

Chapter 1

Random Variables

In the previous chapter, we describe the outcome of a random process by using words like "get a head" which lacks the flavour of mathematics. In this chapter, we will use numbers to describe the outcomes of a random process.

1.1 Random Variables and Expectations

Definitions

- (1) **Random Variable** is a function that maps outcomes of a random process into numbers. The notations are upper case letters such as **X**, **Y**, **Z**.

Toss a coin twice, and let **X** be the total number of heads. The the values of **X** are: **X** = 2 : (H, H) **X** = 1 : (H, T), (T, H) **X** = 0 : (T, T). The event "get at least one head" can be represented as $\mathbf{X} \leq 1$. The probability of this event can be represented as $\mathbf{P}(\mathbf{X} \leq 1)$

- (2) **Discrete Random Variables** are random variables take values that can be enumerated.

X: the number of heads when tossing two coins

Y: the number of points when rolling a die

Z: the number of tosses of a coin till you get the first head.

X, **Y** and **Z** are all discrete random variables. **Z** can take infinite number of values, but those values can be enumerated.

- (3) **Continuous Random Variables** can take continuous values.

X \sim **N**(0, 1).

T: the time you have to wait till the public bus comes.

D: the distance from the center of the board when you throw darts.

X, **T** and **D** are all continuous variables.

- (4) **Probability Distribution** of a random variable describes how the probabilities are assigned to different values of the random variable.

The distribution of a discrete random variable may be given by a table shown below.

Value of X:	x_1	x_2	x_3	\cdots
Probability:	p_1	p_2	p_3	\cdots

The distribution of a continuous random variable can be given by the *density curve* or the *probability density function(pdf)*. For example, the distribution of a normal random variable $\mathbf{X} \sim \mathbf{N}(\mu, \sigma)$ is given by the following probability density function.

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

The expectations of random variables

The distribution of a discrete random variable is given below.

Value of \mathbf{X}:	x_1	x_2	x_3	\cdots
Probability:	p_1	p_2	p_3	\cdots

Find: $\sum p_i$

The **expectation** of \mathbf{X} , which is also the mean is

$$\mu_{\mathbf{x}} = \mathbf{E}(\mathbf{X}) = x_1p_1 + x_2p_2 + x_3p_3 + \cdots = \sum x_i p_i.$$

The notation of the mean of \mathbf{X} is $\mu_{\mathbf{x}}$ instead of $\bar{\mathbf{X}}$. $\mathbf{E}(\mathbf{X})$ is interpreted as the long term average.

The **variance** of \mathbf{X} is the expectation of the random variable $(\mathbf{X} - \mu_{\mathbf{x}})^2$.

Probability:	p_1	p_2	p_3	\cdots
Value of \mathbf{X}:	x_1	x_2	x_3	\cdots
$(\mathbf{X} - \mu_{\mathbf{x}})^2$:	$(x_1 - \mu_{\mathbf{x}})^2$	$(x_2 - \mu_{\mathbf{x}})^2$	$(x_3 - \mu_{\mathbf{x}})^2$	\cdots

$$\mathbf{Var}(\mathbf{X}) = (x_1 - \mu_{\mathbf{x}})^2 + (x_2 - \mu_{\mathbf{x}})^2 + \cdots = \sum (x_i - \mu_{\mathbf{x}})^2$$

The standard deviation of \mathbf{X} : $\sigma_{\mathbf{x}} = \sqrt{\mathbf{Var}(\mathbf{X})}$

The notation of the standard deviation of \mathbf{X} is $\sigma_{\mathbf{x}}$ instead of $\mathbf{S}_{\mathbf{X}}$.

Exercise

\mathbf{X} is the number of heads of tossing a coin twice.

- Calculate the expectation $\mathbf{E}(\mathbf{X})$ of \mathbf{X} . Interpret it.
- Calculate the standard deviation $\sigma_{\mathbf{x}}$ of \mathbf{X} . Interpret it.

1.2 Transforming random variables

Linear transformation of random variables

Suppose the distribution of random variable \mathbf{X} is given below.

