

# Stat 709: Mathematical Statistics

## Lecture 7

Jun Shao

Department of Statistics  
University of Wisconsin  
Madison, WI 53706, USA

## Definitions

- If  $EX^k$  is finite, where  $k$  is a positive integer,  $EX^k$  is called the  $k$ th *moment* of  $X$  or  $P_X$ .
- If  $E|X|^a < \infty$  for some real number  $a$ ,  $E|X|^a$  is called the  $a$ th *absolute moment* of  $X$  or  $P_X$ .
- If  $\mu = EX$  and  $E(X - \mu)^k$  are finite for a positive integer  $k$ ,  $E(X - \mu)^k$  is called the  $k$ th *central moment* of  $X$  or  $P_X$ .
- $E(X - EX)^2$  is called the *variance* of  $X$  or  $P_X$ .
- For random vector  $X = (X_1, \dots, X_k)$ ,  $EX = (EX_1, \dots, EX_k)$
- For random matrix  $M = (M_{ij})$ ,  $EM = (EM_{ij})$
- For random vector  $X$ ,  $\text{Var}(X) = E(X - EX)(X - EX)^T$  is called its *covariance matrix*.  
The  $(i, j)$ th element of  $\text{Var}(X)$ ,  $i \neq j$ , is  $E(X_i - EX_i)(X_j - EX_j)$ , which is called the *covariance* of  $X_i$  and  $X_j$  and is denoted by  $\text{Cov}(X_i, X_j)$ .
- If  $\text{Cov}(X_i, X_j) = 0$ , then  $X_i$  and  $X_j$  are said to be uncorrelated.

## Basic properties

- $[\text{Cov}(X_i, X_j)]^2 \leq \text{Var}(X_i)\text{Var}(X_j), \quad i \neq j$
- For random vector  $X$ ,  $\text{Var}(X)$  is nonnegative definite
- Independence implies uncorrelation, not converse
- If  $Y = c^\tau X$ ,  $c \in \mathcal{R}^k$ , and  $X$  is a random  $k$ -vector,  $EY = c^\tau EX$  and  $\text{Var}(Y) = c^\tau \text{Var}(X)c$ .

## Three useful inequalities

- Cauchy-Schwartz inequality:  $[E(XY)]^2 \leq EX^2 EY^2$  for random variables  $X$  and  $Y$
- Jensen's inequality:  $f(EX) \leq Ef(X)$  for a random vector  $X$  and convex function  $f$  ( $f'' \geq 0$ )
- Chebyshev's inequality: Let  $X$  be a random variable and  $\varphi$  a nonnegative and nondecreasing function on  $[0, \infty)$ ,  $\varphi(-t) = \varphi(t)$ . Then, for each constant  $t \geq 0$ ,

$$\varphi(t)P(|X| \geq t) \leq \int_{\{|X| \geq t\}} \varphi(X) dP \leq E\varphi(X)$$

## Basic properties

- $[\text{Cov}(X_i, X_j)]^2 \leq \text{Var}(X_i)\text{Var}(X_j), \quad i \neq j$
- For random vector  $X$ ,  $\text{Var}(X)$  is nonnegative definite
- Independence implies uncorrelation, not converse
- If  $Y = c^\tau X$ ,  $c \in \mathcal{R}^k$ , and  $X$  is a random  $k$ -vector,  $EY = c^\tau EX$  and  $\text{Var}(Y) = c^\tau \text{Var}(X)c$ .

## Three useful inequalities

- Cauchy-Schwartz inequality:  $[E(XY)]^2 \leq EX^2EY^2$  for random variables  $X$  and  $Y$
- Jensen's inequality:  $f(EX) \leq Ef(X)$  for a random vector  $X$  and convex function  $f$  ( $f'' \geq 0$ )
- Chebyshev's inequality: Let  $X$  be a random variable and  $\varphi$  a nonnegative and nondecreasing function on  $[0, \infty)$ ,  $\varphi(-t) = \varphi(t)$ . Then, for each constant  $t \geq 0$ ,

$$\varphi(t)P(|X| \geq t) \leq \int_{\{|X| \geq t\}} \varphi(X)dP \leq E\varphi(X)$$

## Example 1.18.

If  $X$  is a nonconstant positive random variable with finite mean, then

$$(EX)^{-1} < E(X^{-1}) \quad \text{and} \quad E(\log X) < \log(EX),$$

since  $t^{-1}$  and  $-\log t$  are convex functions on  $(0, \infty)$ .

