Module 2

Random Variable and Its Probability Distribution

2.1 Random Variables

\((\Omega, \mathcal{F}, \mathbb{P})\): A given probability space

Sometimes, we may not be directly interested in the sample space \(\Omega\), rather we may be interested in some numerical aspect of \(\Omega\).

Example:

Random experiment \(E\): Tossing a fair die until one gets a 6 on upper face.

\[ S = \{6\} \cup \{(i, 1)\}: (1, 5, 3) \cup \{(i, 6)\}: (2, 6, 1, 2, 5) \]

\[ \cdots \cup \{(i, \ldots, i=6)\}: (6, 6, 1, 2, 5) \cdots \]

Suppose that we are interested in number of trials required to get a 6. Then we are not directly interested in \(\Omega\). Rather we are interested in the function \(X: S \rightarrow \mathbb{R}\) defined by

\[ X((i, i_2, \ldots, i_n, 6)) = n, \quad i_1, \ldots, i_n \in \{1, \ldots, 5\} \]

\[ = \text{# of trials required to get a 6.} \]

Similarly, sometimes the sample space \(\Omega\) may be quite abstract and may be tedious to deal with. In such situations, also it may be desirable to represent elements of \(\Omega\) by numbers (e.g., for coin tosses, \(\Omega = \{\text{Heads, Tails}\}\) we may treat \(\text{Heads}\) as 1 and \(\text{Tails}\) as 0)

This amounts to assigning a real-valued function on the sample space \(\Omega\) (e.g., \(X(\text{Heads}) = 1\) and \(X(\text{Tails}) = 0\)).

\(\sqrt{2}\)
Definition. Let $(\Omega, \mathcal{B}, P)$ be a given probability space. A real-valued function $X: \Omega \to \mathbb{R}$ is called a random variable if

$$X^{-1}(1-\alpha, \alpha)) \subseteq \{\omega \in \Omega: X(\omega) \leq \alpha\} \in \mathcal{B}, \quad \forall \alpha \in \mathbb{R}.$$ 

Note: If $\mathcal{B} = 2(\Omega)$ then any function $X: \Omega \to \mathbb{R}$ is a random variable. In general not all real-valued functions defined on sample space $\Omega$ are random variables.

Recall that the Borel $\sigma$-field $\mathcal{B}_0 = \text{Smallest } \sigma\text{-field containing all intervals of } \mathbb{R}$.

We state the following theorem without proving it in print.

Theorem. $X: \Omega \to \mathbb{R}$ is a random variable iff $X^{-1}(B) \equiv \{\omega \in \Omega: X(\omega) \in B\} \in \mathcal{B}, \quad \forall B \in \mathcal{B}_0$.

All real-valued functions defined on sample space $\Omega$ which a normal person can think of are generally random variables. However, there are functions (in fact many) defined on sample space $\Omega$ that are not random variables.

For a function $X: \Omega \to \mathbb{R}$ and $B \in \mathcal{B}$, define

$$X^{-1}(B) \equiv \{\omega \in \Omega: X(\omega) \in B\}.$$

Then

- $X^{-1}(B^c) = (X^{-1}(B))^c$
- $X^{-1}(\bigcup_{a \in S} B_a) = \bigcup_{a \in S} X^{-1}(B_a)$
- $X^{-1}(\bigcap_{a \in S} B_a) = \bigcap_{a \in S} X^{-1}(B_a)$

Also note that $\forall A \subseteq B \Rightarrow X^{-1}(A) \subseteq X^{-1}(B)$.

\[2/2\]
Note that if $X: \Omega \to \mathbb{R}$ is a random variable then
\[
X^{-1}(B) = \{ \omega \in \Omega : X(\omega) \in B \}, \quad \forall B \in \mathcal{B}.
\]

Thus one can defined a set function $P_x: \mathcal{B} \to [0,1]$ by
\[
P_x(B) = P(X^{-1}(B))
\]
\[
= P( \{ \omega \in \Omega : X(\omega) \in B \} )
\]
(to be simply written as $P(X \in B)$)

**Question:** Is $(\mathbb{R}, \mathcal{B}, P_x)$ a probability space (i.e., whether $P_x$ is a probability measure on $\mathcal{B}$)? In that case we have
\[
(\mathbb{R}, \mathcal{B}, P_x) \to \text{may be abridged}
\]
\[
\begin{array}{c}
\downarrow \\
(\mathbb{R}, \mathcal{B}_1, P_x) \to \text{deal with real numbers.}
\end{array}
\]

**Theorem (Induced Probability Space/Measure).** $(\mathbb{R}, \mathcal{B}_1, P_x)$ is a probability space, i.e., $P_x$ is a probability measure defined on $\mathcal{B}_1$. $P_x$ is called the probability measure induced by $X$ and $(\mathbb{R}, \mathcal{B}_1, P_x)$ is called the probability space induced by $X$.

