Chapter 3

Discrete Random Variables

In this chapter we study the basics of a random variable. We focus on the discrete random variables as they are relatively easier to interpret. We also study the mean, variance and other statistical properties of random variables.

3.1 Random Variable

Definition 1. A random variable $X$ is a function $X : \Omega \to \mathbb{R}$ that maps an outcome $\xi \in \Omega$ to a number $X(\xi)$ on the real line.

The mapping from $\Omega$ to $\mathbb{R}$ has important practical reasons. Since $\xi \in \Omega$ is simply an outcome, it is not necessarily a number. For example, when tossing a coin, $\xi$ could be “H” or “T”; When voting for a political party, $\xi$ could be “republican” or “democrat”. Mapping these qualitative outcomes to numbers on the real line allows us to analyze the outcomes in a purely mathematical setting.

A pictorial illustration of the definition is shown in Figure 3.1. In this figure, $\xi_1$ and $\xi_2$ are two distinct outcomes of the sample space $\Omega$. They are mapped to two distinct number $X(\xi_1)$ and $X(\xi_2)$ on the real line.

Example. Flip a coin 2 times. The sample space $\Omega$ is

$$\Omega = \{(HH), (HT), (TH), (TT)\}.$$ 

Suppose that $X$ is a random variable that maps an outcome to a number representing the sum of “H”. Then, for the 4 $\xi$’s in the sample space there are only 3 distinct numbers. More precisely, if we let $\xi_1 = (HH), \xi_2 = (HT), \xi_3 = (TH), \xi_4 = (TT)$, then, we have

$$X(\xi_1) = 2, \ X(\xi_2) = 1, \ X(\xi_3) = 1, \ X(\xi_4) = 0.$$ 

Remark. The above example show that the mapping defined by the random variable is not necessarily a one-to-one mapping, because multiple outcomes can be mapped to the same number.
Figure 3.1: Illustration of Definition 3.1.

### 3.2 PMF and CDF

**Definition 2.** The **probability mass function (PMF)** of a random variable $X$ is a function which specifies the probability of obtaining a number $X(\xi) = x$. We denote a PMF as

$$p_X(x) = \mathbb{P}[X = x].$$

$\mathbb{P}[X = x]$ should be interpreted as

$$\mathbb{P}[X = x] = \mathbb{P}\{\xi \mid X(\xi) = x\}.$$  

That is, we seek the probability of the event that $X(\xi)$ equals a number $x$, where the occurrence of the event is specified by the random outcome $\xi$.

**Caution: Difference between $X$ and $x$.**

It is very important to understand the difference between a random variable $X$ and the number $x$. The random variable $X$ is a function, or informally we can call it an object. Since this object is random by nature, the only way to characterize it is by means of a PMF. PMF specifies the probability of obtaining a particular state. A different random variable has a different PMF. The number $x$ is the realization of $X$ in a specific experiment. It specifies what value should $X(\xi)$ take.

It is also important to differentiate a random variable $X$ and an ordinary variable $x$ in algebraic equations. For example, in the equation \(x^2 - 3x + 2 = 0\), the variable $x$ takes two possible values $x = 2$ or $x = 1$. There is nothing random about $x$.  

© 2017 Stanley Chan. All Rights Reserved.
**Example.** Flip a coin twice. Let \( X \) be the number of “H”. Then \( X \) has a PMF

\[
\begin{align*}
p_X(0) &= P[X = 0] = \frac{1}{4}, \\
p_X(1) &= P[X = 1] = \frac{1}{2}, \\
p_X(2) &= P[X = 2] = \frac{1}{4}.
\end{align*}
\]

A pictorial illustration of the PMF is shown in Figure 3.2.

![Figure 3.2: Illustration of a PMF](image)

**Remark.** A PMF should satisfy the condition that

\[
\sum_x p_X(x) = 1.
\]

This result follows from Axiom II that \( P[\Omega] = 1 \). Basically, it states that when summing over all possible outcomes, the probability should be 1.

**Definition 3.** The **cumulative distribution function (CDF)** of a discrete random variable \( X \) is

\[
F_X(x) \overset{\text{def}}{=} P[X \leq x] = \sum_{x' \leq x} p_X(x').
\]

CDF can be considered as an “integration” of the PMF from \(-\infty\) to \( x \), where the variable \( x' \) in the sum is a dummy variable. The usage of the CDF has to do with the integrability of delta functions in the PMF. Historically speaking, delta functions are not considered as proper functions as they are not integrable in the sense of Riemann-Stieltjes. Therefore, in order to describe discrete probability in mathematics, people define probabilities in terms of CDF and take differentiation to obtain PMF. We will return to this point when we discuss continuous random variables.

