Lecture 6: Characteristic functions and inequalities

The mgf's are useful, but they sometimes are not finite.

Another function that can characterise a distribution is the so-called characteristic function, which is always well defined but it is a complex function.

Definition (characteristic functions).

The characteristic function (chf) of the distribution of a random variable X is defined as

$$\phi_X(t) = E(e^{itX}) = E[\cos(tX)] + \imath E[\sin(tX)], \qquad t \in \mathscr{R}$$

where $\imath = \sqrt{-1}$ and $e^{is} = \cos(s) + \imath \sin(s)$.

- If the mgf $M_X(t)$ of X is finite in a neighborhood of 0, then the chf of X is $\phi_X(t) = M_X(\imath t)$, $t \in \mathscr{R}$.
- If $Y = aX + b$ for constants a and b, then the chf of Y in terms of the chf $\phi_X(t)$ of X is $\phi_Y(t) = e^{ibt}\phi_X(at)$, $t \in \mathscr{R}$.
- A chf is uniformly continuous in $\mathscr{R}$.

Example.

Let X be a random variable with pdf

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \qquad x \in \mathscr{R}.$$

When $t > 0$,

$$\begin{aligned}
\phi_X(t) &= E[\cos(tX)] + iE[\sin(tX)] \\
&= \int_{-\infty}^{\infty} \frac{\cos(tx)}{\pi(1+x^2)} dx + i \int_{-\infty}^{\infty} \frac{\sin(tx)}{\pi(1+x^2)} dx \\
&= \frac{2}{\pi} \int_0^{\infty} \frac{\cos(tx)}{1+x^2} dx = e^{-t}
\end{aligned}$$

since $\sin(s)$ is an odd function and $\cos(s)$ is an even function, where the last equality follows from the Fourier transformation.
For $t < 0$,

$$\phi_X(t) = E[\cos(tX)] + iE[\sin(tX)] = E[\cos(-tX)] = e^{t} = e^{-|t|}$$

When $t = 0$, $\phi_X(0) = 1$ for any chf, and hence we obtain that

$$\phi_X(t) = e^{-|t|} \qquad t \in \mathscr{R}$$

Theorem C1.

If a random variable X has finite $E|X|^r$ for a positive integer r, then

$$\left.\frac{d^r \phi_X(t)}{dt^r}\right|_{t=0} = i^r E(X^r)$$

- The proof will be given in the next lecture.
- Thus, like the mgf, the chf can be used to calculate the moments, but we have to first know the existence of moments.
- For the Cauchy distribution, the chf is $e^{-|t|}$.
  Note that this chf is not differentiable at $t = 0$ and previously we showed that the expectation of X does not exist.

Uniqueness and inversion

Similar to the mgf, the chf characterizes the distribution in the sense that there is a one-to-one correspondence between chf and cdf.

We state some inversion formulas.

The proof of the following theorem is omitted.

Theorem C2

Let X be a random variable with cdf F and chf ϕ.
For any real numbers $y < u$,

$$P(y < X < u) + \frac{P(X = y) + P(X = u)}{2} = \lim_{T\to\infty} \int_{-T}^{T} \frac{e^{-ity} - e^{-itu}}{2\pi it} \phi(t) dt$$

Theorem C3.

Let F be a cdf with chf ϕ.

(i) If F is continuous at y and u with $y < u$, then

$$F(u) - F(y) = \lim_{T\to\infty} \int_{-T}^{T} \frac{e^{-ity} - e^{-itu}}{2\pi it} \phi(t) dt$$

(ii) If F is continuous at x, then

$$F(u) = \lim_{y\to-\infty} \lim_{T\to\infty} \int_{-T}^{T} \frac{e^{-ity} - e^{-itu}}{2\pi it} \phi(t) dt$$

(iii) There is a one-to-one correspondence between cdf and chf.

Proof.

(i) follows from Theorem 2C.
(ii)-(iii) follow from the fact that any cdf can only have countably many discontinuous points and, hence, it is determined by the F values at continuity points of F.

If a cdf is continuous, then Theorem 3C(ii) gives an inversion formula.

For a discrete cdf, we have the following result whose proof is similar to that of Theorem 2C and is omitted.

Theorem C4.

Let F be a discrete cdf with chf ϕ.