**Proof.**
\[
P_x(B) = P(X^{-1}(B)) \geq 0, \quad \forall B \in \mathcal{B}_1.
\]
\[
P_x(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1
\]

Let $B_1, B_2, \ldots$ be mutually exclusive. Then
\[
P_x(\bigcup_{i=1}^{\infty} B_i) = P\left( X^{-1}\left( \bigcup_{i=1}^{\infty} B_i \right) \right)
\]
\[
= P\left( \bigcup_{i=1}^{\infty} X^{-1}(B_i) \right)
\]
\[
= \sum_{i=1}^{\infty} P(X^{-1}(B_i))
\]
\[
= \bigcup_{i=1}^{\infty} x^{-1}(A_i) \text{ where } A_i = x^{-1}(B_i)
\]
And $A_i \cap A_j = \emptyset$}
\[
\Rightarrow x^{-1}(A_1 \cap x^{-1}(B)) = \emptyset
\]
Example

A fair coin (heads and tails are equally likely) is tossed two times independently.

Suppose that we are interested in the number of heads in two trials.

\[ \Omega = \{ HH, HT, TH, TT \} \]

\[ P(\{ HH \}) = P(\{ HT \}) = P(\{ TH \}) = P(\{ TT \}) = \frac{1}{4} \]

\[ \mathbb{B} = 2^\Omega. \]

Here any function \( X: \Omega \to \mathbb{R} \) is a random variable. We are interested in the function \( X: \Omega \to \mathbb{R} \), above.

\[ X(\omega) = \# \text{ of heads (H) in } \omega \]

\[ = \begin{cases} 
0, & \text{if } \omega = TT \\
1, & \text{if } \omega \in \{ HT, TH \} \\
2, & \text{if } \omega = HH 
\end{cases} \]

Obviously \( X \) is a random variable with induced probability measure given by

\[ P_X(\{0\}) = P(\{ TT \}) = \frac{1}{4} \]

\[ P_X(\{1\}) = P(\{ HT, TH \}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \]

\[ P_X(\{2\}) = P(\{ HH \}) = \frac{1}{4} \]

\[ P_X(B) = P(X^{-1}(B)) \]

\[ = P(\{ \omega \in \Omega : X(\omega) \in B \}) \]

\[ = \sum_{n \in B} P_X(\{n\}) \quad \text{for } B \in \mathbb{B}. \]

Thus \((\Omega, \mathbb{B}, P_X)\) is a probability space.
Definition: Let $X$ be a random variable. Define the function $F_X: \mathbb{R} \to [0,1]$ by

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R}. $$

The function $F_X$ is called the cumulative distribution function (or simply the distribution function) of random variable $X$.

Note: $F_X(x) = \frac{P(X \leq x)}{\mathbb{P}(X \in A)}$, $X$ in a random variable

This quantity is well defined as $P$ is defined on $\mathbb{P}$. If $X$ was not a random variable then this quantity would not have been defined.

- Domain of $F_X = \mathbb{R}$
- Range of $F_X = [0,1)$

Example (a): Suppose that for some constant $c \in \mathbb{R}$, $P(X = c) = 1$ (we call such a random variable a random variable degenerate at $c$). Then

$$F_X(x) = \begin{cases} 0, & \text{if } x < c \\ 1, & \text{if } x \geq c \end{cases}$$

Here $F_X(x)$ is continuous everywhere except at $x = c$. 

[Diagram image]
(b) Suppose that \( P(x = -1) = P(x = 1) = \frac{1}{2} \). Then

\[
F_X(x) = \begin{cases} 
0, & x < -1 \\
\frac{1}{2}, & -1 \leq x < 1 \\
1, & x \geq 1
\end{cases}
\]

\( F_X \) is continuous everywhere except at \( x = -1 \) and \( x = 1 \).

Theorem: Let \( F_X \) be the distribution function of a random variable \( X \). Then

(a) \( F_X \) is non-decreasing.

(b) \( F_X \) is right continuous.

(c) \( F_X(-\infty) = \lim_{\nu \to \infty} F_X(-\nu) = 0 \) and \( F_X(\infty) = \lim_{\nu \to \infty} F_X(\nu) = 1 \).

Proof: (a) Let \(-\alpha < x < \beta < \alpha\). Then

\( (-\alpha, x] \subseteq (-\alpha, \beta) \)

\( \Rightarrow P_X((-\alpha, x]) \leq P_X((-\alpha, \beta)) \)

\( \Rightarrow F_X(x) \leq F_X(\beta) \).