© 2017 Stanley Chan. All Rights Reserved.
Example. Consider a random variable $X$ with PMF shown in Figure 3.3. The PMF has values

$$p_X(0) = \frac{1}{4}, \quad p_X(1) = \frac{1}{2}, \quad p_X(4) = \frac{1}{4}.$$  

The CDF of $X$ can be computed as

$$F_X(0) = P[X \leq 0] = p_X(0) = \frac{1}{4},$$

$$F_X(1) = P[X \leq 1] = p_X(0) + p_X(1) = \frac{3}{4},$$

$$F_X(4) = P[X \leq 4] = p_X(0) + p_X(1) + p_X(4) = 1.$$  

![Figure 3.3: Illustration of a PMF](image)

In the previous example, we observe several properties of a CDF (of a discrete random variable).

Properties of a discrete CDF:

1. The CDF is a sequence of increasing unit steps.
2. The maximum of the CDF is always 1, i.e., $F_X(+\infty) = 1$.
3. The minimum of the CDF is always 0, i.e., $F_X(-\infty) = 0$.
4. The unit steps have jumps at positions where $p_X(x) > 0$.
5. If the unit step has a jump at $x$, then the height of that jump is $p_X(x)$. Mathematically, we write it as

$$p_X(x) = F_X(x) - F_X(x^-),$$

where $F_X(x^-) = \lim_{h \to 0} F_X(x - h)$. We will discuss this more carefully in Chapter 4.
6. For any other $x$ that does not correspond to a jump, $F_X(x)$ stays as a constant.
7. The unit step always has a solid dot on the left hand side, and an empty dot on the right hand side. We call this right continuous.
3.3 Expectation

**Definition 4.** The *expectation* of a random variable $X$ is

$$E[X] = \sum_x x p_X(x).$$

Expectation is also the mean of the random variable $X$. Intuitively, we can think of $p_X(x)$ as the “percentage” of times that the random variable $X$ attains the value $x$. When this “percentage” multiplies with $x$, we obtain the contribution of each $x$. Summing over all possible values of $x$ then yields the mean.

**Example.** For the random variable $X$ shown in Figure 3.2, the expectation of $X$ is

$$E[X] = (0) \left(\frac{1}{4}\right) + (1) \left(\frac{1}{2}\right) + (2) \left(\frac{1}{4}\right) = 1.$$

**Example.** Consider a game. Flip a coin 3 times. Reward:

- $1 if there are 2 Heads
- $8 if there are 3 Heads
- $0 if there are 0 or 1 Head

The cost to enter the game is $1.5. On average what is the net gain?

To answer this question, we first let $X$ be the number of heads. The PMF of $X$ is

$$p_X(0) = \frac{1}{8}, \quad p_X(1) = \frac{3}{8}, \quad p_X(2) = \frac{3}{8}, \quad p_X(3) = \frac{1}{8}.$$

(Question: How do we obtain these PMF values? This is a simple counting exercise. See combination in Chapter 1.)

![Figure 3.4: PMF of $X$ and $Y$.](image)
We then let $Y$ be the reward. The PMF of $Y$ can be found by “adding” the probabilities of $X$. This yields

$$p_Y(0) = p_X(0) + p_X(1) = \frac{4}{8}, \quad p_Y(1) = p_X(2) = \frac{3}{8}, \quad p_Y(8) = p_X(3) = \frac{1}{8}.$$  

The expectation of $Y$ is

$$\mathbb{E}[X] = \left(0 \left( \frac{4}{8} \right) + (1) \left( \frac{3}{8} \right) + (8) \left( \frac{1}{8} \right) \right) = \frac{11}{8}.$$  

Since the cost of the game is $\frac{12}{8}$, the net gain (on average) is $-\frac{1}{8}$.

