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 $X = (X_1, ..., X_k)$, $EX = (EX_1, ..., EX_k)$
- For random matrix $M = (M_{ij})$, $EM = (EM_{ij})$
- For random vector $X$, $\text{Var}(X) = E(X - EX)(X - EX)^\tau$ 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 \(Y = c^\tau X\), \(c \in \mathbb{R}^k\), 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 \(Y = c^\tau X, \ c \in \mathbb{R}^k\), 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 \(t \geq 0\),
  \[\varphi(t)P(|X| \geq t) \leq \int_{\{|X| \geq t\}} \varphi(X) dP \leq E\varphi(X)\]
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

\[
\int \log \left( \frac{g}{f} \right) f d\nu \geq \int f d\nu = \int \log \left( \frac{g}{f} \right) h d\nu = E \log(g/f)
\]

\[\leq \log(E(g/f)) = \log \left( \int \frac{g}{f} h d\nu \right) = \log \left( \frac{\int g d\nu}{\int f d\nu} \right) \leq 0
\]
**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

$$
\psi_X(t) = E e^{t^\tau X}, \quad t \in \mathbb{R}^k.
$$

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

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

**Remarks**

- If the m.g.f. is finite in a neighborhood of $0 \in \mathbb{R}^k$, 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 $P_X$. 
Moment generating and characteristic functions

Definition 1.5

Let \( X \) be a random \( k \)-vector.

(i) The \textit{moment generating function (m.g.f.)} of \( X \) or \( P_X \) is defined as
\[
\psi_X(t) = Ee^{t^\top X}, \quad t \in \mathbb{R}^k.
\]

(ii) The \textit{characteristic function (ch.f.)} of \( X \) or \( P_X \) is defined as
\[
\phi_X(t) = Ee^{\sqrt{-1}t^\top X} = E[\cos(t^\top X)] + \sqrt{-1} E[\sin(t^\top X)], \quad t \in \mathbb{R}^k
\]

Remarks

- If the m.g.f. is finite in a neighborhood of \( 0 \in \mathbb{R}^k \), 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 \textit{cumulant generating function} of \( X \) or \( P_X \).
Properties of m.g.f. and ch.f.

- If $Y = A^\tau X + c$, where $A$ is a $k \times m$ matrix and $c \in \mathbb{R}^m$, it follows from Definition 1.5 that
  \[
  \psi_Y(u) = e^{c^\tau u} \psi_X(Au) \quad \text{and} \quad \phi_Y(u) = e^{-\sqrt{-1}c^\tau u} \phi_X(Au), \quad u \in \mathbb{R}^m
  \]

- For independent $X_1, \ldots, X_k$,
  \[
  \psi_{\sum_i X_i}(t) = \prod_i \psi_{X_i}(t) \quad \text{and} \quad \phi_{\sum_i X_i}(t) = \prod_i \phi_{X_i}(t), \quad t \in \mathbb{R}^k
  \]

- For $X = (X_1, \ldots, X_k)$ with m.g.f. $\psi_X$ finite in a neighborhood of 0
  \[
  \psi_X(t) = \sum_{(r_1, \ldots, r_k)} \frac{\mu_{r_1, \ldots, r_k} t_1^{r_1} \cdots t_k^{r_k}}{r_1! \cdots r_k!} \quad \mu_{r_1, \ldots, r_k} = E(X_1^{r_1} \cdots X_k^{r_k})
  \]

- Consequently,
  \[
  E(X_1^{r_1} \cdots X_k^{r_k}) = \left. \frac{\partial^{r_1 + \cdots + r_k} \psi_X(t)}{\partial t_1^{r_1} \cdots \partial t_k^{r_k}} \right|_{t=0}
  \]
  \[
  \left. \frac{\partial \psi_X(t)}{\partial t} \right|_{t=0} = EX, \quad \left. \frac{\partial^2 \psi_X(t)}{\partial t \partial t^\tau} \right|_{t=0} = E(XX^\tau)
  \]
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 \( r_1, \ldots, r_k \), 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,

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

Special case of \( k = 1 \):

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

\[
E(X^i) = \psi^{(i)}(0) = \left. \frac{d \psi_X(t)}{dt^i} \right|_{t=0}, \quad \phi_X^{(i)}(0) = (-1)^{i/2} E(X^i)
\]
Example 1.19.

\[ X = N(\mu, \sigma^2) \]

\[ \psi_X(t) = \frac{1}{\sqrt{2\pi}\sigma} \int e^{tx} e^{-(x-\mu)^2 / 2\sigma^2} \, dx \quad \frac{x-\mu}{\sigma} = y \]

\[ = \frac{1}{\sqrt{2\pi}} \int e^{t(\sigma y+\mu)} e^{-y^2 / 2} \, dy = \frac{e^{\mu t + \sigma^2 t^2 / 2}}{\sqrt{2\pi}} \int e^{-(y-\sigma t)^2 / 2} \, dy = e^{\mu t + \sigma^2 t^2 / 2} \]