Value of \mathbf{X}:	x_1	x_2	x_3	\cdots
Probability:	p_1	p_2	p_3	\cdots

\mathbf{Y} is a linear transformation of \mathbf{X} given by $\mathbf{Y} = a\mathbf{X} + b$.

$$\begin{aligned}
 \mathbf{E}(\mathbf{Y}) &= \sum y_i p_i = \sum (ax_i + b)p_i \\
 &= \sum ax_i p_i + \sum bp_i \\
 &= a \sum x_i p_i + b \sum p_i \\
 &= a\mathbf{E}(\mathbf{x}) + b
 \end{aligned}$$

$$\begin{aligned}
 \mathbf{Var}(\mathbf{Y}) &= \sum (y_i - \mu_{\mathbf{Y}})^2 p_i = \sum [(ax_i + b) - (a\mu_X + b)]^2 p_i \\
 &= \sum [ax_i - a\mu_X]^2 p_i \\
 &= a^2 \sum (x_i - \mu_X)^2 p_i = a^2 \mathbf{Var}(\mathbf{X})
 \end{aligned}$$

$$\sigma_Y = \sqrt{\mathbf{Var}(\mathbf{Y})} = \sqrt{a^2 \mathbf{Var}(\mathbf{X})} = |a| \sigma_X$$

As to the relationship between the shape of the distribution of \mathbf{X} and that of \mathbf{Y} , it is easier to consider the continuous case. Multiply \mathbf{X} by a constant only compresses or stretches the probability density curve by a constant. And addition of a constant to \mathbf{X} only shift the probability density curve. Anyway, those linear transformations do not change the shape of the distribution as long as $a > 0$.

If $a > 0$, and the distribution of \mathbf{X} is *symmetric*, *right-skewed*, *left-skewed* or *bell-shaped*, the distribution of \mathbf{Y} is *symmetric*, *right-skewed*, *left-skewed* or *bell-shaped* as well.

The above implies, if $\mathbf{X} \sim \mathbf{N}(\mu, \sigma)$, then $\mathbf{Y} \sim \mathbf{N}(a\mu + b, |a|\sigma)$.

The Baby and the Bathwater

One brand of bathtub comes with a dial to set the water temperature. When the babysafe setting is selected and the tub is filled, the temperature X of the water follows a Normal distribution with a mean of 34°C and a standard deviation of 2°C .

- (a) Define the random variable Y to be the water temperature in degrees Fahrenheit (recall that $F = \frac{9}{5}C + 32$) when the dial is set on babysafe. Find the mean and standard deviation of Y .
- (b) According to Babies R Us, the temperature of a babys bathwater should be between 90°F and 100°F . Find the probability that the water temperature on a randomly selected day when the babysafe setting is used meets the Babies R Us recommendation. Show your work.

1.3 Combining random variables

The Expectation

Suppose the probability model of a random process is given by

S:	O_1	O_2	\cdots
Probability:	p_1	p_2	\cdots

Where $\mathbf{S} = \{O_1, O_2, \cdots\}$ is the sample space.

\mathbf{X} and \mathbf{Y} are two random variables defined on \mathbf{S} , and the mappings are given by

$$\mathbf{X}: \{O_1, O_2, \cdots\} \implies \{x_1, x_2, \cdots\}$$

$$\mathbf{Y}: \{O_1, O_2, \cdots\} \implies \{y_1, y_2, \cdots\}$$

Let $\mathbf{T} = a\mathbf{X} + b\mathbf{Y}$ be a linear combination of \mathbf{X} and \mathbf{Y} , where a and b are constant. Then the distribution of \mathbf{T} is given by

S:	O_1	O_2	\cdots
Probability:	p_1	p_2	\cdots
T:	$ax_1 + by_1$	$ax_2 + by_2$	\cdots

$$\mu_T = \mathbf{E}(\mathbf{T}) = \sum (ax_i + by_i)p_i = a \sum x_i p_i + b \sum y_i p_i = a\mu_X + b\mu_Y$$

In the above deduction, no restriction is applied on the relationship between \mathbf{X} and \mathbf{Y} . Therefore, whether \mathbf{X} and \mathbf{Y} are independent or not, the formula holds.

