§ Random Variable (R.V.)

We often characterize the outcome of a random experiment by a numerical measurement.

*eg.* No. of heads in *n* coin tosses

**Key idea:** Each outcome of the random experiment produces a number.

**Intuitive Definition:** Given (*S*, *F*, *P*), a random variable is a mapping from *S* to the real line:

\[ \mathbf{x} : \mathcal{S} \rightarrow \mathbb{R} \]

To each possible outcome, the R.V. assigns a number.

**N.b.** \( \mathbf{x}(*) \) is fixed (not random). It is the selection of \( w \); that is random.
The mapping $\mathbb{X}(\cdot): \mathcal{S} \rightarrow \mathbb{R}$ is not random, but $\mathbb{X}(w)$ takes on a random value because of the randomness of outcome $w$.

Formal Definition: Given $(\mathcal{S}, \mathcal{F}, P)$, a random variable is a mapping $\mathbb{X}: \mathcal{S} \rightarrow \mathbb{R}$ with the property that if $A \in B(\mathbb{R})$, then $\mathbb{X}^{-1}(A) = \{w \in \mathcal{S} : \mathbb{X}(w) \in A\} \in \mathcal{F}$.

A function $\mathbb{X}: \mathcal{S} \rightarrow \mathbb{R}$ meeting this condition is called a measurable function from $(\mathcal{S}, \mathcal{F})$ to $\mathbb{R}$.

Such functions $\mathbb{X}: \mathcal{S} \rightarrow \mathbb{R}$ have the property that their values (outputs) inherit a prob. measure from the prob. measure of $(\mathcal{S}, \mathcal{F}, P)$.
Random Exp.

\((\mathcal{S}, F, \mathbb{P}) \xrightarrow{\mathbb{X}(\cdot)} (\mathbb{R}, B(\mathbb{R}), \mathbb{P}_x)\)

Measurability of \(\mathbb{X}\) insures that for any \(A \in B(\mathbb{R})\), \(\mathbb{X}^{-1}(A) \in F\), so that \(\mathbb{P}_x(A) = \mathbb{P}(\mathbb{X}^{-1}(A))\).

If \((\mathcal{S}, F, \mathbb{P})\) is chosen with \(\mathcal{S} = \mathbb{R}\), and \(F = B(\mathbb{R})\), then almost all functions \(\mathbb{X} : \mathbb{R} \rightarrow \mathbb{R}\) we will normally encounter are measurable and hence, defines RVs. These include:

1. Continuous ftns
2. polynomials
3. Step functions
4. Indicated ftns \(1_A(\cdot)\)
5. Trig. Function
6. Limits of sequences of measurable ftns
7. Sums (countable) of measurable ftns

In fact, it is difficult to construct a non-measurable function mapping \(\mathcal{S}\) to \(\mathbb{R}\) when we take

\((\mathcal{S}, F, \mathbb{P}) = (\mathbb{R}, B(\mathbb{R}), \mathbb{P})\)
- Given $(S, F, P)$ and a R.V. $X: S \rightarrow A \subset \mathbb{R}$

  Because $X$ is a R.V. (i.e. measurable), all subsets of the form $X^{-1}(F) = \{ w \in S : X(w) \in F \}$, $F \in B(A)$ must be in $\mathcal{F}$ of $(S, F, P)$. So we can compute their prob. and define

  $$P_X(F) = P(X^{-1}(F)) = P(\{ w \in S : X(w) \in F \})$$

  for all $F \in B(A)$

  *n.b. $F \neq \emptyset$*

  Then $P_X(\cdot)$ is a valid prob. measure define on $B(A)$, and $(A, B(A), P_X)$ is itself a valid prob. space

  i.e. $P_X: B(A) \rightarrow \mathbb{R}$ satisfies axioms of prob.

  \[ X: S \rightarrow A \subset \mathbb{R} \]

  \[ (S, F, P) \quad \Rightarrow \quad (A, B(A), P_X) \]

  a valid prob. space.

  *n.b. We already know how to deal with prob. space. We can get $P_X$ from $(S, F, P)$. But often it's easiest to describe directly. We can specify $P_X$ using pdf or pmf.*
We can think of the value a RV takes on as the outcome of a random exp. with \((A, B(A), P_x)\)

So why bother w/ mapping in the first place?

