ST5215 Advanced Statistical Theory, Lecture 6

HUANG Dongming

National University of Singapore

27 Aug 2020

Review

Last time

- Exponential families
- Statistics

Today

- Sufficiency
- Factorization theorem

Recap: Exponential families

Definition

A parametric family $\{P_\theta : \theta \in \Theta\}$ dominated by a σ-finite measure ν on $(\Omega, \mathcal{E})$ is called an *exponential family* iff

$$f_\theta(\omega) = \frac{\mathrm{d}P_\theta}{\mathrm{d}\nu}(\omega) = \exp\left\{[\eta(\theta)]^\top T(\omega) - \xi(\theta)\right\} h(\omega), \qquad \omega \in \Omega, \tag{1}$$

where T is a random p-vector, η is a function from Θ to $\mathcal{R}^p$, h is a nonnegative Borel function on $(\Omega, \mathcal{E})$, and

$$\xi(\theta) = \log\left\{\int_\Omega \exp\{[\eta(\theta)]^\top T(\omega)\} h(\omega)\, \mathrm{d}\nu(\omega)\right\}. \tag{2}$$

- T and h are functions of ω only
- ξ and η are functions of θ only

Recap: The canonical form

Reparametrize the family by $\eta = \eta(\theta)$, so that

$$f_\eta(\omega) = \exp\{\eta^\top T(\omega) - \zeta(\eta)\}h(\omega) \tag{3}$$

where $\zeta(\eta) = \log\left\{\int_\Omega \exp\{\eta^\top T(\omega)\}h(\omega)\ \mathrm{d}\nu(\omega)\right\}$.

- This is the *canonical form* for the family (still not unique)
- η is called the *natural parameter*
- The *natural parameter space*: $\Xi = \{\eta(\theta) : \theta \in \Theta\} \subset \mathcal{R}^p$
- An exponential family in its canonical form is called a *natural exponential family*
- *Full rank*: if Ξ contains an open set

Differential identities of natural exponential families

Let $\mathcal{P}$ be a natural exponential family with p.d.f.

$$f_\eta(x) = \exp\{\eta^\top T(x) - \zeta(\eta)\}h(x) \tag{4}$$

- Let Ξ_f be the set of values of η such that

$$\int |f(\omega)| \exp\left\{\eta^\tau T(\omega)\right\} h(\omega)\ \mathrm{d}\nu(\omega) < \infty.$$

- Define g on Ξ_f by

$$g(\eta) = \int f(\omega) \exp\left\{\eta^\tau T(\omega)\right\} h(\omega)\ \mathrm{d}\nu(\omega).$$

Then

- g is continuous and has continuous derivatives of all orders.
- These derivatives can be computed by differentiation under the integral sign.

This result is related to the dominated convergence theorem. See the reference on LumiNUS: "Files/Readings/Differential"

Statistics

A **statistic** $T(X)$ is a measurable function of sample X.

- $T(X)$ only depends on X
- T is a known function: $T(X)$ is a known value whenever X is known.
- Trivial statistics: X itself, any constant
- Some examples are:
  - sample mean: $\bar{X} = \frac{1}{n} \sum_i X_i$
  - sample variance: $S^2 = \frac{1}{n-1} \sum_i (X_i - \bar{x})^2$
  - order statistics, $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$
  - sample median: middle value of ordered statistics
  - sample minimum: $X_{(1)}$
  - sample maximum: $X_{(n)}$

Sufficient Statistics

Information reduction:

- $\sigma(T(X)) \subset \sigma(X)$ ("$=$" if and only if T is one-to-one)
- Usually $\sigma(T(X))$ simplifies $\sigma(X)$, i.e., a statistic provides a "reduction" of the σ-field

Does such a reduction from $\sigma(X)$ to $\sigma(T(X))$ results in any loss of information concerning the unknown P ?

- If $T(X)$ is fully as informative as X, then statistical analyses can be done using $T(X)$.