The variance and standard deviation

Let's consider the same random variable $\mathbf{T} = a\mathbf{X} + b\mathbf{Y}$ as above and \mathbf{X} and \mathbf{Y} are *independent*.

$$\mathbf{P}(\mathbf{T} = ax_i + by_j) = \mathbf{P}(\mathbf{X} = x_i) \cdot \mathbf{P}(\mathbf{Y} = y_j) \quad (\text{by independence})$$

$$\begin{aligned} \mathbf{Var}(\mathbf{T}) &= \sum_i \sum_j [(ax_i + by_j) - \mu_T]^2 \mathbf{P}(\mathbf{T} = ax_i + by_j) \\ &= \sum_i \sum_j [(ax_i + by_j) - (a\mu_X + b\mu_Y)]^2 \mathbf{P}(\mathbf{X} = x_i) \cdot \mathbf{P}(\mathbf{Y} = y_j) \\ &= \sum_i \sum_j [a^2(x_i - \mu_X)^2 + 2ab(x_i - \mu_X)(y_j - \mu_Y) + b^2(y_j - \mu_Y)^2] \\ &\quad \times \mathbf{P}(\mathbf{X} = x_i) \cdot \mathbf{P}(\mathbf{Y} = y_j) \\ &= \sum_i \sum_j a^2(x_i - \mu_X)^2 \mathbf{P}(\mathbf{X} = x_i) \cdot \mathbf{P}(\mathbf{Y} = y_j) \\ &\quad + \sum_i \sum_j 2ab(x_i - \mu_X)(y_j - \mu_Y) \mathbf{P}(\mathbf{X} = x_i) \cdot \mathbf{P}(\mathbf{Y} = y_j) \\ &\quad + \sum_i \sum_j b^2(y_j - \mu_Y)^2 \mathbf{P}(\mathbf{X} = x_i) \cdot \mathbf{P}(\mathbf{Y} = y_j) \\ &= \sum_i a^2(x_i - \mu_X)^2 \mathbf{P}(\mathbf{X} = x_i) \cdot \sum_j \mathbf{P}(\mathbf{Y} = y_j) \\ &\quad + 2ab \sum_i (x_i - \mu_X) \mathbf{P}(\mathbf{X} = x_i) \cdot \sum_j (y_j - \mu_Y) \mathbf{P}(\mathbf{Y} = y_j) \\ &\quad + \sum_j b^2(y_j - \mu_Y)^2 \mathbf{P}(\mathbf{Y} = y_j) \cdot \sum_i \mathbf{P}(\mathbf{X} = x_i) \\ &= \sum_i a^2(x_i - \mu_X)^2 \mathbf{P}(\mathbf{X} = x_i) + \sum_j b^2(y_j - \mu_Y)^2 \mathbf{P}(\mathbf{Y} = y_j) \\ &= a^2 \mathbf{Var}(X) + b^2 \mathbf{Var}(Y) \end{aligned}$$

In the above deduction we used the following identities.

$$\sum_i \mathbf{P}(\mathbf{X} = x_i) = 1$$

$$\sum_i (x_i - \mu_X) \mathbf{P}(\mathbf{X} = x_i) = 0$$

If \mathbf{X} and \mathbf{Y} are *independent*, $\mathbf{T} = a\mathbf{X} + b\mathbf{Y}$, then $\mathbf{Var}(T) = a^2 \mathbf{Var}(X) + b^2 \mathbf{Var}(Y)$.

Combination of independent normal random variables

If **10% condition** ($n \leq 10\% N$) is met, \mathbf{X} and \mathbf{Y} are independent normal random variable and $\mathbf{T} = a\mathbf{X} + b\mathbf{Y}$, then \mathbf{T} is normally distributed, $\mathbf{T} \sim N(\mu_T, \sigma_T)$, where μ_T and σ_T can be calculated by the previous formulas.

Give Me Some Sugar!

Mr. Starnes likes sugar in his hot tea. From experience, he needs between 8.5 and 9 grams of sugar in a cup of tea for the drink to taste right. While making his tea one morning, Mr. Starnes adds four randomly selected packets of sugar. Suppose the amount of sugar in these packets follows a Normal distribution with mean 2.17 grams and standard deviation 0.08 grams.

Whats the probability that Mr. Starness tea tastes right?

(Pay attention to the difference between $X_1 + X_2 + X_3 + X_4$ and $4X$)

Put a Lid on It!

The diameter C of a randomly selected large drink cup at a fast-food restaurant follows a Normal distribution with a mean of 3.96 inches and a standard deviation of 0.01 inches. The diameter L of a randomly selected large lid at this restaurant follows a Normal distribution with mean 3.98 inches and standard deviation 0.02 inches. For a lid to fit on a cup, the value of L has to be bigger than the value of C , but not by more than 0.06 inches.

