Elementary Probability and Statistics

Probability and Statistics play an important role in various disciplines like engineering, biology, economics and finance etc. These notes will introduce the basics of probability, conditional probability, expectation, law of large numbers, central limit theorem, estimation, hypothesis testing. An application to finance (binomial option pricing) is also discussed.

1 Some interesting facts

Probability challenges intuition: If you take a deep breath, there is a better than 99% chance that you will inhale a molecule that was exhaled in the dying Caesar's last breath. Another less morbid example: in a class of 25 students, there is a better than 50% chance that two students will share the same birthday.

Statistics saves money: Motorola is one of the largest producers of electronics equipment, with annual sales exceeding 17 billion dollars. By implementing a quality improvement plan that makes extensive use of statistics, Motorola saved 2.5 billion dollars in a five-year period.

Statistics in the war: During World War II, Allied intelligence specialists wanted to determine the number of tanks Germany was producing. Traditional spy techniques provided unreliable results, but statisticians obtained accurate estimates by analyzing serial numbers on captured tanks. As one example, records show that Germany actually produced 271 tanks in June 1941. The estimate based on serial numbers was 244, but traditional intelligence methods resulted in the extreme estimate of 1550.

Bets in casinos: In casinos, the house advantage is 5.26% for roulette, 5.9% for blackjack, 1.4% for craps, and 3%-22% for slot machines. Some professional gamblers can systematically win at blackjack by using complicated card-counting techniques. Many casinos react by ejecting card counters or by shuffling the decks more frequently.

Are lie detectors accurate?: The following example is fictional. Through accounting procedures, it is known that 10% of the employees of a store are stealing. The manager would like to fire the thieves, but their only tool is a lie detector, which is 80% accurate: if an employee is a thief, he or she will fail the test with probability 0.8; and if an employee is not a thief, he or she will pass the test with probability 0.8. The manager conducts the test for each employee, and some of them fail the lie detector test. The question then is, "How many of the employees who failed are thieves?" A typical response would be that 80% are thieves. As we will see later on, the fact is that most of these employees are honest.

Another similar phenomenon of this type is the high false-positive rate in medical testing.

Simpson's paradox: This is also a fictional example. The following table shows the shooting percentages of Michael Jordan and Dennis Rodman.

	Michael Jordan			Dennis Rodman		
	Attempt	Made	Percentage	Attempt	Made	Percentage
in Seattle	40	12	30%	4	1	25%
in Utah	20	11	55%	2	1	50%
in NY	20	13	65%	10	6	60%
Total	80	36	45%	16	8	50%

Note that even though Jordan had a better shooting percentage than Rodman every night, Rodman beats Jordan in overall shooting percentage.

2 Basic probability

Probability is the science of *uncertainty* with many applications. We start by introducing some basic concepts and examples.

2.1 Definitions and probability rules

Sample Space, Event: *Sample space* (usually denoted by Ω) is the set of all possible outcomes, and *event* is a subset of the sample space. The event could be a single outcome, or a collection of outcomes.

Example: Tossing a coin once, the sample space consists of two outcomes:

$$\Omega = \{H,\ T\}\,;$$

here H = "heads", T = "tails". In general, for tossing a coin n times, the sample space is

$$\Omega = \{(a_1, a_2, \cdots, a_n); \quad a_i = H \text{ or } T\,\},$$

with the total number of 2^n possible outcomes.

Example: Tossing the coin three times, the sample space is

$$\Omega = \{HHH,\ HHT,\ HTH,\ HTT,\ THH,\ THT,\ TTH,\ TTT\}\,.$$

The following events

$$E_1 = \{\text{the second tosses is heads}\} = \{HHH,\ HHT,\ THH,\ THT\}$$

and

$$E_2 = \{\text{exactly 2 out of 3 tosses are heads}\} = \{HHT,\ THH,\ HTH\}\,,$$

are both subsets of the sample space Ω.

Probability Rules: Probability is the quantitative description of possibility. We will let $\mathbb{P}(E)$ denote the *probability of event* E (i.e., the probability that event E happens, or that the outcome belongs to set E). Naturally, it should satisfy the following rules.

