.
(b) Show that $P_n \ll \nu$ for all $n$ and a measure $\nu$ if and only if $P \ll \nu$ and, when $P \ll \nu$ and $\nu$ is $\sigma$ -finite, $\frac{dP}{d\nu} = \sum_{n=1}^{\infty} a_n \frac{dP_n}{d\nu}$
(c) Derive the Lebesgue p.d.f. of $P$ when $P_n$ is the gamma distribution $\Gamma\left(\alpha, n^{-1}\right)$ (Table 1.2) with $\alpha > 1$ and $a_n$ is proportional to $n^{-\alpha}$
2. Ex 1.6.40 in JS (See Tutorial 2 in Lecture 5)
3. Ex 1.6.46 in JS
Let $X_1$ and $X_2$ be independent random variables having the standard normal distribution. Obtain the joint Lebesgue p.d.f. of $(Y_1, Y_2)$ where $Y_1 = \sqrt{X_1^2 + X_2^2}$ and $Y_2 = X_1/X_2$. Are $Y_i$ 's independent?
4. Ex 1.6.47 in JS
Let $X_1$ and $X_2$ be independent random variables and $Y = X_1 + X_2$ Show that $F_Y(y) = \int F_{X_2}(y-x)dF_{X_1}(x)$
5. Ex 1.6.58(b,c) in JS
Let $X = N_k(\mu, \Sigma)$ with a positive definite $\Sigma$.
(a) Show that $EX = \mu$ and $\text{Var}(X) = \Sigma$
(b) Let $A$ be an $l \times k$ matrix and $B$ be an $m \times k$ matrix. Show that $AX$ and $BX$ are independent if and only if $A\Sigma B^{\tau} = 0$
(c) Suppose that $k = 2$, $X = (X_1, X_2)$, $\mu = 0$, $\text{Var}(X_1) = \text{Var}(X_2) = 1$, and $\text{Cov}(X_1, X_2) = \rho$. Show that $E(\max\{X_1, X_2\}) = \sqrt{(1-\rho)/\pi}$

# Ex 1.6.35 in JS

Let $\{a_n\}$ be a sequence of positive numbers satisfying $\sum_{n=1}^{\infty} a_n = 1$ and let $\{P_n\}$ be a sequence of probability measures on a common measurable space. Define $P = \sum_{n=1}^{\infty} a_n P_n$.
(a) Show that $P$ is a probability measure.
(b) Show that $P_n \ll \nu$ for all $n$ and a measure $\nu$ if and only if $P \ll \nu$ and, when $P \ll \nu$ and $\nu$ is $\sigma$ -finite, $\frac{dP}{d\nu} = \sum_{n=1}^{\infty} a_n \frac{dP_n}{d\nu}$

(c) Derive the Lebesgue p.d.f. of $P$ when $P_n$ is the gamma distribution $\Gamma\left(\alpha, n^{-1}\right)$ (Table 1.2) with $\alpha > 1$ and $a_n$ is proportional to $n^{-\alpha}$

**Proof: (a)**. Checking conditions (i) (ii) is straightforward.
For (iii), suppose $A_i \in \mathcal{F}, i = 1, \ldots,$ and $A_i$'s are disjoint. We have $P(\cup_i A_i) = \sum_{n=1}^{\infty} a_n P_n(\cup_i A_i) = \sum_{n=1}^{\infty} a_n \sum_{i=1}^{\infty} P_n(A_i) = \sum_{i=1}^{\infty} \sum_{n=1}^{\infty} a_n P_n(A_i) = \sum_{i=1}^{\infty} P(A_i)$, where the 3rd equality is due to the Fubini's theorem.
**(b).** "$\Rightarrow$": For any $A \in \mathcal{F}$ s.t. $\nu(A) = 0$, then $P_n(A) = 0$ for all $n$, and thus $P(A) = 0$.
"$\Leftarrow$": For any $A \in \mathcal{F}$ s.t. $\nu(A) = 0$, then $P(A) = 0$ and thus $a_n P_n(A) = 0$ for each $n$. Since $a_n > 0$, we have $P_n(A) = 0$.