1. Insight
2. Can define multiple RVs on single random Exp \((\mathcal{S}, \mathcal{F}, P)\)
3. Generalization to random "objects" e.g. random process.

- We can think of RVs two ways:

  1. **Mathematically**: A "nice" (meas.) ftn mapping \(\mathcal{S}\) of \((\mathcal{S}, \mathcal{F}, P)\) to \(\mathbb{R}\)

  2. **Intuitively**: Something that takes on real values at random

     \(\Rightarrow\) The outcome of a random experiment w/ prob. space

        \((A, B(A), P_x), A \subseteq \mathcal{R}\).

- When we do computations: Concentrate \((A, B(A), P_x)\)
  
  rather than \((\mathcal{S}, \mathcal{F}, P)\). This assume we know \(P_x\).
  
  We can get \(P_x\) from \((\mathcal{S}, \mathcal{F}, P)\), but often its easiest
  
  to describe \(P_x\) directly

- Given \((A, B(A)), A \subseteq \mathcal{R}\), we can specify \(P_x\) using pmf
  
  or pdf \(\Rightarrow (A, B(A), P_x)\)
CDF

Def. Given a RV $X$ on $(\mathcal{S}, \mathcal{F}, P)$ and having prob. measure $P_x$ on $(A, B(A), \mathbb{P})$, the cumulative distribution function (CDF) of $X$ is

$$F_x(\alpha) = P_x((\mathbb{R}, \mathbb{R}_x]$$

$$= P_x(\{x: x \leq \alpha\}), \alpha \in \mathbb{R}$$

$$= P(\{w: X(w) \leq \alpha\})$$

Notation: We write $\{X \leq \alpha\}$ to describe the event

$$\{w: X(w) \leq \alpha\} \in \mathcal{F}$$

So $\{X \leq \alpha\} = \{w: X(w) \leq \alpha\} \in \mathcal{F}$ in $(\mathcal{S}, \mathcal{F}, P)$

or $\{X \leq \alpha\} = \{x: x \leq \alpha\} \in B(A)$ in $(A, B(A), \mathbb{P})$

$F_x(\cdot)$ specifies $P_x(\cdot)$:

Fact: It is possible to construct any set $F \in B(\mathbb{R})$ from a countable sequence of unions and complements (and intersection) of intervals of the form $(-\infty, x_n]$

Hence, the CDF is sufficient to specify $P_x$ on $(A, B(A), \mathbb{P})$ for any $A \in B(\mathbb{R})$ [We often take $A = \mathbb{R}$]

* We will always specify $F_x(\alpha), \forall \alpha \in \mathbb{R}$, regardless of range space $A$. 
Property of CDF: \( F_X(x) = P(\{X \leq x\}) \)

1. \( F_X(\pm \infty) = 1 \) and \( F_X(\mp \infty) = 0 \)

2. If \( x_1 < x_2 \), \( F_X(x_1) \leq F_X(x_2) \)

3. \( P(\{X > x\}) = 1 - F_X(x) = 1 - P_X(\{X \leq x\}) \)

4. If \( x_1 < x_2 \), \( P(\{x_1 < X \leq x_2\}) = F_X(x_2) - F_X(x_1) \)

5. \( P(\{X = x_0\}) = F_X(x_0) - F_X(x_0^-) \)

where \( F_X(x^-) = \lim_{\varepsilon \to 0} F_X(x_0 - \varepsilon) \)

- We say that a R.V. \( X \) is (absolutely) **continuous** if \( F_X(x) \) is a continuous fin of \( x \) for all \( x \in \mathbb{R} \).

- We say \( X \) is **discrete** if \( F_X(x) \) is a staircase fin.
• We say \( X \) is mixed if \( F_X(x) \) has discontinuities, but is not a staircase fn.

![Graph of \( F_X(x) \)]

**PDF**

**Defn:** The probability density function (pdf) of R.V. \( X \) is defined as the derivative of its CDF:

\[
f_X(x) = \frac{d}{dx} F_X(x)
\]

**n.b.**

\[
f_X(x^-) = \frac{d}{dx} F_X(x^-) = \lim_{\Delta x \to 0} \frac{F_X(x^- + \Delta x) - F_X(x^-)}{\Delta x}
\]

It does not work if CDF has jumps.!!

