Chapter 2: Random Variables

Example 1. Tossing a fair coin twice:

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

Define for any \( \omega \in \Omega \), \( X(\omega) \)=number of heads in \( \omega \). \( X(\omega) \) is a random variable.

Definition 1. A random variable (RV) is a function \( X: \Omega \rightarrow \mathbb{R} \).

Definition 2 (Cumulative distribution function (CDF)).

\[ F(\chi) = P(X \leq \chi). \tag{1} \]

Example 2. The cumulative distribution function of \( x \) is as (Figure 1)

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

Lemma 1. Properties of CDF

\[ \lim_{x \to -\infty} F_X(x) = 0 \quad \tag{2} \]
\[ \lim_{x \to +\infty} F_X(x) = 1, \quad \tag{3} \]
(2) $F_X(x)$ is non-decreasing:

$$x_1 \leq x_2 \implies F_X(x_1) \leq F_X(x_2)$$

(4) $F_X(x)$ is continuous from the right

$$\lim_{\epsilon \to 0} F_X(x + \epsilon) = F_X(x), \epsilon > 0$$

(5)

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) + P(X = a)$$

(6)

$$= F_X(b) - F_X(a) + P(X = a)$$

(7)

$$P(X = a) = \lim_{\epsilon \to 0} F_X(a) - F_X(a - \epsilon), \epsilon > 0$$

Definition 3. If random variable $X$ has finite or countable number of values, $X$ is called discrete.

Example 3. Non-countable example: $\mathbb{R}$.

A set $S$ is countable if you can find a bijection of $f$.

$$f : S \to N.$$ 

Definition 4. $X$ is continuous if $F_X(x)$ is continuous.

Definition 5 (Probability density function(pdf)).

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (x \text{ is continuous}).$$


By definition,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x - \mu)^2}{2\sigma^2}}$$
Therefore,

\[ F_X(a) = P(x \leq a) = \int_{-\infty}^{a} f_X(x)dx, \]

\[ = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{a} \frac{-(x-\mu)^2}{2\sigma^2} dx. \]

We should always have:

\[ \int_{-\infty}^{+\infty} f_X(x)dx = 1. \]

**Definition 6** (mean, variance of a RV X). For the continuous case:

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

\[ V(X) = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f_X(x)dx. \]

For the discrete case:

\[ E(X) = \mu = \sum_{i=-\infty}^{+\infty} x_i P(X = x_i), \]

\[ V(X) = \sigma^2 = \sum_{i=-\infty}^{+\infty} (x_i - \mu)^2 P(X = x_i). \]

**Example 5.** X is uniformly distributed in (0, 1].

\[ F_X(x) = \begin{cases} 
0 & x < 0, \\
\int_{0}^{x} 1dx = x & 0 \leq x < 1, \\
1 & x \geq 1.
\end{cases} \]

\[ E(X) = \int_{0}^{1} X \times 1dx = \frac{1}{2}, \]

\[ V(X) = \int_{0}^{1} (X - \frac{1}{2})^2 \times 1dx = \frac{1}{12}. \]

**Lemma 2** (Probability Density Functions).

(1) Uniform X uniform over [a,b]:

\[ f_X(x) = \begin{cases} 
\frac{1}{b-a} & \text{if } a \leq x \leq b \\
0 & \text{otherwise}
\end{cases} \]  \[ (10) \]

\footnote{Figure from Wikipedia: \url{https://en.wikipedia.org/wiki/Uniform_distribution_(continuous)}}
(2) Gaussian distribution:

\[
f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},
\]

(11)

(3) Exponential distribution:

\[
f_X(x) = \begin{cases} 
\frac{1}{\mu} e^{-\frac{x}{\mu}} & \text{if } x \geq 0 \\
0 & \text{if } x < 0
\end{cases}
\]

(12)

(4) Rayleigh Distribution:

\[
f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, x \geq 0,
\]

(13)

---

\(^2\) Figure from Wikipedia: [https://en.wikipedia.org/wiki/Normal_distribution](https://en.wikipedia.org/wiki/Normal_distribution)

\(^3\) Figure from Wikipedia: [https://en.wikipedia.org/wiki/Exponential_distribution](https://en.wikipedia.org/wiki/Exponential_distribution)

\(^4\) Figure from Wikipedia: [https://en.wikipedia.org/wiki/Rayleigh_distribution](https://en.wikipedia.org/wiki/Rayleigh_distribution)
(5) **Laplacian Distribution:**

\[ f_X(x) = \frac{1}{\sqrt{2\sigma}} e^{-\frac{\sqrt{2}x}{\sigma}}. \]  

\[ \sigma = 0.5 \]
\[ \sigma = 1 \]
\[ \sigma = 2 \]
\[ \sigma = 3 \]
\[ \sigma = 4 \]
1 Example of Discrete Random Variable

1.1 Bernoulli RV

flipping a coin, \( P(H) = p, P(T) = 1 - p \), if head occurs \( X = 1 \), if tail occurs \( X = 0 \), \( P(X = 0) = 1 - p, P(X = 1) = p \). The CDF of a bernoulli RV is as Figure 8.