# Ex 1.6.35 in JS (Cont.)

If $P \ll \nu$ and $\nu$ is $\sigma$-finite, by Fubini's theorem,
$P(A) = \sum a_n P_n(A) = \sum a_n \int I_A \frac{\mathrm{d}P_n}{\mathrm{d}\nu}\ \mathrm{d}\nu = \int I_A \sum a_n \frac{\mathrm{d}P_n}{\mathrm{d}\nu}\ \mathrm{d}\nu.$
So $\frac{\mathrm{d}P}{\mathrm{d}\nu} = \sum a_n \frac{\mathrm{d}P_n}{\mathrm{d}\nu}$ by the uniqueness of R-N derivative.

**(c).** Using part (b), the p.d.f. of $P$ is given by

$$
\begin{aligned}
f(x) &= \sum_n c n^{-\alpha} \frac{n^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-nx) \\
&= \frac{c}{\Gamma(\alpha)} x^{\alpha-1} \sum_n \exp(-nx) \\
&= \frac{c}{\Gamma(\alpha)} \frac{x^{\alpha-1}}{e^x - 1},
\end{aligned}
$$

where $c = (\sum_n n^{-\alpha})^{-1}$ □

# Ex 1.6.46 in JS

Let $X_1$ and $X_2$ be independent random variables having the standard normal distribution. Obtain the joint Lebesgue p.d.f. of $(Y_1, Y_2)$ where $Y_1 = \sqrt{X_1^2 + X_2^2}$ and $Y_2 = X_1/X_2$. Are $Y_i$ 's independent?

We need the following result:

## Proposition

- *Let $X$ be a random $k$-vector with a Lebesgue density $f_X$ and let $Y = g(X)$, where $g$ is a Borel function from $\left(\mathcal{R}^k, \mathcal{B}^k\right)$ to $\left(\mathcal{R}^k, \mathcal{B}^k\right)$.*
- *Let $A_1, \ldots, A_m$ be disjoint sets in $\mathcal{B}^k$ such that $\mathcal{R}^k - (A_1 \cup \cdots \cup A_m)$ has Lebesgue measure 0 and $g$ on $A_j$ is one-to-one with a non-vanishing Jacobian, i.e., the determinant $\mathrm{Det}(\partial g(x)/\partial x) \neq 0$ on $A_j, j = 1, \ldots, m$.*
- *Let $h_j$ be the inverse function of $g$ on $A_j, j = 1, \ldots, m$*

*Then $Y$ has the following Lebesgue density:*

$$f_Y(y) = \sum_{j:1\leq j\leq m, y\in g(A_j)} \left|\mathrm{Det}\left(\partial h_j(y)/\partial y\right)\right| f_X\left(h_j(y)\right)$$

- Let $A_1 = \{(x_1, x_2) : x_1 > 0, x_2 > 0\}$, $A_2 = \{(x_1, x_2) : x_1 > 0, x_2 < 0\}$ $A_3 = \{(x_1, x_2) : x_1 < 0, x_2 > 0\}$, and $A_4 = \{(x_1, x_2) : x_1 < 0, x_2 < 0\}$.
- $\mathcal{R}^2 - (A_1 \cup A_2 \cup A_3 \cup A_4) = \{(0,0)\}$ has measure 0
- On each $A_i$, the function $(y_1, y_2) = \left(\sqrt{x_1^2 + x_2^2}, x_1/x_2\right)$ is one-to-one with

$$\left|\text{Det}\left(\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}\right)\right| = \left|\text{Det}\begin{pmatrix} \frac{y_2}{\sqrt{1+y_2^2}} & \frac{y_1}{\sqrt{1+y_2^2}} - \frac{y_1 y_2^2}{\left(1+y_2^2\right)^{3/2}} \\ \frac{1}{\sqrt{1+y_2^2}} & -\frac{y_1 y_2}{\left(1+y_2^2\right)^{3/2}} \end{pmatrix}\right| = \frac{y_1}{1+y_2^2}.$$

- Since the joint Lebesgue density of $(X_1, X_2)$ is

$$\frac{1}{2\pi} e^{-\left(x_1^2 + x_2^2\right)/2}$$

and $x_1^2 + x_2^2 = y_1^2$, the joint Lebesgue density of $(Y_1, Y_2)$ is

$$\sum_{j:1 \le j \le m, y \in g(A_j)} \frac{1}{2\pi} e^{-\left(x_1^2 + x_2^2\right)/2} \left|\text{Det}\left(\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}\right)\right| = (y_1 e^{-y_1^2/2})(\frac{1}{\pi}\frac{1}{1+y_2^2}).$$

- This p.d.f. is the product of p.d.f.'s of $Y_1$ and $Y_2$, so they are independent.