**n.b.** We will broaden the defn. of derivative to include Dirac delta functions as the derivatives of step discontinuities

\[
\delta(x) = 0, \forall x \neq 0, \text{ and } \int_{-\infty}^{\infty} \delta(x) dx = \int_{-\infty}^{\infty} \delta(x) dx = 1.
\]
**Example:** Roll a fair die. Consider the R.V. $X$ to be the numerical outcome

$$F_X(x) = \frac{1}{6}1_{[1, \infty)}(x) + \frac{1}{6}1_{[2, \infty)}(x) + \cdots + \frac{1}{6}1_{[6, \infty)}(x)$$

$$f_X(x) = \frac{dF_X(x)}{dx} = \frac{1}{6}\delta(x-1) + \frac{1}{6}\delta(x-2) + \cdots + \frac{1}{6}\delta(x-6)$$

**Properties of PDF:**

1. $f_X(x) \geq 0$, $\forall x \in \mathbb{R}$
2. $F_X(x) = \int_{-\infty}^{x} f_X(\omega) d\omega$
3. $\int_{-\infty}^{\infty} f_X(\omega) d\omega = 1$ ($F_X(+\infty) - F_X(-\infty) = 1 - 0 = 1$)
4. $P(\{X_1 < X \leq X_2\}) = \int_{x_1}^{x_2} f_X(\omega) d\omega = F_X(x_2) - F_X(x_1)$

We will often describe a R.V. $X$ by giving its pdf or cdf, ignoring the underlying $(S, F, P)$. But remember, $(S, F, P)$ is there!
Ex. 1. A R.V. $Z$ is called Gaussian if it has pdf

$$f_Z(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad \mu \in \mathbb{R}, \quad \sigma > 0$$

![Gaussian PDF](image)

**Note:**

$$F_Z(x) = \int_{-\infty}^{x} f_Z(t)\,dt = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$\Phi(*)$ cannot be written in closed form, but can be numerically calculated and is widely tabulated.

$$\mathbb{P}(\{a < Z < b\}) = F_Z(b) - F_Z(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

We often write $Z \sim N(\mu, \sigma)$ to describe a R.V. with Gaussian distribution.

Ex. 2 (Uniform Dist.) R.V. $Z$ is uniformly distributed on Ex. $[x_1, x_2]$ if it has pdf

$$f_Z(x) = \frac{1}{x_2-x_1} 1_{[x_1, x_2]}(x)$$

$$F_Z(x) = \frac{x-x_1}{x_2-x_1} 1_{[x_1, x_2]}(x) + 1_{(x_2, \infty)}(x)$$

![Uniform PDF](image)

We often write $Z \sim U[x_1, x_2]$ to describe R.V.s with uniform distribution.
\textbf{Ex. 3. (Binomial Distributed R.V.)}

Discrete R.V. taking on values \{0, 1, 2, \ldots, n\} \subset \mathbb{N} with pmf

\[ P(k) = \binom{n}{k} p^k (1-p)^{n-k}. \]

The cdf of this R.V. is

\[ F_X(x) = P(X \leq x) = \sum_{k=0}^{m} \binom{n}{k} p^k (1-p)^{n-k}. \]

pick \( m \) s.t. \( m \leq x < m+1 \), \( m \in \{0, 1, 2, \ldots, n\} \)

\[ F_X(x) \text{ is a staircase fn that jumps up by height } P(k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ at } x = k, \text{ for } k = 0, 1, \ldots, n \]

\[ \therefore f_X(x) = \frac{dF_X(x)}{dx} = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta(x-k) \]

(Other prob. distribution will be discussed in a separate topic.)
• **Conditional Distributions**:

Given $(S, F, P)$ with $X$ defined on it. Given $A, M \in F$, we know:

$$P(A|IM) = \frac{P(ANM)}{P(M)}$$

Now take $A = \{X \leq x\} = \{w: X(w) \leq x\}$

Then we have conditional CDF of $X$ conditioned on $M$:

$$F_X(x|M) = P(\{X \leq x\}|M) = \frac{P(\{X \leq x\} \cap M)}{P(M)}$$

• **Defn of $F_X(x|M)$ is just like $F_X(x)$, except with**.
  
  prob. measurement $P(\cdot)$ replaced by prob. measure $P(\cdot|M)$

  Because $P(\cdot|M)$ is a valid prob. measure, $F_X(x|M)$ is a valid CDF with all assoc. properties.