![Figure 8: Cumulative distribution function of Bernoulli Random Variable](image)

1.2 Binomial distribution

Tossing a die \( n \) times, \( P(H) = p, P(T) = 1 - p \). \( X \) is number of heads, \( x \in \{0, 1, \ldots, n\} \).

\[
P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.
\]
Remark 1. Let $Y_i \in \{0, 1\}$ denote the outcome of tossing the die the $i$th time

$$X = Y_1 + Y_2 + \cdots + Y_n.$$ 

i.e., a Binomial RV can be thought of as the sum of $n$ independent Bernoulli RV.

Example 6 (Random graph). Each edge exists with probability $p$, $X$ is the number of neighbor of node $1$ (Figure 9).

$$Y_i = \begin{cases} 
1, & \text{if node 1 is connected to } i+1, \\
0, & \text{otherwise.} 
\end{cases}$$

$$X = Y_1 + Y_2 + \cdots + Y_{n-1}.$$ 

So $X$ follows the Binomial distribution.

![Figure 9: Random Graphs](image)

Example 7 (BSC). Suppose we are transmitting a file of length $n$. Consider a BSC where the probability of error is $p$ and the probability of receiving the correct bit is $1-p$. (Figure 10) What is

![Figure 10: Binary Symmetric Channel with probability of error $P_e = p$.](image)
the probability that we have \( k \) errors?

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

Let \( X \) represent the number of errors, what is \( E(X) \)

\[
E(X) = \sum_{k=0}^{n} k P(X = k),
\]

\[
= \sum_{k=0}^{n} k \binom{n}{k} p^k (1 - p)^{n-k},
\]

\[
= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1 - p)^{n-k},
\]

\[
= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1 - p)^{n-k+1},
\]

\[
= np.
\]

Binomial theorem:

\[
(x + y)^n = \sum_{k=1}^{n} \binom{n}{k} x^k y^{n-k}
\]

\[
(p + 1 - p)^{n-1} = \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1 - p)^{n-k+1}
\]

\[
= 1.
\]

**Theorem 1.** For any two RVs \( X_1 \) and \( X_2 \), \( Y = X_1 + X_2 \),

\[
E(Y) = E(X_1) + E(X_2). \tag{15}
\]

It does not matter whether \( X_1 \) and \( X_2 \) are independent or not.

1.3 Geometric distribution

You keep tossing a coin until you observe a Head. \( X \) is the number of times you have to toss the coin.

\[
X \in \{1, 2, \ldots\},
\]

\[
P(X = k) = (1 - p)^k p.
\]

**Example 8** (Binary erasure channel). Suppose you have a BEC channel with feedback. When you get a erasure, you ask the sender to retransmit. (Figure 11)

Suppose you pay one dollar for each retransmission. Let \( X \) be the amount of money you pay per transmission.

\[
E(X) = \frac{1}{1 - p},
\]

\[
= \frac{1}{0.9} \approx 1.11 \text{"}. 
\]
Figure 11: Binary Erasure Channel with probability of erasure \( P_e = 0.1 \).

For geometric distribution,

\[ P(H) \approx \frac{1}{E(X)}, \]

which \( E(X) \) is the number of coin flips on average.

**Proof.**

\[
E(X) = \sum_{k=1}^{\infty} k P(X = k),
\]

(16)

\[
= \sum_{k=1}^{\infty} k(1-p)^{k-1}p,
\]

(17)

\[
= p \sum_{k=1}^{\infty} k(1-p)^{k-1}.
\]

(18)

Recall that for \(|x| < 1\),

\[
\sum_{k=0}^{\infty} x^k = \frac{1}{1-x},
\]

(19)

\[
\frac{d}{dk} \sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2},
\]

(20)

\[
\sum_{k=1}^{\infty} k(1-p)^k = \frac{1}{p^2}.
\]

(21)

So,

\[
E(X) = \frac{1}{p^2},
\]

(22)

\[
= \frac{1}{p}.
\]

(23)
1.4 Poisson distribution

Suppose a server receives $\lambda$ searches per second on average. The probability that the server receives $k$ searches for this second is

$$ P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots, \infty. \quad (24) $$

For poisson distribution

$$ \sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \quad (25) $$

$$ = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \quad (26) $$

$$ = 1. \quad (27) $$

$$ E(X) = \lambda. \quad (28) $$

**Example 9** (Interpretation of poisson distribution as an arrival experiment).

Suppose average of arrival customers per second is $\lambda$. Suppose server goes down if $X \geq 100$. We want to find the probability of $P(X = k)$.

$$ P(\text{server going down}) = P(X \geq 100). \quad (29) $$

We divide the one second to $n$ intervals, each length of the interval is $\frac{1}{n}$ second. The probability $p$ of getting requests in small interval is $\frac{\lambda}{n}$.

