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 \( X: \Omega \rightarrow \mathbb{R} \).

Example 2. Let \( w \) be the temperature in °F at 3:00 pm on Thursday afternoon. Let \( X \) be the r.v. which the temperature in °C. Then

\[ X = \frac{5}{9} (w - 32) \]

Definition 2 (Cumulative distribution function (CDF)).

\[ F(x) = P(X \leq x). \] \hspace{1cm} (1)

Example 3. 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

(1)
\[
\lim_{x \to -\infty} F_X(x) = 0 \\
\lim_{x \to +\infty} F_X(x) = 1,
\]

(2) $F_X(x)$ is non-decreasing:

\[ x_1 \leq x_2 \implies F_X(x_1) \leq F_X(x_2) \]

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

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

(4)
\[
P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) + P(X = a)
\]

\[
= F_X(b) - F_X(a) + P(X = a)
\]

(5)
\[
P(X = a) = \lim_{\epsilon \to 0} F_X(a) - F_X(a - \epsilon), \epsilon > 0
\]

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

Remark 1. A set $S$ is countable if its elements can be indexed, i.e., we can find a injective function from $S$ to the natural numbers

Example 4. Non-countable example: $\mathbb{R}$.

Example 5. Countable example: The number of tosses we need till get a Head

Lemma 2. If $X$ is continuous, then $F_X(x)$ is continuous.

Definition 4 (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,
\[
F_X(a) = P(x \leq a) = \int_{-\infty}^{a} f_X(x)dx,
\]

\[
= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{a} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.
\]
We should always have:

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

**Definition 5** (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 7.** 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 3** (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)

(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: It is the probability distribution of the waiting time between events in a Poisson process in which events occur continuously and independently at a constant average

\[1\text{Figure from Wikipedia: } \text{https://en.wikipedia.org/wiki/Uniform_distribution_(continuous)}\]

\[2\text{Figure from Wikipedia: } \text{https://en.wikipedia.org/wiki/Normal_distribution}\]
rate (check Poisson process in later lectures)

\[ f_X(x) = \begin{cases} 
\lambda e^{-\lambda} & \text{if } x \geq 0 \\
0 & \text{if } x < 0
\end{cases} \quad (12) \]

**The mean:**

\[
\mathbb{E}[X] = \int_{x=0}^{\infty} x f(x) \, dx
\]

\[
= \int_{x=0}^{\infty} x \lambda \exp(-\lambda x) \, dx
\]

\[
= \lambda \int_{x=0}^{\infty} x \exp(-\lambda x) \, dx
\]

\[
= \lambda \left( \left[ -\frac{1}{\lambda} x \exp(-\lambda x) \right]_{x=\infty}^{x=0} + \int_{x=0}^{\infty} \frac{1}{\lambda} \exp(-\lambda x) \, dx \right)
\]

\[
= \lambda \left( 0 + \frac{1}{\lambda^2} \right)
\]

\[
= \frac{1}{\lambda}
\]

**Homework:** Find the variance of the exponential distribution.

**Answer:**

\[
\mathbb{E}[X^2] = \int_{x=0}^{\infty} x^2 f(x) \, dx
\]

\[
= \int_{x=0}^{\infty} x^2 \lambda \exp(-\lambda x) \, dx
\]

\[
= \lambda \int_{x=0}^{\infty} x^2 \exp(-\lambda x) \, dx
\]

\[
= \lambda \left( \left[ -\frac{1}{\lambda^2} x^2 \exp(-\lambda x) \right]_{x=\infty}^{x=0} + \int_{x=0}^{\infty} \frac{1}{\lambda} x \exp(-\lambda x) \, dx \right)
\]

\[
= \lambda \left( 0 + \frac{1}{\lambda} \left( \frac{1}{\lambda} \mathbb{E}[X] \right) \right)
\]

\[
= \lambda \left( \frac{1}{\lambda^3} \right)
\]

\[
= \frac{1}{\lambda^2}
\]
(4) Rayleigh Distribution:

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

(13)

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

(14)

\[^3\text{Figure from Wikipedia:}\ https://en.wikipedia.org/wiki/Exponential_distribution}\n\[^4\text{Figure from Wikipedia:}\ https://en.wikipedia.org/wiki/Rayleigh_distribution\]
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.

\[ F(x) \]

\[ \begin{align*}
F(x) & = 1 - p \\
F(x) & = 1 \\
\end{align*} \]

Figure 8: Cumulative distribution function of Bernoulli Random Variable

---

5Figure from Wikipedia: [https://en.wikipedia.org/wiki/Laplace_distribution](https://en.wikipedia.org/wiki/Laplace_distribution)
1.2 Binomial distribution

Tossing a coin 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 2.** Let \( Y_i \in \{0, 1\} \) denote the outcome of tossing the coin the ith 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 8** (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 9** (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 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 $E(X)$

$$E(X) = \sum_{k=0}^{n} kP(X = k),$$
$$= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k},$$
$$= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k},$$
$$= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-k+1},$$
$$= np.$$  

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 $X_2$, $Y = X_1 + X_2$,

$$E(Y) = E(X_1) + E(X_2).$$  \hspace{1cm} (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.

$$X \in \{1, 2, \ldots \},$$
$$P(X = K) = (1-p)^{K-1}p.$$
Example 10 (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.