**Defn** The conditional density function (continual pdf) of $X$ conditioned on $M \in F$ is

$$f_X(x|M) = \frac{dF_X(x|M)}{dx}$$

**N.b.** Because $F_X(x|M)$ is a valid cdf, $f_X(x|M)$ is a valid pdf having all properties of a pdf.
In general, we must know the underlying structure of the random experiment to determine \( F_x(x|\mathcal{M}) \) or \( f_x(x|\mathcal{M}) \). But sometimes we can define \( \mathcal{M} \in \mathcal{F} \) in terms of the R.V. \( X \).

\[ e.g.1. M = \{ X \leq a \} \]

\[ 2. M = \{ b \leq X \leq a \}, \quad b < a \]

1. Let \( M = \{ X \leq a \} \)

\[
F_x(x|M) = \frac{P(\{ X \leq x \} \cap \{ X \leq a \})}{P(\{ X \leq a \})}
\]

Two cases:

(a) if \( x > a \) : \( \{ X \leq x \} \cap \{ X \leq a \} = \{ X \leq a \} \)

\[ \Rightarrow F_x(x|M) = 1. \]

(b) if \( x \leq a \) : \( \{ X \leq x \} \cap \{ X \leq a \} = \{ X \leq x \} \)

\[ \Rightarrow F_x(x|M) = \frac{F_x(x)}{F_x(a)} \]

\[ \therefore F_x(x|\{ X \leq a \}) = \begin{cases} \frac{F_x(x)}{F_x(a)}, & x \leq a \\ 1, & x > a \end{cases} \]
The conditional pdf of $Z$ cond. on $\{X \leq a\}$ is

$$f_Z(x | \{X \leq a\}) = \frac{dF_Z(x | \{X \leq a\})}{dx}$$

$$= \left\{ \begin{array} {cl}
\frac{f_Z(x)}{F_Z(a)}, & x \leq a \\
0, & x > a
\end{array} \right.$$  

2. Now consider $M = \{ b < X \leq a \}, \quad a > b$.

$$F_Z(x | M) = \frac{P(\{X \leq x\} \cap \{b < X \leq a\})}{P(\{b < X \leq a\})}$$

Three distinct regions:

(a) $x > a$ (b) $b < x \leq a$ (c) $x \leq b$

Analyzing, we get

$$F_Z(x | \{b < X \leq a\}) = \left\{ \begin{array} {cl}
1 & x > a \\
\frac{F_Z(x) - F_Z(b)}{F_Z(a) - F_Z(b)} & b < x \leq a \\
0 & x \leq b
\end{array} \right.$$
The corresponding pdf is
\[ f_X(x|\{b < x \leq a\}) = \frac{f_X(x)}{F_X(a) - F_X(b)} 1_{(b,a]} \]

Total Prob. Law

Given R.V. \( X \) on \((S,F,P)\), let \( \{A_1, A_2, \ldots, A_n\} \) be a partition of \( S \).

\[ P(\{ X \leq x \}) = P(\{ X \leq x | A_1 \} P(A_1) + P(\{ X \leq x | A_2 \} P(A_2) + \cdots \\
\cdots + P(\{ X \leq x | A_n \} P(A_n) \)

\[ F_X(x) = F_X(x|A_1)P(A_1) + F_X(x|A_2)P(A_2) + \cdots \]

\[ f_X(x) = \frac{dF_X(x)}{dx} = f_X(x|A_1)P(A_1) + f_X(x|A_2)P(A_2) + \cdots \]
Bayes' Formula

1. \( P(A | \{ \mathbf{x} \leq x \} ) = \frac{F_x(x \mid A) \cdot P(A)}{F_x(x)} \)

2. \( P(A | \{ x_1 < \mathbf{x} \leq x_2 \} ) = \frac{F_x(x_2 \mid A) - F_x(x_1 \mid A)}{F_x(x_2) - F_x(x_1 \mid A)} \cdot P(A) \)

3. \( P(A | \{ \mathbf{x} = x \} ) \) (limit case of 2)

\[
P(A | \mathbf{x} = x) = \frac{P(\{\mathbf{x} = x\} \mid A) \cdot P(A)}{P(\{\mathbf{x} = x\})} = \frac{0}{0}
\]

If \( \mathbf{x} \) is continuous, \( P(\{\mathbf{x} = x\}) = 0 \)

So \( P(A | \mathbf{x} = x) \) is indeterminate, but clearly \( P(A | \mathbf{x} = x) \) makes sense. How can we define it?