Let  $f$  and  $g$  be positive integrable functions on a measure space with a  $\sigma$ -finite measure  $\nu$ .

If  $\int f d\nu \geq \int g d\nu > 0$ , we want to show that

$$\int f \log \left( \frac{f}{g} \right) d\nu \geq 0.$$

Let  $h = f / \int f d\nu$ , a p.d.f. w.r.t.  $\nu$ .

Let  $Y = g/f$  be a random variable with  $h$  as its p.d.f.

By Jensen's inequality,  $E \log(g/f) \leq \log(E(g/f))$ .

The result follows from

$$\begin{aligned} \int \log \left( \frac{g}{f} \right) f d\nu / \int f d\nu &= \int \log \left( \frac{g}{f} \right) h d\nu = E \log(g/f) \\ &\leq \log(E(g/f)) = \log \left( \int \frac{g}{f} h d\nu \right) = \log \left( \frac{\int g d\nu}{\int f d\nu} \right) \leq 0 \end{aligned}$$

## Definition 1.5

Let  $X$  be a random  $k$ -vector.

- (i) The *moment generating function* (m.g.f.) of  $X$  or  $P_X$  is defined as

$$\psi_X(t) = Ee^{t^\tau X}, \quad t \in \mathcal{R}^k.$$

- (ii) The *characteristic function* (ch.f.) of  $X$  or  $P_X$  is defined as

$$\phi_X(t) = Ee^{\sqrt{-1}t^\tau X} = E[\cos(t^\tau X)] + \sqrt{-1} E[\sin(t^\tau X)], \quad t \in \mathcal{R}^k$$

## Remarks

- If the m.g.f. is finite in a neighborhood of  $0 \in \mathcal{R}^k$ , then
  - moments of  $X$  of any order are finite,
  - $\phi_X(t)$  can be obtained by replacing  $t$  in  $\psi_X(t)$  by  $\sqrt{-1}t$
- If  $0 < \psi_X(t) < \infty$ , then  $\kappa_X(t) = \log \psi_X(t)$  is called the *cumulant generating function* of  $X$  or  $P_X$ .

## Definition 1.5

Let  $X$  be a random  $k$ -vector.

- (i) The *moment generating function* (m.g.f.) of  $X$  or  $P_X$  is defined as

$$\psi_X(t) = Ee^{t^\tau X}, \quad t \in \mathcal{R}^k.$$

- (ii) The *characteristic function* (ch.f.) of  $X$  or  $P_X$  is defined as

$$\phi_X(t) = Ee^{\sqrt{-1}t^\tau X} = E[\cos(t^\tau X)] + \sqrt{-1} E[\sin(t^\tau X)], \quad t \in \mathcal{R}^k$$

## Remarks

- If the m.g.f. is finite in a neighborhood of  $0 \in \mathcal{R}^k$ , then
  - moments of  $X$  of any order are finite,
  - $\phi_X(t)$  can be obtained by replacing  $t$  in  $\psi_X(t)$  by  $\sqrt{-1}t$
- If  $0 < \psi_X(t) < \infty$ , then  $\kappa_X(t) = \log \psi_X(t)$  is called the *cumulant generating function* of  $X$  or  $P_X$ .