Whats the probability that a randomly selected large lid will fit on a randomly chosen large drink cup?

1.4 Binomial random variables

Binomial settings

Toss a die 5 times. Let \mathbf{X} be the number of times with points no bigger than 2. in this example, the following conditions are met.

- (1) **Binary.** The outcome for each trial can either be a "success" or a "failure". In this example, "Success" means "throw the die, and it turn out to be either 1 or 2".
- (2) **Independent.** The trials are independent with each other. In this example, all the tosses are independent with each other.
- (3) **Number.** The number of trials are known before hand. In this example, we know in advance that we will toss 5 times.
- (4) **Success.** The probability of success is a constant through out all the trials. In this example the probability of success of each toss is $\frac{1}{3}$.

For easy memorization, we can take the first alphabet of the for conditions, which is "BINS". If those conditions are met, then it is a **binomial setting**. The number of "success" \mathbf{X} is a **binomial random variable**. \mathbf{X} follows a binomial distribution and denoted as

$$\mathbf{X} \sim \mathbf{B}(n, p).$$

Where n is the number of trials, and p is the probability of "success".

Exercise

Here are three scenarios involving chance behavior. In each case, determine whether or not the given random variable has a binomial distribution

- (a) Genetics says that children receive genes from each of their parents independently. Each child of a particular set of parents has probability 0.25 of having type O blood. Suppose these parents have 5 children. Let \mathbf{X} = the number of children with type O blood.
- (b) Shuffle a deck of cards. Turn over the top card. Put the card back in the deck, and shuffle again. Repeat this process 10 times. \mathbf{Y} = the number of aces you observe.
- (c) Shuffle a deck of cards. Turn over the first 10 cards, one at a time. Let \mathbf{Z} = the number of aces you observe.
- (d) Shuffle a deck of cards. Turn over the top card. Put the card back in the deck, and shuffle again. Repeat this process until you get an ace. Let \mathbf{W} = the number of cards required.

Binomial distribution

Toss a die 5 times. Let \mathbf{X} be the number of times with points no bigger than 2. Let's calculate the probability when \mathbf{X} takes different values, as shown in figure 1.1.

