Lecture 2
Transformations and Expectations

- Lecture 1 introduced $\{S, \mathcal{F}, P\} \xrightarrow{X} \{\mathbb{R}, \mathcal{B}, \mu\}$.
- Lecture 2 deals with a transformation of X, $g(X)$: $\{S, \mathcal{F}, P\} \xrightarrow{X} \{\mathbb{R}, \mathcal{B}, \mu\} \xrightarrow{g(X)} \{\mathbb{R}, \mathcal{B}, \mu\}$.

1. Distributions of Functions of a Random Variable

- Let X be a random variable and let $g(\cdot)$ be a real measurable function on $\mathbb{R}$. In order that $Y = g(X)$ is a well defined random variable, some mild conditions have to be imposed on g.
- The same applies if X is a random vector on $\mathbb{R}^k$ and $g(\cdot)$ is a real function on $\mathbb{R}^k$. The condition is that for any $y \in \mathbb{R}$,

$$\{s \in S : g(X(s)) \leq y\} \in \mathcal{F}.$$

  Then $Y = g(X)$ is a random variable and we say that $g(\cdot)$ is *measurable*.
- Let X be a random variable with CDF $F(\cdot)$. Let $g(\cdot)$ be a measurable function on $\mathbb{R}$. Let $H(\cdot)$ be the CDF of $Y = g(X)$.

The question we now ask is how H and F are related.

Distribution of a function of a random variable

Theorem 1: Let $g(\cdot)$ be invertible and let this inverse $g^{-1}(y)$ be monotonically increasing. Then $H(y) = F\left(g^{-1}(y)\right)$. □

Proof:

$$\begin{aligned} H(y) &= \Pr(Y \leq y) \\ &= \Pr(g(X) \leq y) \\ &= \Pr(g^{-1}(g(X)) \leq g^{-1}(y)) \\ &= \Pr(X \leq g^{-1}(y)) \\ &= F(g^{-1}(y)). \end{aligned}$$

Theorem 2: Let $g\left(\cdot\right)$ be invertible and let this inverse $g^{-1}\left(y\right)$ be monotonically decreasing. Then
$H\left(y\right) = 1 - F\left(g^{-1}\left(y\right)\right) + P\left(X = g^{-1}\left(y\right)\right).$ □

Proof:

$$
\begin{aligned}
H(y) &= \Pr(Y \leq y) \\
&= \Pr(g(X) \leq y) \\
&= \Pr(g^{-1}(g(X)) \geq g^{-1}(y)) \\
&= \Pr(X \geq g^{-1}(y)) \\
&= 1 - \Pr(X < g^{-1}(y)) \\
&= 1 - \Pr(X \leq g^{-1}(y)) + \Pr(X = g^{-1}(y)) \\
&= 1 - F(g^{-1}(y)) + \Pr(X = g^{-1}(y)).
\end{aligned}
$$

Theorem 3: If $F\left(\cdot\right)$ is continuous, the conclusion of the above theorem becomes $H\left(y\right) = 1 - F\left(g^{-1}\left(y\right)\right).$ □

Now, consider the relationship between the *densities* of X and $Y = g(X)$.
If F is continuous and $g^{-1}(y)$ is differentiable, then H is continuous. Let f be the density of F ($f(x) = dF(x)/dx$), and let $h(y)$ be the density of H ($h(y) = dH(y)/dy$). Then

$$h(y) = \frac{dH(y)}{dy} = \frac{dF(g^{-1}(y))}{dy} = f(g^{-1}(y))\frac{dg^{-1}(y)}{dy}$$

if $g^{-1}(y)$ is monotonically increasing, and

$$h(y) = \frac{dH(y)}{dy} = \frac{d\{1 - F(g^{-1}(y))\}}{dy} = -f(g^{-1}(y))\frac{dg^{-1}(y)}{dy}$$

if $g^{-1}(y)$ is monotonically decreasing. Thus we have:

Theorem 4: Let $F(\cdot)$ be continuous with density $f(\cdot)$ and let $g^{-1}(\cdot)$ be monotonic and differentiable. Then $H(\cdot)$ is continuous with density

$$h(y) = f\left(g^{-1}(y)\right)\left|\frac{dg^{-1}(y)}{dy}\right| = f(x)\left|\frac{dx}{dy}\right|.$$

$\square$

Example 1. Let X be a random variable with density

$$f(x) = \begin{cases} e^{-x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

Let $Y = g(X) = \sqrt{X}$. Calculate the density $h(y)$ of Y.

Solution: Note that $g^{-1}(y) = x = y^2$ for $y \geq 0$.

$$h(y) = f(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| = \begin{cases} f(y^2)|2y| = e^{-y^2}|2y| & \text{for } y \geq 0 \\ 0 & \text{for } y < 0 \end{cases} .$$

Thus

$$h(y) = \begin{cases} (2y)(e^{-y^2}) & \text{for } y \geq 0 \\ 0 & \text{for } y < 0 \end{cases} .$$

Example 2. Let X be a random variable with density

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = g(X) = 2X$. Calculate the density $h(y)$ of Y.

Solution: Note that $g^{-1}(y) = x = \frac{1}{2}y$ for $0 < y < 2$.

$$\begin{aligned} h(y) &= f(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| \\ &= \begin{cases} f(\frac{1}{2}y)|\frac{1}{2}| = 2(\frac{1}{2}y)\frac{1}{2} = \frac{1}{2}y & \text{for } 0 < y < 2 \\ 0 & \text{elsewhere} \end{cases} . \end{aligned}$$

Example 3. Let X be a random variable with density

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = g(X) = 8X^3$. Calculate the density $h(y)$ of Y.

Solution: Note that $g^{-1}(y) = x = (\frac{1}{8}y)^{1/3} = \frac{1}{2}y^{1/3}$ for $0 < y < 8$.

$$\begin{aligned} h(y) &= f(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| \\ &= \begin{cases} \begin{array}{l} f(\frac{1}{2}y^{1/3})|\frac{1}{6}y^{-2/3}| \\ = 2(\frac{1}{2}y^{1/3})(|\frac{1}{6}y^{-2/3}|) \\ = \frac{1}{6}y^{-1/3} \end{array} & \text{for } 0 < y < 8 \\ 0 & \text{elsewhere} \end{cases} . \end{aligned}$$

Remark. From the above examples simple rules may be observed. For example, in Example 3.

(a) Verify that the transformation $Y = g(X) = 8X^3$ maps $A = \{x : 0 < x < 1\}$ onto $B = \{y : 0 < y < 8\}$ and that the transformation is one-to-one, i.e., $g(X)$ is monotonic and thus invertible.

(b) Determine $h(y)$ on this set B by substituting $\frac{1}{2}y^{1/3}$ for x in $f(x)$ and then multiply this result by the derivative of $\frac{1}{2}y^{1/3}$.

Example 4. Let X have the uniform distribution with density $f(x) = 1$, $0 < x < 1$; $f(x) = 0$ elsewhere. Find the density of $Y = -2\ln X$. What is the density $h(y)$ of Y?

Solution: $g^{-1}(y) = x = e^{-y/2}$, $0 < y < \infty$.

$$\begin{aligned} h(y) &= f(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| \\ &= \begin{cases} \begin{array}{l} f(e^{-y/2})\lvert -\frac{1}{2}e^{-y/2}\rvert \\ = 1 \cdot (\frac{1}{2}e^{-y/2}) \\ = \frac{1}{2}e^{-y/2} \end{array} & \text{for } 0 < y < \infty \\ 0 & \text{elsewhere.} \end{cases} \end{aligned}$$

In the above four examples, the transformation function was monotonic. Now the question is that what if $g^{-1}(\cdot)$ is not monotonic? In this case, we can not use the above theorems. Here is an example:

Example 5. Let X be a continuously distributed random variable with density $f(x) = e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$. Let $Y = g(X) = X^2$. Derive the density $h(y)$ of Y.

Solution: Note that $g(x) = x^2$ is not monotonic and not invertible. Then we have to use the following general idea.

$$
\begin{aligned}
H(y) &= \Pr(Y \leq y) \\
&= \Pr(X^2 \leq y) \\
&= \Pr(-\sqrt{y} \leq X \leq \sqrt{y}) \\
&= F(\sqrt{y}) - F(-\sqrt{y})
\end{aligned}
$$

$$\begin{aligned} h(y) &= \frac{dH(y)}{dy} = \frac{d}{dy}\{F(\sqrt{y}) - F(-\sqrt{y})\} \\ &= f(\sqrt{y})\frac{d}{dy}(\sqrt{y}) - f(-\sqrt{y})\frac{d}{dy}(-\sqrt{y}) \\ &= \frac{1}{2\sqrt{y}}(f(\sqrt{y}) + f(-\sqrt{y})) \\ &= \frac{1}{2\sqrt{y}}\left(\frac{e^{-\frac{1}{2}y}}{\sqrt{2\pi}} + \frac{e^{-\frac{1}{2}y}}{\sqrt{2\pi}}\right) \\ &= \frac{e^{-\frac{1}{2}y}}{\sqrt{2\pi}\sqrt{y}} \end{aligned}$$

Thus

$$h(y) = \begin{cases} \frac{e^{-\frac{1}{2}y}}{\sqrt{2\pi}\sqrt{y}} & \text{for } y > 0 \\ 0 & \text{elsewhere} \end{cases}.$$

In the same way, Examples 1-4 can be solved:

Example 1. Let X be a random variable with density

$$f(x) = \begin{cases} e^{-x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

Let $Y = g(X) = \sqrt{X}$. Calculate the density $h(y)$ of Y.