A direct calculation shows that

\[ E(X) = \psi_X'(0) = \mu \]

\[ E(X)^2 = \psi_X''(0) = \sigma^2 + \mu^2 \]

\[ E(X)^3 = \psi_X^{(3)}(0) = 3\sigma^2 \mu + \mu^3 \]

\[ E(X)^4 = \psi_X^{(4)}(0) = 3\sigma^4 + 6\sigma^2 \mu^2 + \mu^4 \]

If \( \mu = 0 \), then \( E(X)^p = 0 \) when \( p \) is an odd integer

\[ E(X)^p = (p-1)(p-3) \cdots 3 \cdot 1 \sigma^p \] when \( p \) is an even integer

The cumulant generating function of \( X \) is

\[ \kappa_X(t) = \log \psi_X(t) = \mu t + \sigma^2 t^2 / 2 \]

\( \kappa_1 = \mu \), \( \kappa_2 = \sigma^2 \), and \( \kappa_r = 0 \) for \( r = 3, 4, \ldots \).
Example 1.19 (continued): A random variable $X$ has finite $E(X^k)$ for $k = 1, 2..., \text{ but } \psi_X(t) = \infty, \text{ for any } t \neq 0$

$P_n$: the probability measure for $N(0, n^2)$ with p.d.f. $f_n, n = 1, 2, \ldots$
$P = \sum_{n=1}^{\infty} 2^{-n}P_n$ 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 $k$,

$$E(X^k) = \int x^k dP = \int \sum_{n=1}^{\infty} x^k 2^{-n} dP_n = \sum_{n=1}^{\infty} 2^{-n} \int x^k dP_n$$

$$= \sum_{n=1}^{\infty} 2^{-n}(k-1)(k-3)\cdots 1n^k < \infty$$

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 $t \in \mathbb{R}^k$, 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 $X_i, i = 1, \ldots, k$, be independent random variables and $X_i$ have the gamma distribution $\Gamma(\alpha_i, \gamma)$ (Table 1.2), $i = 1, \ldots, k$.

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

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

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

From Table 1.2, the gamma distribution $\Gamma(\alpha_1 + \cdots + \alpha_k, \gamma)$ 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 $t \in \mathbb{R}^k$, 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 $X_i, i = 1, \ldots, k$, be independent random variables and $X_i$ have the gamma distribution $\Gamma(\alpha_i, \gamma)$ (Table 1.2), $i = 1, \ldots, k$.

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

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

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

From Table 1.2, the gamma distribution $\Gamma(\alpha_1 + \cdots + \alpha_k, \gamma)$ 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 $t \in \mathbb{R}^k$, 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 $X_i$, $i = 1, \ldots, k$, be independent random variables and $X_i$ have the gamma distribution $\Gamma(\alpha_i, \gamma)$ (Table 1.2), $i = 1, \ldots, k$.

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

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

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

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

Can two random variables with different distributions have the same moments of any order?

\[ X_1 \text{ has pdf } f_1(x) = \frac{1}{\sqrt{2\pi x}} e^{-(\log x)^2/2}, \quad x \geq 0 \]

\[ X_2 \text{ has pdf } f_2(x) = f_1(x)[1 + \sin(2\pi \log x)], \quad x \geq 0 \]

For any positive integer \( n \),

\[ E(X_1^n) = \frac{1}{\sqrt{2\pi}} \int_0^\infty x^{n-1} e^{-(\log x)^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{ny-y^2/2} dy = e^{n^2/2} \]

\[ E(X_2^n) = E(X_1^n) + \frac{e^{n^2/2}}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-s^2/2} \sin(2\pi s) ds = E(X_1^n) \]

This shows that \( X_1 \) and \( X_2 \) have the same moments of order \( n = 1, 2, \ldots \), but they have different distributions.

\[ M_X(t) = \int_0^\infty \frac{e^{tx}}{\sqrt{2\pi x}} e^{-(\log x)^2/2} dx = \infty, \quad t > 0 \]

\[ M_X(t) = \int_0^\infty \frac{e^{tx}}{\sqrt{2\pi x}} e^{-(\log x)^2/2} dx \leq \int_0^\infty \frac{1}{\sqrt{2\pi x}} e^{-(\log x)^2/2} dx = 1, \quad t \leq 0 \]
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.

⇒ If $X$ and $-X$ have the same distribution, then by Theorem 1.6, $\phi_X(t) = \phi_{-X}(t)$. But $\phi_{-X}(t) = \phi_X(-t)$. Then $\phi_X(t) = \phi_X(-t)$.

Note that $\sin(-t\tau X) = -\sin(t\tau X)$ and $\cos(t\tau X) = \cos(-t\tau X)$.

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

⇐ 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, $\phi_X(t) = \phi_{-X}(t)$.
But $\phi_{-X}(t) = \phi_X(-t)$.
Then $\phi_X(t) = \phi_X(-t)$.
Note that $\sin(-t^\tau X) = -\sin(t^\tau X)$ 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, $\phi_X(t) = \phi_{-X}(t)$. But $\phi_{-X}(t) = \phi_X(-t)$. Then $\phi_X(t) = \phi_X(-t)$.

Note that $\sin(-t^\tau X) = -\sin(t^\tau X)$ 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.