1. $0 \le \mathbb{P}(E) \le 1$ for all events E.
2. $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$. Here $\emptyset$ is the null set (a set that contains no element).
3. $\mathbb{P}(E) + \mathbb{P}(E^c) = 1$. Here E^c is the *complement* of E. Intuitively, E^c is the event that E does not happen.
4. $\mathbb{P}(E_1 \text{ or } E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2)$ if E_1 and E_2 are *disjoint* events (i.e. E_1 and E_2 has no common outcome; or, E_1 and E_2 cannot happen together). If we define

$$E_1 \cup E_2 \doteq \{\omega \in \Omega;\ \ \omega \in E_1 \ \text{ or } \ \omega \in E_2\}$$

and

$$E_1 \cap E_2 \doteq \{\omega \in \Omega;\ \ \omega \in E_1 \ \text{ and } \ \omega \in E_2\},$$

it is easy to see that

$$E_1 \cup E_2 = \{E_1 \text{ or } E_2 \text{ happens}\}; \quad E_1 \cap E_2 = \{\text{both } E_1 \text{ and } E_2 \text{ happen}\}.$$

Therefore, we can rewrite this rule as

$$\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) \quad \text{if } E_1 \cap E_2 = \emptyset.$$

5. $\mathbb{P}(E_1 \text{ and } E_2) = \mathbb{P}(E_1) \cdot \mathbb{P}(E_2)$ if event E_1 is *independent* of event E_2. Intuitively, "E_1 and E_2 are independent" means that the knowledge of event E_1 does not affect the probability of E_2, and vice versa. This rule can also be written as

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1) \cdot \mathbb{P}(E_2) \quad \text{if } E_1 \text{ and } E_2 \text{ are independent.}$$

Example: A coin is said to be "fair" if it has a 50% chance to be heads (H) and 50% chance to be tails (T). Also, the successive tosses are assumed to be independent. When a fair coin is tossed 3 times,

1. What is the probability that the outcome is HHH? How about HTT?
   Solution: Both probabilities are $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$.
2. What is the probability of at least 2 heads?
   Solution: The probability is $\mathbb{P}(HHH) + \mathbb{P}(HHT) + \mathbb{P}(HTH) + \mathbb{P}(THH) = \frac{1}{8} \times 4 = \frac{1}{2}$.
3. What is the probability of at least 1 Heads?
   Solution: The probability is $1 - \mathbb{P}(\text{no heads}) = 1 - \mathbb{P}(TTT) = 1 - \frac{1}{8} = \frac{7}{8}$.
4. If the coin is tossed n times, what is the probability of at least 1 Tails?
   Solution: The probability is $1 - \mathbb{P}(\text{no heads}) = 1 - \frac{1}{2^n}$.
5. If the coin is tossed 10 times with 1 heads and 9 tails, what is the probability that the next toss is a heads? Is this probability going to be bigger than, less than or equal 50%.
   Solution: Equal 50%, since the tosses are independent.

Example: What is the probability that in a class of size 25 that at least two students share the same birthday?

Solution: The sample space is

$$\Omega = \{(01/01,\cdots,01/01,01/01),((01/01,\cdots,01/01,01/02),\cdots,(12/31,\cdots,12/31,12/31)\}$$

which contains 365^{25} in total. It is assumed that each day of the year is equally likely, and that the birthdays of different students are independent. In this case each outcome has a probability of $\frac{1}{365^{25}}$. However,

$$\mathbb{P}(\text{at least two students share the same birthday}) = 1-\mathbb{P}(\text{no two students share the same birthday})$$

and the event "no two students share the same birthday" contains exactly

$$365 \cdot 364 \cdot \cdots \cdot (365-24)$$

outcomes (since the first student can have any day, the second must only avoid the birthday of the first student, and so on). Therefore, the probability that at least 2 students share the same birthday is

$$1 - \frac{365 \cdot 364 \cdot \cdots \cdot (365-24)}{365^{25}} = .57.$$

Example (The Flippant Juror): A three-man jury has two members (say, two wise guys) each of whom independently has probability p of making the correct decision and a third member who flips a coin for each decision (majority rules). A one-man jury has probability p to make the correct decision. Which jury has the better probability of making the correct decision?

Solution: The probability of the three-man jury making a correct decision is

$$\begin{aligned}&\mathbb{P}(\text{all three are correct}) + \mathbb{P}(\text{only the two wise guys are correct})\\ &\qquad + \mathbb{P}(\text{only one wise guy and the flippant juror are correct})\\ &= p \times p \times \frac{1}{2} + p \times p \times \frac{1}{2} + 2 \times p \times (1-p) \times \frac{1}{2} = p^2 + p(1-p) = p.\end{aligned}$$

Therefore, the three-man jury and the one-man jury has the same probability to make the correct decision.

2.2 Conditional probability

The *conditional probability* of event E given that event F happened is defined as

$$\mathbb{P}(E \mid F) \doteq \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

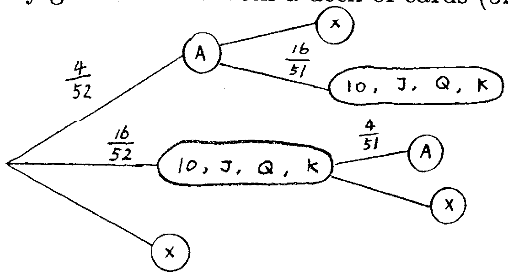
provided $\mathbb{P}(F) > 0$. This definition directly yields

General Multiplication Rule: For an arbitrary pair of events E, F,

$$\mathbb{P}(E \cap F) = \mathbb{P}(F) \cdot \mathbb{P}(E|F) = \mathbb{P}(E) \cdot \mathbb{P}(F|E).$$

Exercise: Suppose E and F are independent. Then $\mathbb{P}(E \mid F) = \mathbb{P}(E)$. This fits intuition well.