(i) If F has a jump at x, then

$$F(x) - F(x-) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-itx} \phi(t) dt$$

(ii) If D is the set of all discontinuous points of F, then

$$\sum_{x\in D} [F(x) - F(x-)] = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\phi(t)|^2 dt$$

Theorem 4C(ii) gives a sufficient and necessary condition for a cdf to be continuous.

We know that $e^{-|t|}$ is a chf, and

$$\int_{-T}^{T} (e^{-|t|})^2 dt \le \int_{-\infty}^{\infty} e^{-2|t|} dt = 2\int_{0}^{\infty} e^{-2t} dt = 1 \qquad T > 0$$

Hence, the corresponding cdf must be continuous.

In fact, in this case we know that the cdf corresponding to $e^{-|t|}$ has a pdf.

Theorem C5.

Suppose that $X_1, X_2, ...$ is a sequence of random variables with chf's $\phi_{X_n}(t)$, and $\lim_{n\to\infty} \phi_{X_n}(t) = \phi(t)$, $t \in \mathscr{R}$, where $\phi(t)$ is continuous in t. Then ϕ must be the chf of a random variable X and for all x at which $F_X(x)$ is continuous,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

The proof is omitted.

Symmetric random variables

A random variable X is symmetric about 0 if $-X$ has the same distribution as X.

Theorem C6.

A random variable X is symmetric about 0 iff its chf ϕ_X is real-valued function.

Proof.

- If X and $-X$ have the same distribution, then $\phi_X(t) = \phi_{-X}(t)$.
  But $\phi_{-X}(t) = \phi_X(-t)$.
  Then $\phi_X(t) = \phi_X(-t)$.
  Note that $\sin(-tX) = -\sin(tX)$ and $\cos(tX) = \cos(-tX)$.
  Hence $E[\sin(tX)] = 0$ and, thus, ϕ_X is real-valued.
- If ϕ_X is real-valued, then $\phi_X(t) = E[\cos(tX)]$ and
  $\phi_{-X}(t) = \phi_X(-t) = \phi_X(t)$.
  By Theorem C3, X and $-X$ must have the same distribution.

Inequalities

Inequalities are useful for statistical theory.

Theorem 3.6.1 (Chebychev's inequality)

Let X be a random variable and let $g(x)$ be a nonnegative function. For any $r > 0$,

$$P(g(X) \geq r) \leq \frac{E[g(X)]}{r}$$

Proof.

Assuming that X has a pdf (if X has a pmf, we replace integral by summation), we have

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{\infty} g(x) f_X(x) dx \geq \int_{-\infty}^{\infty} I(\{g(x) \geq r\}) g(x) f_X(x) dx \\ &\geq r \int_{-\infty}^{\infty} I(\{g(x) \geq r\}) f_X(x) dx = r \int_{\{g(x) \geq r\}} f_X(x) dx \\ &= r P(g(X) \geq r) \end{aligned}$$

where $I(A)$ is the indicator function of the set A.

Different forms of Chebychev's inequality

- If g is nondecreasing, then another form of Chebychev's inequality is, for $\varepsilon > 0$,
$$P(X \geq \varepsilon) \leq \frac{E[g(X)]}{g(\varepsilon)}$$
- Suppose that X has expectation μ and variance σ^2.
For $g(x) = (x-\mu)^2/\sigma^2$, we have
$$P(|X-\mu| \geq t\sigma) = P\left(\frac{(X-\mu)^2}{\sigma^2} \geq t^2\right) \leq \frac{1}{t^2} E\frac{(X-\mu)^2}{\sigma^2} = \frac{1}{t^2}$$
- If X has a finite kth moment with an integer k, then, for $t > 0$,
$$P(|X-\mu| \geq t) \leq \frac{E|X-\mu|^k}{t^k}$$
- If X has a finite mgf $M_X(t)$ for $t \in (-h,h)$, then, for $r > 0$ and $t > 0$,
$$P(X \geq r) \leq \frac{E(e^{tX})}{e^{tr}} = \frac{M_X(t)}{e^{tr}}, \qquad P(X \leq -r) \leq \frac{E(e^{-tX})}{e^{tr}} = \frac{M_X(-t)}{e^{tr}}$$
$$P(|X| \geq r) \leq \frac{M_X(t) + M_X(-t)}{e^{tr}}$$

Chebychev's inequality is useful, but sometimes it is too loose because it does not require much from the distribution of X except some moment conditions.