**Approach:** Consider \( \lim_{\Delta x \to 0} \)

\[
\mathbf{x} \to (x) P(A | \{ x < \mathbf{x} \leq x + \Delta x \})
\]

\[
= \lim_{\Delta x \to 0} \left[ \frac{F_x(x + \Delta x \mid A) - F_x(x \mid A)}{F_x(x + \Delta x) - F_x(x)} \right] \cdot P(A)
\]

\[
= \lim_{\Delta x \to 0} \left[ \frac{F_x(x + \Delta x \mid A) - F_x(x \mid A)}{\Delta x} \right] P(A) = \frac{f_x(x \mid A)}{f_x(x)} \cdot P(A)
\]

\[
\therefore \quad P(A | \{ \mathbf{x} = x \} ) = \frac{f_x(x \mid A)}{f_x(x)} \cdot P(A)
\]
**Bayes' Theorem**

From above,

\[ P(A | \{Z = x \}) = \frac{f_Z(x | A)}{f_Z(x)} P(A) \]

and the total probability:

\[ P(A) = \int_{-\infty}^{\infty} P(A | Z = \alpha) f_Z(\alpha) d\alpha. \]

\[ f_Z(x | A) = \frac{P(A | \{Z = x \}) f_Z(x)}{\int_{-\infty}^{\infty} P(A | \{Z = \alpha \}) f_Z(\alpha) d\alpha} \]

This is the continuous version of Bayes' Theorem.

**Functions of a R.V.**

Assume \( Z \) is a R.V. on \( (\mathcal{S}, \mathcal{F}, P) \).

Now form mapping \( Y = g(Z) \)

where \( g : \mathbb{R} \rightarrow \mathbb{R} \)

\[ Y = g(Z) : \mathcal{S} \rightarrow \mathcal{Y} \]
So \( Y(*) = g(\mathcal{X}*) \); \( \mathcal{X} \rightarrow \mathcal{R} \) appears to be a R.V.

Is \( Y \) a R.V.?

Recall from defn of R.V. \( \Rightarrow Y \) is a R.V. if \( Y(w) = g(\mathcal{X}(w)) \) is measurable w.r.t. \((\mathcal{X}, \mathcal{F}, P)\)

Any \( g(*) \) we will encounter, \( Y(w) = g(\mathcal{X}(w)) \) will be measurable w.r.t. \((\mathcal{X}, \mathcal{F}, P)\), and hence, \( Y \) will be a valid R.V.

The Distribution of \( Y = g(\mathcal{X}) \)

Given \( \mathcal{X} \) on \((\mathcal{X}, \mathcal{F}, P)\) with cdf \( F_{\mathcal{X}}(x) \) and a fin \( g(*) \),

We define \( Y = g(\mathcal{X}) \), what is \( F_Y(y) = ? \)

I. From the mapping and definition of CDF/PDF
II. The Direct pdf Method
III. Monte Carlo Simulation.
I. From Mapping and definition of CDF/PDF

Four Examples

Ex. 1 Consider the "generic" fn $g(\cdot)$

\[ g(x) \in (a, b) \]
\[ Y = g(X), \text{ } X \text{ a R.V. w/ } F_X(x) \]

What is $F_Y(y)$, $F_Y(y) = P(Y \leq y)$

Consider $Y_1 = Y_1 = g(x_1)$

\[ F_Y(y_1) = P(\{ Y \leq y_1 \}) = P(\{ X \leq x_1 \}) = F_X(x_1) \]

Consider $Y_2 = Y_2 = g(x_2') = g(x_2) = g(x_2'')$

\[ F_Y(y_2) = P(\{ Y \leq y_2 \}) = P(\{ X \leq x_2' \cup \{ x_2'' < X \leq x_2'' \} \}) \]
\[ = P(\{ X \leq x_2' \}) + P(\{ x_2'' < X \leq x_2'' \}) \]
\[ = F_X(x_2') + [F_X(x_2'') - F_X(x_2'')] \]