Solution:

$$H(y) = \Pr(Y \leq y) = \Pr(\sqrt{X} \leq y) = \Pr(X \leq y^2) = F(y^2)$$

$$\begin{aligned} h(y) &= \frac{dH(y)}{dy} = \frac{dF(y^2)}{dy} = f(y^2)\frac{d}{dy}(y^2) = f(y^2)(2y) \\ &= \begin{cases} (e^{-y^2})(2y) & \text{for } y \geq 0 \\ 0 & \text{elsewhere} \end{cases} \end{aligned}$$

Example 2. Let X be a random variable with density

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = g(X) = 2X$. Calculate the density $h(y)$ of Y.

Solution:

$$H(y) = \Pr(Y \le y) = \Pr(2X \le y) = \Pr\left(X \le \frac{1}{2}y\right) = F\left(\frac{1}{2}y\right)$$

$$\begin{aligned} h(y) &= \frac{dH(y)}{dy} = \frac{dF\left(\frac{1}{2}y\right)}{dy} = f\left(\frac{1}{2}y\right)\frac{d}{dy}\left(\frac{1}{2}y\right) = \frac{1}{2}f\left(\frac{1}{2}y\right) \\ &= \begin{cases} \frac{1}{2}y & \text{for } 0 < y < 2 \\ 0 & \text{elsewhere} \end{cases} \end{aligned}$$

Example 3. Let X be a random variable with density

$$f(x) = \begin{cases} 2x & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = g(X) = 8X^3$. Calculate the density $h(y)$ of Y.

Solution:

$$H(y) = \Pr(Y \leq y) = \Pr(8X^3 \leq y) = \Pr\left(X \leq \frac{1}{2}y^{1/3}\right) = F\left(\frac{1}{2}y^{1/3}\right)$$

$$\begin{aligned} h(y) &= \frac{dH(y)}{dy} = \frac{dF\left(\frac{1}{2}y^{1/3}\right)}{dy} = f\left(\frac{1}{2}y^{1/3}\right)\frac{d}{dy}\left(\frac{1}{2}y^{1/3}\right) \\ &= f\left(\frac{1}{2}y^{1/3}\right)\left(\frac{1}{6}y^{-2/3}\right) \\ &= \begin{cases} \frac{1}{6}y^{-1/3} & \text{for } 0 < y < 8 \\ 0 & \text{elsewhere} \end{cases} \end{aligned}$$

Example 4. Let X have the uniform distribution with density $f(x) = 1$, $0 < x < 1$; $f(x) = 0$ elsewhere. Find the density of $Y = -2\ln X$.

Solution:

$$\begin{aligned} H(y) &= \Pr(Y \le y) = \Pr(-2\ln X \le y) = \Pr\left(\ln X \ge -\frac{1}{2}y\right) \\ &= \Pr\left(X \ge e^{-\frac{1}{2}y}\right) = 1 - \Pr\left(X < e^{-\frac{1}{2}y}\right) = 1 - F\left(e^{-\frac{1}{2}y}\right) \end{aligned}$$

$$\begin{aligned} h(y) &= \frac{dH(y)}{dy} = \frac{d}{dy}\{1 - F\left(e^{-\frac{1}{2}y}\right)\} = -f(e^{-\frac{1}{2}y})\frac{d}{dy}(e^{-\frac{1}{2}y}) \\ &= -f(e^{-\frac{1}{2}y})(-\frac{1}{2})(e^{-\frac{1}{2}y}) \\ h(y) &= \begin{cases} \frac{1}{2}e^{-\frac{1}{2}y} & \text{for } 0 < y < \infty \\ 0 & \text{elsewhere} \end{cases} \end{aligned}$$

Summary: Suppose X is a continuous random variable (c.r.v.). What is the pdf of Y given pdf $f_X(x)$ of X? The basic idea is first to find the distribution function of Y and then its probability density by differentiation.

Step 1: Find the expression of $F_Y(y)$ using $F_X(x)$:

$$F_Y(y) = \Pr(Y \le y) = \Pr[g(X) \le y] = \Pr[X \in g^{-1}(y)]$$

where

$$g^{-1}(y) = \{x \in \mathbb{R} : g(x) \le y\},$$

i.e., $g^{-1}(y)$ is a subset in Ω_X that contains all x's satisfying the inequality $g(x) \le y$.

Step 2:

$$f_Y(y) = F'_Y(y).$$

Step 3: Always check if $f_Y(y)$ is a density (i.e. check $f_Y(y) \ge 0$, $\int_{-\infty}^{\infty} f_Y(y)dy = 1$).

Question for you: Suppose a c.r.v. X has a pdf

$$f_X(x) = 1 \text{ for } -\frac{1}{2} < x < \frac{1}{2}; \text{ and zero elsewhere.}$$

Find the pdf for the following new random variables Y.
(a) $Y = a + bX,\ b \neq 0$
(b) $Y = X^2$
(c) $Y = |X|$

Remark: It is very important to identify first the possible values of Y (i.e, the support of Y). For this purpose, it is extremely useful to draw a picture of $Y = g(X)$.

Solution for (a) $Y = a + bX,\ b \neq 0$

$$F_Y(y) = \Pr[Y \leq y] = \Pr[a + bX \leq y] = \Pr[bX \leq y - a].$$

Case (i) $b > 0$: We have

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) = \Pr[bX \leq y - a] = \Pr[X \leq (y-a)/b] \\ &= F_X\left(\frac{y-a}{b}\right) \end{aligned}$$

It follows that

$$F_Y(y) = F_X\left(\frac{y-a}{b}\right) = F_X(z)$$

where $z = (y-a)/b$. Then

$$f_Y(y) = F'_Y(y) = F'_X(z)\frac{dz}{dy} = f_X\left(\frac{y-a}{b}\right)\frac{1}{b} = 1 \cdot \frac{1}{b}$$

for $-\frac{1}{2} < \frac{y-a}{b} < \frac{1}{2}$ (i.e., $a - \frac{b}{2} < y < a + \frac{b}{2}$). It follows that

$$f_Y(y) = \begin{cases} \frac{1}{b} & a - \frac{b}{2} < y < a + \frac{b}{2} \\ 0 & \text{otherwise} \end{cases}.$$

Case (ii) $b < 0$: We have

$$\Pr(Y \leq y) = 1 - F_X\left(\frac{y-a}{b}\right).$$

It follows that $F_Y(y) = 1 - F_X(z)$, where $z = (y-a)/b$. By differentiation, we obtain

$$f_Y(y) = 0 - F_X'(z)\frac{dz}{dy} = -f_X(z)\frac{1}{b} = -f_X\left(\frac{y-a}{b}\right)\frac{1}{b}.$$

It follows that

$$f_Y(y) = \begin{cases} -\frac{1}{b} & a + \frac{b}{2} < y < a - \frac{b}{2} \\ 0 & \text{otherwise} \end{cases}$$

Solution for (b) $Y = X^2$.
Observe that Y always takes nonnegative value between $[0, \frac{1}{4})$. Let $y \geq 0$, then

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) = \Pr(X^2 \leq y) = \Pr(-y^{1/2} \leq X \leq y^{1/2}) \\ &= F_X(y^{1/2}) - F_X(-y^{1/2}). \end{aligned}$$

By differentiation,

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}\left[F_X(y^{1/2}) - F_X(-y^{1/2})\right] \\ &= f_X(y^{1/2})\frac{1}{2y^{1/2}} + f_X(-y^{1/2})\frac{1}{2y^{1/2}}. \end{aligned}$$

Thus we have

$$f_Y(y) = \begin{cases} \frac{1}{y^{1/2}} & 0 \leq y < \frac{1}{4} \\ 0 & \text{otherwise} \end{cases}.$$

Solution for (c) $Y = |X|$.

Observe that Y always takes nonnegative value between $[0, \frac{1}{2})$. Let $y \geq 0$, then

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) = \Pr(|X| \leq y) = \Pr(-y \leq X \leq y) \\ &= F_X(y) - F_X(-y). \end{aligned}$$

By differentiation,

$$f_Y(y) = \frac{d}{dy}\left[F_X(y) - F_X(-y)\right] = f_X(y) + f_X(-y).$$

Thus we have

$$f_Y(y) = \begin{cases} 2 & 0 \leq y < \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}.$$

Question for you: Suppose a c.r.v. X has a pdf

$$f_X(x) = \frac{1}{2}\alpha e^{-\alpha|x|}, \text{ for } -\infty < x < \infty.$$

where $\alpha > 0$. This is called **the double exponential (or Laplace) distribution.** Find the pdf for the following new random variables Y:
(a) $Y = |X|$
(b) $Y = X^2$.

Solution for (a) $Y = |X|$:

(a) Y is nonnegative, $y \in [0, \infty)$:

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) = \Pr(|X| \leq y) = \Pr(-y \leq X \leq y) \\ &= F_X(y) - F_X(-y). \end{aligned}$$

By differentiation,

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}\left[F_X(y) - F_X(-y)\right] = f_X(y) + f_X(-y) \\ &= \frac{1}{2}\alpha e^{-\alpha|y|} + \frac{1}{2}\alpha e^{-\alpha|y|} \\ &= \alpha e^{-\alpha|y|}. \end{aligned}$$

Thus we have

$$f_Y(y) = \begin{cases} \alpha e^{-\alpha|y|} & y \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

This is an **exponential(**α). The absolute value of an double exponential random variable follows an exponential distribution.