Definition

Let X be a sample from an unknown population $P \in \mathcal{P}$, where $\mathcal{P}$ is a family of populations. A statistic $T(X)$ is said to be *sufficient* for $P \in \mathcal{P}$ iff the conditional distribution of X given T is known (does not depend on P).

Sufficient Statistics (Cont.)

- If $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, then we also say $T(X)$ is *sufficient* for θ iff the conditional distribution of X given T does not depend on θ
- Sufficiency depends on the given family $\mathcal{P}$:
  If
  $$\mathcal{P}_0 \subset \mathcal{P} \subset \mathcal{P}_1, \tag{5}$$
  and T is sufficient for $P \in \mathcal{P}$, then
  - T is also sufficient for $P \in \mathcal{P}_0$
  - T is not necessarily sufficient for $P \in \mathcal{P}_1$
    - Let $\mathcal{P} = \{N(\mu, \sigma^2) : \mu \in \mathcal{R}, \sigma = 1\}$. Then the sample mean $\bar{X}$ is sufficient for $P \in \mathcal{P}$.
    - Let $\mathcal{P}_1 = \{N(\mu, \sigma^2) : \mu \in \mathcal{R}, \sigma > 0\}$. Then $\mathcal{P} \subset \mathcal{P}_1$ but $\bar{X}$ is not sufficient for $P \in \mathcal{P}_1$

Generating fake data

- Suppose T is sufficient for $\{P_\theta : \theta \in \Theta\}$
- Let $Q_t(B) = P_\theta(X \in B \mid T = t)$
- Then $P_\theta(X \in B \mid T) = Q_T(B)$
- By tower property (a.k.a., *tower law, law of total expectation, smoothing property*)

$$P_\theta(X \in B) = E_\theta[P_\theta(X \in B \mid T)] = E_\theta Q_T(B)$$

Now we use a random number generator to construct "fake" data $\tilde{X}$ s. t. $\tilde{X} \sim Q_t(\cdot)$ if the observed T equals to t, i.e.,

$$\tilde{X} \mid T = t \sim Q_t$$

- By tower property,

$$P_\theta(\tilde{X} \in B) = E_\theta[P_\theta(\tilde{X} \in B \mid T)] = E_\theta Q_T(B) = P_\theta(X \in B)$$

This shows that $\tilde{X} \sim P_\theta$, so the distributions for $\tilde{X}$ are the same as distributions for real X regardless of the value of θ

Example: Sum of Bernoulli trials

Let $X = (X_1, \ldots, X_n)$ and $X_1, \ldots, X_n$ be i.i.d. from the Bernoulli distribution with p.d.f. (w.r.t. the counting measure)

$$f_\theta(z) = \theta^z (1-\theta)^{1-z} I_{\{0,1\}}(z), \qquad z \in \mathcal{R}, \qquad \theta \in (0,1). \tag{6}$$

- $\mathcal{P} = \{\prod_{i=1}^n f_\theta(x_i) : \theta \in (0,1)\}$
- Take $T(X) = \sum_{i=1}^n X_i$: the number of 1's in X
- We will show $T(X)$ is sufficient for θ
- Once we know $T(X)$, other information in X is about the positions of these 1's
    - but such information is not useful for estimating θ the probability of getting a 1; *redundant for* θ

- Compute the marginal p.d.f. of T

$$P(T = t) = \binom{n}{t}\theta^t(1-\theta)^{n-t} I_{\{0,1,\dots,n\}}(t) \tag{7}$$

and the conditional distribution of X given T

$$P(X = x \mid T = t) = \frac{P(X = x, T = t)}{P(T = t)}, \tag{8}$$

where $x = (x_1, \dots, x_n) \in \{0,1\}^n$ and $t \in \{0, 1, \dots, n\}$.
  - If $t \neq \sum_{i=1}^n x_i$, then $P(X = x, T = t) = 0$.
  - If $t = \sum_{i=1}^n x_i$, then

$$\begin{aligned} P(X = x, T = t) &= P(X = x) \\ &= \prod_{i=1}^n P(X_i = x_i) \\ &= \theta^t(1-\theta)^{n-t} \prod_{i=1}^n I_{\{0,1\}}(x_i) \end{aligned}$$

Then

$$P(X = x \mid T = t) = \frac{1}{\binom{n}{t}} \frac{\prod_{i=1}^n I_{\{0,1\}}(x_i)}{I_{\{0,1,\dots,n\}}(t)} \tag{9}$$

How to find a sufficient statistic?