Example (Probability trees): What is the probability of getting a "blackjack" if you are randomly given 2 cards from a deck of cards (52 total)?



$$\frac{4}{52}\cdot\frac{16}{51}+\frac{16}{52}\cdot\frac{4}{51}\approx .048$$

Example: A box contain 3 cards. One card is red on both sides (RR), one card is green on both sides (GG), and one card is red on one side and green on the other (RG). One card is selected from the box at random, and the color on one side is observed. If this side is green, what is the probability that the other side of the card is also green?

$$\mathbb{P}(\text{the other side is green}\,|\,\text{one side is green}) = \frac{\mathbb{P}(\text{both sides are green})}{\mathbb{P}(\text{one side is green})}$$

but $\mathbb{P}(\text{both sides are green}) = \frac{1}{3}$, and

$$\begin{aligned}\mathbb{P}(\text{one side is green}) &= \mathbb{P}(\text{the card drawn is GG}) + \\ &\quad\ \mathbb{P}(\text{the card drawn is RG } \textit{and} \text{ you see the green}) \\ &= \frac{1}{3}+\frac{1}{3}\frac{1}{2}=\frac{1}{2}.\end{aligned}$$

Therefore, $\mathbb{P}(\text{the other side is green}\mid\text{one side is green}) = \frac{2}{3}$.

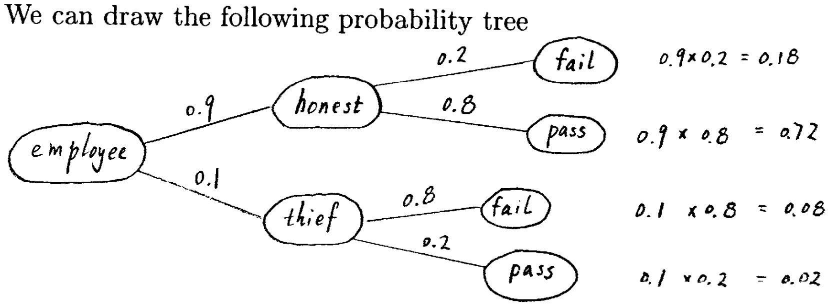
Are lie detectors accurate (cont.): We need to compute the conditional probability of

$$\mathbb{P}(\text{the employee is a thief}\mid\text{the employee fails the test}).$$

By definition, it equals

$$\frac{\mathbb{P}(\text{the employee is a thief and fails the test})}{\mathbb{P}(\text{the employee fails the test})}$$

We can draw the following probability tree



It follows that the numerator is $0.8\cdot 0.1 = 0.08$ and the denominator is $0.8\cdot 0.1+0.2\cdot 0.9 = 0.26$. Therefore, the conditional probability under consideration is

$$\frac{0.08}{0.26} = 31\%.$$

Most of the employees who fail the test are indeed honest!

3 Random variable

Loosely speaking, *random variables* (rv) are just random (uncertain) numerical outcomes. For example, suppose we use X_n to record the outcome from the n-th coin toss, and let $X_n = 1$ if it is a Heads and $X_n = 0$ if it is a Tails. Then X_n is a random variable. If we toss the coin 10 times, the total number of Heads is $X_1 + X_2 + \cdots + X_{10}$, which is also a random variable. Other examples include: the stock index tomorrow, the weight of a person randomly selected.

Random variables that can only take finitely (or more precisely, countably) many possible values are said to be *discrete random variables*; otherwise, they are said to be *continuous random variables.* In the first example above, X_n can only take two possible values $\{0, 1\}$, so it is a discrete random variable. The weight of a person is, however, a continuous random variable. As for the case of stock index, even though the stock index is documented in the unit of 0.01, it is very convenient to treat it as a continuous random variable.

3.1 The distribution of a discrete random variable

Consider a discrete random variable X which can only take finitely many possible values. Its distribution can be described in a table.