Solution for (b) $Y = X^2$.

$$\begin{aligned} F_Y(y) &= \Pr(Y \le y) = \Pr(X^2 \le y) = \Pr(-y^{1/2} \le X \le y^{1/2}) \\ &= F_X(y^{1/2}) - F_X(-y^{1/2}). \end{aligned}$$

By differentiation,

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}\left[F_X(y^{1/2}) - F_X(-y^{1/2})\right] \\ &= f_X(y^{1/2})\frac{1}{2y^{1/2}} + f_X(-y^{1/2})\frac{1}{2y^{1/2}} \\ &= \frac{1}{2}\alpha\frac{1}{y^{1/2}}e^{-\alpha y^{1/2}},\ \text{for } y > 0, \end{aligned}$$

and $f_Y(y) = 0$ if $y \le 0$. This is a special case of the so-called **Weibull distribution** $(\beta = \frac{1}{2}, \delta = \alpha^{-2}, \gamma = 0)$.
A Weibull distribution has a pdf

$$f_X(x) = \begin{cases} \frac{\beta}{\delta}\left(\frac{x-\gamma}{\delta}\right)^{\beta-1}\exp\left[-\left(\frac{x-\gamma}{\delta}\right)^{\beta}\right] & x > \gamma \\ 0 & \text{otherwise} \end{cases}.$$

Example. Suppose X has a pdf

$$f_X(x) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \text{ for } -\infty < x < \infty.$$

where $\sigma > 0$. This is called the normal distribution with mean μ and variance σ^2. Find the pdf for $Y = \exp(X)$. The distribution of Y is called the **lognormal distribution**.
Solution: Note that $Y \in (0, \infty)$.

$$F_Y(y) = \Pr(Y \le y) = \Pr(e^X \le y) = \Pr(X \le \ln y) = F_X(\ln y).$$

By differentiation,

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_X(\ln y) = f_X(\ln y)\frac{1}{y} \\ &= \frac{1}{(2\pi)^{1/2}\sigma}\frac{1}{y} \exp\left[-\frac{(\ln y-\mu)^2}{2\sigma^2}\right], \text{ for } 0 < y < \infty. \end{aligned}$$

Example [Probability Integral Transform]: Suppose X has a continuous distribution $F_X(x)$ which is strictly monotonically increasing. Find the pdf of $Y = F_X(X)$; that is,

$$Y = \int_{-\infty}^{X} f_X(x)dx.$$

Remark: The CDF $F_X(x) = \Pr(X \leq x)$ is not a random variable when x is a realization. However, $F_X(X)$ is a random variable because it is a function of X.

Solution: The support of $Y = F_X(X)$ is $y \in [0,1]$. Then

$$F_Y(y) = \Pr(Y \leq y) = \Pr[F_X(X) \leq y].$$

Because $F_X(x)$ is strictly increasing, its inverse function denoted as $F_X^{-1}(y)$ exists and is also strictly increasing. For any real-value x; we have

$$\begin{aligned} F_X^{-1}[F_X(x)] &= x, \\ F_X^{-1}[F_X(y)] &= y. \end{aligned}$$

By applying the inverse function operation, we obtain

$$\begin{aligned}
F_Y(y) &= \Pr(Y \leq y) \\
&= \Pr[F_X(X) \leq y] \\
&= \Pr[F_X^{-1}[F_X(X)] \leq F_X^{-1}(y)] \\
&= \Pr[X \leq F_X^{-1}(y)] \\
&= F_X[F_X^{-1}(y)] \\
&= y.
\end{aligned}$$

By differentiation, the pdf is

$$f_Y(y) = \begin{cases} 1 & y \in [0,1] \\ 0 & \text{otherwise} \end{cases} .$$

This is a uniform distribution on [0, 1].

Remark: The transformation $Y = F_X(X)$ is called the probability integral transformation (PIT) of X.
Rosenblatt, M. (1952), "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, 23, 470-472.

The result that $F_X(X)$ has a uniform distribution on $[0,\ 1]$ is useful for:

1. evaluating density forecasts. e.g., F.X. Diebold, T.A. Gunther, and A.S. Tay (1998, IER).
2. simulating random numbers from a certain distribution (Discussion 1).

1. Evaluating density forecasts

This is a goodness-of-fit test for distributional models. To check whether a probability model $F_0(x)$ is correctly specified for $F(x)$, one can check if $F_0(X)$ is $U[0\ 1]$ using observed data.

The hypotheses of interest:

$H_0 : X$ has the distribution $F_0(x)$

$H_1 : X$ has a distribution different from $F_0(x)$

Question for you: Explain how you would test for the hypothesis using PIT.

2. Simulating random numbers (CB Section 5.6.1)

The property of $Y = F_X(X) \sim U[0\ 1]$ can be used to generate random variables.

To generate an observation on X from a population with CDF $F_X(\cdot)$, we can generate a number u from the uniform distribution $U[0\ 1]$, and then solve for x in the equation $F_X(x) = u$.

Example [Simulating an exponentially distributed random variable, CB Example 5.6.3]: The random variable X follows an exponential distribution if $f_X(x) = e^{-x}$ for $x \geq 0$, and 0 otherwise. The CDF of X is

$$F_X(x) = \begin{cases} 1 - e^{-x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} .$$

Defining $Y = F_X(X) = 1 - e^{-X}$, we know Y is $U[0\ 1]$. Now generate $Y \sim U[0\ 1]$ from the computer. Then

$$\begin{aligned} X &= -\ln(1-Y) \sim \text{Exponential}(1) \\ \text{or} \quad X &= -\ln Y \sim \text{Exponential}(1) \end{aligned}$$

See CB Example 2.1.4 and CB Example 5.6.3. See also CB Equation (5.6.5) and CB Exercise 5.49.

Question for you:

- Explain how you can generate random numbers from an **exponential** distribution
$$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda}, \qquad x > 0.$$
- Generate 1000 observations from exponential distribution with $\lambda = 1, \frac{1}{3}, \frac{1}{5}$. [Do **not** use the built-in command such as EXPRND in Matlab.]
- Compute the sample mean and the sample variance of your random sample.
- Draw the histogram. Can you confirm the results in CB Figure 2.3.1? See also CB Figure 5.6.1.
- Show analytically the mean and variance of your exponential distribution.

[Turn in your answers together with your computer code and the output (figure) in PDF via email to the TA. Indicate which software you use, e.g., Matlab, Gauss, or R. Include your name in the file and in the file name.]

Question for you:

- How would you generate random numbers from a **mixture of two normal** distribution distributions, (say) $\mathcal{N}(-3,1)$ with probability 0.5 and $\mathcal{N}(3,1)$ with probability 0.5?
- Use a computer to draw 1000 observations.
- Compute the sample mean and the sample variance of your random sample.
- Draw the histogram.
- Show analytically the mean and variance of your mixture normal distribution.

Question for you:

- How would you generate random numbers from a **lognormal** distribution with the density

$$f_X(x) = \frac{1}{2\sqrt{2\pi}} \frac{1}{x} \exp\left[-\frac{(\ln x - 1)^2}{8}\right], \text{ for } x > 0.$$

- Use a computer to draw 1000 observations.
- Compute the sample mean and the sample variance of your random sample.
- Draw the histogram.
- Show analytically the mean and variance of your log normal distribution.

Discrete case

How to generate random numbers from a discrete probability distribution?

Suppose X is a discrete random variable taking on values $x_1 < x_2 < \cdots < x_k$. We can write

$$\Pr\left[F_X(x_{i-1}) < U \leq F_X(x_i)\right] = F_X(x_i) - F_X(x_{i-1}) = \Pr(X = x_i)$$

for $i = 1, 2, \ldots, k$, where we define $x_0 = -\infty$ and $F_X(x_0) = 0$.Thus, we can first generate a uniform random number U, and then set

$$X = x_i \ \text{ if } F_X(x_{i-1}) < U \leq F_X(x_i).$$

The new random numbers generated in this way will follow a discrete probability distribution $F_X(x)$.

Suppose X is a discrete random variable (d.r.v.). How to find the pmf $f_Y(y)$ given the pmf $f_X(x)$ of X?

$$f_Y(y) = \sum_{x:g(x)=y} f_X(x) = \sum_{x \in g^{-1}(y)} f_X(x)$$

where the summation is over all possible x's whose $g(x) = y$. Intuitively, $Y = g(X)$ is a transformation from the sample space $\mathcal{X}$ to a new sample space $\mathcal{Y} = \{y : y = g(x),\ x \in \mathcal{X}\}$. The pmf of Y can be defined as an induced probability function using the pmf $f_X(x)$, and this is exactly the formula

$$\begin{aligned} f_Y(y) &= \Pr(Y = y) = \Pr(X \in \{x \in \mathcal{X} : g(x) = y\}) \\ &= \sum_{x:g(x)=y} f_X(x) = \sum_{x \in g^{-1}(y)} f_X(x), \end{aligned}$$

for $y \in \mathcal{Y}$; and $f_Y(y) = 0$ for $y \notin \mathcal{Y}$. See CB page 48.