# Ex 1.6.47 in JS

Let $X$ and $Y$ be independent random variables and $Z = X + Y$. Show that $F_Z(t) = \int F_Y(t-x)dF_X(x)$

**Remark**. The notation $\int \mathrm{d}F_X$ is the same as $\int \mathrm{d}P_X$. The former one is historically used for Riemann–Stieltjes integral.
**Proof:** Since $X$ and $Y$ are independent, their joint distribution is given by the product measure $P_X \times P_Y$.

$$
\begin{aligned}
F_{X+Y}(t) &= P(X + Y \le t) \\
&= \int I_{(-\infty,t)}(x+y)\,\mathrm{d}(P_X \times P_Y)(x,y) \\
&= \int \left( \int I_{(-\infty,t)}(x+y)\,\mathrm{d}P_Y(y) \right)\,\mathrm{d}P_X(x) \\
&= \int F_Y(t-x)dF_X(x),
\end{aligned}
$$

where the 3rd equality follows from Fubini's theorem.

# Ex 1.6.58 in JS

Let $X = N_k(\mu, \Sigma)$ with a positive definite $\Sigma$.
(b) Let $A$ be an $l \times k$ matrix and $B$ be an $m \times k$ matrix. Show that $AX$ and $BX$ are independent if and only if $A\Sigma B^\tau = 0$
(c) Suppose that $k = 2, X = (X_1, X_2), \mu = 0, \mathrm{Var}(X_1) = \mathrm{Var}(X_2) = 1$, and $\mathrm{Cov}(X_1, X_2) = \rho$. Show that $E(\max\{X_1, X_2\}) = \sqrt{(1-\rho)/\pi}$

**Proof:** We will make use of the following properties of normal random vectors:

- A linear transformation of a normal random vector is still a normal random vector
- Two components of a normal random vector are independent if and only if they are uncorrelated

These can be proved by using ch.f.
Formally, if $X \sim N(\mu, \Sigma)$ is a normal random $n$-vector, then

- For any $C \in \mathcal{R}^{m \times n}$, $CX \sim N(C\mu, C\Sigma C^\top)$.
- $X_i \perp X_j$ iff $\Sigma_{ij} = 0$

# Ex 1.6.58 in JS (Cont.)

Part (b):

- Let $C = \begin{pmatrix} A \\ B \end{pmatrix}$, then $CX \sim N(C\mu, C\Sigma C^\top)$
- We have $AX \perp\!\!\!\perp BX \Leftrightarrow C\Sigma C^\top[1,2] = \mathbf{0}$, which is $A\Sigma B^\top = \mathbf{0}$.

Part (c):

- Note that

$$|X - Y| = \max\{X, Y\} - \min\{X, Y\} = \max\{X, Y\} + \max\{-X, -Y\}$$

- Since the mean is 0, the random vector $(X, Y)$ is symmetric about 0, so the distribution of $\max\{X, Y\}$ and $\max\{-X, -Y\}$ are the same
- $E|X - Y| = 2E(\max\{X, Y\})$
- $X - Y \sim N(0, 2 - 2\rho)$ because $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X, Y) = 2 - 2\rho$
- $E|X - Y| = \sqrt{2/\pi}\sqrt{2 - 2\rho} = 2\sqrt{(1 - \rho)/\pi}$
- Hence $E(\max\{X, Y\}) = \sqrt{(1 - \rho)/\pi}$

# Ex 1.6.86 in JS

Let $X$ and $Y$ be integrable random variables on $(\Omega, \mathcal{F}, P)$ and $\mathcal{A} \subset \mathcal{F}$ be a $\sigma$-field. Show that $E[YE(X \mid \mathcal{A})] = E[XE(Y \mid \mathcal{A})]$, assuming that both integrals exist.