- Finding a sufficient statistic by means of the definition is not convenient.
  - It involves guessing a statistic T that might be sufficient and computing the conditional distribution of X given T.
- For families of populations having p.d.f.'s, a simple way of finding sufficient statistics is to use the *factorization theorem*.

Factorization Theorem

Theorem

Suppose that X is a sample from $P \in \mathcal{P}$ and $\mathcal{P}$ is a family of probability measures on $(\mathcal{R}^n, \mathcal{B}^n)$ dominated by a σ-finite measure ν.Then $T(X)$ is sufficient for $P \in \mathcal{P}$ **if and only if** *there are nonnegative Borel functions*

- *$h(x)$ (which does not depend on P) on $(\mathcal{R}^n, \mathcal{B}^n)$, and*
- *$g_P(t)$ (which depends on P) on the range of T*

such that

$$\frac{\mathrm{d}P}{\mathrm{d}\nu}(x) = g_P(T(x))h(x). \tag{10}$$

- Intuition: h is known and the unknown part g_P involve T only
- Application: the T statstic in an exponential family

$$f_\theta(\omega) = \frac{\mathrm{d}P_\theta}{\mathrm{d}\nu}(\omega) = \exp\left\{[\eta(\theta)]^\top T(\omega) - \xi(\theta)\right\} h(\omega), \omega \in \Omega, \tag{11}$$

is sufficient for θ

Example: Varying supports

Suppose $X_1, \ldots, X_n$ are i.i.d. r.v.s. from the uniform distribution on $(\theta, \theta+1)$. The common marginal density is

$$f_\theta(x) = \begin{cases} 1, & x \in (\theta, \theta+1) \\ 0, & \text{otherwise} \end{cases}$$

- Write the density using indicator functions: $f_\theta = 1_{(\theta,\theta+1)}$
- The joint density is

$$p_\theta(x) = \prod_{i=1}^{n} 1_{(\theta,\theta+1)}(x_i)$$

- Note that the density equals one if and only if $x_{(n)} = \max_i\{x_i\} < \theta + 1$ and $x_{(1)} = \min_i\{x_i\} > \theta$, so

$$p_\theta(x) = 1_{(\theta,\infty)}\left(x_{(1)}\right) 1_{(-\infty,\theta+1)}\left(x_{(n)}\right)$$

- By the factorization theorem, $T = \left(X_{(1)}, X_{(n)}\right)$ is sufficient.

Example: Truncation families

- Suppose $\phi(x)$ be a positive Borel function on $(\mathcal{R}, \mathcal{B})$ such that $\int_a^b \phi(x)\,\mathrm{d}x < \infty$ for all pairs of a and b that $a < b$.
- Let θ be the vector (a, b), $\Theta = \{(a, b) \in \mathcal{R}^2 : a < b\}$, and

$$f_\theta(x) = c(\theta)\phi(x) I_{(a,b)}(x), \tag{12}$$

  where $c(\theta) = [\int_a^b \phi(x)\,\mathrm{d}x]^{-1}$.
- $\{f_\theta : \theta \in \Theta\}$ is called a truncation family.
- It is parametric and is dominated by Lebesgue measure

Suppose $X_1, \ldots, X_n$ are i.i.d. sampled from f_θ

- The joint p.d.f. of $X = (X_1, \ldots, X_n)$ is

$$\prod_{i=1}^n f_\theta(x_i) = [c(\theta)]^n \left[\prod_{i=1}^n I_{(a,b)}(x_i)\right] \left[\prod_{i=1}^n \phi(x_i)\right] \tag{13}$$

- $\prod_{i=1}^n I_{(a,b)}(x_i) = I_{(a,\infty)}(x_{(1)}) I_{(-\infty,b)}(x_{(n)})$.
- So $T(X) = (X_{(1)}, X_{(n)})$ is sufficient for $\theta = (a, b)$.