Example (Binomial, CB Example 2.1.1.):
A discrete random variable X has its pmf of the form

$$f_X(x) = \Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n,$$

where n is a positive integer and $0 \leq p \leq 1$. Let $Y = g(X) = n - X$. Here $\mathcal{X} = \{0, 1, \ldots, n\}$, and $\mathcal{Y} = \{y : y = g(x),\ x \in \mathcal{X}\} = \{0, 1, \ldots, n\}$. Then Y also has a binomial distribution with the following pmf because

$$\begin{aligned} f_Y(y) &= \Pr(Y = y) = \sum_{x \in g^{-1}(y)} f_X(x) = f_X(n-y) \\ &= \binom{n}{n-y} p^{n-y} (1-p)^{n-(n-y)} = \binom{n}{y} (1-p)^y p^{n-y}, \end{aligned}$$

2. Mathematical Expectations

Definition: Let $g(\cdot)$ be a measurable real function on $\mathbb{R}$ and let X be a random variable in $\mathbb{R}$ with distribution function F. Then the mathematical expectation of $g(X)$ is defined by
$\mathbb{E}g(x) = \int g(x)\,dF(x)$. $\square$

Remark: If $F(\cdot)$ is continuous with density f, then $dF(x) = f(x)dx$, so that the Stieltjes integral can be written as a Riemann integral: $\mathbb{E}g(x) = \int g(x)\,f(x)dx$. If $F(\cdot)$ is discrete, $\mathbb{E}g(x) = \sum_{x \in C} g(x)\,f(x)$ where C is a countable subset of $\mathbb{R}$. $\square$

The following theorem follows easily from the definition of mathematical expectation.

Theorem: Let $g_1(\cdot)$ be measurable real functions on $\mathbb{R}^k$ and let X be a random vector in $\mathbb{R}^k$. We have:
(i) $\mathbb{E}cg_1(X) = c\mathbb{E}g_1(X)$ for any constant $c \in \mathbb{R}$
(ii) $|\mathbb{E}g_1(X)| \leq \mathbb{E}|g_1(X)|$
Let $g_2(\cdot)$ be measurable real functions on $\mathbb{R}^m$ and let Y be random vectors in $\mathbb{R}^m$. Then:
(iii) $\mathbb{E}[\alpha g_1(X) + \beta g_2(X)] = \alpha\mathbb{E}g_1(X) + \beta\mathbb{E}g_2(X)$ for any real constants α, β
(iv) If X and Y are independent, then
$\mathbb{E}g_1(X)\, g_2(Y) = \mathbb{E}g_1(X) \cdot \mathbb{E}g_2(Y)$
(v) If $\Pr(g_1(X) \leq g_2(Y)) = 1$, then $\mathbb{E}g_1(X) \leq \mathbb{E}g_2(Y)$. $\square$

Remark: (v) A random variable is bounded then its mathematical expectation is also bounded.

Moments of distribution

Let X be a random variable with distribution function $F(x)$.
Definition: The m-th moment of X is defined as
$\mathbb{E}X^m = \int x^m dF(x)$ provided that $\mathbb{E}|X|^m < \infty$.
Definition: The first $(m = 1)$ moment of X, $\mathbb{E}X$, is called the *mean* of (the distribution of) X. This is usually denoted by μ_X.
Definition: The *variance* of X is $\mathbb{E}(X - \mu_X)^2$. It is usually denoted by σ_X^2 or by $Var(X)$. The square root of σ_X^2 (thus σ_X) is called the *standard deviation* of X.
Exercise: Show that $\sigma_X^2 = \mathbb{E}(X^2) - \mu_X^2$.

Definition [CB 2.2.1, Expected Value of $g(X)$]: Suppose X is a r.v. with pmf or pdf $f_X(x)$: Then the expected value or mean of a measurable function $g(X)$ is defined as

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) dF_X(x) = \begin{cases} \sum_x g(x) f_X(x) & \text{drv} \\ \int_{\mathbb{R}} g(x) f_X(x) dx & \text{crv} \end{cases}$$

provided that the integral or sum exists.

Remarks: (i) $g(X)$ is a r.v. because X is a r.v. (ii) If $\mathbb{E}|g(X)| = \infty$, we say that $\mathbb{E}[g(X)]$ does not exist.
We now consider some examples of expectations.

Case I. When $g(X) = X$:
The mean of a random variable X is defined as

$$\begin{aligned} \mu_X &= \mathbb{E}(X) = \int_{\mathbb{R}} x dF_X(x) \\ &= \begin{cases} \sum_x x f_X(x) & \text{drv} \\ \int_{\mathbb{R}} x f_X(x) dx & \text{crv} \end{cases} \end{aligned}$$

where the summation is over all possible x's.
Remark: The mean μ_X is also called the expected value of X, or the first moment of X. It is a measure of central tendency for the distribution of X. It can be viewed as a "location" parameter.

Remark: It is said that the expectation X *exists* for a continuous distribution if and only if

$$\int_{\mathbb{R}} |x| f_X(x) dx < \infty.$$

Whenever X is a bounded random variable, that is, whenever there are numbers a and b $(-\infty < a < b < \infty)$ such that $\Pr(a \leq X \leq b) = 1$, then μ_X must exist.

Example (CB Example 2.2.4): Suppose X has a pdf

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2} \qquad \text{for } -\infty < x < \infty.$$

This is called a Cauchy distribution. Then

$$\int_{-\infty}^{\infty} |x| f_X(x) dx = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx = \infty. \quad \text{(Show this. CB p. 57)}$$

Therefore, the expectation does not exist for the Cauchy distribution.

Theorem: Suppose $\mathbb{E}(X^2)$ exists. Then

$$\mu_X = \arg\min_b \mathbb{E}(X-b)^2.$$

Proof:

$$\frac{d\mathbb{E}(X-b)^2}{db} = 0$$

Then

$$\frac{d\left[\int_{-\infty}^{\infty}(x-b)^2 f(x)dx\right]}{db} = -2\int_{-\infty}^{\infty}(x-b)f(x)dx = 0$$

so we have

$$b = \frac{\int_{-\infty}^{\infty} xf(x)dx}{\int_{-\infty}^{\infty} f(x)dx} = \mu_X.$$

Remark (CB Example 2.2.6):

$$\begin{aligned}
\mathbb{E}(X-b)^2 &= \mathbb{E}(X-\mathbb{E}X+\mathbb{E}X-b)^2 \\
&= \mathbb{E}(X-\mathbb{E}X)^2+(\mathbb{E}X-b)^2+2\mathbb{E}\left[(X-\mathbb{E}X)(\mathbb{E}X-b)\right] \\
&= \mathbb{E}(X-\mathbb{E}X)^2+(\mathbb{E}X-b)^2 \\
&= Var(X)+(bias)^2
\end{aligned}$$

as $\mathbb{E}(X-\mathbb{E}X)(\mathbb{E}X-b)=(\mathbb{E}X-b)\times\mathbb{E}(X-\mathbb{E}X)=0$. We have no choice over the first term but the second term can be made zero by choosing $b=\mathbb{E}X=\mu_X$. Hence

$$\min_b \mathbb{E}(X-b)^2=\mathbb{E}(X-\mu_X)^2.$$

[Remark: Does $X=\mu_X$ has the largest probability to occur? Is $\Pr(X=\mu_X)$ the largest? No. μ_X is not necessarily the mode.]

Example (CB Exercise 2.18):

$$Median_X = \arg\min_b \mathbb{E}|X-b| = \int_{-\infty}^{\infty} |x-b| f(x)dx.$$

Proof:

$$\mathbb{E}|X-b| = \int_{-\infty}^{\infty} |x-b| f(x)dx = \int_{-\infty}^{b} -(x-b)f(x)dx + \int_{b}^{\infty} (x-b)f(x)dx$$

Then

$$\frac{d\mathbb{E}|X-b|}{db} = \int_{-\infty}^{b} f(x)dx - \int_{b}^{\infty} f(x)dx = 0$$

We have $\int_{-\infty}^{b} f(x)dx = \int_{b}^{\infty} f(x)dx = 0.5$, and thus b is the median.
This is mimimum since

$$\frac{d^2\mathbb{E}|X-b|}{db^2} = 2f(b) > 0.$$

Case II. When $g(X) = (X - \mu_X)^2$:
The variance of a random variable X is defined as

$$
\begin{aligned}
\sigma_X^2 &= \mathbb{E}(X - \mu_X)^2 = \int_{\mathbb{R}} (x - \mu_X)^2 dF_X(x) \\
&= \begin{cases} \sum_x (x - \mu_X)^2 f_X(x) & \text{drv} \\ \int_{\mathbb{R}} (x - \mu_X)^2 f_X(x) dx & \text{crv} \end{cases}
\end{aligned}
$$

where the summation is over all possible x's. The standard deviation of X is $\sigma_X = \sqrt{\sigma_X^2}$.

Case III. When $g(X) = X^k$ and $g(X) = (X - \mu_X)^k$:
The kth moment of a random variable X is defined as

$$\mathbb{E}(X^k) = \int_{\mathbb{R}} x^k dF_X(x) = \begin{cases} \sum_x x^k f_X(x) & \text{drv} \\ \int_{\mathbb{R}} x^k f_X(x)dx & \text{crv} \end{cases} .$$

Similarly, the kth central moment of X is defined as

$$\mathbb{E}(X-\mu_X)^k = \int_{\mathbb{R}} (x-\mu_X)^k dF_X(x) = \begin{cases} \sum_x (x-\mu_X)^k f_X(x) & \text{drv} \\ \int_{\mathbb{R}} (x-\mu_X)^k f_X(x)dx & \text{crv} \end{cases} .$$

Remark: What is the relationship between uncentered moments and centered moments? Using the binomial formula,

$$\mathbb{E}(X-\mu_X)^k = \mathbb{E}\sum_{i=0}^{k} \left(\begin{array}{c} k \\ i \end{array} \right) X^i(-\mu_X)^{k-i} = \sum_{i=0}^{k} \left(\begin{array}{c} k \\ i \end{array} \right) \mathbb{E}X^i(-\mu_X)^{k-i}$$

Thus, the k-th central moment is a linear combination of the first k uncentered moments. Similarly, we have

$$\mathbb{E}X^k = \mathbb{E}(X - \mu_X + \mu_X)^k = \sum_{i=0}^{k} \left(\begin{array}{c} k \\ i \end{array} \right) \mathbb{E}(X - \mu_X)^i(\mu_X)^{k-i}$$

Remark: Various moments of a distribution is an important class of expectations. They can intuitively describe the features of a distribution. In econometrics, however, primary interest is in the first moment and the second moment of X.