**Property 1.** The expectation of a random variable $X$ has the following properties

(a) **Function.** For any function $g$,

$$\mathbb{E}[g(X)] = \sum_x g(x)p_X(x).$$

(b) **Linearity.** For any function $g$ and $h$,

$$\mathbb{E}[g(X) + h(X)] = \mathbb{E}[g(X)] + \mathbb{E}[h(X)].$$

(c) **Scale.** For any constant $c$,

$$\mathbb{E}[cX] = c\mathbb{E}[X].$$

(d) **DC Shift.** For any constant $c$,

$$\mathbb{E}[X + c] = \mathbb{E}[X] + c.$$  

Property 1(a) is simply a change of variable. The proof can be seen from Figure 3.5. When we have a function $Y = g(X)$, the PMF of $Y$ will have impulses moved from $x$ (the horizontal axis) to $g(x)$ (the vertical axis). The PMF values (i.e., the probabilities, or the height of the stems), however, are not changed. If the mapping $g(X)$ is many-to-one, then multiple PMF values will add to the same position. Therefore, when we compute $\mathbb{E}[g(X)]$, we compute the expectation along the vertical axis.

Property 1(b) states that when we have two functions $g(X)$ and $h(X)$, the expectation of the sum is the sum of the expectation. However, note that $g(X)$ and $h(X)$ are two different functions of the *same* random variable $X$. That is, while the functional outputs are different, the randomness are the same. We will study the more general case $g(X) + h(Y)$ when we discuss two random variables. Property 1(c) and Property 1(d) are easy to verify.
Figure 3.5: By letting \( g(X) = Y \), the PMFs are not changed. However, the value at which the PMF takes changes.

### 3.4 Moment and Variance

**Definition 5.** The \( k \)-th moment of a random variable \( X \) is

\[
\mathbb{E}[X^k] = \sum_x x^k p_X(x).
\]

Following the definition of moment, we see that the expectation \( \mathbb{E}[X] \) is the first moment, and \( \mathbb{E}[X^2] \) is the second moment.

**Example.** Flip a coin 3 times. Let \( X \) be the number of heads. Then,

\[
p_X(0) = \frac{1}{8}, \quad p_X(1) = \frac{3}{8}, \quad p_X(2) = \frac{3}{8}, \quad p_X(3) = \frac{1}{8}.
\]

The second moment \( \mathbb{E}[X^2] \) is

\[
\mathbb{E}[X^2] = (0)^2 \left( \frac{1}{8} \right) + (1)^2 \left( \frac{3}{8} \right) + (2)^2 \left( \frac{3}{8} \right) + (3)^2 \left( \frac{1}{8} \right) = 3.
\]

**Example.** Consider a random variable \( X \) with PMF

\[
p_X(k) = \frac{1}{2^k}, \quad k = 1, 2, \ldots.
\]

The first moment of \( X \) is

\[
\mathbb{E}[X] = \sum_{k=1}^{\infty} k \left( \frac{1}{2} \right)^k = \frac{1}{2} \sum_{k=1}^{\infty} k \left( \frac{1}{2} \right)^{k-1} = \frac{1}{2} \left( \frac{1}{(1 - \frac{1}{2})^2} \right) = 2.
\]
**Definition 6.** The variance of a random variable $X$ is

$$\text{Var}[X] = \mathbb{E}[(X - \mu_X)^2],$$

where $\mu_X = \mathbb{E}[X]$ is the expectation of $X$.

We denote $\sigma_X^2 = \text{Var}[X]$. $\sigma_X$ is called the standard deviation of $X$.

**Property 2.** The variance of a random variable $X$ has the following properties

(a) **Moment.**

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

(b) **Scale.** For any constant $c$,

$$\text{Var}[cX] = c^2\text{Var}[X].$$

(c) **DC Shift.** For any constant $c$,

$$\text{Var}[X + c] = \text{Var}[X].$$

**Proof of Property 2(a).**

$$\text{Var}[X] = \mathbb{E}[(X - \mu_X)^2] = \mathbb{E}[X^2 - 2X\mu_X + \mu_X^2]$$

$$= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mu_X + \mu_X^2$$

$$= \mathbb{E}[X^2] - \mu_X^2.$$

**Proof of Property 2(b).** Note that $\mathbb{E}[cX] = c\mu_X$. Thus,

$$\text{Var}[cX] = \mathbb{E}[(cX - c\mu_X)^2] = c^2\mathbb{E}[(X - \mu_X)^2] = c^2\text{Var}[X].$$

**Proof of Property 2(c).**

$$\text{Var}[X + c] = \mathbb{E}[(X + c - \mathbb{E}[X + c])^2] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \text{Var}[X].$$
Property 2(c) states that when $X$ is shifted by a DC of $c$, the variance is unchanged, although the mean is shifted.