X	x_1	x_2	$\cdots$	x_k
Probability	p_1	p_2	$\cdots$	p_k

Here

$$p_i = \mathbb{P}(X = x_i), \qquad \sum_{i=1}^{k} p_i = 1.$$

Expectation & Variance: The *expectation* (mean) of the random variable X is defined as

$$\mu = \mathbb{E}(X) \doteq \sum_{i=1}^{k} x_i \cdot \mathbb{P}(X = x_i) = \sum_{i=1}^{k} x_i p_i.$$

The *variance* of X is defined as the expectation of the random variable $(X - \mu)^2$, or

$$\sigma^2 = \mathrm{Var}(X) \doteq \mathbb{E}\left[(X - \mu)^2\right] = \sum_{i=1}^{k} (x_i - \mu)^2 p_i = \sum_{i=1}^{k} x_i^2 p_i - \mu^2 = \mathbb{E}\left(X^2\right) - (\mathbb{E}X)^2$$

The *standard deviation* of X is

$$\sigma = \sqrt{\sigma^2} = \sqrt{\mathrm{Var}(X)}$$

Independence: Two discrete random variables X and Y are said to be *independent* if

$$\mathbb{P}(X = x_i,\ Y = y_j) = \mathbb{P}(X = x_i) \cdot \mathbb{P}(Y = y_j); \qquad \forall\ i,\ j.$$

The simplest example of discrete random variable is the *Bernoulli random variable*, which can be described as

X	1	0
Probability	p	$1-p$

$0 \le p \le 1.$

For a Bernoulli random variable X, we have

$$\mathbb{E}(X) = p, \qquad \mathrm{Var}(X) = p(1-p)$$

Proposition 1. We have the following rules for mean and variance:

(1) $\mathbb{E}(aX+b) = a\mathbb{E}(X) + b$ for all rv X and constants a, b.

(2) $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$ for all rv X and constants a, b.

(3) $\mathbb{E}(aX+bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ for all rv X, Y and constants a, b.

(4) $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$ for all *independent* rv X, Y.

(5) $\mathrm{Var}(aX+bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y)$ for all *independent* rv X, Y and constants a, b.

Proof: Suppose $\mathbb{P}(X = x_i) = p_i$ and $\sum p_i - 1$. We have

$$\mathbb{E}(aX+b) = \sum (ax_i + b)p_i = a\sum x_i p_i + b\sum p_i = a\mathbb{E}X + b.$$

To prove (3), let $\mathbb{P}(Y = y_j) = q_j$ and $\sum q_j = 1$. It follows that

$$\begin{aligned}
\mathbb{E}(aX+bY) &= \sum_{i,j} (ax_i + by_j)\mathbb{P}(X = x_i,\ Y = y_j) \\
&= a\sum_{i,j} x_i \mathbb{P}(X = x_i,\ Y = y_j) + b\sum_{i,j} \mathbb{P}(X = x_i,\ Y = y_j) \\
&= a\sum_i x_i p_i + b\sum_j y_j q_j = a\mathbb{E}(X) + b\mathbb{E}(Y).
\end{aligned}$$

As to (4), we have

$$\mathbb{E}(XY) = \sum_{i,j} x_i y_j \mathbb{P}(X = x_i,\ Y = y_j) = \sum_{i,j} x_i y_j p_i q_j = \sum_i x_i p_i \cdot \sum_j y_j q_j = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

The proof for (2) and (5) are left as exercise. □

Example: Suppose a box contains 2 red balls, 1 green ball and 1 black ball. What is the distribution of the number of red balls among two balls that are randomly picked out of the box. Let X denote this number. Then (verify!)

X	0	1	2
Probability	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

and therefore

$$\mathbb{E}(X) = 1, \qquad \mathrm{Var}(X) = \frac{1}{3}.$$

Now suppose one gets \$5 for each red ball, and loses \$2 for every non-red ball, and let Y be the net gain. What are the expectation and variance of Y?

In this case, we have

$$Y = 5X - 2(2 - X) = 7X - 4$$

Therefore, $\mathbb{E}(Y) = 7\mathbb{E}(X) - 4 = 3$ and $\mathrm{Var}(Y) = 7^2 \cdot \mathrm{Var}(X) = \frac{49}{3}$.

Binomial distribution

The binomial distribution is another important discrete distribution. Suppose we flip an unfair coin (probability of Heads is p, and probability of Tails is $q = 1 - p$) n times, and assume that the flips are independent. Let

$$X_i = \begin{cases} 1 & ; \quad \text{if the } i\text{-th toss is a Heads} \\ 0 & ; \quad \text{if the } i\text{-th toss is a Tails} \end{cases}, \qquad 1 \leq i \leq n.$$

Then the total number of Heads is

$$S = X_1 + X_2 + \cdots + X_n.$$

Note that $\{X_1, X_2, \cdots, X_n\}$ is a sequence of independent Bernoulli rv with the same distribution

$$\mathbb{P}(X_i = 1) = p, \quad \mathbb{P}(X_i = 0) = q = 1 - p. \quad \forall\, i = 1, 2, \cdots, n;$$

(such a sequence is said to be an iid sequence, here iid means "independent identically and distributed").