When evaluating expectations of nonlinear functions of X, we can proceed in two ways:

1. Directly
$$\mathbb{E}g(X) = \int_{\mathbb{R}} g(x) f_X(x) dx$$
2. Indirectly
$$\mathbb{E}g(X) = \mathbb{E}Y = \int_{\mathbb{R}} y f_Y(y) dy$$

Moments of exponential distribution

$$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda}, \qquad 0 \le x < \infty, \lambda > 0.$$

CB Example 2.2.2 (Mean):

$$\begin{aligned} \mathbb{E}X &= \int_0^\infty x \frac{1}{\lambda} e^{-x/\lambda} dx \\ &= -xe^{-x/\lambda}\Big|_0^\infty + \int_0^\infty e^{-x/\lambda} dx \\ &= \int_0^\infty e^{-x/\lambda} dx = \lambda \end{aligned}$$

CB Example 2.3.3 (Variance):

$$\begin{aligned} VarX &= \mathbb{E}(X-\lambda)^2 = \int_0^\infty (x-\lambda)^2 \frac{1}{\lambda} e^{-x/\lambda} dx \\ &= \int_0^\infty (x^2 - 2\lambda x + \lambda^2) \frac{1}{\lambda} e^{-x/\lambda} dx = \cdots = \lambda^2 \end{aligned}$$

Moments of binomial distribution

$$\Pr(X = x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n,\ 0 \le p \le 1.$$

CB Example 2.2.3 (Mean):

$$\mathbb{E}X = \sum_{x=0}^{n} x \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x} = np$$

CB Example 2.3.5 (Variance):

$$VarX = \mathbb{E}X^2 - (np)^2 = np(1-p)$$

Skewness

The third central moment $\mathbb{E}(X-\mu_X)^3$ is a measure of "skewness (i.e., lack of symmetry)" or asymmetry of the distribution for X. It has been used to measure financial crashes because when more negative large values than positive large values occur, $\mathbb{E}(X-\mu_X)^3$ will be large and negative.
Skewness is defined as

$$S = \frac{\mathbb{E}(X-\mu_X)^3}{\sigma_X^3}.$$

Assuming that $\mu_X = 0$, then positive (negative) skewness indicates a higher (lower) probability of experiencing large gains than large losses of the same magnitude.

Example [Capital Asset Pricing Model with Skewness]: Kraus and Litzenberger (1976) use a utility function defined over mean, standard deviation, and the third root of skewness.

Kurtosis

The fourth central moment $\mathbb{E}(X - \mu_X)^4$ is a measure of of how heavy the tail of a distribution is. Heavy tails mean that it is more likely for extreme events to occur.
Kurtosis is defined as

$$K = \frac{\mathbb{E}(X - \mu_X)^4}{\sigma_X^4}.$$

The kurtosis of a distribution is a measure of the extent to which a distribution is peaked or fat.

Example: Suppose X is a normal random variable with mean μ and variance σ^2. That is, its pdf

$$f_X(x) = \frac{1}{(2\pi)^{1/2}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

Then $S = 0$ and $K = 3$.

Remarks

- The excess kurtosis of a random variable X is defined as $K - 3$. Thus, an excess kurtosis implies that extraordinary gains or losses are more common than a normal distribution predicts.
- A distribution with kurtosis larger than that of $\mathcal{N}(\mu, \sigma^2)$ is called a leptokurtic distribution. Both skewness and kurtosis are closely related to the tails of the distribution of X.
- Recent studies find evidence that skewness and kurtosis of financial time series are time-varying. See (e.g.) Hansen (1994, *IER*), Harvey and Siddique (2000, *Jounal of Finance*). These findings are important for modelling the distribution of financial time series, which is important in derivative pricing and risk management.

3. Quantiles

Definition [α-quantile]: Suppose X has a cdf $F_X(x)$. Then the α-quantile of the distribution $F_X(x)$ for $\alpha \in (0,1)$ is defined as $Q(\alpha)$, which satisfies the following equation:

$$\Pr[X \leq Q(\alpha)] = \alpha,$$

or equivalently

$$\begin{aligned} F_X[Q(\alpha)] &= \alpha, \\ Q(\alpha) &= \inf\{q : F_X(q) > \alpha\}. \end{aligned}$$

When $F_X(x)$ is strictly increasing, we have

$$Q(\alpha) = F_X^{-1}(\alpha).$$

Remark: Suppose $f_X(x) = F_X'(x)$ exists almost everywhere. Then

$$\int_{-\infty}^{Q(\alpha)} f_X(x)dx = \alpha.$$

Special cases:

1. Median.

When $\alpha = 0.5$, $Q(\alpha) = m$ is called the median. Here, we have

$$\int_{-\infty}^{m} f_X(x)dx = \frac{1}{2}.$$

What is the interpretation for the median? What is the difference between the mean and the median? The median is the 0.5-quantile. It is the cutoff point that divides the population that $F_X(x)$ represents in half if $F_X(x) = \alpha$ has a solution. The mean is an excellent measure of location for symmetric or nearly symmetric distributions. But it can be misleading when used to measure the location of highly skewed data. In contrast, the median is a more robust measure of the central tendency of a distribution in the sense that it is not much affected by a few outliers.

The median m is the optimal solution for minimizing the mean absolute error, namely,

$$m = \arg\min_{a \in \mathbb{R}} \mathbb{E}|X - a|.$$

See CB Exercise 2.18

2. VaR

In the finance literature, when X is a return on the portfolio over a certain time period, with a small (e.g., $\alpha = 0.01$) is usually called the Value-at-Risk. Intuitively, $Q(\alpha)$ is the threshold level that actual loss will exceed with probability α. It has been used by Bank of International Settlements (BIS) to set the level of bank risk capital.

Example [J.P. Morgan RiskMetrics]: The value-at-risk (VaR) at level α, $Q_t(\alpha)$, of a portfolio over a certain time horizon is defined as

$$\Pr\left[X_t < Q_t(\alpha) | I_{t-1}\right] = \alpha,$$

where I_{t-1} is the information available at time $t-1$.It is the threshold which actual loss will exceed with probability α. This forms a basis for determining the suitable level of risk capital to prevent extreme adverse market events which would otherwise cause bankruptcy.

3. Other concept: Lower Partial Moments

$$LPM(k;c) = \mathbb{E}(X^k | X < c), \qquad k = 1, 2, \ldots$$

These lower partial moments describe the behavior of the left tail of the distribution of X, which is more closely associated with losses or risk.

Why is this a useful concept? The measure

$$\mathbb{E}\left[X_t | X_t < Q_t(\alpha), I_{t-1}\right]$$

is called the *expected shortfall* in financial risk management. This is the expected loss given that the "violation" has occurred.

4. Some Important Inequalities (CB 3.6, 4.7)

Theorem [Chebychev's Inequality]: Let X be a random variable with mean μ and variance σ^2, and let g be a strictly positive and increasing measurable real function on $(0, \infty)$ such that $\mathbb{E}g(X) < \infty$. Then we have several versions of Chebychev's inequality:

(1) For every constant $b > 0$,

$$\mathbb{E}g(x) \geq b \Pr(g(X) \geq b)$$

(2) For every constant $c > 0$,

$$\begin{aligned} \Pr(|X - \mu| \geq c) &\leq \sigma^2/c^2, \\ \Pr(|X - \mu| \leq c) &\geq 1 - \sigma^2/c^2. \end{aligned}$$

(3)

$$\begin{aligned} \Pr(|X - \mu| \geq k\sigma) &\leq 1/k^2, \\ \Pr(|X - \mu| \leq k\sigma) &\geq 1 - 1/k^2. \end{aligned}$$

$\square$

Proof. (1)

$$
\begin{aligned}
\mathbb{E}g(x) &= \int_{-\infty}^{\infty} g(x)dF(x) = \int_{-\infty}^{\infty} g(x)f(x)dx \\
&= \int_{-\infty}^{g^{-1}(b)} g(x)f(x)dx + \int_{g^{-1}(b)}^{\infty} g(x)f(x)dx \\
&\geq \int_{g^{-1}(b)}^{\infty} g(x)f(x)dx \quad [g^{-1}(b) \leq x < \infty \Longrightarrow b \leq g(x) < \infty] \\
&\geq \int_{g^{-1}(b)}^{\infty} bf(x)dx = b\int_{g^{-1}(b)}^{\infty} f(x)dx \\
&= b\Pr\left(X \geq g^{-1}(b)\right) = b\Pr\left(g\left(X\right) \geq b\right)
\end{aligned}
$$

(2) Let $g(X) = (X-\mu)^2$ and $b = c^2$. (3) Let $k = c/\sigma$.
Remark: (1) is also known as Markov inequality.