More useful probability inequality can be derived when we know something about the distribution of X.

Cauchy-Schwartz's inequality

This is another simple but very useful inequality.

If X and Y are random variables with $E(X^2) < \infty$ and $E(Y^2) < \infty$, then the following Cauchy-Schwartz's inequality holds:

$$[E(XY)]^2 \leq E(X^2)E(Y^2)$$

with equality holds iff $P(X = cY) = 1$ for a constant c.

In fact, we also have

$$[E|XY|]^2 \leq E(X^2)E(Y^2)$$

Proof of Cauchy-Schwartz's inequality

Let $a^2 = E(X^2) < \infty$ and $b^2 = E(Y^2) < \infty$.

For any $t > 0$ and $s > 0$,

$$2\sqrt{st} \le s + t \qquad \text{because} \;\; (\sqrt{s} - \sqrt{t})^2 \ge 0$$

Letting $s = X^2/a^2$ and $t = Y^2/b^2$, we obtain

$$\frac{|XY|}{ab} \le \frac{X^2}{2a^2} + \frac{Y^2}{2b^2} \qquad \text{hence} \qquad \frac{E|XY|}{ab} \le \frac{E(X^2)}{2a^2} + \frac{E(Y^2)}{2b^2} = 1$$

which means

$$[E|XY|]^2 \le a^2b^2 = E(X^2)E(Y^2)$$

The other inequality follows since

$$[E(XY)]^2 \le [E|XY|]^2 \le E(X^2)E(Y^2)$$

We next consider what happens if the equality holds.
If $[E|XY|]^2 = E(X^2)E(Y^2)$, then

$$\frac{E|XY|}{ab} = \frac{E(X^2)}{2a^2} + \frac{E(Y^2)}{2b^2} \quad \text{i.e.,} \quad E\left(\frac{|X|}{a} - \frac{|Y|}{b}\right)^2 = 0$$

It follows from a result established previously that

$$P\left(\frac{|X|}{a}-\frac{|Y|}{b}=0\right)=1 \quad \text{i.e.,} \quad P\left(|X|=\frac{a}{b}|Y|\right)=1$$

Finally, consider the situation where $[E(XY)]^2 = E(X^2)E(Y^2)$.
Since $[E(XY)]^2 \leq [E|XY|]^2$, this implies

$$[E(XY)]^2 = [E|XY|]^2 = E(X^2)E(Y^2)$$

and, by the early proof, $P(|X| = \frac{a}{b}|Y|) = 1$.
From $[E(XY)]^2 = [E|XY|]^2$, we must have $\pm E(XY) = E|XY|$.
Suppose $E(XY) = E|XY|$ (the proof for $-E(XY) = E|XY|$ is similar).
Since $|XY| - XY \geq 0$, $E|XY| - E(XY) = 0$ implies $P(|XY| = XY) = 1$.
Combining this with the early result, we must have $P(X = \frac{a}{b}Y) = 1$.

Cauchy-Schwartz's inequality is a special case of the following result.

- Hölder's inequality: If p and q are positive constants satisfying $p > 1$ and $p^{-1} + q^{-1} = 1$ and X and Y are random variables, then

$$E|XY| \leq (E|X|^p)^{1/p}(E|Y|^q)^{1/q}$$

Using Hölder's inequality, we can obtain the following two inequalities.

- Liapounov's inequality: If r and s are constants satisfying $1 \le r \le s$ and X is a random variable, then

$$(E|X|^r)^{1/r} \le (E|X|^s)^{1/s}$$

- Minkowski's inequality: If $p \ge 1$ is a constant and X and Y are random variables, then

$$(E|X+Y|^p)^{1/p} \le (E|X|^p)^{1/p} + (E|Y|^p)^{1/p}$$

Convex set and function

- A set $A \subset \mathscr{R}^k$ is convex iff $x \in A$, $y \in A$, and $t \in (0,1)$ imply