(b) Since \( F_X \) is monotone, \( \lim_{\nu \to \infty} F_X(\nu) = 1 \). Therefore

\( \boxed{\frac{1}{2}} \)
\[
\lim_{h \to 0} F_X(x + h) = \lim_{h \to 0} F_X(x + \frac{h}{n}) = \lim_{h \to 0} P_X((-\infty, x + \frac{h}{n})) = P_X((-\infty, x)) = F_X(x).
\]

(c) \quad F_X(-\infty) = \lim_{h \to 0} F_X(-h) = \lim_{h \to 0} P_X((-\infty, -h)) = P_X((-\infty, -\infty)) = P_X(\emptyset) = 0

(d) \quad F_X(\infty) = \lim_{h \to 0} F_X(h) = \lim_{h \to 0} P_X((-\infty, h)) = P_X((-\infty, h)) = P_X(\infty) = 1

Remark: Since any distribution function \( F_X \) is monotone, it has only countable number of discontinuities. Moreover, it has only jump discontinuities (i.e., \( F_X(a+) \) and \( F_X(a-) \) exist but \( F_X(a+) \neq F_X(a-) \) for \( a \neq \infty \).

(ii) A distribution function \( F_X \) is continuous at \( a \) if \( a \in \mathbb{R} \) if and only if \( F_X(a) = F_X(a+) = F_X(a-) \).
(iii) \[ P(x < b^-) = \lim_{n \to \infty} P(x \leq b - \frac{1}{n}) \quad (\text{Continuity of probability measures}) \]

Then

\[ P(x < b) = F_x(b^-), \quad \forall b \in \mathbb{R} \]

(iv) Let \(-\infty < a < b < \infty\). Since \(b - a \to 0\), \(\exists n \in \mathbb{N}\) such that

\[ a < b - \frac{1}{n} \]

\[ \Rightarrow a < b - \frac{1}{n}, \quad \forall n \geq n_0 \]

\[ \Rightarrow F_x(a) \leq F_x(b - \frac{1}{n}), \quad \forall n \geq n_0 \]

\[ \Rightarrow F_x(a) \leq \lim_{n \to \infty} F_x(b - \frac{1}{n}) \]

\[ F_x(a) \leq F_x(b^-), \quad \forall a < b \]

or

\[ P(x \leq a) \leq P(x < b), \quad \forall a < b \]

(v) For \(a \in \mathbb{R}\),

\[ P(x = a) = P(x \leq a) - P(x < a) = F_x(a) - F_x(a-) \]

\[ \Rightarrow P(x = a) = F_x(a) - F_x(a^-), \quad \forall a \in \mathbb{R}. \]

Thus, a distribution function \(F_x\) is continuous at \(a \in \mathbb{R}\) if \(\forall b \in \mathbb{R}\), \(P(x = a) = 0\).
(VI) For $a, b \in \mathbb{R}$

$$1 - P(X \leq a) = P(X > a) = 1 - F_X(a)$$

$$P(a < x < b) = P(x < b) - P(x \leq a) = F_X(b) - F_X(a) \quad a < b$$

$$P(a \leq x < b) = P(x < b) - P(x \leq a) = F_X(b) - F_X(a), \quad a < b$$

$$P(a < x \leq b) = P(x \leq b) - P(x < a) = F_X(b) - F_X(a), \quad a < b$$

$$P(a \leq x \leq b) = P(x \leq b) - P(x < a) = F_X(b) - F_X(a), \quad a < b$$

(VII) All the distribution functions that we will encounter in this course will be differentiable everywhere except possibly on a countable set E.

**Exercise:** Let $D$ be the set of discontinuity points of a distribution function $F$. For each $n \in \mathbb{N}$, define

$$D_n = \{ x \in \mathbb{R} : F(x) - F(x^-) \geq \frac{1}{n} \}, \quad n \in \mathbb{N}$$

Show that each $D_n$ $(n \in \mathbb{N})$ is finite. Hence show that $D$ is countable.

We state the following theorem without providing a proof:

**Theorem:** Let $X$ be a random variable defined on probability space $(\Omega, \mathcal{F}, P)$. Given the distribution function $F_X$ of $X$, one can determine the probability measure induced by $X$:

$$F_X(B) = P(X^{-1}(B)) = P(X \in B), \quad B \in \mathcal{B}$$

Thus it suffices to study the distribution function $F_X$.

(I) Given a function $F : \mathbb{R} \to \mathbb{R}$ that is non-decreasing, right continuous, and for which $F(\infty) = 1$ and $F(\infty) = 1$, there exists a random variable $X$ on some probability space $(\Omega, \mathcal{F}, P)$ such that the distribution function of $X$ is $F_X$. 