Exercise

Suppose $X_1, \dots, X_n$ are i.i.d. r.v.s with common marginal Lebesgue density

$$f_\theta(x) = \begin{cases} (\theta+1)x^\theta, & x \in (0,1) \\ 0, & \text{otherwise} \end{cases}$$

for some unknown $\theta \in (-1, \infty)$.
Find a sufficient statistic for this model.
Solution:

- The joint density p_θ is

$$p_\theta(x) = \prod_{i=1}^n f_\theta(x_i) = \prod_{i=1}^n (\theta+1)x_i^\theta = (\theta+1)^n \left(\prod_{i=1}^n x_i\right)^\theta, \quad x \in (0,1)^n$$

  with $p_\theta(x) = 0$ if $x \notin (0,1)^n$.
- Taking $g_\theta(t) = (\theta+1)^n t^\theta$ and $h = 1_{(0,1)^n}$, from the factorization theorem, $T = \prod_{i=1}^n X_i$ is sufficient.

Proof of Factorization Theorem

We require two lemmas.

Lemma (S, Exercise 1.35)

Let $\{c_i\}$ be a sequence of positive numbers satisfying $\sum_{i=1}^{\infty} c_i = 1$ and let $\{P_i\}$ be a sequence of probability measures on a common measurable space. Define $Q = \sum_{i=1}^{\infty} c_i P_i$.

1. *Q is a probability measure;*
2. *Let ν be a σ-finite measure. Then $P_i \ll \nu$ for all i if and only if $Q \ll \nu$. When $Q \ll \nu$,*

$$\frac{\mathrm{d}Q}{\mathrm{d}\nu} = \sum_{i=1}^{\infty} c_i \frac{\mathrm{d}P_i}{\mathrm{d}\nu}. \tag{14}$$

Lemma (2.1)

If a family $\mathcal{P}$ is dominated by a σ-finite measure, then $\mathcal{P}$ is dominated by a probability measure $Q = \sum_{i=1}^{\infty} c_i P_i$ where $P_i \in \mathcal{P}$ and $c_i \geq 0$.

Factorization Theorem

Theorem

Suppose that X is a sample from $P \in \mathcal{P}$ and $\mathcal{P}$ is a family of probability measures on $(\mathcal{R}^n, \mathcal{B}^n)$ dominated by a σ-finite measure ν. Then $T(X)$ is sufficient for $P \in \mathcal{P}$ **if and only if** *there are nonnegative Borel functions*

- *$h(x)$ (which does not depend on P) on $(\mathcal{R}^n, \mathcal{B}^n)$, and*
- *$g_P(t)$ (which depends on P) on the range of T*

such that

$$\frac{\mathrm{d}P}{\mathrm{d}\nu}(x) = g_P(T(x))h(x). \tag{15}$$

"$\Rightarrow$":

Assume T is sufficient, and want to show $\frac{\mathrm{d}P}{\mathrm{d}\nu}(x) = g_P(T(x))h(x)$

- Note that $P(A) = \int P(A \mid T)\,\mathrm{d}P$
- Let Q be the probability measure in Lemma 2.1
- Suppose we can show $P(A \mid T) = E_Q(I_A \mid T)$, Q-a.s.
  Let $\frac{\mathrm{d}P}{\mathrm{d}Q}(x)$ be the Radon-Nikodym derivative of P with respect to Q on the space $(\mathcal{R}^n, \sigma(T), Q)$, denoted by $g_P(T(x))$, we have

$$
\begin{aligned}
P(A) = \int P(A \mid T)\,\mathrm{d}P \qquad &= \int E_Q(I_A \mid T) g_P(T(x))\,\mathrm{d}Q \\
&= \int E_Q(I_A\, g_P(T(x)) \mid T)\,\mathrm{d}Q \\
&= \int I_A\, g_P(T(x))\,\mathrm{d}Q \\
&= \int_A g_P(T(x)) \frac{\mathrm{d}Q}{\mathrm{d}\nu}(x)\,\mathrm{d}\nu(x),
\end{aligned}
$$

for all $A \in \mathcal{B}^n$. Then equation (9) holds with $h(x) = \frac{\mathrm{d}Q}{\mathrm{d}\nu}(x)$.