The random variable S can take value in the set $\{0, 1, 2, \cdots, n\}$. Its distribution is easy to compute. Recall

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}, \qquad \text{with convention } 0! = 1.$$

Given any integer $0 \leq x \leq n$, the event

$$\{S = x\} = \{x \text{ Heads and } n - x \text{ Tails}\}$$

contains $\binom{n}{x}$ single sequence of outcomes (why?), and each of these sequence has a probability $p^x q^{n-x}$ (why?). Therefore

$$\mathbb{P}(X = x) = \binom{n}{x} \cdot p^x q^{n-x}; \qquad x = 0, 1, 2, \cdots, n.$$

Definition: A random variable X is said to have a *Binomial distribution with parameters n and p*, denoted by $X \sim \text{Binomial}(n; p)$, if the distribution of X is

X	0	1	$\cdots$	x	$\cdots$	n
Probability	q^n	npq^{n-1}	$\cdots$	$\binom{n}{x} \cdot p^x q^{n-x}$	$\cdots$	p^n

In order to compute the expectation and variance of a Binomial distribution, one can either go with brute force, or as follows.

It is by definition that the total number of Heads $S \sim \text{Binomial}(n; p)$. Therefore it suffices to compute the expectation and variance of the random variable S (note that the expectation and variance only depend on the distribution, regardless of the specific underlying random variable). However,

$$\mathbb{E}(S) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) = np$$

and by independence,

$$\text{Var}(S) = \text{Var}(X_1) + \text{Var}(X_2) + \cdots + \text{Var}(X_n) = np(1-p) = npq.$$

In other words, **Binomial**$(n; p)$ **has mean** np **and variance** npq.

Example: A typical American roulette wheel has 38 equally likely numbers. If the player's number comes up, he is paid 35 times his stake and gets his original stake back; otherwise he loses his stake. A player bets one dollar each round. What is the distribution of the number of his winning rounds if he bets n rounds in total, and what is his expected gain after n rounds.

Solution: Let X denote the total number of winning rounds, then $X \sim \text{Binomial}(n; \frac{1}{38})$. His total gain is

$$Y = 35X - (n - X) = 36X - n.$$

Therefore

$$\mathbb{E}(Y) = 36\mathbb{E}(X) - n = 36 \cdot \frac{1}{38}n - n = -\frac{1}{19}n.$$

3.2 The distribution of a continuous random variable

Let X be a continuous random variable taking value on real line $\mathbb{R}$. We will consider only those X for which a probability density function $f(x)$ exists.

Definition: A function f is said to be the *probability density function* (pdf) of the rv X if

(1) $f(x) \geq 0$ for all $x \in \mathbb{R}$.

(2) $\int_{-\infty}^{\infty} f(x)\, dx = 1$

(3) $\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)\, dx$ for all $-\infty \leq a \leq b \leq \infty$.
In other words, the probability is the area under the probability density function.

The function $F(x) \doteq \mathbb{P}(X \leq x)$ for $x \in \mathbb{R}$ is said to be the *cumulative distribution function* (cdf) of the random variable X. It is easy to see that

$$F(x) = \int_{-\infty}^{x} f(y)\, dy; \qquad f(x) = F'(x).$$

Exercise: Show that

$$\mathbb{P}(X = a) = 0 = \mathbb{P}(X = b); \;\; \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < x \leq b).$$

Expectation & Variance: The *expectation* (mean) of X is defined as

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx.$$

In general, the expectation of the random variable $Y = g(X)$, where g is an arbitrary function, is

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f(x)\, dx.$$

The *variance* of X is defined as

$$\sigma^2 = \mathrm{Var}(X) = \mathbb{E}\left[(X-\mu)^2\right] = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

Exercise: Verify the last equality.

Proof: Indeed, the variance equals

$$\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = \int_{-\infty}^{\infty} x^2 f(x)\, dx - 2\mu \int_{-\infty}^{\infty} x f(x)\, dx + \mu^2 \int_{-\infty}^{\infty} f(x)\, dx = \mathbb{E}(X^2) - (\mathbb{E}X)^2.$$

Proposition 2. The rules for expectation and variance given by Proposition 1 also hold for continuous random variables.

Exponential random variable

A random variable X is said to be an *exponential random variable with rate* λ, if its probability density funciton is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & ; \quad x \geq 0 \\ 0 & ; \quad x < 0 \end{cases}$$

The mean and variance of X are (exercise!)

$$\mathbb{E}(X) = \frac{1}{\lambda}, \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

The exponential distribution has the so-called *memoryless property*: for any $t \geq 0$, $h \geq 0$,

$$\mathbb{P}(X \geq t+h \mid X \geq t) = \frac{\mathbb{P}(X \geq t+h)}{\mathbb{P}(X \geq t)} = \frac{\int_{t+h}^{\infty} \lambda e^{-\lambda x}\, dx}{\int_{t}^{\infty} \lambda e^{-\lambda x}\, dx} = \frac{e^{-\lambda(t+h)}}{e^{-\lambda t}} = e^{-\lambda h} = \mathbb{P}(X \geq h).$$

Exercise: Suppose $X_1, X_2, \cdots, X_n$ is a sequence of iid exponential random variables with rate λ. Show that $X = \min(X_1, X_2, \cdots, X_n)$ is an exponential random variable with rate $n\lambda$.