[9/2]
Example

Let

\[
F(x) = \begin{cases} 
0, & x < 0 \\
\frac{2}{3}, & 0 \leq x < 1 \\
\frac{1}{2}, & 1 \leq x < 2 \\
\frac{2}{3}, & 2 \leq x < 3 \\
1, & x \geq 3 
\end{cases}
\]

Clearly \( F \) is right continuous, \( \uparrow \) \( F(-1) = 0 \) and \( F(1) = \frac{2}{3} \).

The \( F \) is a distribution function of some random variable \( X \).

\[ D = \text{Set of discontinuity points of } F \]
\[ = \{ 1, 2, 3 \} \] \( (\text{there are jump discontinuities}) \)

\[ \Rightarrow \text{Set of jump points of } F \]

For \( x \in \{ 1, 2, 3 \} \) \( (\text{points where } F \text{ is continuous}) \)

\[ P(x=1) = F(1) - F(1-) = 0 \]
\[ P(x=2) = F(2) - F(2-) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6} \]
\[ P(x=3) = F(3) - F(3-) = 1 - \frac{2}{3} = \frac{1}{3} \]
\[ P(x<3) = F(3) = \frac{2}{3} \]
\[ P(x \geq \frac{1}{2}) = 1 - F(\frac{1}{2}-) = 1 - \frac{1}{6} = \frac{5}{6} \]
\[ P(2 < x \leq 4) = F(4) - F(2) = 1 - \frac{2}{3} = \frac{1}{3} \]
\[ P(1 \leq x < 2) = F(2) - F(1-) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6} \]
\[ P(2 \leq x < 3) = F(3) - F(2-) = 1 - \frac{1}{2} = \frac{1}{2} \]

Note that

\[ \text{Num of jumps} = \frac{P(x=1) + P(x=2) + P(x=3)}{10/2} \]
Remark. Let $F$ be the distribution function of a random variable $X$ and let $D$ be the set of discontinuity points of $F$ (no that $D$ in a countable set). Then

$$
\text{Sum of jumps of } F = \sum_{x \in D} (F(x) - F(x^-)) = \sum_{x \in D} P(X = x) = P(X \in D) \in [0,1].
$$

2.2. Random Variables of Discrete Type

(5.0, p. 9): A given probability space $X: \Omega \to \mathbb{R}$: a given random variable with distribution function $F_X$.

Definition. The random variable $X$ is said to be of discrete type (or simply a discrete type random variable) if there exists a countable set $S_X$ such that $P(X = x) = F_X(x) - F_X(x^-) > 0$, $x \in S_X$ and

$$
\sum_{x \in S_X} P(X = x) = 1 \quad \text{(i.e., } X \text{ has a sole or isolated mass concentrated on a countable set } S_X ).
$$

The set $S_X$ is called the range or range of the discrete random variable $X$.

Remark (1). Let $X$ be a random variable with range $S_X$. Then

$$
P(x \in x) = F_X(x) - F_X(x^-) > 0, \quad x \in S_X
$$

and

$$
P(x \notin S_X) = 0, \quad x \notin S_X.
$$

Thus the range $S_X$ of a discrete type random $X \in S_X$.
Variable $X$ is nothing but the \emph{set of discontinuity points of the distribution function}.

(iii) A random variable $X$ is of discrete type if

$$\iff \text{sum of jumps of } F_X \left( \sum_{x \in S_x} F_{x}(x) - F_{x}(x-1) \right)$$

equals 1.

**Definition.** Let $X$ be a discrete type random variable with distribution function $F_X$. Define the function $f_X : \mathbb{N} \to \mathbb{R}$ by

$$f_X(x) = \begin{cases} p(x \in x), & \text{if } x \in S_x \\ 0, & \text{if } x \notin S_x \end{cases}$$

The function $f_X$ is called the \emph{probability mass function (p.m.f.)} of random variable $X$.

**Remark.** (i) For $A \in \mathcal{B}_0$, we have

$$p(X \in A) = p(X \in A \cap S_x) \quad (p(x \in S_x) = 1)$$

$$= \sum_{x \in A \cap S_x} p(x \in x)$$

$$= \sum_{x \in A \cap S_x} f_X(x).$$

Thus the \emph{joint} of a discrete type RV $X$ determines the \emph{induced probability measure} $\mu_{F_X} = \{p(x \in A) : A \in \mathcal{B}_0\}$.

(ii) Let $X$ be a discrete type random variable. Then

$$F_X(x) = p(X \leq x)$$

$$= \sum_{y \leq x} p(X \leq y)$$

$$= \sum_{y \in S_x : y \leq x} f_X(y).$$
\[ \sum \{ F(x) - 1 \} \]

This \( F_x \) is a step function with jumps at \( a \in S_x \).