Theorem [Hölder's Inequality, CB 4.7.2]: Let X and Y be random variables and let p and q be constants such that $1 < p < \infty$, $1 < q < \infty$, $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$|\mathbb{E}XY| \le \mathbb{E}\,|XY| \le \{\mathbb{E}\,|X|^p\}^{1/p} \cdot \{\mathbb{E}\,|Y|^q\}^{1/q}.$$

□

Proof. The function $\ln(z)$ is concave. For every $a > 0, b > 0$, and p, q specified above, we have

$$\ln\left(\frac{1}{p}a + \frac{1}{q}b\right) \ge \frac{1}{p}\ln a + \frac{1}{q}\ln b.$$

(Draw a figure to see this.)

Consequently,

$$\left(\frac{1}{p}a + \frac{1}{q}b\right) \geq \exp\left(\frac{1}{p}\ln a + \frac{1}{q}\ln b\right) = a^{\frac{1}{p}} b^{\frac{1}{q}}.$$

Now let $a = \frac{|X|^p}{\mathbb{E}|X|^p}$ and $b = \frac{|Y|^q}{\mathbb{E}|Y|^q}$. Then

$$\begin{aligned}
\left(\frac{1}{p}\frac{|X|^p}{\mathbb{E}\,|X|^p} + \frac{1}{q}\frac{|Y|^q}{\mathbb{E}\,|Y|^q}\right) &\geq \exp\left(\frac{1}{p}\ln\frac{|X|^p}{\mathbb{E}\,|X|^p} + \frac{1}{q}\ln\frac{|Y|^q}{\mathbb{E}\,|Y|^q}\right) \\
&= \left(\frac{|X|^p}{\mathbb{E}\,|X|^p}\right)^{\frac{1}{p}}\left(\frac{|Y|^q}{\mathbb{E}\,|Y|^q}\right)^{\frac{1}{q}} \\
&= \frac{|X||Y|}{\{\mathbb{E}\,|X|^p\}^{1/p}\{\mathbb{E}\,|Y|^q\}^{1/q}}.
\end{aligned}$$

Taking mathematical expectations we get

$$\mathbb{E}\left(\frac{1}{p}\frac{|X|^p}{\mathbb{E}\,|X|^p}+\frac{1}{q}\frac{|Y|^q}{\mathbb{E}\,|Y|^q}\right)=1\geq\frac{\mathbb{E}|X||Y|}{\{\mathbb{E}\,|X|^p\}^{1/p}\{\mathbb{E}\,|Y|^q\}^{1/q}}$$

which proves the theorem. (Also see the proof in CB p.187)

Remark: For $p=q=2$, we get the Cauchy-Schwarz inequality (CB 4.7.3).

Theorem [Liapounov's Inequality]: Let X be a random variable. For $p \geq 1$ we have

$$\mathbb{E}\,|X| \leq \{\mathbb{E}\,|X|^p\}^{1/p}\,.$$

$\square$

Proof. Let $Y \equiv 1$ in the Hölder's Inequality.

Theorem [Jensen's Inequality, CB Theorem 4.7.7]: If $\varphi(\cdot)$ is a convex measurable real function on $\mathbb{R}$ and X is a random variable such that $\mathbb{E}|X| < \infty$, $\mathbb{E}\left|\varphi\left(X\right)\right| < \infty$, then $\varphi\left(\mathbb{E}X\right) \leq \mathbb{E}\varphi\left(X\right)$. $\square$
Proof. We prove only for discrete X. Let $x_j \in \mathbb{R}$, $j = 1, 2, \ldots, n$,

$$\Pr(X = x_j) = \lambda_j > 0, \qquad \sum_{j=1}^{n} \lambda_j = 1.$$

Then

$$\mathbb{E}X = \sum_{j=1}^{n} x_j \Pr(X = x_j) = \sum_{j=1}^{n} \lambda_j x_j.$$

Since $\varphi(\cdot)$ is convex

$$\varphi\left(\mathbb{E}X\right) = \varphi\left(\sum_{j=1}^{n} \lambda_j x_j\right) \leq \sum_{j=1}^{n} \lambda_j \varphi\left(x_j\right) = \mathbb{E}\varphi\left(X\right).$$

5. Moment Generating Function

CB Section 2.3

Moment Generating Function

When we are interested in various moments of X, the mathematical manipulations become quite involved as we calculate higher order moments. Fortunately, it is convenient to use the moment generating functions (mgf).
Note: This corresponds to the case of $g(X) = \exp(tX)$, where t is a real number.

Definition [CB 2.3.6]: The mgf of a r.v. X is defined as

$$M_X(t) = \mathbb{E}[\exp(tX)].$$

If the expectation exists for t in some neighborhood of 0, then we say that $M_X(t)$ exists for t in a small neighborhood of 0. That is, there exists some $\varepsilon > 0$ such that for all $t \in (-\varepsilon\ \varepsilon)$, $\mathbb{E}[\exp(tX)]$ exists. If the expectation does not exist for any small neighborhood of 0, then we say that $M_X(t)$ does not exist. $\square$

Question for you:

Question 1: Does the existence of the mgf $M_X(t)$ imply the existence of an infinite set of moments?

Answer: Yes, the existence of $M_X(t)$ for $t \in (-\varepsilon\ \varepsilon)$ implies that derivatives of $M_X(t)$ of all orders exist at $t = 0$. Why? Recall

$$e^{tx} = \sum_{k=0}^{\infty} \frac{(tx)^k}{k!}.$$

See Hamilton (1994, p. 716). Then the series expansion for $M_X(t)$ is

$$M_X(t) = \mathbb{E}[\exp(tX)] = \mathbb{E}\sum_{k=0}^{\infty} \frac{(tX)^k}{k!} = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathbb{E}\left(X^k\right).$$

If, for some k such that $\mathbb{E}(X^k) = \infty$; then $M_X(t)$ does not exist. Therefore, if $M_X(t)$ exists for $t \in (-\varepsilon\ \varepsilon)$, then all moments of X must exist. The expansion above thus provides a way to characterize all moments from the mgf uniquely.

Theorem (CB Theorem 2.3.7): If the moment generating function $M_X(t)$ exists for t in some neighborhood of 0, then for $k = 1, 2, \ldots$

$$\left.\frac{d^k M_X(t)}{dt^k}\right|_{t=0} = \mathbb{E}(X^k).$$

That is, the kth moment is equal to the kth derivative of $M_X(t)$ evaluated at $t = 0$.

Proof: Assuming that we can differentiate under the integral sign (see CB Section 2.4), we have

$$\begin{aligned} \frac{dM_X(t)}{dt} &= \frac{d}{dt}\int_{-\infty}^{\infty} e^{tX} f_X(x)dx \\ &= \int_{-\infty}^{\infty} \frac{d}{dt} e^{tX} f_X(x)dx \\ &= \int_{-\infty}^{\infty} x e^{tX} f_X(x)dx \\ &= \mathbb{E}(Xe^{tX}). \end{aligned}$$

Thus
$$\left.\frac{dM_X(t)}{dt}\right|_{t=0} = \mathbb{E}(Xe^{tX})\big|_{t=0} = \mathbb{E}(X).$$
Proceeding analogously,
$$\left.\frac{d^kM_X(t)}{dt^k}\right|_{t=0} = \mathbb{E}(X^ke^{tX})\Big|_{t=0} = \mathbb{E}(X^k).$$

Every moment of X can be computed from $M_X(t)$, provided $M_X(t)$ exists for $t \in (-\varepsilon\ \varepsilon)$.

CB Example 2.3.8 (Gamma mgf):

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \qquad 0 < x < \infty,\ \alpha > 0,\ \beta > 0$$

$$\begin{aligned} M_X(t) &= \int_0^\infty e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x(1-\beta t)} \\ &= \begin{cases} \left(\frac{1}{1-\beta t}\right)^\alpha & \text{if } t < \frac{1}{\beta} \\ \infty & \text{if } t \geq \frac{1}{\beta} \end{cases} \end{aligned}$$

See CB Example 2.3.8 for the derivation.
Note that if $t \geq 1/\beta$, the mgf does not exist because $\int_0^\infty x^{\alpha-1} e^{-x(1-\beta t)/\beta} dx = \infty$.
The mean of the gamma distribution is given by

$$\mathbb{E}(X) = \left.\frac{dM_X(t)}{dt}\right|_{t=0} = \left.\frac{\alpha\beta}{(1-\beta t)^{\alpha+1}}\right|_{t=0} = \alpha\beta.$$

Other moments can be calculated in a similar manner. Exercise: What is $Var(X)$?

Question for you:

Question 2: Does the existence of an infinite set of moments imply the existence of the mgf $M_X(t)$?

Answer: Yes and No.

1. No. It is possible that the mgf may not exist.
   1.1 The mgf is a weighted sum of all moments. If the mgf is finite, the all moments must be finite. This was Question 1.
   1.2 Even if all moments are finite, the sum, mgf, may not be finite. (Although all moments are finite, the mgf does not exist for a lognormal distribution. We will show this in Lecture 3. CB Exercise 2.36.)
2. Yes, if the distribution has a bounded support.

Another important role of the mgf

In addition to using the mgf to compute moments, we can also use the mgf to characterize the distribution

1. if the mgf exists [CB Theorem 2.3.11(b)], or
2. if the distribution has a bounded support [CB Theorem 2.3.11(a)].

We now turn to each of these cases.