## Properties of m.g.f. and ch.f.

- If  $Y = A^T X + c$ , where  $A$  is a  $k \times m$  matrix and  $c \in \mathcal{R}^m$ , it follows from Definition 1.5 that

$$\psi_Y(u) = e^{c^T u} \psi_X(Au) \quad \text{and} \quad \phi_Y(u) = e^{\sqrt{-1}c^T u} \phi_X(Au), \quad u \in \mathcal{R}^m$$

- For independent  $X_1, \dots, X_k$ ,

$$\psi_{\sum_i X_i}(t) = \prod_i \psi_{X_i}(t) \quad \text{and} \quad \phi_{\sum_i X_i}(t) = \prod_i \phi_{X_i}(t), \quad t \in \mathcal{R}^k$$

- For  $X = (X_1, \dots, X_k)$  with m.g.f.  $\psi_X$  finite in a neighborhood of 0

$$\psi_X(t) = \sum_{(r_1, \dots, r_k)} \frac{\mu_{r_1, \dots, r_k} t_1^{r_1} \cdots t_k^{r_k}}{r_1! \cdots r_k!} \quad \mu_{r_1, \dots, r_k} = E(X_1^{r_1} \cdots X_k^{r_k})$$

- Consequently,

$$E(X_1^{r_1} \cdots X_k^{r_k}) = \left. \frac{\partial^{r_1 + \cdots + r_k} \psi_X(t)}{\partial t_1^{r_1} \cdots \partial t_k^{r_k}} \right|_{t=0}$$

$$\left. \frac{\partial \psi_X(t)}{\partial t} \right|_{t=0} = EX, \quad \left. \frac{\partial^2 \psi_X(t)}{\partial t \partial t^T} \right|_{t=0} = E(XX^T)$$

## Properties of m.g.f. and ch.f.

- If  $\psi_X$  is not finite and  $E|X_1^{r_1} \cdots X_k^{r_k}| < \infty$  for some nonnegative integers  $r_1, \dots, r_k$ , then

$$\left. \frac{\partial^{r_1 + \cdots + r_k} \phi_X(t)}{\partial t_1^{r_1} \cdots \partial t_k^{r_k}} \right|_{t=0} = (-1)^{(r_1 + \cdots + r_k)/2} E(X_1^{r_1} \cdots X_k^{r_k})$$

In particular,

$$\left. \frac{\partial \phi_X(t)}{\partial t} \right|_{t=0} = \sqrt{-1} EX, \quad \left. \frac{\partial^2 \phi_X(t)}{\partial t \partial t^\tau} \right|_{t=0} = -E(XX^\tau)$$

- Special case of  $k = 1$ :

$$\psi_X(t) = \sum_{i=0}^{\infty} \frac{E(X^i)t^i}{i!} \quad \text{if } \psi(t) < \infty$$

$$E(X^i) = \psi^{(i)}(0) = \left. \frac{d\psi_X^i(t)}{dt^i} \right|_{t=0}, \quad \phi_X^{(i)}(0) = (-1)^{i/2} E(X^i)$$

**Example:** A random variable  $X$  has finite  $E(X^k)$  for  $k = 1, 2, \dots$ , but  $\psi_X(t) = \infty$ , for any  $t \neq 0$

$P_n$ : the probability measure for  $N(0, n^2)$  with p.d.f.  $f_n$ ,  $n = 1, 2, \dots$

$P = \sum_{n=1}^{\infty} 2^{-n} P_n$  is a probability measure with Lebesgue p.d.f.

$\sum_{n=1}^{\infty} 2^{-n} f_n$  (Exercise 35)

Let  $X$  be a random variable having distribution  $P$ .

It follows from Fubini's theorem that  $X$  has finite moments of any order; for even  $k$ ,

$$\begin{aligned} E(X^k) &= \int x^k dP = \int \sum_{n=1}^{\infty} x^k 2^{-n} dP_n = \sum_{n=1}^{\infty} 2^{-n} \int x^k dP_n \\ &= \sum_{n=1}^{\infty} 2^{-n} (k-1)(k-3)\dots 1 n^k < \infty \end{aligned}$$

and  $E(X^k) = 0$  for odd  $k$ .

By Fubini's theorem again, for any  $t \neq 0$ ,

$$\psi_X(t) = \int e^{tx} dP = \sum_{n=1}^{\infty} 2^{-n} \int e^{tx} dP_n = \sum_{n=1}^{\infty} 2^{-n} e^{n^2 t^2 / 2} = \infty$$

## Theorem 1.6. (Uniqueness)

Let  $X$  and  $Y$  be random  $k$ -vectors.