Note that the function \( f_x(\cdot) \) of a discrete-type random variable \( X \) satisfies:

1. \( f_x(\alpha) \geq 0 \quad \forall \alpha \in S_x \)
2. \( f_x(\alpha) = 0 \quad \forall \alpha \in S_x^c \)

and (c) \( \sum_{\alpha \in S_x} f_x(\alpha) = 1 \).

We state the following theorem without proving the proof.

**Theorem**

Given a function \( g : \mathbb{R} \rightarrow \mathbb{R} \) and a countable set

S satisfying:

1. \( g(\alpha) > 0 \quad \forall \alpha \in S \)
2. \( g(\alpha) = 0 \quad \forall \alpha \in S^c \)

and (c) \( \sum_{\alpha \in S} g(\alpha) = 1 \),

there exists a discrete-type random variable \( X \) on

some probability space \((\Omega, \mathcal{F}, P)\) such that \( X \) has

of \( g \) in \( \mathbb{Q} \).

**Example.** A fair die (all outcomes are equally likely) is tossed repeatedly and independently until a \( 6 \) is observed. Then

\[ S = \{ \alpha_1, \ldots, \alpha_6 \} : \text{new, } \alpha_1, \ldots, \alpha_5, \alpha_6 = 6 \],

\[ p(\alpha_1, \ldots, \alpha_6) = \left( \frac{5}{6} \right)^{a_1} \left( \frac{1}{6} \right), \text{new, } \alpha_1, \ldots, \alpha_5, \alpha_6 = 6 \]

Let

\[ X = \# \text{ of tosses required to get at 6.} \]

Since \( S \) is discrete, any function \( X: \Omega \rightarrow \mathbb{R} \) is a random variable (as \( \theta = \theta(n) \))
We have \( X : \mathbb{N} \to \mathbb{R} \) given by

\[
X(\ldots, x_n, \ldots) = y, \quad y \in \mathbb{N}, \quad 2x_1/\ldots, 5y \leq x_n \leq y/2.
\]

Also,

\[
1(x) = P(x=1) = \begin{cases} \left( \frac{5}{6} \right)^{x-1} \left( \frac{1}{6} \right) & \lambda = 1, 2, 3, \ldots, \\ 0 & \text{otherwise} \end{cases}
\]

\( S_x = \{ 0, 1, 2, 3, \ldots, y \} \), and

\[
\sum_{x \in S_x} b_x(x) = 1.
\]

Thus, \( X \) is a discrete type random variable with range \( S_x = \{ 0, 1, 2, 3, \ldots, y \} \) and distribution function

\[
F_X(x) = \begin{cases} 0 & x < 1 \\
\frac{x}{6} & 1 \leq x < 2 \\
\frac{2}{6} & 2 \leq x < 3 \\
\frac{3}{6} & 3 \leq x < 4 \\
\frac{4}{6} & 4 \leq x < 5 \\
\frac{5}{6} & 5 \leq x < 6 \\
\frac{6}{6} & x \geq 6 \\
1 & 1 \leq x \leq y \\
\end{cases}
\]

2.3. Continuous Type and Absolutely Continuous Type

**Random Variables**

**Definition.** (1) A random variable \( X \) is said to be of continuous type (or simply a continuous type random variable) if its distribution function \( F_X \) is continuous at every \( \mathbb{R} \).

Let \( X \) be a random variable (of discrete or continuous type). The set \( S_x = \{ x \in \mathbb{R} : F_X(x+t) - F_X(x) = 0, \quad x \in \mathbb{R}, \quad t \geq 0 \} \)
Called support of distribution of \( X \).

(III) A random variable \( X \) is said to be of absolutely continuous type (or simply an absolutely continuous type random variable) if there exists a non-negative integrable function \( f_X : \mathbb{R} \to [0, \infty) \) such that for any \( x \in \mathbb{R} \):

\[
F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t) \, dt, \quad x \in \mathbb{R}.
\]

The function \( f_X \) is called a probability density function (p.d.f.) of \( f_X \).

Remark: (i) For discrete type random variables support and range are the same.

(ii) From fundamental theorem of calculus we know that the definite integral

\[
F_X(x) = \int_{-\infty}^{x} f_X(t) \, dt, \quad x \in \mathbb{R}
\]

is a continuous function on \( \mathbb{R} \). Thus every random variable of absolutely continuous type is also of continuous type.

(iii) There are random variables of continuous type that are not of absolutely continuous type.