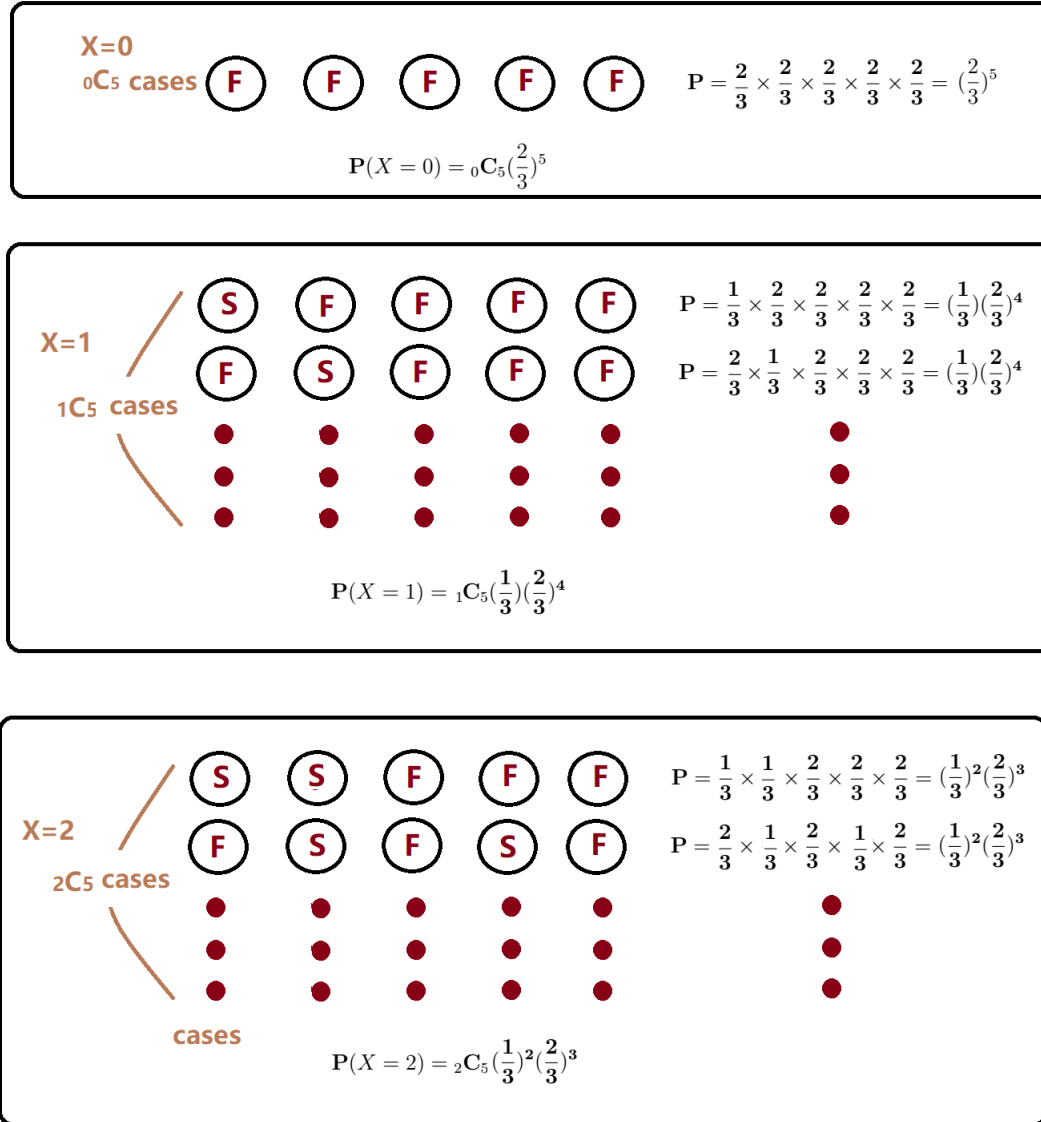


Figure 1.1: The distribution of binomial distribution

Generalize the results in figure 1.1, we have the binomial probability formula for $X \sim B(n, p)$:

$$P(X = k) = {}_kC_n p^k (1 - p)^{(n-k)}, \quad k = 0, 1, \dots, n$$

The binomial random variable $X \sim N(n, p)$ is a discrete random variable and takes values $\{0, 1, 2, \dots, n\}$. The distribution is symmetric and unimodal(one peak), with mode around $\frac{n+1}{2}$.

Mean ,Variance and standard deviation

$X \sim N(n, p)$, means X is the number of success among n trials with the probability of success for each trial p . Define a new random variable X_i as the number of success of the i th trial. The distribution of X_i , $\mathbf{Var}(X_i)$ and $\mathbf{E}(X_i)$ are given below.

Calculate $\mathbf{E}(X_i)$ and $\mathbf{Var}(X_i)$.

X_i	0	1
Probability:	$1 - p$	p

$$\mathbf{E}(X_i) = p \quad \mathbf{Var}(X_i) = p(1 - p).$$

X can be expressed as the sum of X_i s.

$$X = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^n X_i$$

According to formulas in **1.3**,

$$\mathbf{E}(X) = \sum_{i=1}^n np, \quad \mathbf{Var}(X) = \sum_{i=1}^n p(1 - p) = np(1 - p), \quad \sigma_X = \sqrt{np(1 - p)}.$$

Applying the torque

A machine fastens plastic screw-on caps onto containers of motor oil. If the machine applies more torque than the cap can withstand, the cap will break. Both the torque applied and the strength of the caps vary. The capping-machine torque T follows a Normal distribution with mean 7 inch-pounds and standard deviation 0.9 inch-pounds. The cap strength C (the torque that would break the cap) follows a Normal distribution with mean 10 inch-pounds and standard deviation 1.2 inch-pounds.

Now this machine fastened plastic screw-on caps onto 10 containers, What is the probability that no more than 1 caps are broken?

10% condition

In binomial setting, the trials have to be independent, and this strongly restrict the application of binomial distribution. In fact, when those trials are "roughly independent", binomial distribution can be applied and give a relative good result. When those trials are "roughly independent"?

