1. Distribution Function

**Definition 1** A random variable is a real valued function defined on probability space \((S, \Gamma, P)\). In other words, A random variable assigns a real number to any outcome in \(S\).

Induced probability space is \((S_X, \Gamma_X, P_X)\) such that

1. \(S_X = \{X(s)|s \in S\}\)
2. \(\Gamma_X = \{X(A)|A \in \Gamma\}\)
3. \(P_X(A_X) = P(X^{-1}(A_X))\) for any \(A_X \in \Gamma_X\)
   where \(X^{-1}(A_X) = \{s|X(s) \in A_X\}\)

(e.g) Consider \(S = \{HH, HT, TH, TT\}\), all possible outcomes from two tosses of a coin.
\[
\begin{align*}
\Gamma &= \{\emptyset, \{HH\}, \{HT, TH\}, \{TT\}, \{HH, HT, TH\}, \{HH, TT\}, \{HT, TH, TT\}, S\} \\
P(\{HH\}) &= .25, \ P(\{HT, TH\}) = .5, \ P(\{TT\}) = .25 \\
X & \text{ is the number of heads.} \\
S_X &= \{0, 1, 2\} \\
\Gamma_X &= \{\emptyset, \{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, S_X\} \\
P_X(\{0\}) &= .25, P_X(\{1\}) = .5, P_X(\{2\}) = .25, P_X(\{0, 1\}) = .75, \ldots
\end{align*}
\]

We use the notation \(Pr\) as

\[
\begin{align*}
Pr(X \in A_X) &= P_X(A_X) = P(X^{-1}(A_X)) \\
Pr(a < X \leq b) &= P_X(\{x|a < x \leq b\}) = P(\{s|a < X(s) \leq b\})
\end{align*}
\]

**Definition 2** Let \(X\) be a random variable defined on \((S, \Gamma, P)\). The cumulative distribution function (cdf) of \(X\) is a function \(F_X : \mathbb{R} \rightarrow [0, 1]\) that satisfies

\[
F(t) = Pr(X \leq t)
\]

(e.g) With previous example,

\[
F_X(t) = \begin{cases} 
0 & \text{when } t < 0 \\
.25 & \text{when } 0 \leq t < 1 \\
.75 & \text{when } 1 \leq t < 2 \\
1 & \text{when } t \geq 2
\end{cases}
\]

and

\[
Pr(0 < X \leq 2) = F_X(2) - F_X(0) = .75
\]

**Definition 3** Let \(X\) be a discrete random variable. The probability mass function (pmf) of \(X\), \(p_X : S_X \rightarrow [0, 1]\) is defined by

\[
p_X(x) = Pr(X = x)
\]
The support of \( X \) is the set whose element has positive probability mass. (e.g) With previous example, \( p_X(0) = .25, p_X(1) = .5, p_X(2) = .25 \)

The support of \( X = \{0, 1, 2\} = S_X \)

**Definition 4** Let \( X \) be a continuous random variable. The probability density function (pdf) of \( X \)

\[ f_X : S_X \to \mathbb{R} \]

defined by

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

Note that this is not complete because \( F_X \) may not be differentiable at all points. If \( F_X \) is not differentiable at \( x \), \( f_X(x) \) has an arbitrary number. The support of \( X \) is a closure of the set whose element has positive probability density.

A pdf \( f_X \) satisfies

\[ \Pr(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx \]

Consider a new random variable \( Y = g(X) \) for some function \( g : \text{strictly increasing} \). What is the distribution of \( Y \)?

\[ F_Y(y) = \Pr(Y \leq y) = \Pr(g(X) \leq y) = \Pr(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) \]  \( (1) \)

Based on this, we have the following theorem

**Theorem 5 (Transformation Theorem)** Let \( g : \mathbb{R} \to \mathbb{R} \) be a monotone function. Define \( Y = g(X) \). If \( X \) is a discrete random variable,

\[ p_Y(y) = p_X(g^{-1}(y)) \]

and if \( X \) is a continuous random variable,

\[ f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}}{dy}(y) \right| \]