Theorem [Uniqueness of mgfs, CB Theorem 2.3.11(b)]:
Consider two r.v. X and Y. Suppose their mgf's $M_X(t)$ and $M_Y(t)$ exist in a neighborhood of 0. Then X and Y have the same $M_X(t)$ and $M_Y(t)$ for all $t \in (-\varepsilon\ \varepsilon)$, if and only if $F_X(u) = F_Y(u)$ for all $u \in \mathbb{R}$.

Remark: The theorem says that given some $M_X(t)$, suppose we can find some distribution $F_X(x)$ that corresponds to $M_X(t)$. Then $F_X(x)$ must be the only distribution that generates $M_X(t)$.

Remark: The proof relies on the theory of Laplace transforms (see, e.g., Feller 1971). The defining equation of $M_X(t)$, when pmf or pdf $f_X(x)$ exists, is given by

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx,$$

which defines a Laplace transform. A key fact about Laplace transforms is their uniqueness. If $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx$ is valid for all t with $|t| < h$, where h is some positive number, then given $M_X(t)$, there is only one function $f_X(x)$ that satisfies the Laplace transform. Given this fact, the theorem is reasonable. However, the formal proof is rather technical and provides no additional insight. See CB 2.6.1. p. 66.

Question 3: We have seen that if the mgf $M_X(t)$ exists in a neighborhood of 0, it characterizes an infinite set of moments. Thus, a natural question is whether we can use the infinite set of moments to uniquely characterize a distribution function? In other words, suppose $\mathbb{E}(X^r) = \mathbb{E}(Y^r)$ for all $r > 0$, are X and Y identically distributed?

Answer: No in general. Characterizing the set of moments is not enough to determine a distribution uniquely because there may be two distinct random variables having the same moments. Example: See CB Example 2.3.10.

Answer: Yes for some special cases. The set of infinite moments cannot uniquely characterize a distribution. However, there exists a special case in which we can use moments to characterize the distribution. This arises when the random variables have bounded support (see, e.g., Billingsley 1995, Section 30). CB p. 65. See the next Theorem.

Theorem [CB 2.3.11(a)]: Let $F_X(x)$ and $F_Y(y)$ be two cdfs both of which have bounded support. Then $F_X(x) = F_Y(x)$ for all x if and only if $\mathbb{E}X^r = \mathbb{E}Y^r$ for all integers $r = 0, 1, 2, \ldots$.

Summary remarks: The existence of all moments is not equivalent to the existence of the moment generating function. Intuitively, the mgf is a weighted sum of all moments. If the mgf is finite, then all moments must be finite. If some moment does not exist, then the mgf does not exist. It is possible that all moments are finite but their weighted sum is infinite (i.e., the mgf does not exist). An example is the log-normal random variable.

In Section 7, we will provide a solution to this difficulty, by introducing the characteristic function, which maintains the nice features of the mgf and exists for all distributions (no matter whether moments exist).

Next, we consider a convergence theorem for mgfs that is very useful in investigating the asymptotic or limiting behavior of some distributions.

Convergence of mgfs

Theorem [Convergence of mgfs; CB Theorem 2.3.12]:
Suppose $\{X_n, n = 1, 2, \ldots\}$ is a sequence of random variables, each with mgf $M_n(t)$ and cdf $F_n(x)$. Furthermore, suppose that

$$\lim_{n\to\infty} M_n(t) = M_X(t)$$

for all t in a neighborhood of 0 and $M_X(t)$ is a mgf of some random variable X with cdf $F_X(x)$. Then

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

for all x where $F_X(x)$ is continuous.

Remarks:

- Let $\{X_n\}$ be a sequence of r.v.'s with distribution functions $\{F_n(x)\}$. Let X be a r.v. with F_X. If $F_n(x) \to F_X(x)$ as $n \to \infty$ for all continuity points (i.e., all x's where $F_X(x)$ is continuous), then we say X_n converges in distribution to the r.v. X, denoted $X_n \xrightarrow{d} X$.
- Implications: F_n and M_n are generally unknown, but if we can show they converge to some well-known limits F_X and M_X, then we can use these well-known limits as the approximations to the unknown F_n and M_n.

Example (mgf of binomial distribution, CB Example 2.3.9):

$$\Pr(X = x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n,\ 0 \le p \le 1.$$

$$\begin{aligned} M_X(t) &= \sum_{x=0}^{n} e^{tx} \left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \left(\begin{array}{c} n \\ x \end{array} \right) \left(pe^t\right)^x (1-p)^{n-x} \\ &= \left[pe^t + (1-p)\right]^n \end{aligned}$$

Example (mgf of Poisson distribution, CB Example 2.3.13):

$$\Pr(Y = y) = \frac{e^{-\lambda}\lambda^y}{y!}, \qquad y = 0, 1, 2\ldots,\ \lambda > 0.$$

$$M_Y(t) = \sum_{y=0}^{\infty} e^{ty}\frac{e^{-\lambda}\lambda^y}{y!} = e^{-\lambda}\sum_{y=0}^{\infty}\frac{\left(e^t\lambda\right)^y}{y!} = e^{-\lambda}e^{\left(e^t\lambda\right)} = e^{\lambda(e^t-1)}.$$

Recall the series expansion (Hamilton 1994, p. 716):

$$\begin{aligned}\exp(a) &= \sum_{k=0}^{\infty}\frac{a^k}{k!}\\ \exp(e^t\lambda) &= \sum_{y=0}^{\infty}\frac{(e^t\lambda)^y}{y!}.\end{aligned}$$

Question for you:

Poisson approximation of binomial distribution (CB Example 2.3.13)

Let $np = \lambda$. Then

$$M_X(t) \to M_Y(t), \qquad \text{as } n \to \infty.$$

Proof:

$$\begin{aligned} M_X(t) &= \left[pe^t + (1-p)\right]^n \\ &= \left[1 + p(e^t - 1)\right]^n \\ &= \left[1 + \frac{1}{n}(e^t - 1)(np)\right]^n \\ &= \left[1 + \frac{1}{n}(e^t - 1)\lambda\right]^n \\ &\to e^{\lambda(e^t - 1)} = M_Y(t). \end{aligned}$$

Remark: See CB Figure 2.3.3.

Theorem [CB 2.3.15]: For any constants a and b, the mgf of the random variable $aX + b$ is given by

$$M_{aX+b}(t) = e^{bt} M_X(at).$$

Proof: CB page 68.

6. Cumulant Generating Function

CB Section 2.6

Cumulant Generating Function

Definition: Let the mgf of a r.v. X is defined as $M_X(t) = E[\exp(tX)]$. Then the cumulant generating function is the function

$$S(t) = \log M_X(t).$$

This function generates the cumulants of X, which are defined as the coefficients in the Taylor series of the cumulant generating function.

Example (CB Exercise 2.32):

$$\begin{aligned} \frac{d}{dt} S(t)\Big|_{t=0} &= \mathbb{E}X \\ \frac{d^2}{dt^2} S(t)\Big|_{t=0} &= VarX \end{aligned}$$

Remark (Advanced): It is particularly useful for the Edgeworth expansion, e.g., in the analysis of bootstrap. See Hall, Peter. (1992, Chapter 2) *The Bootstrap and Edgeworth Expansion.*

7. Characteristic Function

CB Section 2.6

Characteristic Function

Motivation: For some distributions (e.g., Cauchy distributions), the mgf does not exist.
As the name suggests, characteristic functions characterize distributions in the sense that two distributions are equal if and only if their characteristic functions are equal.

Definition [Characteristic Function]: The characteristic function of a r.v. X with cdf $F_X(x)$ is defined as

$$\varphi_X(t) = \mathbb{E}e^{itX} = \int_{-\infty}^{\infty} e^{itX} dF_X(x),$$

where $i = \sqrt{-1}$. $\square$

Properties of Characteristic Function:

- The characteristic function is bounded, that is,

$$|\varphi_X(t)| < \varphi_X(0) < 1 \qquad \text{for } -\infty < t < \infty.$$

  Question for you: Show this.

- The characteristic function $\varphi_X(t)$ is continuous over $(-\infty\ \infty)$.
- $\varphi_X(-t) = \varphi_X^*(t)$, where $\varphi_X^*(t)$ denotes the complex conjugate of $\varphi_X(t)$.
- Suppose $Y = aX + b$, where a and b are any real number, then

$$\varphi_Y(t) = e^{itb}\varphi_X(at).$$

- If $X_1, X_2, \ldots, X_n$ are mutually independent, then

$$\varphi_{X_1+X_2+\cdots+X_n}(t) = \varphi_{X_1}(t)\cdots\varphi_{X_n}(t).$$

Remarks:
(i) $\varphi_X(t) = M_X(it)$.
(ii) $\varphi_X(t)$ is more general than $M_X(t)$ in the sense that $\varphi_X(t)$ always exists for all distributions and all t while $M_X(t)$ may not exist for some distributions or for some t. Why?

$$\begin{aligned} |\varphi_X(t)| &= \left| \int_{-\infty}^{\infty} e^{itX} dF_X(x) \right| \\ &\le \int_{-\infty}^{\infty} \left| e^{itX} \right| dF_X(x) = \int dF_X(x) = 1 < \infty, \end{aligned}$$

because

$$\begin{aligned} e^{itx} &= \cos(tx) + i \sin(tx), \\ \left| e^{itx} \right|^2 &= \cos^2(tx) + \sin^2(tx) = 1. \end{aligned}$$

Theorem [Uniqueness of characteristic function]: Suppose two r.v. X and Y with $\varphi_X(t)$ and $\varphi_Y(t)$. Then X and Y are identically distributed if and only if $\varphi_X(t) = \varphi_Y(t)$ for all $t \in \mathbb{R}$. $\square$

Remark. If F_1 and F_2 are distributions with the same characteristic functions, then F_1 and F_2 are equal. Thus to each characteristic function corresponds to a unique distribution.