Showing $P(A \mid T) = E_Q(I_A \mid T)$

For any $B \in \sigma(T)$, by Lemma S,

$$\begin{aligned} Q(A \cap B) &= \sum_{j=1}^{\infty} c_j P_j(A \cap B) \\ &= \sum_{j=1}^{\infty} c_j \int_B P_j(A \mid T)\, \mathrm{d}P_j \\ (\because P(A \mid T) \text{ doesn't depend on } P \in \mathcal{P}\ ) &= \sum_{j=1}^{\infty} c_j \int_B P(A \mid T)\, \mathrm{d}P_j \\ (\because \text{Fubini's theorem}\ ) &= \int_B \sum_{j=1}^{\infty} c_j P(A \mid T)\, \mathrm{d}P_j \\ (\because \sum_j c_j = 1\ \ ) &= \int_B P(A \mid T)\, \mathrm{d}Q, \end{aligned}$$

Hence, $P(A \mid T) = E_Q(I_A \mid T)$ Q-a.s.

Factorization Theorem

Theorem

Suppose that X is a sample from $P \in \mathcal{P}$ and $\mathcal{P}$ is a family of probability measures on $(\mathcal{R}^n, \mathcal{B}^n)$ dominated by a σ-finite measure ν. Then $T(X)$ is sufficient for $P \in \mathcal{P}$ **if and only if** *there are nonnegative Borel functions*

- *$h(x)$ (which does not depend on P) on $(\mathcal{R}^n, \mathcal{B}^n)$, and*
- *$g_P(t)$ (which depends on P) on the range of T*

such that

$$\frac{\mathrm{d}P}{\mathrm{d}\nu}(x) = g_P(T(x))h(x). \tag{16}$$

"$\Longleftarrow$":

Suppose that Equation (9) holds.

- Let Q be the probability measure in Lemma 2.1
- Want to show for all $A \in \sigma(X)$ and $P \in \mathcal{P}$,

$$P(A \mid T) = E_Q(I_A \mid T) \quad P\text{-a.s.} \ , \tag{17}$$

  because $E_Q(I_A \mid T)$ does not vary with $P \in \mathcal{P}$.

- By Chain rule, $\frac{\mathrm{d}P}{d\nu} = \frac{\mathrm{d}P}{dQ}\frac{dQ}{d\nu}$ ν-a.e.
- Hence

$$\frac{\mathrm{d}P}{dQ} = \frac{\mathrm{d}P}{d\nu} \Big/ \frac{dQ}{d\nu} = \frac{\mathrm{d}P}{d\nu} \Big/ \sum_{i=1}^{\infty} c_i \frac{dP_i}{d\nu} = g_P(T) \Big/ \sum_{i=1}^{\infty} c_i g_{P_i}(T), \tag{18}$$

  Q-a.s., where the second equality follows from Lemma S.

- So $\mathrm{d}P/dQ$ is a Borel function of T.

Showing $P(A \mid T) = E_Q(I_A \mid T)$

For any $B \in \sigma(T)$, we calculate $P(A \cap B)$ as follows

$$\begin{aligned}
\int_B E_Q(I_A \mid T)\,\mathrm{d}P &= \int_B E_Q(I_A \mid T)\frac{\mathrm{d}P}{dQ}dQ \\
(\because\ \mathrm{d}P/dQ \text{ is a Borel function of } T\ ) &= \int_B E_Q\left(I_A\frac{\mathrm{d}P}{dQ} \mid T\right)dQ \\
(\because \text{Def. of conditional expectation }) &= \int_B I_A\frac{\mathrm{d}P}{dQ}dQ \\
&= \int_B I_A\,\mathrm{d}P \\
(\because \text{Def. of conditional expectation }) &= \int_B P(A \mid T)\,\mathrm{d}P
\end{aligned}$$

So $P(A \mid T) = E_Q(I_A \mid T)$, P-a.s.