**Example.** Flip a coin with probability $p$ to get a Head. Let $X$ be a random variable denoting the outcome. The PMF of $X$ is

$$p_X(0) = 1 - p, \quad p_X(1) = p.$$  

The expectation of $X$ is

$$
\mathbb{E}[X] = (0)p_X(0) + (1)p_X(1) = (0)(1 - p) + (1)(p) = p.
$$

The second moment is

$$
\mathbb{E}[X^2] = (0)^2p_X(0) + (1)^2p_X(1) = p.
$$

The variance is

$$
\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1 - p).
$$

### 3.5 Common Discrete Random Variables

**Bernoulli Random Variable**

**Definition 7.** Let $X$ be a Bernoulli random variable. Then, the PMF of $X$ is

$$p_X(0) = 1 - p, \quad p_X(1) = p,$$

where $0 < p < 1$ is called the Bernoulli parameter. We write

$$X \sim \text{Bernoulli}(p)$$

to say that $X$ is drawn from a Bernoulli distribution with a parameter $p$.

Bernoulli random variable is the simplest type of discrete random variables. It is used to describe random events that involve two states, e.g., coin flip (H or T), binary bit (1 or 0), true or false, yes or no, present or absent, etc.

The parameter $p$ in a Bernoulli random variable controls the probability of obtaining 1. In a coin flip event, $p$ is usually $\frac{1}{2}$, meaning that the coin is fair. However, for biased coins $p$ is not necessarily $\frac{1}{2}$. For other situations such as binary bits (0 or 1), the probability of obtaining 1 could be very different from the probability of obtaining 0.

In the present chapter, we assume that $p$ is known. Our goal is to study the random variable $X$ generated from Bernoulli($p$). The inverse problem of estimating $p$ for a given $X$ is known as an estimation problem which we will discuss later.
Proposition 1. If $X \sim \text{Bernoulli}(p)$, then

\[
\begin{align*}
\mathbb{E}[X] &= p, \\
\mathbb{E}[X^2] &= p, \\
\text{Var}[X] &= p(1 - p).
\end{align*}
\]

Proof. See Example for Property 2. \qed

Binomial Random Variable

Definition 8. Let $X$ be a Binomial random variable. Then, the PMF of $X$ is

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

where $0 < p < 1$ is the Binomial parameter, and $n$ is the total number of states. We write

\[
X \sim \text{Binomial}(n, p)
\]

to say that $X$ is drawn from a Binomial distribution with a parameter $p$ of size $n$.

Remark: Binomial from Bernoulli. A binomial random variable can be obtained by summing a sequence of Bernoulli random variables. That is, if $I_1, \ldots, I_n$ is a sequence of Bernoulli random variables with $I_j \sim \text{Bernoulli}(p)$ for all $i = 1, \ldots, n$, then the resulting variable

\[
X = I_1 + I_2 + \ldots + I_n
\]
is a Binomial random variable of size $n$ and parameter $p$. The proof of this statement will be discussed when we introduce moment generating function.

Proposition 2. If $X \sim \text{Binomial}(n, p)$, then

\[
\begin{align*}
\mathbb{E}[X] &= np, \\
\mathbb{E}[X^2] &= np(np + (1 - p)), \\
\text{Var}[X] &= np(1 - p).
\end{align*}
\]

Proof. Knowing that a binomial random variable is a sequence of Bernoulli random variable makes the proof of the above statements easy. Let $I_1, \ldots, I_n$ be a sequence of Bernoulli random variables with $I_j \sim \text{Bernoulli}(p)$ for all $i = 1, \ldots, n$. Let

\[
X = I_1 + I_2 + \ldots + I_n.
\]
Then, we can show that
\[ E[X] = E[I_1 + I_2 + \ldots + I_n] = p + p + \ldots + p = np. \]
Similarly, the variance and second moment are
\[ \text{Var}[X] = \text{Var}[I_1 + \ldots + I_n] = \text{Var}[I_1] + \ldots + \text{Var}[I_n] = np(1-p), \]
\[ E[X^2] = \text{Var}[X] + E[X]^2 = np(1-p) + (np)^2. \]

**Geometric Random Variable**

**Definition 9.** Let \( X \) be a Geometric random variable. Then, the PMF of \( X \) is
\[ p_X(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots, \]
where \( 0 < p < 1 \) is the Geometric parameter. We write
\[ X \sim \text{Geometric}(p) \]
to say that \( X \) is drawn from a Geometric distribution with a parameter \( p \).

Physically, a geometric random variable defines a sequence of Bernoulli trials with \( k-1 \) consecutive failures followed by one success. In telephone networks, geometric random variables is often used to model \( k-1 \) misses followed by a successful connection.