![Figure 12: one second divided into n intervals.](image)

Now we can consider it to be Bernoulli distribution with $p$.

$$ P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad (29) $$

$$ = \binom{n}{k} \left( \frac{\lambda}{n} \right)^k (1 - \frac{\lambda}{n})^{n-k} \quad (30) $$

$$ \approx \frac{n^k}{k!} \left( \frac{p}{1-p} \right)^k (1-p)^n, \quad (31) $$

$$ = \frac{1}{k!} (np)^k e^{-np}, \quad (32) $$

$$ = \frac{1}{k!} \lambda^k e^{-\lambda}, \quad (33) $$

$$ = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (34) $$
We get (31) because of
\[
\binom{n}{k} = \frac{1}{k!} n(n-1) \ldots (n-k+1),
\] (35)
\[
\approx \frac{n^k}{k!}, \text{ (k is a constant and n goes to infinity).}
\] (36)

This means we can approximate Binomial(n,p) by Poisson with \( \lambda = np \) (if n is very large).

2 Two Random Variables

Example 10. Let \( X \) and \( Y \) be Bernoulli Random Variable. If \( Y = 0 \), we know \( X \) must equal to

<table>
<thead>
<tr>
<th></th>
<th>Y=0</th>
<th>Y=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>X=0</td>
<td>( \frac{1}{2} )</td>
<td>( \frac{1}{4} )</td>
</tr>
<tr>
<td>X=1</td>
<td>0</td>
<td>( \frac{1}{4} )</td>
</tr>
</tbody>
</table>

Table 1: Joint probability mass function of \( X \) and \( Y \).

0, so \( X \) and \( Y \) are dependent.

\[
P(X = 0) = \frac{3}{4},
\]
\[
P(X = 1) = \frac{1}{4}.
\]

Here is an example which \( X \) and \( Y \) are independent, but they have the same marginal distribution.

<table>
<thead>
<tr>
<th></th>
<th>Y=0</th>
<th>Y=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>X=0</td>
<td>( \frac{1}{8} )</td>
<td>( \frac{7}{8} )</td>
</tr>
<tr>
<td>X=1</td>
<td>( \frac{1}{5} )</td>
<td>( \frac{1}{5} )</td>
</tr>
</tbody>
</table>

Table 2: Joint probability mass function of \( X \) and \( Y \).

2.1 Marginalization

You have the joint distribution \( P_{X,Y}(x,y) \).

\[
P_X(x_0) = \sum_y P_{X,Y}(x_0,y),
\] (37)
\[
P_Y(y_0) = \sum_x P_{X,Y}(x,y_0).
\] (38)

Definition 7. If \( X \) and \( Y \) are continuous random variables, then the joint CDF:

\[
F_{X,Y}(x,y) = P(X \leq x, Y \leq y).
\] (39)
Given joint CDF $F_{X,Y}(x, y)$,

$$F_X(x_0) = F_{X,Y}(x_0, +\infty). \quad (40)$$

**Definition 8.** When the CDF is differentiable, the joint pdf is defined as

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}, \quad (41)$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy, \quad (42)$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx. \quad (43)$$

**Definition 9.** $X$ and $Y$ are independent if and only if

$$F_{X,Y}(x, y) = F_X(x)F_Y(y), \quad (44)$$

$$f_{X,Y}(x, y) = f_X(x)f_Y(y). \quad (45)$$

**Definition 10.** Conditional CDF of marginal distribution is

$$F_{X,Y}(x|y) = P(X \leq x|Y \leq y), \quad (46)$$

$$= \frac{F_{X,Y}(X \leq x, Y \leq y)}{P(Y \leq y)}. \quad (47)$$

**Example 11.** $X$ and $Y$ are 2 random variables given by the joint pdf

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2\sigma^2(1-\rho^2)}(x^2 + y^2 - 2\rho xy)\right].$$

What is $f_X(x)$?

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy,$n

$$= \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2(1-\rho^2)}(x^2 + y^2 - 2\rho xy)\right] dy,$n

$$= \frac{\exp\left[-\frac{x^2}{2\sigma^2(1-\rho^2)}\right]}{2\pi\sigma^2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2(1-\rho^2)}(y^2 - 2\rho xy + \rho x^2 - \rho^2 x^2)\right] dy,$n

$$= \frac{\exp\left[-\frac{x^2 + \rho^2 x^2}{2\sigma^2(1-\rho^2)}\right]}{2\pi\sigma^2\sqrt{\sigma\rho^2}} \int_{-\infty}^{\infty} e^{\frac{-(y-\rho x)}{\sqrt{2\sigma^2(1-\rho^2)}}} dy,$n

$$= \frac{\exp\left[-\frac{(1-\rho^2)x^2}{2\sigma^2(1-\rho^2)}\right]}{\sqrt{2\pi\sigma}} \sqrt{2\pi} \int_{-\infty}^{\infty} e^{\frac{-(y-\rho x)}{\sqrt{2\sigma^2}}} dy.$n

**Because**

$$\frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-\frac{(x-\rho x)^2}{2\sigma^2}} dx = 1.$$
So if $\rho = 0$,

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}.$$  

Similarly,

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{y^2}{2\sigma^2}}.$$  

We can have

$$f_{X,Y}(x,y) = f_X(x)f_Y(y).$$

So $X$ and $Y$ are independent. If $\rho \neq 0$, $X$ and $Y$ are not independent.