So, to determine $F_Y(y)$, such an analysis must be done for each $y \in \mathbb{R}$. See some special cases in the following example.
Example \( Y = g(X) = aX + b, \ a, b \in \mathbb{R} \)

**Two cases**

(i) \( a > 0 \)

\[
F_Y(y) = P(Y \leq y) = P(\{X \leq \frac{y-b}{a}\}) = F_X\left(\frac{y-b}{a}\right)
\]

and

\[
f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X\left(\frac{y-b}{a}\right)}{dy} = \frac{1}{a}f_X\left(\frac{y-b}{a}\right)
\]

(ii) \( a < 0 \)

\[
F_Y(y) = P(Y \leq y) = P(\{X \geq \frac{y-b}{a}\}) = 1 - F_X\left(\frac{y-b}{a}\right)
\]

\[
f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{-1}{a}f_X\left(\frac{y-b}{a}\right)
\]

\[\text{pdf can't be negative, but } a < 0, \text{ so it's ok!}\]

**N.b.** Combine both cases

\[
f_Y(y) = \frac{1}{|a|}f_X\left(\frac{y-b}{a}\right)
\]
Ex. b. \( Y = g(X) = X^2 \)

Note immediately that \( F_Y(y) = 0, y < 0 \) for \( y > 0 \).

\[
F_Y(y) = P \left( \{ X \leq y \} \right)
= P \left( \{-\sqrt{y} \leq X \leq +\sqrt{y}\} \right)
= F_X(\sqrt{y}) - F_X(-\sqrt{y})
\]

\[
= \left[ F_X(\sqrt{y}) - F_X(-\sqrt{y}) \right] 1_{(0,\infty)}(y)
\]

and

\[
f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{2\sqrt{y}} \left[ f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right] 1_{(0,\infty)}(y)
\]

Ex. 2. Suppose \( g(x) \) is constant across some interval \([x_0, x_1]\).

\[
F_Y(y) = P \left( \{ X \leq y \} \right)
\]

(i) \( y = y_1 \):

\[
P(\{ X = y_1 \}) = P(\{ x_0 < X < x_1 \})
= F_X(x_1) - F_X(x_0)
\]

(ii) \( y < y_1 \):

\[
F_Y(y) = P(\{ X \leq y_1 \}) = P(\{ x \leq x_1 \})
= P(\{ x \leq g^{-1}(y) \}) = F_X(g^{-1}(y))
\]

(iii) \( y > y_1 \):

\[
F_Y(y) = P(\{ X \leq y_\beta \}) = P(\{ x \leq x_\beta \})
= P(\{ x \leq g^{-1}(y) \}) = F_X(g^{-1}(y))
\]

This gives

\[
F_Y(y) = \begin{cases} 
F_X(g^{-1}(y)), & y < y_1 \\
F_X(x_1), & y = y_1 \left( P(\{ X \leq y_1 \}) \right) \\
F_X(g^{-1}(y)), & y > y_1 
\end{cases}
\]
\[ \{ \varepsilon \leq y_i \} = \{ \varepsilon = y_i \} \cup \{ \varepsilon < y_i \} \]

\[ P(\{ \varepsilon = y_i \}) = F_x(x_i) - F_x(x_{i-1}) \]

**Ex3.** \( g(x) \) has the form of a "generalized quantizer"

\[ g(x) = g(x_i) = y_i, \quad x_{i-1} < x \leq x_i \]

where \( x_1 < x_2 < \cdots < x_n < \cdots \)

The cdf \( F_\varepsilon(y) \) is defined by the prob.

\[ P(\{ \varepsilon = y_k \}) = P(\{ x_{k-1} < \varepsilon \leq x_k \}) \]

\[ = F_x(x_k) - F_x(x_{k-1}) \]
Ex 4  Assume \( g(x) \) has jump discontinuity at \( x_0 \).
\[ g(x_0^-) \neq g(x_0^+) \]

Assume:
\[ g(x) \leq g(x_0^-), \quad x < x_0 \]
\[ g(x) \geq g(x_0^+), \quad x > x_0 \]

for \( g(x_0^-) \leq y \leq g(x_0^+) \)
\[ F_Y(y) = P(\{ Y \leq y \}) = P(\{ X \leq x_0^+ \}) = F_X(x_0^+) \]

\[ \therefore F_Y(y) \text{ in this case appears as follows:} \]

n.b.
Looking at examples 1~4 gives insights into how to handle general \( g(x) \)
The Direct pdf Method: Determine $f_Y(y)$ from $f_X(x)$

Suppose $g: \mathbb{R} \to \mathbb{R}$ has an inverse $g^{-1}(\cdot)$ such that $y = g(x), \ x = g^{-1}(y)$ and assume that $\frac{dx}{dy} = \frac{dg^{-1}(x)}{dy}$ exists.