Proof: We have for all $x \geq 0$ that

$$\mathbb{P}(X > x) = \mathbb{P}(X_1 > x,\, X_2 > x,\, \cdots,\, X_n > x) = \mathbb{P}(X_1 > x)\cdot\mathbb{P}(X_2 > x)\cdots\mathbb{P}(X_n > x) = e^{-n\lambda x}.$$

Therefore, the cdf of X is

$$F(x) = \mathbb{P}(X \leq x) = 1 - \mathbb{P}(X > x) = 1 - e^{-n\lambda x}$$

and its pdf is

$$f(x) = F'(x) = n\lambda e^{-n\lambda x}, \quad \forall\, x \geq 0.$$

Uniform distribution

Let a, b be two given real numbers such that $a < b$. We say a random variable X is *uniformly distributed on interval* (a, b) if the pdf is

$$f(x) = \begin{cases} \frac{1}{b-a} & ; \quad x \in (a, b) \\ 0 & ; \quad x \notin (a, b). \end{cases}$$

Standard normal (Guassian) distribution

The normal distribution is the single most important distribution in probability and statistics. A random variable Z is said to be a *standard normal* random variable if the pdf of Z is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \text{or equivalently} \quad \mathbb{P}(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx := \Phi(z).$$

There is no explicit form for $\Phi(z)$, the cumulative distribution function of standard normal. However, the function can be numerically evaluated, and a table is attached to the end of the lecture notes.

Exercise: Show that $\mathbb{P}(Z \geq z) = 1 - \Phi(z) = \Phi(-z) = \mathbb{P}(Z \leq -z)$ for all z.

Exercise: Show that for a standard normal random variable Z, we have

$$\mathbb{E}(Z) = 0, \quad \text{Var}(Z) = 1.$$

We usually the use notation $Z \sim N(0, 1)$ to indicate that Z is a standard normal random variable.

General normal (Guassian) distributions

A random variable X is said to be a *normal* random variable if the pdf of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \forall\, x \in \mathbb{R},$$

for some constants $\mu \in \mathbb{R}$, $\sigma > 0$. We usually use the notation $X \sim N(\mu, \sigma^2)$. The standard normal is the special case with $\mu = 0$, $\sigma = 1$.

Exercise: Show that the above normal random variable X has $\mathbb{E}(Z) = \mu$, $\text{Var}(Z) = \sigma^2$.

Normal distribution has some very important properties. One of them is the following standardization.

Proposition 3. Suppose $X \sim N(\mu, \sigma^2)$. Then $aX + b \sim N(a\mu + b, a^2\sigma^2)$ for arbitrary constants $a \neq 0, b$. In particular,

$$Z \doteq \frac{X - \mu}{\sigma}$$

is a standard normal random variable.

Proof: We will give proof for $a > 0$, the case $a < 0$ is similar and thus omitted. Let $Y = aX + b$. It follows that the cumulative distribution function of Y is

$$F(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\left(X \leq \frac{y-b}{a}\right) = \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}a\sigma} e^{-\frac{[t-(a\mu+b)]^2}{2(a\sigma)^2}}\, dt;$$

here the last equality follows from a change of variable $x = \frac{t-b}{a}$. Therefore, the pdf for Y is

$$f(y) = F'(y) = \frac{1}{\sqrt{2\pi}a\sigma} e^{-\frac{[y-(a\mu+b)]^2}{2(a\sigma)^2}}, \quad \text{or } Y \sim N(a\mu + b, a^2\sigma^2).$$

Example: Suppose $X \sim N(1, 4)$. Find the probability $\mathbb{P}(0 < X < 3)$ using the table of standard normal distribution.

Solution: Note that the table gives the value for all $\Phi(x)$ with $x \geq 0$. Since $Z = \frac{X-1}{2}$ is a standard normal, we have

$$\begin{aligned}
\mathbb{P}(0 < X < 3) &= \mathbb{P}\left(\frac{0-1}{2} < \frac{X-1}{2} < \frac{3-1}{2}\right) = \mathbb{P}(-0.5 < Z < 1) \\
&= \mathbb{P}(Z < 1) - \mathbb{P}(Z \leq -0.5) \\
&= \Phi(1) - \Phi(-0.5).
\end{aligned}$$

Checking the table, we have

$$\Phi(1) = 0.8413 \quad \text{and} \quad \Phi(-0.5) = 1 - \Phi(0.5) = 1 - 0.6915 = 0.3085.$$

Hence $\mathbb{P}(0 < X < 3) = 0.8413 - 0.3085 = 0.5328$.

We will finish this section by stating the following theorem without proof.