(iv) If \( X \) is of continuous type (or of absolutely continuous type) then

\[
P(X \leq x) = F_X(x) - F_X(x^-) = 0, \quad x \in \mathbb{R}.
\]

It general for any countable set \( C \)

\[
P(X \in C) = \sum_{x \in C} P(X = x) = 0.
\]
(II) If \( x \) is of continuous type, then
\[
    p(x < a) = P(x \leq a) = F(x), \quad a \in \mathbb{R}
\]
\[
    p(x \geq b) = 1 - P(x < a) = 1 - F(x), \quad b \in \mathbb{R}
\]
and, for \(-\infty < a < b \),
\[
    p(a < x < b) = p(a \leq x < b) = P(a < x \leq b) = P(a \leq x \leq b) = F(b) - F(a).
\]

\[
    = \int_{\infty}^{b} b \cdot x \cdot h(x) \, dx - \int_{-\infty}^{a} b \cdot x \cdot h(x) \, dx
\]

\[
    = \int_{a}^{b} b \cdot x \cdot h(x) \, dx.
\]

In general, for any \( B \in \mathcal{B} \), one has
\[
    P(x \in B) = \int_{B} b \cdot x \cdot h(x) \, dx = \int_{\mathbb{R}} b \cdot x \cdot h(x) \, dx.
\]

(VI) Let \( b \cdot x \cdot h(x) \) be the p.d.f. of an absolutely continuous type random variable \( x \) and let \( E \) be any countable set. Define \( g : \mathbb{R} \to \mathbb{R} \) as
\[
    g(x) = \begin{cases} \frac{1}{c_{x}} & \text{if } x \neq E, \\ c_{x} & \text{if } x \in E, \end{cases}
\]
where \( c_{x} > 0 \) are arbitrary. Then
\[
    F_{E}(x) = \int_{-\infty}^{x} b \cdot x \cdot h(x) \, dx = \int_{\mathbb{R}} b \cdot x \cdot h(x) \, dx.
\]

Thus \( g \) is also a p.d.f. of \( x \). Thus the p.d.f. of a continuous type random variable may not be unique.

Theorem. Let \( x \) be a continuous type random variable with distribution function \( F(x) \) that is differentiable everywhere except on a countable set \( E \). Suppose that
\[
    \int_{E} b \cdot x \cdot h(x) \, dx = 0.
\]

Then \( x \) is of absolutely continuous type with a p.d.f.
\[
    b \cdot x \cdot h(x) = \begin{cases} F'(x) & \text{if } x \in E, \\ 0 & \text{if } x \in E^{c} \end{cases}
\]

16/2.
Let \( f(x) \) be the p.d.f. of an absolutely continuous type random variable \( X \). Then

(i) \( f(x) \geq 0, \forall x \in \mathbb{R} \)

and (ii) \( \int_{-\infty}^{\infty} f(x) \, dx = F(x; a) - F(x; 0) = 1 \).

Now we state the following theorem without its proof.

**Theorem.** Given a real-valued function \( g: \mathbb{R} \to \mathbb{R} \) satisfying:

(i) \( g(x) \geq 0, \forall x \in \mathbb{R} \)

and (ii) \( \int_{-\infty}^{\infty} g(x) \, dx = 1 \),

there exists a random variable \( X \) of absolutely continuous type on some probability space \( \Omega \) such that the p.d.f. of \( X \) is \( f(x) \).

**Note:** The random variables of absolutely continuous type can be studied using their pdfs alone.

**Example.** Let \( X \) be a random variable with distribution function

\[
F(x) = \begin{cases} 
0, & x < 0 \\
\frac{x}{4}, & 0 \leq x < 1 \\
\frac{x}{2}, & 1 \leq x < 2 \\
\frac{3x}{8}, & 2 \leq x < \frac{5}{2} \\
1, & x \geq \frac{5}{2}.
\end{cases}
\]

\( D = \text{set of discontinuity points of } F \)

\[ = \{1, \frac{5}{2} \} \]

\[ \text{[13/2]} \]
$$D+q \Rightarrow x \text{ is not of Continuous (and hence absolutely Continuous) type}$$

Sum of jumps = \( \sum_{x \in D} (F(x)-F(x^-)) \)

= \( \frac{1}{3} - \frac{1}{4} \) + \( \frac{3}{4} - \frac{2}{3} \) + \( 1 - \frac{15}{16} \)

= \( \frac{11}{48} < 1 \)

\( \Rightarrow \) \( x \) is not of discrete type.

Thus \( x \) is neither of discrete type nor of Continuous type (and hence nor of absolutely Continuous type). \( \therefore \) 1.