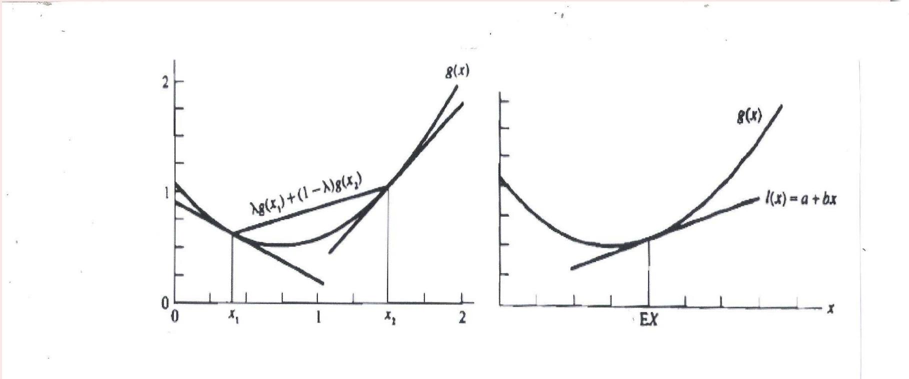
$$tx + (1-t)y \in A$$

- A function g from a convex $A \subset \mathscr{R}^k$ to $\mathscr{R}$ is convex iff $x \in A$, $y \in A$, and $t \in (0,1)$ imply

$$g(tx + (1-t)y) \le t\,g(x) + (1-t)g(y)$$

  and g is strictly convex iff the previous inequality holds with $\le$ replaced by $<$.

- If g is twice differentiable on a convex A, then a necessary and sufficient condition for g to be convex (or strictly convex) is that the $k \times k$ second order partial derivative matrix $\partial^2 g / \partial x \partial x'$ is nonnegative definite (or positive definite).
- If g is convex, then $-g$ is concave.
- The following is a very useful inequality in statistics.



Jensen's inequality

If g is a convex function on a convex $A \subset \mathscr{R}$ and X is a random variable with $P(X \in A) = 1$, then

$$g(E(X)) \leq E[g(X)]$$

provided that the expectations exist. If g is strictly convex, then $\leq$ in the previous inequality can be replaced by $<$ unless $P(g(X) = c) = 1$ for a constant c.

Jensen's inequality also holds for a convex g defined on $A \subset \mathscr{R}^k$ and a random vector X defined on $\mathscr{R}^k$ introduced in Chapter 4.

Proof of Jensen's inequality

Let $l(x) = a + bx$ be the tangent line to $g(x)$ at $g(E(X))$ (see the figure). Since g is convex, $g(x) \geq a + bx$ for all x and, hence,

$$E[g(X)] \geq E(a + bX) = a + bE(X) = l(E(X)) = g(E(X)).$$

If $E[g(X)] = g(E(X))$, then $P(g(X) = a + bX) = 1$, which cannot occur if g is strictly convex and $g(X)$ is not a constant.

Examples

- The function $g(x) = x^{-1}$ is strictly convex. Hence,

$$(EX)^{-1} < E(X^{-1})$$

unless $P(X = c) = 1$ for a constant c.

- The function $g(x) = -\log x$ is strictly convex ($\log x$ is strictly concave). Then

$$-\log(EX) < -E(\log X) \qquad \text{i.e.,} \qquad E(\log X) < \log(EX)$$

unless $P(X = c) = 1$ for a constant c.

- Let f and g be positive functions satisfying $0 < \int_{-\infty}^{\infty} g(x)dx \leq \int_{-\infty}^{\infty} f(x)dx = 1$. We want to show that

$$\int_{-\infty}^{\infty} f(x) \log \frac{g(x)}{f(x)} dx \leq 0$$

Note that f is a pdf.
Let $X \sim f$ and $Y = \frac{g(X)}{f(X)}$.

By Jensen's inequality,

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)\log\frac{g(x)}{f(x)}dx &= E\left(\log\frac{g(X)}{f(X)}\right) = E(\log Y) \\
&\leq \log(EY) = \log\left(E\frac{g(X)}{f(X)}\right) \\
&= \log\left(\int_{-\infty}^{\infty}\frac{g(x)}{f(x)}f(x)dx\right) = \log\left(\int_{-\infty}^{\infty}g(x)dx\right) \\
&\leq \log\left(\int_{-\infty}^{\infty}f(x)dx\right) = \log(1) \\
&= 0
\end{aligned}$$

where the last inequality follows from the fact that $\log x$ is increasing. Also, the strict $<$ holds unless $P(f(X) = g(X)) = 1$.