(e.g) With previous example, let \( g(x) = x^2 \) and define \( Y = g(X) \), then

\[ p_Y(0) = p_X(g^{-1}(0)) = p_X(0) = .25 \]
\[ p_Y(1) = p_X(g^{-1}(1)) = p_X(1) = .5 \]
\[ p_Y(4) = p_X(g^{-1}(4)) = p_X(2) = .25 \]

(e.g) Let \( f_X(x) = .5 \) on \([0,2]\) and 0 otherwise. Let \( g(x) = x^3 \) and define \( Y = g(X) \).

\[ f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}}{dy}(y) \right| = \frac{1}{2} \left| \frac{dy}{dx}(y) \right| = \frac{1}{6} y^{-\frac{2}{3}} \]

on \([0,8]\) and 0 otherwise.
Proof.
Discrete case is straightforward. Using (1), prove first for \( g \): strictly increasing.

\[
f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(g^{-1}(y))}{dy} = f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy}
\]

Next, note that for \( g \): strictly decreasing,

\[
F_Y(y) = \Pr(Y \leq y) = \Pr(g(X) \leq y) = \Pr(X \geq g^{-1}(y)) = 1 - \Pr(X \leq g^{-1}(y)) = 1 - F_X(g^{-1}(y))
\]

So,

\[
f_Y(y) = \frac{dF_Y(y)}{dy} = -\frac{dF_X(g^{-1}(y))}{dy} = -f_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy}
\]

Combining two, we have the desired result.

What if \( g \) is not monotone? When \( X \) is discrete,

\[
p_Y(y) = \sum_{x: g(x) = y} p_X(x)
\]

When \( X \) is continuous, we can partition \( \mathbb{R} \) into \( C_1, \ldots, C_n \) on each of which \( g \) is monotone, and find an inverse function on each \( C_i \), say, \( h_1, \ldots, h_n \). Then,

\[
f_Y(y) = \sum_{i=1}^n f_X(h_i(y)) \left| \frac{dh_i(y)}{dy} \right|
\]

2. Expectation

**Definition 6** The expectation of random variable \( X \) is defined by

\[
E[X] = \sum_{x \in S_X} xp_X(x)
\]

if \( X \) is discrete, and

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

if \( X \) is continuous.

Note that, of course, expectation is defined only when it indeed exists.

(e.g) Non-existence of expectation.

\( f_X(x) = x^{-2} \) on \([1, \infty)\) and 0 otherwise.

\[
\int_{-\infty}^{\infty} xf_X(x)dx = \log x \bigg|_{1}^{\infty} = \infty
\]
What’s the expectation of \( g(X) \)? Recall Transformation Theorem. Let \( Y = g(X) \), and for simplicity assume \( g \) is strictly increasing.

\[
E[g(X)] = E[Y] = \sum_{y \in S_Y} yp_Y(y) = \sum_{y \in S_Y} yp_X(g^{-1}(y)) = \sum_{g(x) \in S_Y} g(x)p_X(x) = \sum_{x \in S_X} g(x)p_X(x)
\]

\[
E[g(X)] = E[Y] = \int_{-\infty}^{\infty} yf_Y(y)dy = \int_{-\infty}^{\infty} yf_X(g^{-1}(y)) \frac{dg^{-1}(y)}{dy}dy = \int_{-\infty}^{\infty} yf_X(g^{-1}(y)) dg^{-1}(y) = \int_{-\infty}^{\infty} g(x)f_X(x)dx
\]

It can be easily extended to the case in which \( g \) is not monotone, by partitioning \( \mathbb{R} \).

We call \( E[X^m] \) the \( m \)-th moment of \( X \). Especially, \( E[X] \) is called the mean of \( X \), and \( E[(X - E[X])^2] \) is called the variance of \( X \). Here we have very useful function.