- (i) If  $\phi_X(t) = \phi_Y(t)$  for all  $t \in \mathcal{R}^k$ , then  $P_X = P_Y$ .
- (ii) If  $\psi_X(t) = \psi_Y(t) < \infty$  for all  $t$  in a neighborhood of 0, then  $P_X = P_Y$ .

### Proof

See the textbook.

### Example 1.20

Let  $X_i$ ,  $i = 1, \dots, k$ , be independent random variables and  $X_i$  have the gamma distribution  $\Gamma(\alpha_i, \gamma)$  (Table 1.2),  $i = 1, \dots, k$ .

From Table 1.2,  $X_i$  has the m.g.f.  $\psi_{X_i}(t) = (1 - \gamma t)^{-\alpha_i}$ ,  $t < \gamma^{-1}$ ,  $i = 1, \dots, k$ .

Then, the m.g.f. of  $Y = X_1 + \dots + X_k$  is equal to

$$\psi_Y(t) = \prod_i \psi_{X_i}(t) = \prod_i (1 - \gamma t)^{-\alpha_i} = (1 - \gamma t)^{-(\alpha_1 + \dots + \alpha_k)}, \quad t < \gamma^{-1}.$$

From Table 1.2, the gamma distribution  $\Gamma(\alpha_1 + \dots + \alpha_k, \gamma)$  has the m.g.f.  $\psi_Y(t)$  and, hence, is the distribution of  $Y$  (by Theorem 1.6).

## Theorem 1.6. (Uniqueness)

Let  $X$  and  $Y$  be random  $k$ -vectors.

- (i) If  $\phi_X(t) = \phi_Y(t)$  for all  $t \in \mathcal{R}^k$ , then  $P_X = P_Y$ .
- (ii) If  $\psi_X(t) = \psi_Y(t) < \infty$  for all  $t$  in a neighborhood of 0, then  $P_X = P_Y$ .

## Proof

See the textbook.

## Example 1.20

Let  $X_i$ ,  $i = 1, \dots, k$ , be independent random variables and  $X_i$  have the gamma distribution  $\Gamma(\alpha_i, \gamma)$  (Table 1.2),  $i = 1, \dots, k$ .

From Table 1.2,  $X_i$  has the m.g.f.  $\psi_{X_i}(t) = (1 - \gamma t)^{-\alpha_i}$ ,  $t < \gamma^{-1}$ ,  $i = 1, \dots, k$ .

Then, the m.g.f. of  $Y = X_1 + \dots + X_k$  is equal to

$$\psi_Y(t) = \prod_i \psi_{X_i}(t) = \prod_i (1 - \gamma t)^{-\alpha_i} = (1 - \gamma t)^{-(\alpha_1 + \dots + \alpha_k)}, \quad t < \gamma^{-1}.$$

From Table 1.2, the gamma distribution  $\Gamma(\alpha_1 + \dots + \alpha_k, \gamma)$  has the m.g.f.  $\psi_Y(t)$  and, hence, is the distribution of  $Y$  (by Theorem 1.6).

## Theorem 1.6. (Uniqueness)

Let  $X$  and  $Y$  be random  $k$ -vectors.

- (i) If  $\phi_X(t) = \phi_Y(t)$  for all  $t \in \mathcal{R}^k$ , then  $P_X = P_Y$ .
- (ii) If  $\psi_X(t) = \psi_Y(t) < \infty$  for all  $t$  in a neighborhood of 0, then  $P_X = P_Y$ .

## Proof

See the textbook.

## Example 1.20

Let  $X_i$ ,  $i = 1, \dots, k$ , be independent random variables and  $X_i$  have the gamma distribution  $\Gamma(\alpha_i, \gamma)$  (Table 1.2),  $i = 1, \dots, k$ .

From Table 1.2,  $X_i$  has the m.g.f.  $\psi_{X_i}(t) = (1 - \gamma t)^{-\alpha_i}$ ,  $t < \gamma^{-1}$ ,  $i = 1, \dots, k$ .