**Proposition 3.** If \( X \sim \text{Geometric}(p) \), then
\[ E[X] = \frac{1}{p}, \]
\[ E[X^2] = \frac{2}{p^2} - \frac{1}{p}, \]
\[ \text{Var}[X] = \frac{1 - p}{p^2}. \]

**Proof.** We will prove the mean and leave the second moment and variance as exercise.
\[ E[X] = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p \left( \sum_{k=1}^{\infty} k(1-p)^{k-1} \right) \]
\[ = p \left( \frac{1}{(1 - (1-p))^2} \right) \overset{(a)}{=} \frac{1}{p}, \]
where \((a)\) follows from the Geometric series identities in Chapter 1.

© 2017 Stanley Chan. All Rights Reserved.
Poisson Random Variable

**Definition 10.** Let $X$ be a Poisson random variable. Then, the PMF of $X$ is

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots,$$

where $\lambda > 0$ is the Poisson rate. We write

$$X \sim \text{Poisson}(\lambda)$$

to say that $X$ is drawn from a Poisson distribution with a parameter $\lambda$.

Poisson random variables are often used to model arrival processes such as photon arrivals, electron mission, telephone call arrivals. The parameter $\lambda$ in the Poisson distribution determines the rate of the arrival.

**Proposition 4.** If $X \sim \text{Poisson}(\lambda)$, then

$$\mathbb{E}[X] = \lambda,$$

$$\mathbb{E}[X^2] = \lambda + \lambda^2,$$

$$\text{Var}[X] = \lambda.$$

**Proof.** We will prove the mean and leave the second moment and variance as exercise.

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

$$= \lambda e^{-\lambda} \sum_{k'=0}^{\infty} \frac{\lambda^{k'}}{k'!}$$

$$= \lambda e^{-\lambda} e^\lambda = \lambda.$$

**Example.** In a telephone call center, the average phone call arrival rate is $\lambda$ (calls / sec). Let $X$ be the actual number of calls / sec. We want to find $\mathbb{P}[X > 4]$ and $\mathbb{P}[X \leq 5]$.

$$\mathbb{P}[X > 4] = 1 - \mathbb{P}[X \leq 4] = 1 - \sum_{k=0}^{4} \frac{\lambda^k}{k!} e^{-\lambda}.$$

$$\mathbb{P}[X \leq 5] = \sum_{k=0}^{5} \frac{\lambda^k}{k!} e^{-\lambda}.$$
Proposition 5. Poisson Approximation to Binomial. For small $p$ and large $n$,

\[
\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda},
\]

where $\lambda \overset{\text{def}}{=} np$.

This result is extremely useful for numerical computation. Below is an example.

Example. Consider an optical communication system. The bit arrival rate is $10^9$ bit/sec, and the probability of having one error bit is $10^{-9}$. Suppose we want to find the probability of having 5 error bit in one second.

Let $X$ be the number of error bits. In one second, there are $10^9$ bits. Since we do not know the location of these 5 bits, we have to enumerate all possibilities. This leads to a binomial distribution. Using the Binomial distribution, we know that the probability of having $k$ error bits is

\[
P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} = \left(\frac{10^9}{k}\right) (10^{-9})^k (1 - 10^{-9})^{10^9 - k}.
\]

This quantity is difficult to calculate in floating point arithmetic.

Using Poisson to Binomial approximation, we can see that the same probability can be approximated by

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

where $\lambda = np = 10^9 (10^{-9}) = 1$. Putting $k = 5$ yields $P[X = 5] \approx 0.003$.

Proof of Proposition 3.5. Let $\lambda = np$. Then,

\[
\binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}
= \frac{\lambda^k}{k!} \frac{n(n-1) \cdots (n-k+1)}{n \cdot n \cdots n} \left(1 - \frac{\lambda}{n}\right)^{n-k}
= \frac{\lambda^k}{k!} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{n-k} \left(1 - \frac{\lambda}{n}\right)^n
\rightarrow 1 \text{ as } n \rightarrow \infty
\rightarrow 1 \text{ as } n \rightarrow \infty
= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n.
\]
We claim that \((1 - \frac{\lambda}{n})^n \to e^{-\lambda}\). This can be proved by noting that
\[
\log(1 + x) \approx x, \quad x \ll 1.
\]
It then follows that \(\log \left(1 - \frac{\lambda}{n}\right) \approx -\frac{\lambda}{n}\). Hence, \((1 - \frac{\lambda}{n})^n \approx e^{-\lambda}\)