We have:

\( P(x=1) = F(1) - F(1^-) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \)

\( P(x=2) = F(1) - F(1^-) = \frac{2}{4} - 1 = \frac{1}{12} \)

\( P(x=x) = F(x^-) - F(x^-) \geq 1 - \frac{15}{16} = \frac{1}{16} \)

for \( x \notin \{1, 2, 5/4\} \)

\( P(x=x) = F(x^-) - F(x^-) = 0 \)

\( P(1 < x \leq \frac{5}{4}) = F(\frac{5}{4}) - F(1) = 1 - \frac{1}{3} = \frac{2}{3} \)

\( P(1 < x < \frac{5}{4}) = F(\frac{5}{4}) - F(1) = \frac{15}{16} - \frac{1}{2} = \frac{3}{4} \)

\( P(1 < x < \frac{5}{4}) = F(\frac{5}{4}) - F(1) = \frac{15}{16} - \frac{1}{2} = \frac{1}{16} \)

\( P(x = 2) = 1 - F(2^-) = 1 - \frac{2}{3} = \frac{1}{3} \)

\( P(x > 2) = 1 - F(2) = 1 - \frac{3}{4} = \frac{1}{4} \)
Example. Let \( X \) be a random variable with distribution function

\[
F(x) = \begin{cases} 
0, & x < 0 \\
\frac{x}{2}, & 0 \leq x < 1 \\
\frac{x}{2}, & 1 \leq x < 2 \\
1, & x \geq 2 
\end{cases}
\]

Clearly \( F \) is continuous everywhere.

\( \Rightarrow X \) is of continuous type.

Moreover \( F \) is differentiable everywhere except at 0, 1 and 2.

\[
F'(x) = \begin{cases} 
0, & x < 0 \\
\frac{1}{2}, & 0 \leq x < 1 \\
\frac{1}{2}, & 1 \leq x < 2 \\
0, & x \geq 2 
\end{cases}
\]

(Note: here although \( F' \) is not defined on \( [0,1] \) or \( [1,2] \) or \( [2,\infty) \), the Lebesgue integral is defined)

\[
\int_{-\infty}^{\infty} F'(x) \, dx = \frac{1}{2} \int_{0}^{1} \, dx + \frac{1}{2} \int_{1}^{2} \, dx = 1
\]

\( \Rightarrow X \) is of absolutely continuous type with \( f.d.f. \)

\[
f(x) = \begin{cases} 
x, & 0 \leq x < 1 \\
\frac{1}{2}, & 1 \leq x < 2 \\
0, & \text{otherwise}
\end{cases}
\]
Theorem Let $F$ be the distribution function of a random variable $X$. Then $F$ can be decomposed as

$$F(x) = \alpha F_d(x) + (1-\alpha) F_c(x),$$

where $\alpha \epsilon [0,1]$, $F_d$ is distribution function of some discrete type random variable (say $X_d$) and $F_c$ is distribution function of some continuous type random variable (say $X_c$).

Proof. Let

$$D = \text{set of discontinuity points of } F.$$

Case I $D = \emptyset$

Take $\alpha = 0$, $F_c \equiv F$ and $F_d$ as distribution function of any discrete type random variable.

Case II $D \neq \emptyset$

Then $D$ is countable. For simplicity assume that $D = \{a_1, \ldots, a_n\}$ is finite where $-\infty < a_1 < a_2 < \ldots < a_n < \infty$.

Let

$$p_i = P(X=a_i) = F(a_i)-F(a_{i-1}), \quad i = 1, \ldots, n$$

and $\alpha = \sum_{i=1}^n p_i$.

So that $\alpha \epsilon (0,1)$, and $p_i > 0$, $i = 1, \ldots, n$.

If $\alpha = 1$, take $F_d \equiv F$ and $F_c$ to be distribution function of any continuous type random variable.

Now suppose that $\alpha \epsilon (0,1)$. Define $F_d: \mathbb{R} \to [0,1]$ by

$$F_d(x) = \begin{cases} \sum_{a_i < x} \frac{p_i}{\alpha} & 1 < 0, \\ a_1 < x < a_2, \\ \vdots \\ \sum_{a_{i-1} < x} \frac{p_i}{\alpha} & a_i < x < a_{i+1}, \quad i = 1, \ldots, n-1 \\ 1 & x \geq a_n. \end{cases}$$
Then \( F_d \uparrow \), \( F_d \) is right continuous, \( F_d(-a) = 0 \) and \( F_d(a) = 1 \).

The set of discontinuity points of \( F_d \) is \( \mathcal{D} = \{ a_1, \ldots, a_n \} \) and

\[
\sum_{x \in \mathcal{D}} (F_d(x+)-F_d(x-)) = \sum_{i=1}^{n} \frac{b_i}{a_i} > 1
\]

\( \Rightarrow \) \( F_d \) is a distribution function of some discrete type random variable \( \{ \lambda \} \).