**Proof:** We split the proof in 3 parts: (1) $Y$ is bounded; (2) $X, Y \geq 0$; (3) $X, Y$ are general r.v.s

**(1). Suppose $Y$ is bounded**. Both $YE(X \mid \mathcal{A})$ and $XE(Y \mid \mathcal{A})$ are integrable.

Using tower property and the fact that $E[E(X \mid \mathcal{A})] = EX$, we obtain that

$$\begin{aligned} E[YE(X \mid \mathcal{A})] &= E\{E[YE(X \mid \mathcal{A}) \mid \mathcal{A}]\} \\ &= E[E(X \mid \mathcal{A})E(Y \mid \mathcal{A})] \\ &= E\{E[XE(Y \mid \mathcal{A}) \mid \mathcal{A}]\} \\ &= E[XE(Y \mid \mathcal{A})] \end{aligned}$$

**(2).Suppose $X, Y \geq 0$.**

- Let $Y_n = \min\{Y, n\}, n = 1, 2, \ldots$. Then $Y_n \leq n$.
- Then $0 \leq Y_1 \leq Y_2 \leq \cdots \leq Y$ and $\lim_n Y_n = Y$.
- By the properties of conditional expectation, $0 \leq E\,(Y_1 \mid \mathcal{A}) \leq E\,(Y_2 \mid \mathcal{A}) \leq \cdots$ a.s. and $\lim_n E\,(Y_n \mid \mathcal{A}) = E(Y \mid \mathcal{A})$ a.s.
- Since $Y_n$ is bounded, $E[Y_n E(X \mid \mathcal{A})] = E[XE(Y_n \mid \mathcal{A})]$ by part (1)
- Since $X \geq 0$, $E(X \mid \mathcal{A}) \geq 0$. We have

$$\begin{aligned} E[YE(X \mid \mathcal{A})] &= \lim_n E[Y_n E(X \mid \mathcal{A})] \\ &= \lim_n E[XE(Y_n \mid \mathcal{A})] \\ &= E[XE(Y \mid \mathcal{A})], \end{aligned}$$

where the first and the third equations are because $X \geq 0$, $E(X \mid \mathcal{A}) \geq 0$, and the monotone convergence theorem.

(3) We now consider general $X, Y$.
Let $f_+$ and $f_-$ denote the positive and negative parts of a function $f$
Note that

$$E\{[XE(Y \mid \mathcal{A})]_+\} = E\{X_+[E(Y \mid \mathcal{A})]_+\} + E\{X_-[E(Y \mid \mathcal{A})]_-\}$$

and

$$E\{[XE(Y \mid \mathcal{A})]_-\} = E\{X_+[E(Y \mid \mathcal{A})]_-\} + E\{X_-[E(Y \mid \mathcal{A})]_+\}$$

Since $E[XE(Y \mid \mathcal{A})]$ exists, without loss of generality we assume that

$$E\{[XE(Y \mid \mathcal{A})]_+\} = E\{X_+[E(Y \mid \mathcal{A})]_+\} + E\{X_-[E(Y \mid \mathcal{A})]_-\} < \infty$$

Then, both

$$E[X_+E(Y \mid \mathcal{A})] = E\{X_+[E(Y \mid \mathcal{A})]_+\} - E\{X_+[E(Y \mid \mathcal{A})]_-\}$$

and

$$E[X_-E(Y \mid \mathcal{A})] = E\{X_-[E(Y \mid \mathcal{A})]_+\} - E\{X_-[E(Y \mid \mathcal{A})]_-\}$$

are well defined and their difference is also well defined.

Applying the result established in (2), we obtain that

$$\begin{aligned} E\left[X_+ E(Y \mid \mathcal{A})\right] &= E\left\{E\left(X_+ \mid \mathcal{A}\right)\left[E(Y \mid \mathcal{A})\right]_+\right\} - E\left\{E\left(X_+ \mid \mathcal{A}\right)\left[E(Y \mid \mathcal{A})\right]_-\right\} \\ &= E\left[E\left(X_+ \mid \mathcal{A}\right) E(Y \mid \mathcal{A})\right] \end{aligned}$$

where the last equality follows from the linearity of integrals. Similarly,

$$\begin{aligned} E\left[X_- E(Y \mid \mathcal{A})\right] &= E\left\{E\left(X_- \mid \mathcal{A}\right)\left[E(Y \mid \mathcal{A})\right]_+\right\} - E\left\{E\left(X_- \mid \mathcal{A}\right)\left[E(Y \mid \mathcal{A})\right]_-\right\} \\ &= E\left[E\left(X_- \mid \mathcal{A}\right) E(Y \mid \mathcal{A})\right] \end{aligned}$$

By the linearity of integrals,

$$\begin{aligned} E[XE(Y \mid \mathcal{A})] &= E\left[X_+ E(Y \mid \mathcal{A})\right] - E\left[X_- E(Y \mid \mathcal{A})\right] \\ &= E\left[E\left(X_+ \mid \mathcal{A}\right) E(Y \mid \mathcal{A})\right] - E\left[E\left(X_- \mid \mathcal{A}\right) E(Y \mid \mathcal{A})\right] \\ &= E[E(X \mid \mathcal{A})E(Y \mid \mathcal{A})]\} \end{aligned}$$