Then it can be shown that

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

where $x(y) = g^{-1}(y)$

**Proof**

1. Assume $y = g(x)$ is monotonically increasing.

   $$y + \Delta y = g(x + \Delta x) \leq g(x) + g'(x) \Delta x$$

   $$\Rightarrow P(\{ x + \Delta x \leq y + \Delta y \}) = P(\{ y \leq x + \Delta x \})$$

   $$\Rightarrow f_X(x) \Delta x = f_Y(y) \Delta y$$

   $$\Rightarrow f_Y(y) = f_X(x) \frac{\Delta x}{\Delta y}$$

   In the limit $\Delta x \to 0, \ \Delta y \to 0$, we get

   $$f_Y(y) = f_X(x(y)) \left| \frac{dx(y)}{dy} \right|$$

A similar argument holds for $g(x)$ monotonically decreasing. $\Rightarrow \frac{dx}{dy} < 0$.

**In general,** $f_Y(y) = f_X(x(y)) \left| \frac{dx(y)}{dy} \right|$

(for monotonic)
Moments of R.V.

**Defn:** The mean or expected value of R.V. \(X\) is defined as:

\[
E\{X\} = \int_{-\infty}^{\infty} x f_X(x) \, dx
\]

**n.b.** This defn. also works for discrete R.V.s taking on values \(x_k\), if we write pdf as:

\[
f_X = \sum_k p_k \delta(x-x_k) \quad \text{where } p_k = P(X=x_k)
\]

Then:

\[
E\{X\} = \int_{-\infty}^{\infty} x \sum_k p_k \delta(x-x_k) \, dx
\]

\[
= \sum_k p_k \int_{-\infty}^{\infty} x \delta(x-x_k) \, dx
\]

\[
= \sum_k p_k x_k
\]

So for discrete R.V.

\[
E\{X\} = \sum_k p_k x_k = \sum_k p(x_k) x_k.
\]

**Defn:** Let \(X\) be a R.V. on \((S,F,P)\), and let \(M \in F\). Then the conditional mean of \(X\) conditioned on \(M\) is:

\[
E\{X \mid M\} = \int_{-\infty}^{\infty} x f_{X \mid M}(x \mid M) \, dx
\]
Now suppose we have \((S, F, P)\) and \(X\), and \(Y = g(X)\) \((g: \mathbb{R} \to \mathbb{R})\).

What is \(E\{Y\}\)?

By defn: \(E\{Y\} = \int_{-\infty}^{\infty} y f_{Y}(y) \, dy\)

It appears we must find \(f_{Y}(y)\) first. Fortunately, this is not necessary.

**FACT**: Let \(X\) be a R.V. and \(Y = g(X)\), then

\[
E\{Y\} = E\{g(X)\} = \int_{-\infty}^{\infty} g(x) f_X(x) \, dx
\]

i.e. \(\int_{-\infty}^{\infty} g(x) f_X(x) \, dx = \int_{-\infty}^{\infty} y f_{Y}(y) \, dy\)

**Basic Idea of Proof**:

Assume \(g(x)\) monotonic increasing

\(y = g(x), \quad \frac{dx}{dy} = 1 \frac{dx}{dy} \quad > 0\)

So we have

\[
\int_{-\infty}^{\infty} y f_{Y}(y) \, dy = \int_{-\infty}^{\infty} g(x) \underbrace{f_X(x)}_{f_Y(y)} \, dx
\]

\[
= \int_{-\infty}^{\infty} g(x) f_X(x) \, dx = E\{Y\}
\]

This can be generalized to arbitrary functions \(g(\cdot)\).

We accept this as true

\[
E\{g(X)\} = \bar{g}(X) = \int_{-\infty}^{\infty} g(x) f_X(x) \, dx
\]
Linearity of Expectation

\( g_1(x), g_2(x) \) are two fns of \( x \).
\( \alpha, \beta \in \mathbb{R} \) are two real numbers

\[
E \left\{ \alpha g_1(x) + \beta g_2(x) \right\} = \alpha E\{g_1(x)\} + \beta E\{g_2(x)\}
\]

Proof: Exercise.