Tutorial

1. (Stein's identity). Suppose the distribution of X has density in an exponential family whose support is $(-\infty, \infty)$. If g is any differentiable function such that $E\left|g'(X)\right| < \infty$, then
$$E\left\{\left[\frac{h'(X)}{h(X)} + \sum_{i=1}^{p} \eta_i T_i'(X)\right] g(X)\right\} = -Eg'(X),$$
where η_i's are the coordinates of $\eta(\theta)$ and T_i's are the coordinates of $T(X)$. When $p = 1$ and $h'(X) = 0$ (e.g., the normal family with fixed σ), the identity becomes $E\left\{(X-\mu)g(X)\right\} = \sigma^2 Eg'(X)$,
2. Show that if two r.v.s X and Y are independent, then their characteristic functions ϕ_X and ϕ_Y satisfy $\phi_X(t)\phi_Y(t) = \phi_{X+Y}(t)$ for all $t \in \mathcal{R}$.
3. Find an example of two r.v.s X and Y such that X and Y are not independent but their characteristic functions ϕ_X and ϕ_Y satisfy $\phi_X(t)\phi_Y(t) = \phi_{X+Y}(t)$ for all $t \in \mathcal{R}$
4. Let X be an integrable random variable on the probability space $(\Omega, \mathcal{F}, P)$, $\mathcal{A}$ and $\mathcal{A}_0$ be σ-fields satisfying $\mathcal{A}_0 \subset \mathcal{A} \subset \mathcal{F}$. Show that $E\left[E(X \mid \mathcal{A}) \mid \mathcal{A}_0\right] = E\left(X \mid \mathcal{A}_0\right) = E\left[E\left(X \mid \mathcal{A}_0\right) \mid \mathcal{A}\right]$ a.s.
5. Let X be an integrable random variable on the probability space $(\Omega, \mathcal{F}, P), \mathcal{A}$ be a sub-σ-field of $\mathcal{F}$, and Y be another random variable satisfying $\sigma(Y) \subset \mathcal{A}$ and $E|XY| < \infty$. Show that $E(XY \mid \mathcal{A}) = YE(X \mid \mathcal{A})$ a.s.

Ex 1

(Stein's identity). Suppose the distribution of X has density in an exponential family whose support is $(-\infty, \infty)$. If g is any differentiable function such that $E\,|g'(X)| < \infty$, then

$$E\left\{\left[\frac{h'(X)}{h(X)} + \sum_{i=1}^{p} \eta_i T_i'(X)\right] g(X)\right\} = -Eg'(X),$$

where η_i's are the coordinates of $\eta(\theta)$ and T_i's are the coordinates of $T(X)$.
When $p = 1$ and $h'(X) = 0$ (e.g., the normal family with fixed σ), the identity becomes $E\left\{(X-\mu)g(X)\right\} = \sigma^2 Eg'(X)$,

Remark. This exercise is taken from Lemma 5.15 of Chapter 1 in Theory of Point Estimation, 2nd Ed.
The idea behind the proof is simple: integration by parts. However, I think a rigorous proof needs extra conditions on the density $f(x)$ and $g(x)$ such as $\lim_{x\to\pm\infty} f(x)g(x) = 0$.
We leave the proof of the general case for next tutorial and only focus on the normal family.

Ex 1: normal case

(Stein's lemma). Suppose the distribution of X has density in a univariate normal density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-(x-\mu)^2/2\sigma^2\right),$$

and g is a differentiable function such that $E\left|g'(X)\right| < \infty$ and $\lim_{x\to\pm\infty} f(x)g(x) = 0$, then $E\left\{(X-\mu)g(X)\right\} = \sigma^2 Eg'(X)$,
Proof:

- $f'(x) = -\frac{x-\mu}{\sigma^2} f(x)$
- For any $a, b \in \mathcal{R}$, using integration by parts, we have

$$\int_a^b f'(x)g(x)\,\mathrm{d}x = g(b)f(b) - g(a)f(a) - \int_a^b f(x)g'(x)\,\mathrm{d}x \quad (19)$$

- Note that here we have use the smoothness of f and the assumption that $E\left|g'(X)\right| < \infty$, which implies $g'(x)$ is integrable on any close interval
- Then we take $a \to -\infty$ and $b \to \infty$.