Proposition 4. Suppose $\{X_1, X_2, \cdots, X_n\}$ is a sequence of *independent* normal random variables with $X_i \sim N(\mu_i, \sigma_i^2)$. Then

$$\sum_i X_i \sim N\left(\sum_i \mu_i, \sum_i \sigma_i^2\right).$$

4 Law of large numbers and central limit theorem

Law of large numbers (LLN) and central limit theorem (CLT) are two fundamental results concerned with the asymptotic behavior of summations of iid random variables.

Theorem 1. Suppose $\{X_1, X_2, \cdots\}$ is a sequence of iid random variables with $\mathbb{E}(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$. Let $S_n \doteq \sum_{i=1}^n X_i$. Then, as $n \to \infty$,

$$\frac{S_n}{n} \to \mu \qquad \text{(LLN)}$$

and

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \Rightarrow N(0,1) \qquad \text{(CLT)}$$

Remark: The convergence "$\Rightarrow$" in the central limit theorem means that, as $n \to \infty$, the distribution of $\frac{S_n - n\mu}{\sqrt{n}\sigma}$ converges to that of a standard normal. In other words, when n is large, the distribution of $\frac{S_n - n\mu}{\sqrt{n}\sigma}$ is approximately $N(0,1)$.

Example: Suppose a fair coin is flipped 100 times and assume that the flips are independent. What (approximately) is the probability of having between 45 and 55 Heads?

Solution: Let

$$X_i = \begin{cases} 1 & ; \quad \text{if the } i\text{-th toss is a Heads} \\ 0 & ; \quad \text{if the } i\text{-th toss is a Tails} \end{cases}.$$

Then $\{X_1, X_2, \cdots, X_{100}\}$ is a sequence of iid Bernoulli random variables. The total number of Heads is

$$X = X_1 + X_2 + \cdots + X_{100},$$

and $X \sim$ Binomial(100; 0.5). We could compute the desired probability using

$$\sum_{x=45}^{55} P(X = x) = \frac{1}{2^{100}} \sum_{x=45}^{55} \binom{100}{x}.$$

This is an extremely tedious computation. Instead, one can get an good (i.e., approximate but accurate) answer by using the CLT. We know $E(X_i) = 1/2$ and $V(X_i) = 1/4$ so by the CLT

$$\frac{1}{\sqrt{100 \cdot 1/4}} \sum_{i=1}^{100} (X_i - 1/2) \sim \text{Normal}(0,1).$$

Or

$$\frac{X - 50}{5} \sim \text{Normal}(0,1).$$

Now if Z is a standard normal then

$$\begin{aligned} \mathbb{P}(45 \le X \le 55) &= \mathbb{P}\left(\frac{45-50}{5} \le \frac{X-50}{5} \le \frac{55-50}{5}\right) \\ &= \mathbb{P}(-1 \le Z \le 1) \\ &= \Phi(1) - \Phi(-1) \\ &= .6826. \end{aligned}$$

The last number comes from our table.

5 Some challenging probability problems

An experiment in personal taste for money: (a) A box contains 10 black balls and 10 white balls, identical except for color. You choose "black" or "white". One ball is drawn at random, and if the color matches your choice, you get \$100, otherwise nothing. How much money are you willing to pay to play this game? The game will be played just once.

(b) A friend of yours has available many black and many white balls, and he puts them into the box to suit himself. Like in (a), you choose "black" or "white". One ball is drawn at random, and if the color matches your choice, you get \$100, otherwise nothing. How much money are you willing to pay to play this game? The game will be played just once.

Solution: No one can say what amount is appropriate for you to pay for either game. Even though your expected value in game (a) is \$50, you may not be willing to pay anything near \$50 to play it. The possible loss of \$20 or \$30 may mean too much to you. Let us suppose you decided to offer \$1 to play game (a).

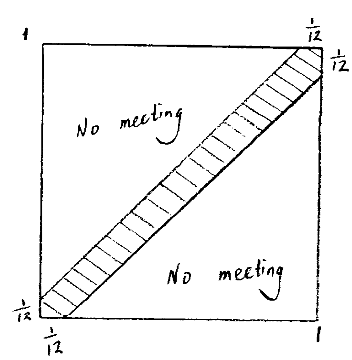
What we can say is that you should be willing to pay at least as much to play game (b) as game (a). You can always choose your color at random by tossing a coin and thus ensure that you have a fifty-fifty chance of being right, and therefore an expectation of \$50. Furthermore, if you have some information about your friend's preference, you can take advantage of that to improve your chances.

Should I switch: The host of this game will present you three boxes. You are told that one of these three boxes is a prize-box that contains some handsome cash prize, while the other two are empty. First you will randomly pick one box. The host, who knows exactly which box is the prize-box, will then open an empty box from the remaining two. Now you are given a choice: (a) you can stick with your original pick, (b) you can switch to the other box. If you pick out the prize-box, you win. Will you go with (a) or (b)? Or do you think it does not make a difference?