Defn: The variance of a R.V. \( X \) is defined as

\[
\text{Var}(X) = \sigma^2 = E\{(X - \overline{X})^2\}, \quad E\{X\} = \overline{X}
\]

\[
= \int_{-\infty}^{\infty} (x - \overline{X})^2 f_{X}(x) \, dx
\]

\( \sigma = \sqrt{\text{Var}(X)} \) is called standard deviation of \( X \).

n.b. \( \text{Var}(X) = E\{(X - \overline{X})^2\} \)

\[
= E\{x^2 - 2x\overline{X} + \overline{X}^2\}
\]

\[
= E\{x^2\} - 2\overline{X} E\{x\} + \overline{X}^2
\]

\[
= E\{x^2\} - 2 \overline{X} \overline{X} + \overline{X}^2
\]

\[
= E\{x^2\} - (E\{x\})^2
\]
Moments

Given a R.V. \( X \), we can define a family of expected value called moments.

**Defn:** The \( n \)-th (non-central) moment of \( X \) is

\[
M_n = E\{X^n\} = \int_{-\infty}^{\infty} x^n f_X(x) \, dx
\]

\( n = 0, 1, 2, \ldots \)

\( n \)-b. \( M_1 = E\{X\} = \text{mean of } X \)

\( M_2 = E\{X^2\} = \text{mean-square value of } X \)

**Defn:** The \( n \)-th central moment of \( X \) is

\[
\mu_n = E\{(X-\overline{X})^n\}, \quad n = 2, 3, 4, \ldots
\]

\( n \)-b. \( \mu_1 = 0 \)

\( \mu_2 = E\{(X-\overline{X})^2\} = \sigma_x^2 = \text{Var}(X) \)

**Fact:** Given the non-central moments, \( m_0, m_1, \ldots, m_n \),

we can compute the central moment \( \mu_n \) as

\[
\mu_n = E\{(X-\overline{X})^n\} = E\left\{ \sum_{k=0}^{n} \binom{n}{k} \overline{X}^k (-\overline{X})^{n-k} \right\}
\]

\[
= \sum_{k=0}^{n} \binom{n}{k} E\{X^k\} (-\overline{X})^{n-k}
\]

\[
= \sum_{k=0}^{n} \binom{n}{k} m_k (-m)^{n-k}.
\]
Similarly, the binomial theorem can be used to show
\[ M_n = E\{X^n\} = E\{[\bar{X} - \bar{X} + \bar{X}]^n\} \]
\[ = \sum_{k=0}^{n} \binom{n}{k} \mu_k m_1^{n-k} \]

**Important Moment Relationships:**

- \( \mu_0 = m_0 = 1 \)
- \( m_1 = \bar{X} \)
- \( \mu_1 = 0 \)
- \( \mu_2 = \sigma_x^2 = \text{Var}(\bar{X}) = m_2 - m_1^2 \)
- \( m_3 = m_3 - 3m_1m_2 + 2m_1^3 \)
  
\( (m_3 = m_3 + 3m_1\sigma_x^2 + m_1^3) \)

**n.b.:** First moment provides a measure of central tendency.
Second central moment provides a measure of dispersion.
n.b.

Dimensionless measure of dispersion.

\[ V(x) = \frac{\sigma(x)}{E(x)} \times 100\% \quad \text{Coefficient of Variation} \]

Central moment of odd order are zero for symmetric pdf.

\[ \beta_1 = \frac{E\{(x-\bar{x})^3\}}{\sigma(x)^3} \quad \text{Coefficient of Skewness} \]

The fourth central moment provides a measure of peakedness (called kurtosis) of a pdf.

\[ \beta_2 = \frac{E\{(x-\bar{x})^4\}}{\sigma(x)^4} \quad \text{Coefficient of Kurtosis} \]

Final Notes:

1. We often describe a R.V. by \( m_1 = E(x) \) and \( m_2 = \text{Var}(x) \) (mean & variance). These two moments do not (in general) characterize the pdf of a R.V.
2. If we know \( m_n \) for \( n=1,2,\ldots \) we can uniquely characterize the pdf of a R.V.