Switching $X$ and $Y$, we also conclude that

$$E[YE(X \mid \mathcal{A})] = E[E(X \mid \mathcal{A})E(Y \mid \mathcal{A})]$$

Hence, $E[XE(Y \mid \mathcal{A})] = E[YE(X \mid \mathcal{A})]$

# Problem 10

Which of the following parametrizations are identifiable? (Prove or disprove.)

a. $X_1, \ldots, X_p$ are independent with $X_i \sim \mathcal{N}\left(\alpha_i + \nu, \sigma^2\right)$

$$\theta = \left(\alpha_1, \alpha_2, \ldots, \alpha_p, \nu, \sigma^2\right) \tag{6}$$

and $P_\theta$ is the distribution of $\mathbf{X} = (X_1, \ldots, X_p)$

b. Same as (a) with $\alpha = (\alpha_1, \ldots, \alpha_p)$ restricted to

$$\left\{(a_1, \ldots, a_p) : \sum_{i=1}^{p} a_i = 0\right\} \tag{7}$$

c. $X$ and $Y$ are independent $\mathcal{N}\left(\mu_1, \sigma^2\right)$ and $\mathcal{N}\left(\mu_2, \sigma^2\right), \theta = (\mu_1, \mu_2)$ and we observe $Y - X$

d. $X_{ij}, i = 1, \ldots, p; j = 1, \ldots, b$ are independent with $X_{ij} \sim \mathcal{N}\left(\mu_{ij}, \sigma^2\right)$ where $\mu_{ij} = \nu + \alpha_i + \lambda_j, \theta = \left(\alpha_1, \ldots, \alpha_p, \lambda_1, \ldots, \lambda_b, \nu, \sigma^2\right)$ and $P_\theta$ is the distribution of $X_{11}, \ldots, X_{pb}$

e. Same as (d) with $(\alpha_1, \ldots, \alpha_p)$ and $(\lambda_1, \ldots, \lambda_b)$ restricted to the sets where $\sum_{i=1}^{p} \alpha_i = 0$ and $\sum_{j=1}^{b} \lambda_j = 0$

# Problem 10: Solution

a. No.
   - Fixed any $\theta = \left(\alpha_1, \alpha_2, \ldots, \alpha_p, \nu, \sigma^2\right)$.
   - For any $a \neq 0$, take $\tilde{\theta}_a = \left(\alpha_1 - a, \alpha_2 - a, \ldots, \alpha_p - a, \nu + a, \sigma^2\right)$.
   - Then $P_\theta$ and $P_{\tilde{\theta}_a}$ are the same.
b. Yes.
   - Suppose $P_\theta = P_{\tilde{\theta}}$, where $\theta = \left(\alpha_1, \alpha_2, \ldots, \alpha_p, \nu, \sigma^2\right)$ and $\tilde{\theta} = \left(\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_p, \tilde{\nu}, \tilde{\sigma}^2\right)$ both are in the set in the equation (7).
   - Since two normal distributions are the same if and only if they have the same means and the same variances, we have $\alpha_i + \nu = \tilde{\alpha}_i + \tilde{\nu}$ and $\sigma^2 = \tilde{\sigma}^2$.
   - Summing the first equation over all $i$, we have $p\nu = p\tilde{\nu}$.
   - Therefore $\nu = \tilde{\nu}$ and $\alpha_i = \tilde{\alpha}_i$ for all $i$.
c. No.
   - The population $P_{(\mu_1, \mu_2)}$ of $Y - X$ is $\mathcal{N}\left(\mu_1 - \mu_2, \sigma^2\right)$
   - But for any $a \neq 0$, $P_{(\mu_1, \mu_2)}$ and $P_{(\mu_1 + a, \mu_2 + a)}$ are the same
d. No. Because $\theta = \left(\alpha_1, \ldots, \alpha_p, \lambda_1, \ldots, \lambda_b, \nu, \sigma^2\right)$ and $\tilde{\theta} = \left(\alpha_1 + a, \ldots, \alpha_p + a, \lambda_1, \ldots, \lambda_b, \nu - a, \sigma^2\right)$ are the same.
e. Yes. Use the same argument as in (b).