Ex 2

Show that if two r.v.s X and Y are independent, then their characteristic functions ϕ_X and ϕ_Y satisfy $\phi_X(t)\phi_Y(t) = \phi_{X+Y}(t)$ for all $t \in \mathcal{R}$.
Proof: Use the independence between $e^{\sqrt{-1}t^\top X}$ and $e^{\sqrt{-1}t^\top X}$.

Ex 3

Find an example of two r.v.s X and Y such that X and Y are not independent but their characteristic functions ϕ_X and ϕ_Y satisfy $\phi_X(t)\phi_Y(t) = \phi_{X+Y}(t)$ for all $t \in \mathcal{R}$
Proof:

- Let $X = Y$ be a random variable having the Cauchy distribution with $\phi_X(t) = \phi_Y(t) = e^{-|t|}$.
- X and Y are not independent.
- Using the result of Exercise 5 in Tutorial Aug 20, the characteristic function of $X + Y = 2X$ is

$$\begin{aligned}\phi_{X+Y}(t) &= E\left(e^{\sqrt{-1}t(2X)}\right) = \phi_X(2t) = e^{-|2t|} \\ &= e^{-|t|}e^{-|t|} = \phi_X(t)\phi_Y(t)\end{aligned}$$

Ex 4

Let X be an integrable random variable on the probability space $(\Omega, \mathcal{F}, P)$, $\mathcal{A}$ and $\mathcal{A}_0$ be σ-fields satisfying $\mathcal{A}_0 \subset \mathcal{A} \subset \mathcal{F}$. Show that $E\left[E(X \mid \mathcal{A}) \mid \mathcal{A}_0\right] = E\left(X \mid \mathcal{A}_0\right) = E\left[E\left(X \mid \mathcal{A}_0\right) \mid \mathcal{A}\right]$ a.s.

Proof:

1. Since $E\left(X \mid \mathcal{A}_0\right)$ is measurable from $(\Omega, \mathcal{A}_0)$ to $(\mathcal{R}, \mathcal{B})$ and $\mathcal{A}_0 \subset \mathcal{A}$, $E\left(X \mid \mathcal{A}_0\right)$ is measurable from $(\Omega, \mathcal{A})$ to $(\mathcal{R}, \mathcal{B})$ and, thus, $E\left(X \mid \mathcal{A}_0\right) = E\left[E\left(X \mid \mathcal{A}_0\right) \mid \mathcal{A}\right]$ a.s.
2. For any $A \in \mathcal{A}_0 \subset \mathcal{A}$

$$\int_A E\left[E(X \mid \mathcal{A}) \mid \mathcal{A}_0\right] \, \mathrm{d}P = \int_A E(X \mid \mathcal{A}) \, \mathrm{d}P = \int_A X \, \mathrm{d}P,$$

where the first equality is because $E\left[E(X \mid \mathcal{A}) \mid \mathcal{A}_0\right]$ is measurable from $(\Omega, \mathcal{A}_0)$ to $(\mathcal{R}, \mathcal{B})$. We conclude that $E\left[E(X \mid \mathcal{A}) \mid \mathcal{A}_0\right] = E\left(X \mid \mathcal{A}_0\right)$ a.s.

Ex 5

Let X be an integrable random variable on the probability space $(\Omega, \mathcal{F}, P), \mathcal{A}$ be a sub-σ-field of $\mathcal{F}$, and Y be another random variable satisfying $\sigma(Y) \subset \mathcal{A}$ and $E|XY| < \infty$. Show that
$E(XY \mid \mathcal{A}) = YE(X \mid \mathcal{A})$ a.s.
Proof: Note that both sides are measurable w.r.t. $(\Omega, \mathcal{A})$. So we only need to show that for any $A \in \mathcal{A}$,

$$\int_A YE(X \mid \mathcal{A}) \, \mathrm{d}P = \int_A XY \, \mathrm{d}P$$

We use the *canonical method:* first assume Y is an indicator function, then a simple function, then a nonnegative function, and finally a general function.