**Definition 7** A moment generating function (mgf) of \( X \) is

\[ M_X(t) = E[e^{tX}] \]

Note also that this is defined only when \( E[e^{tX}] \) exists on \(-h < t < h\) for some \( h > 0 \). The mgf has the following properties.

1. \( \frac{d^n}{dt^n} M_X(0) = E[X^n] \)
2. The mgf of a random variable is unique up to the distribution. In other words,

\[ M_X(t) = M_Y(t) \quad \forall t \in (-h, h) \Leftrightarrow F_X(t) = F_Y(t) \quad \forall t \in \mathbb{R} \]

**Theorem 8 (Markov’s Inequality)** For any nonnegative function \( u(\cdot) \), if \( E[u(X)] \) exists,

\[
\Pr[u(X) \geq a] \leq \frac{E[u(X)]}{a}
\]

**Theorem 9 (Chebyshev’s Inequality)** Let \( \mu \) and \( \sigma^2 \) be the mean and variance of random variable \( X \), respectively. Then,

\[
\Pr(|X - \mu| \geq b) \leq \frac{\sigma^2}{b^2}
\]

\[
\Pr(|X - \mu| \geq b\sigma) \leq \frac{1}{b^2}
\]

\[
\Pr \left( \left| \frac{X - \mu}{\sigma} \right| \geq b \right) \leq \frac{1}{b^2}
\]
Theorem 10 (Jensen’s Inequality) Let \( g(\cdot) \) be a convex function on interval \( I \). If the support of \( X \) is a subset of \( I \),
\[
E[g(X)] \geq g(E[X])
\]
Note that \( g \) is convex if \( g'' > 0 \). Note also that \(-g\) is convex if \( g \) is concave. So if \( g \) is concave, we have
\[
E[g(X)] \leq g(E[X])
\]
3. Random Vector
A random vector is a vector of random variables.
\[
X = (X_1, \cdots, X_n)
\]
(e.g) Consider \( S = \{1, 2, 3, 4, 5, 6\} \), all possible outcomes from a cast of a dice.
\begin{itemize}
  \item \( X_1 = 1 \) if we get odd number, and 0 otherwise.
  \item \( X_2 = 1 \) if we get 1, 2, or 3, and 0 otherwise.
  \item \( X(1) = (1, 1), \ X(2) = (0, 1) \)
\end{itemize}
\( F_X(x) \), or \( F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) \) which is called joint cdf, is defined by
\[
F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \Pr(X_1 \leq x_1, \cdots, X_n \leq x_n)
\]
(e.g) With previous example,
\( F_X(1, 1) = 1, \ F_X(1, 0) = .5, \ F_X(0, 1) = .5, \ F_X(0, 0) = \frac{1}{3} \)
Let \( X = (X_1, \cdots, X_n) \) be a random vector. A marginal distribution of \( X_i \) is individual distribution of the random variable \( X_i \) which is given by
\[
F_{X_i}(x_i) = \lim_{x_j \to \infty, \forall j \neq i} F_X(x)
\]
Joint pmf or pdf and marginal pmf or pdf can be obtained by the same way as in the random variable case.
(e.g) With previous example, \( p_X(1, 1) = \frac{1}{6}, \ p_X(0, 0) = \frac{1}{3} \)
\[
F_{X_1}(0) = .5, \ F_{X_1}(1) = 1
\]
\[
p_{X_1}(0) = 1, \ p_{X_1}(1) = .5
\]
Definition 11 Let \( (X_1, X_2) \) be a random vector. Conditional pmf of \( X_2 \) given \( X_1 = x_1 \) is given by
\[
p_{X_2|X_1=x_1}(x_2) = \frac{p_{X_1,X_2}(x_1,x_2)}{p_{X_1}(x_1)}
\]
and conditional pdf of \( X_2 \) given \( X_1 = x_1 \) is given by
\[
f_{X_2|X_1=x_1}(x_2) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)}
\]