Define \( F_c : \mathbb{R} \to [0, 1] \) by

\[
F_c(x) = \frac{F(x) - cF_d(x)}{1-c}, \quad x \in \mathbb{R}
\]

For \(-c < x < c < c,\)

\[
F_d(y) - F_d(x) = \sum_{i : x < a_i \leq y} \frac{b_i}{a_i} - \sum_{i : a_i \leq x} \frac{b_i}{a_i}
\]

\[
= \sum_{i : \lambda < a_i \leq \gamma} \frac{b_i}{a_i}
\]

\[
F(y) - F(x) = \int_{x < x \leq y} P(X = a_i)
\]

\[
\geq \sum_{i : \lambda < a_i \leq \gamma} P(X = a_i)
\]

\[
= \sum_{i : \lambda < a_i \leq \gamma} \frac{b_i}{a_i} = c (F_d(y) - F_d(x-))
\]

\( \Rightarrow \) \( F_c(y) - F_c(x) = \frac{[F(y) - F(x)] - c(F(y) - F(x-))}{1-c} \geq 0 \)

\( \Rightarrow \) \( F_c \uparrow \)

hence that for \( \lambda \neq \{ a_1, a_2, \ldots, a_n \} \)

\[
F_c(x) - F_c(x-) = \frac{F(x) - F(x-)}{1-c} - c (F_d(x) - F_d(x-)) \geq 0
\]

(as \( F \) and \( F_d \) are continuous on \( \mathbb{R}^c \))

And

\[
F_c(a_i) - F_c(a_i-) = \frac{F(a_i) - F(a_i-)}{1-c} - c (F_d(a_i) - F_d(a_i-)) \geq 0
\]
\[ \frac{b c}{1-x} = 0, \quad c = 1 \ldots n \quad \left( \frac{F(a-1) - F(a)}{F(a) - F(a+1)} = \frac{b c}{1-x} \right) \]

\( \Rightarrow F_c \text{ is continuous on } \mathbb{R} \)

\[ F_c(-\infty) = \frac{F(0-1) - x F(0)}{1-x} = 0 \]

\[ F_c(\infty) = \frac{F(\alpha) - x F(\alpha)}{1-x} = 1 \]

\( \Rightarrow F_c \text{ is a distribution function of a random variable of continuous type (say } X_c) \)

**Example** Let \( X \) be a random variable with distribution function

\[ F(x) = \begin{cases} 
0, & x < 0 \\
\frac{x}{4}, & 0 \leq x < 1 \\
\frac{x}{3}, & 1 \leq x < 2 \\
\frac{2x}{5}, & 2 \leq x < \frac{5}{2} \\
1, & x \geq \frac{5}{2} 
\end{cases} \]

\[ D = \text{Net of discontinuities} \text{ of } F \]

\[ = \{ \frac{1}{4}, \frac{5}{2} \} \]

\[ k_1 = P(X > 1) = F(1) - F(1-) = \frac{2}{4} - \frac{1}{4} = \frac{1}{4} \]

\[ k_2 = P(X > 2) = F(2) - F(2-) = \frac{2}{5} - \frac{1}{3} = \frac{7}{15} \]

\[ k_3 = P(X > \frac{5}{2}) = F(\frac{5}{2}) - F(\frac{5}{2}-) = 1 - \frac{5}{1} = 1 - \frac{1}{16} = \frac{15}{16} \]

\[ \alpha = k_1 + k_2 + k_3 = \frac{11}{16} \]
\[
F_{d}(x) = \begin{cases} 
0 & x < 0.5 \\
\frac{4}{11} & 0.5 \leq x < 1 \\
\frac{9}{11} & 1 \leq x < 2 \\
1 & x \geq 2 
\end{cases} 
\]

\[
F_{e}(x) = \frac{F(x) - \alpha F_{d}(x)}{1 - \alpha}
\]

\[
F(x) = \begin{cases} 
0 & x < 0 \\
\frac{12x}{3\pi} & 0 \leq x < 1 \\
\frac{1}{3\pi} \left(4x - 1\right) & 1 \leq x < 2 \\
\frac{2}{3\pi} \left(9x - 4\right) & 2 \leq x \leq \frac{5}{2} \\
1 & x > \frac{5}{2}
\end{cases}
\]

We have

\[
F(x) = \alpha F_{d}(x) + (1 - \alpha) F_{e}(x), \quad \alpha = \frac{11}{48}
\]

where \(\alpha\) is a distribution function of a discrete type random variable and \(F_{d}\) (given by (2.5.2)) is a distribution function of a continuous type random variable. Here \(F_{e}\) is differentiable everywhere except possibly at \(0, \frac{1}{2}, \frac{5}{2}\) with

\[
F_{e}'(x) = \begin{cases} 
\frac{12}{3\pi} & 0 \leq x < 1 \\
\frac{16}{3\pi} & 1 \leq x < 2 \\
\frac{18}{3\pi} & 2 \leq x \leq \frac{5}{2} \\
0 & x > \frac{5}{2}
\end{cases}
\]

And \(\int_{-\infty}^{\infty} F_{e}'(x) \, dx = 1\)

\(\Rightarrow F_{e}\) is in fact a d.f. of a random variable of absolutely continuous type.