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

$$= \frac{1}{0.9} \approx 1.11 \$. $$

For geometric distribution,

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

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

**Intuition:** You have to make $E(X)$ trials, and in these $E(X)$ trials, the success happens once at the last trial.

**Proof.**

$$E(X) = \sum_{k=1}^{\infty} kP(X = k), \quad (16)$$

$$= \sum_{k=1}^{\infty} k(1 - p)^{k-1}p, \quad (17)$$

$$= p \sum_{k=1}^{\infty} k(1 - p)^{k-1}. \quad (18)$$

Recall that for $|x| < 1$,

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1 - x}, \quad (19)$$

$$\frac{d}{dk} \sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1 - x)^2}, \quad (20)$$

$$\sum_{k=1}^{\infty} k(1 - p)^k = \frac{1}{p^2}. \quad (21)$$
So,

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

\[
= \frac{1}{p}. \tag{23}
\]

\[
(\text{22})
\]

\section*{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

\[
P(X = k) = C \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots, \infty. \tag{24}
\]

To find \( C \):

\[
\sum_{k=0}^{\infty} P(X = k) = 1
\]

\[
C \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = 1
\]

\[
C = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}
\]

\[
C = e^{-\lambda}.
\]

Then the pdf of the poisson distribution for an average of \( \lambda \) arrivals per time unit is:

\[
P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots, \infty. \tag{25}
\]

The mean is:

\[
E(X) = \lambda.
\]

\textbf{Example 11} (Interpretation of poisson distribution as an arrival experiment). \newline
Suppose average of arrival customers per second is \( \lambda \). Suppose server goes down if \( X \geq 100 \). We want to find the probability of \( P(X = k) \).

\[
P(\text{server going down}) = P(X \geq 100).
\]

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.

Now we can consider it to be Bernoulli distribution with \( p \).

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

or

\[
P(X = k) = \left( \frac{n}{k} \right) \left( \frac{\lambda}{n} \right)^k \left( \frac{1}{n} \right)^{n-k},
\]

\[
\approx \frac{n^k}{k!} \left( \frac{p}{1-p} \right)^k (1-p)^n,
\]

or

\[
P(X = k) = \frac{1}{k!} (np)^k e^{-np},
\]

\[
= \frac{1}{k!} \lambda^k e^{-\lambda},
\]

\[
= \frac{\lambda^k}{k!} e^{-\lambda}.
\]

We get (28) because of

\[
\binom{n}{k} = \frac{1}{k!} n(n-1) \ldots (n-k+1),
\]

or

\[
\approx \frac{n^k}{k!}, \text{ (k is a constant and n goes to infinity).}
\]

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

2 Two Random Variables

Example 12. 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>3/4</td>
<td>1/4</td>
</tr>
<tr>
<td>X=1</td>
<td>0</td>
<td>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 a 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>⅓</td>
<td>⅔</td>
</tr>
<tr>
<td>X=1</td>
<td>⅔</td>
<td>⅓</td>
</tr>
</tbody>
</table>

Table 2: Joint probability mass function of $X$ and $Y$.

### 2.1 Marginalization

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

\[
P_X(x) = \sum_y P_{X,Y}(x_0,y), \tag{34}
\]

\[
P_Y(y) = \sum_x P_{X,Y}(x,y_0). \tag{35}
\]

**Definition 6.** If $X$ and $Y$ are continuous random variables, then the joint CDF:

\[
F_{X,Y}(x,y) = P(X \leq x, Y \leq y). \tag{36}
\]

Given joint CDF $F_{X,Y}(x,y)$,

\[
F_X(x_0) = F_{X,Y}(x_0, +\infty). \tag{37}
\]

**Definition 7.** When the CDF is differentiable, the joint pdf is defined as

\[
f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}, \tag{38}
\]

\[
f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy, \tag{39}
\]

\[
f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx. \tag{40}
\]

**Definition 8.** $X$ and $Y$ are independent if and only if

\[
F_{X,Y}(x,y) = F_X(x)F_Y(y), \tag{41}
\]

\[
f_{X,Y}(x,y) = f_X(x)f_Y(y). \tag{42}
\]

**Definition 9.** Conditional CDF of marginal distribution is

\[
F_{X,Y}(x|y) = P(X \leq x|Y \leq y), \tag{43}
\]

\[
F_{X,Y}(x|Y \leq y) = \frac{F_{X,Y}(X \leq x, Y \leq y)}{P(Y \leq y)}. \tag{44}
\]

**Example 13.** $X$ 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 $f_X(x)$?

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

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

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

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

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

Because

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

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.