Suppose, we have a bag of balls, the total number is denoted as N , and 20% of them are black, the others are white. Randomly take 2 out of those N balls. Let \mathbf{C}_1 be the color of the first ball, and \mathbf{C}_2 the color of the second ball. Take a look at the following conditional probabilities.

$$\begin{aligned}
 N = 5, \quad & P(\mathbf{C}_2 = \text{black} \mid \mathbf{C}_1 = \text{black}) = 0, \quad P(\mathbf{C}_2 = \text{black} \mid \mathbf{C}_1 = \text{white}) = \frac{1}{4} \\
 N = 10, \quad & P(\mathbf{C}_2 = \text{black} \mid \mathbf{C}_1 = \text{black}) = \frac{1}{9}, \quad P(\mathbf{C}_2 = \text{black} \mid \mathbf{C}_1 = \text{white}) = \frac{2}{9} \\
 N = 20, \quad & P(\mathbf{C}_2 = \text{black} \mid \mathbf{C}_1 = \text{black}) = \frac{3}{19}, \quad P(\mathbf{C}_2 = \text{black} \mid \mathbf{C}_1 = \text{white}) = \frac{4}{19} \\
 N = 40, \quad & P(\mathbf{C}_2 = \text{black} \mid \mathbf{C}_1 = \text{black}) = \frac{7}{39}, \quad P(\mathbf{C}_2 = \text{black} \mid \mathbf{C}_1 = \text{white}) = \frac{8}{39} \\
 N = 80, \quad & P(\mathbf{C}_2 = \text{black} \mid \mathbf{C}_1 = \text{black}) = \frac{15}{79}, \quad P(\mathbf{C}_2 = \text{black} \mid \mathbf{C}_1 = \text{white}) = \frac{16}{79}
 \end{aligned}$$

The difference between $P(\mathbf{C}_2 = \text{black} \mid \mathbf{C}_1 = \text{white})$ and $P(\mathbf{C}_2 = \text{black} \mid \mathbf{C}_1 = \text{black})$ decreases from $\frac{1}{4}$ to $\frac{1}{79}$ as N increases from 5 to 80. This can be interpreted as "the probability that $\mathbf{C}_1 = \text{black}$ becomes less and less influenced by the values of \mathbf{C}_2 ". Loosely speaking, "as N increases, \mathbf{C}_1 and \mathbf{C}_2 become more independent". Thus comes the following.

Let N be the population size and n the sample size. If $n \leq 10\% N$, then the individuals in the sample are roughly independent.

" $n \leq 10\% N$ " is called **the 10% condition**.

Bad Flash Drives

A supplier inspects an SRS of 10 flash drives from a shipment of 10,000 flash drives. Suppose that (unknown to the supplier) 2% of the flash drives in the shipment are defective.

Find the probability that no more than 2 of the flash drives are defective.

Approximation of binomial distribution

In the formula of the binomial distribution

$$\mathbf{P}(X = k) = {}_k\mathbf{C}_n p^k (1 - p)^{(n-k)},$$

as n increase, the difficulty of computation of $\mathbf{P}(X = k)$ increase rapidly. However, the binomial distribution can be approximated by normal distribution if **the large counts condition** is met.

If $X \sim N(n, p)$, $np \leq 10$ and $n(1 - p) \leq 10$, then the distribution of X can be approximated by $N(\mu_X, \sigma_X)$ where $\mu_X = np$, $\sigma_X = \sqrt{np(1 - P)}$.

The condition " $np \leq 10$ and $n(1 - p) \leq 10$ " is called **large counts condition**.

Checking for Survey Error

One way of checking the effect of undercoverage, nonresponse, and other sources of error in a sample survey is to compare the sample with known facts about the population. About 12% of American adults identify themselves as black. Suppose we take an SRS of 1500 American adults and let X be the number of blacks in the sample.

- (a) Show that X is approximately a binomial random variable.
- (b) Check the conditions for using a Normal approximation in this setting.
- (c) Use a Normal distribution to estimate the probability that the sample will contain between 165 and 195 blacks

1.5 Geometry random variables

Take a look at the following settings.

- (1) Roll a die till you get double. \mathbf{X} is the number of rolls
- (2) Toss a coin till you get a head. \mathbf{Y} is the number of tosses.

In the above setting, the outcomes of each outcome is binary, either "success" or "failure". The trials are independent. The probability of "success" among all trials is a constant. However, we don't know the number of trials in advance, and the number of trials are defined as random variables \mathbf{X} and \mathbf{Y} in the above two setting. \mathbf{X} and \mathbf{Y} are called **Geometric random variables**

Deduce the formulas

Suppose \mathbf{X} is a *geometric random variable* with probability of success p .

Find $\mathbf{P}(X = k)$ and $\mathbf{E}(X)$.