Then, the m.g.f. of  $Y = X_1 + \dots + X_k$  is equal to

$$\psi_Y(t) = \prod_i \psi_{X_i}(t) = \prod_i (1 - \gamma t)^{-\alpha_i} = (1 - \gamma t)^{-(\alpha_1 + \dots + \alpha_k)}, \quad t < \gamma^{-1}.$$

From Table 1.2, the gamma distribution  $\Gamma(\alpha_1 + \dots + \alpha_k, \gamma)$  has the m.g.f.  $\psi_Y(t)$  and, hence, is the distribution of  $Y$  (by Theorem 1.6).

## Symmetry

A random vector  $X$  is symmetric about 0 iff  $X$  and  $-X$  have the same distribution

Claim:  $X$  is symmetric about 0 iff its ch.f.  $\phi_X$  is real-valued.

$\Rightarrow$  If  $X$  and  $-X$  have the same distribution, then by Theorem 1.6,

$$\phi_X(t) = \phi_{-X}(t).$$

$$\text{But } \phi_{-X}(t) = \phi_X(-t).$$

$$\text{Then } \phi_X(t) = \phi_X(-t).$$

$$\text{Note that } \sin(-t^T X) = -\sin(t^T X) \text{ and } \cos(t^T X) = \cos(-t^T X)$$

Hence  $E[\sin(t^T X)] = 0$  and, thus,  $\phi_X$  is real-valued.

$\Leftarrow$  If  $\phi_X$  is real-valued, then  $\phi_X(t) = E[\cos(t^T X)]$  and

$$\phi_{-X}(t) = \phi_X(-t) = \phi_X(t).$$

By Theorem 1.6,  $X$  and  $-X$  must have the same distribution.

## Symmetry

A random vector  $X$  is symmetric about 0 iff  $X$  and  $-X$  have the same distribution

**Claim:**  $X$  is symmetric about 0 iff its ch.f.  $\phi_X$  is real-valued.

$\Rightarrow$  If  $X$  and  $-X$  have the same distribution, then by Theorem 1.6,

$$\phi_X(t) = \phi_{-X}(t).$$

$$\text{But } \phi_{-X}(t) = \phi_X(-t).$$

$$\text{Then } \phi_X(t) = \phi_X(-t).$$

$$\text{Note that } \sin(-t^\tau X) = -\sin(t^\tau X) \text{ and } \cos(t^\tau X) = \cos(-t^\tau X)$$

Hence  $E[\sin(t^\tau X)] = 0$  and, thus,  $\phi_X$  is real-valued.

$\Leftarrow$  If  $\phi_X$  is real-valued, then  $\phi_X(t) = E[\cos(t^\tau X)]$  and

$$\phi_{-X}(t) = \phi_X(-t) = \phi_X(t).$$

By Theorem 1.6,  $X$  and  $-X$  must have the same distribution.

## Symmetry

A random vector  $X$  is symmetric about 0 iff  $X$  and  $-X$  have the same distribution

**Claim:**  $X$  is symmetric about 0 iff its ch.f.  $\phi_X$  is real-valued.

$\Rightarrow$  If  $X$  and  $-X$  have the same distribution, then by Theorem 1.6,

$$\phi_X(t) = \phi_{-X}(t).$$

$$\text{But } \phi_{-X}(t) = \phi_X(-t).$$

$$\text{Then } \phi_X(t) = \phi_X(-t).$$

$$\text{Note that } \sin(-t^\tau X) = -\sin(t^\tau X) \text{ and } \cos(t^\tau X) = \cos(-t^\tau X)$$

Hence  $E[\sin(t^\tau X)] = 0$  and, thus,  $\phi_X$  is real-valued.

$\Leftarrow$  If  $\phi_X$  is real-valued, then  $\phi_X(t) = E[\cos(t^\tau X)]$  and

$$\phi_{-X}(t) = \phi_X(-t) = \phi_X(t).$$

By Theorem 1.6,  $X$  and  $-X$  must have the same distribution.