Theorem [Convergence in characteristic functions]: Let $\{X_n\}$ be a sequence of r.v.'s with distribution functions $F_n(x)$ and characteristic functions $\{\varphi_n(t)\}$. Let X be a random variable with distribution function $F_X(x)$ and characteristic function $\varphi_X(t)$. Let $n \to \infty$.
(i) If $F_n(x) \to F_X(x)$ for all continuity points $x \in \mathbb{R}$, then for every $t \in \mathbb{R}$, $\varphi_n(t) \to \varphi_X(t)$.
(ii) Further, if for every $t \in \mathbb{R}, \varphi_n(t) \to \varphi_X(t)$ and $\varphi_X(t)$ is continuous at $t = 0$, then $F_n(x) \to F_X(x)$ for all continuity points $x \in \mathbb{R}$.

Remark: Convergence in distribution is equivalent to convergence of characteristic functions. We use the above theorem to prove a central limit theorem. The characteristic function is useful because sometimes it is more convenient to study the behaviors of characteristic functions $\varphi_n(t)$ than distribution functions. For example, in finance, a general class of popular continuous-time models called affine jump diffusion models (see, e.g., Duffie, Pan and Singleton 2000) is well known for not having a closed form for its (conditional) probability density function, which makes the maximum likelihood estimation unfeasible. However, this class of models has a closed form expression for its (conditional) characteristic function. Thus, inference and estimation are based on the characteristic function. For another example, the popular affine term structure model has a closed form for the conditional characteristic function, but has no closed form of the conditional transition density function.

Characteristic functions are convenient for deriving the moments of a distribution. Suppose that $\varphi(t)$ is m times differentiable and that

$$\int |x|^m dF(x) < \infty.$$

See CB Theorem 2.4.3. Then, we may each time differentiate under the integral (See the next section). Thus:

$$\begin{aligned} \varphi^{(1)}(t) &\equiv \frac{d}{dt}\varphi(t) = \int \frac{d}{dt} e^{itx} dF(x) = \int ixe^{itx} dF(x) \\ &\Longrightarrow \varphi^{(1)}(0) = i \int x dF(x) \end{aligned}$$

$$\begin{aligned}
\varphi^{(2)}(t) &\equiv \frac{d^2}{dt^2}\varphi(t) = \int i^2 x^2 e^{itx} dF(x) \\
&\implies \varphi^{(2)}(0) = i^2 \int x^2 dF(x)
\end{aligned}$$

$$\vdots$$

$$\begin{aligned}
\varphi^{(m)}(t) &\equiv \frac{d^m}{dt^m}\varphi(t) = \int i^m x^m e^{itx} dF(x) \\
&\implies \varphi^{(m)}(0) = i^m \int x^m dF(x)
\end{aligned}$$

Thus:

Theorem: Let $\int |x|^m \, dF(x) < \infty$ and $\varphi^{(m)}(t)$ be the m-th derivative of $\varphi(t)$. Then

$$\frac{\varphi^{(m)}(0)}{i^m} = \int x^m dF(x) = \mathbb{E}x^m.$$

□

Remark:

$$\begin{aligned} \mathbb{E}x^1 &= -i\varphi^{(1)}(0) \\ \mathbb{E}x^2 &= -\varphi^{(2)}(0) \\ \mathbb{E}x^3 &= i\varphi^{(3)}(0) \\ &\vdots \end{aligned}$$

8. Appendix: Differentiating Under an Integral Sign

CB Section 2.4

Differentiating Under an Integral Sign

Why do we need to care?

Example: Characteristic functions are convenient for deriving the moments of a distribution. Under what conditions the following is true?

$$
\begin{aligned}
\frac{d}{dt}\varphi(t) &= \frac{d}{dt}\int_{-\infty}^{\infty} e^{itx} dF(x) \\
&= \int_{-\infty}^{\infty} \frac{\partial}{\partial t} e^{itx} dF(x) = \int_{-\infty}^{\infty} ixe^{itx} dF(x)
\end{aligned}
$$

$$
\begin{aligned}
\frac{d^m}{dt^m}\varphi(t) &= \frac{d^m}{dt^m}\int_{-\infty}^{\infty} e^{itx} dF(x) \\
&= \int_{-\infty}^{\infty} \frac{\partial^m}{\partial t^m} e^{itx} dF(x) = \int i^m x^m e^{itx} dF(x)
\end{aligned}
$$

- Suppose that (i) $\varphi(t)$ is m times differentiable, and also suppose that (ii)

$$\int |x|^m dF(x) < \infty. \tag{1}$$

  Then, we may differentiate under the integral (for each $m = 1, 2, \ldots$). Why?

- Note that the conditions (i) and (ii) are the two conditions of CB Theorem 2.4.3. The second condition (ii) in (1) is to have **the Lebesque's Dominated Convergence Theorem**.

Under (ii)

$$\int_{-\infty}^{\infty} x^m dF(x) < \int |x|^m dF(x) < \infty,$$

$$\begin{aligned}
\left.\frac{d^m}{dt^m}\varphi(t)\right|_{t=0} &= \left.\frac{d^m}{dt^m}\int_{-\infty}^{\infty} e^{itx} dF(x)\right|_{t=0} \\
&= \left.\int_{-\infty}^{\infty} \frac{\partial^m}{\partial t^m} e^{itx} dF(x)\right|_{t=0} \\
&= \left.\int_{-\infty}^{\infty} i^m x^m e^{itx} dF(x)\right|_{t=0} \\
&= i^m \int_{-\infty}^{\infty} x^m dF(x)
\end{aligned}$$

Two cases when we may interchange the order of integration and differentiation.

Case 1. When we have the integral of a differentiable function over a finite range (CB Theorem 2.4.1).

Case 2. If we have the integral of a differentiable function over a infinite range, it is when the Lebesque's Dominated Convergence Theorem holds (CB Theorems 2.4.2, 2.4.3, 2.4.4).

Case 1. When we have the integral of a differentiable function over a finite range.

If we have the integral of a differentiable function over a **finite** range, then

$$\frac{d}{d\theta}\int_a^b f(x,\theta)dx = \int_a^b \frac{\partial}{\partial\theta} f(x,\theta)dx.$$

(Why? It is clear from the Leibnitz's Rule in CB Theorem 2.4.1.)

Case 2. If we have the integral of a differentiable function over a infinite range, it is when the Lebesque's Dominated Convergence Theorem holds.

To understand the conditions under which the following may hold

$$\frac{d}{d\theta}\int_{-\infty}^{\infty} f(x,\theta)dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(x,\theta)dx,$$

recall what the differentiation is

$$\frac{\partial}{\partial\theta} f(x,\theta) := \lim_{\delta\to 0} \frac{f(x,\theta+\delta)-f(x,\theta)}{\delta}.$$

Therefore,

$$\begin{aligned}
\frac{d}{d\theta}\int_{-\infty}^{\infty} f(x,\theta)dx &= \lim_{\delta\to 0}\int_{-\infty}^{\infty}\left[\frac{f(x,\theta+\delta)-f(x,\theta)}{\delta}\right]dx \\
\int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f(x,\theta)dx &= \int_{-\infty}^{\infty}\lim_{\delta\to 0}\left[\frac{f(x,\theta+\delta)-f(x,\theta)}{\delta}\right]dx
\end{aligned}$$

Thus, the question of whether interchanging the order of differentiation and integration

$$\frac{d}{d\theta}\int_{-\infty}^{\infty} f(x,\theta)dx = \int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}f(x,\theta)dx$$

is really a question of whether limits and integration can be interchanged

$$\lim_{\delta\to 0}\int_{-\infty}^{\infty}\left[\frac{f(x,\theta+\delta)-f(x,\theta)}{\delta}\right]dx = \int_{-\infty}^{\infty}\lim_{\delta\to 0}\left[\frac{f(x,\theta+\delta)-f(x,\theta)}{\delta}\right]d$$

Limits of Expectations and Expectations of Limits

Theorem [Dominated Convergence Theorem]: If $|g_n(x)| \le G(x)$ for $n \ge 1$, $\int_{-\infty}^{\infty} G(x)dF(x) < \infty$, and $\lim_{n\to\infty} g_n(x) = g(x)$ except for $x \in N$, where N is a set with probability zero, then

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} g_n(x)dF(x) = \int_{-\infty}^{\infty} \lim_{n\to\infty} g_n(x)dF(x).$$

Remark: CB Theorems 2.4.2, 2.4.3, 2.4.4 are all corollaries of the Dominated Convergence Theorem.

9. Conclusions

Conclusions

In this chapter, we introduce the concept of random variables. To describe the probability measure of a random variable, we introduce the cdf. Depending on whether a random variable is discrete or continuous, we introduce the pmf and pdf. Methods to derive the probability distribution of a transformation of a random variable are also developed.

We define a class of moments, which is a summary characteristic of a probability distribution. Economic intuitions and applications for some moments are provided. We also discuss the moment generating function and the characteristic function. Both of them can be differentiated to generate moments (when moments exist), and more importantly, both of them can be used to uniquely characterize the probability distribution of a random variable.