Ex 5: Indicator

Suppose $Y = aI_B$, where $a \in \mathcal{R}$ and $B \in \mathcal{A}$.
Then $A \cap B \in \mathcal{A}$ and

$$\begin{aligned}\int_A XY \,\mathrm{d}P &= a \int_A XI_B \,\mathrm{d}P \\ &= a \int_{A\cap B} X \,\mathrm{d}P \\ &= a \int_{A\cap B} E(X \mid \mathcal{A}) \,\mathrm{d}P \\ &= a \int_A I_B E(X \mid \mathcal{A}) \,\mathrm{d}P \\ &= \int_A YE(X \mid \mathcal{A}) \,\mathrm{d}P\end{aligned}$$

Ex 5: Simple function

Suppose $Y = \sum_{i=1}^{k} a_i I_{B_i}$, where $B_i \in \mathcal{A}$.
Then

$$\begin{aligned}
\int_A YX \, \mathrm{d}P &= \sum_{i=1}^{k} \int_A a_i I_{B_i} X \, \mathrm{d}P \\
&= \sum_{i=1}^{k} \int_A a_i I_{B_i} E(X \mid \mathcal{A}) \, \mathrm{d}P \\
&= \int_A Y E(X \mid \mathcal{A}) \, \mathrm{d}P
\end{aligned}$$

Ex 5: Nonnegative function

Suppose that $X \geq 0$ and $Y \geq 0$.

- Since Y is $\mathcal{A}$-measurable, there exists a sequence of increasing simple functions Y_n such that $\sigma(Y_n) \subset \mathcal{A}$, $Y_n \leq Y$ and $\lim_n Y_n = Y$
- $XY_n \uparrow XY$
- $Y_n E(X \mid \mathcal{A}) \uparrow YE(X \mid \mathcal{A})$

$$\begin{aligned}
\int_A YX \,\mathrm{d}P &= \lim_n \int_A Y_n X \,\mathrm{d}P \\
&= \lim_n \int_A Y_n E(X \mid \mathcal{A}) \,\mathrm{d}P \\
&= \int_A YE(X \mid \mathcal{A}) \,\mathrm{d}P,
\end{aligned}$$

where the 1st and 3rd equalities are due to the monotone convergence theorem.

Ex 5: General function

For general X and Y, consider X_+, X_-, Y_+, and Y_-.

- Since $\sigma(Y) \subset \mathcal{A}$, so are $\sigma(Y_+)$ and $\sigma(Y_-)$
- Since $E|XY| < \infty$, both $E(X_+Y_+)$ and $E(X_+Y_-)$ are finite

$$\begin{aligned}\int_A X_+Y \,\mathrm{d}P &= \int_A X_+Y_+ \,\mathrm{d}P - \int_A X_+Y_- \,\mathrm{d}P \\ &= \int_A Y_+E(X_+ \mid \mathcal{A}) \,\mathrm{d}P - \int_A Y_-E(X_+ \mid \mathcal{A}) \,\mathrm{d}P \\ &= \int_A YE(X_+ \mid \mathcal{A}) \,\mathrm{d}P\end{aligned}$$

- Similarly, $\int_A X_-Y \,\mathrm{d}P = \int_A YE(X_- \mid \mathcal{A}) \,\mathrm{d}P$

$$\begin{aligned}\int_A XY \,\mathrm{d}P &= \int_A X_+Y \,\mathrm{d}P - \int_A X_-Y \,\mathrm{d}P \\ &= \int_A YE(X_+ \mid \mathcal{A}) \,\mathrm{d}P - \int_A YE(X_- \mid \mathcal{A}) \,\mathrm{d}P \\ &= \int_A YE(X \mid \mathcal{A}) \,\mathrm{d}P\end{aligned}$$