The hurried duelers: Duels in the town of Discretion are rarely fatal. There, each contestant comes at a random moment between 5am and 6am on the appointed day and leaves exactly 5 minutes later, honor served, unless his opponent arrives within the time interval and then they fight. What fraction of duels lead to violence?

Solution: Let X and Y denote the time of the arrivals. The shaded region below shows the arrival times for which the duelists meet. The probability that they do not meet is $(\frac{11}{12})^2$, and so the fraction of duels in which they meet is

$$1-\left(\frac{11}{12}\right)^2=\frac{23}{144}\approx\frac{1}{6}.$$



Gambler's ruin problem: Player A has a dollars, and player B has b dollars. Each play gives one of the players \$1 from the other. Both players have a fifty-fifty chance to win each play. They play until one is bankrupt. What is that chance that player M wins?

Solution: Let X_i the money player A gets in the i-th play, and thus

$$S_n = a + \sum_{i=1}^{n} X_i, \qquad S_0 = a$$

will be the total amount of money player A has after n-th play. When S_n becomes 0, player A goes bankrupt, and player B goes bankrupt when S_n is $a+b$. By assumption

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}.$$

In order to solve the problem we define

$$P_x \doteq \mathbb{P}(\text{player } A \text{ wins when player } A \text{ has } x \text{ dollars to start with}); \qquad 0 \le x \le a+b.$$

Immediately $P_0 = 0$ and $P_{a+b} = 1$. For all $0 < x < a+b$, we have

$$\begin{aligned} P_x &= \mathbb{P}(\text{player } A \text{ wins after starting with } x \,|\, X_1 = 1)\mathbb{P}(X_1 = 1) \\ &\quad + \mathbb{P}(\text{player } A \text{ wins after starting with } x \,|\, X_1 = -1)\mathbb{P}(X_1 = -1) \\ &= \frac{1}{2}P_{x+1} + \frac{1}{2}P_{x-1}; \qquad \forall 0 < x < a+b. \end{aligned}$$

This implies that $\frac{1}{2}P_{x+1} - \frac{1}{2}P_x = \frac{1}{2}P_x - \frac{1}{2}P_{x-1}$, and so

$$P_1 - P_0 = P_2 - P_1 = \cdots = P_{a+b} - P_{a+b-1}.$$

But $P_{a+b} = 1$ and $P_0 = 0$. It follows that

$$P_x = \frac{x}{a+b}.$$

In particular, the probability asked for is

$$P_a = \frac{a}{a+b}.$$

What's this probability: Suppose X and Y are two independent, identically distributed continuous random variable. We also assume that X and Y are both *symmetric*, in other words, their probability density functions are even-symmetric. Find the probability $\mathbb{P}(X > 0, X+Y > 0)$.

Solution: The probability is $\frac{3}{8}$. Indeed,

$$\begin{aligned} \mathbb{P}(X > 0, X+Y > 0) &= \mathbb{P}(X > 0, Y > 0) + \mathbb{P}(X > 0, Y < 0, X+Y > 0) \\ &= \frac{1}{4} + \mathbb{P}(Y < 0, X+Y > 0) \\ &= \frac{1}{4} + \mathbb{P}(X < 0, X+Y > 0) \\ &= \frac{1}{4} + \mathbb{P}(X+Y > 0) - \mathbb{P}(X > 0, X+Y > 0) \\ &= \frac{1}{4} + \frac{1}{2} - \mathbb{P}(X > 0, X+Y > 0). \end{aligned}$$

Or,

$$2\mathbb{P}(X > 0, X+Y > 0) = \frac{3}{4} \quad \Rightarrow \quad \mathbb{P}(X > 0, X+Y > 0) = \frac{3}{8}.$$

6 Estimation with confidence

We will use the following example to illustrate the main idea of *estimation* and the associated *confidence interval.*

Example: Suppose that the distribution of income per family in the city has a standard deviation $\sigma = 2000$ dollars. Our interest is to estimate the average income per family. To this end, a random sample of size $n = 400$ families are selected. Let X_i denote the income for the i-th family in the sample.

A natural candidate to estimate the *population mean* (in this case, the average income per family in the city) is the *sample mean*, which is denoted by

$$\bar{X} \doteq \frac{1}{n}(X_1 + X_2 + \cdots + X_n), \qquad n = 400.$$

Also a natural question to ask is that "How far is our estimate $\bar{X}$ away from the true population mean (say, μ)?" This question can be rephrased as, for example, "What is the probability that the estimate is correct within \$100?" (of course, you can replace \$100 with any amount you like). The answer of this question follows readily from central limit theorem. Indeed, CLT implies that approximately

$$\bar{X} \sim \; N(\mu, \frac{\sigma}{\sqrt{n}}) = N(\mu, 100)$$

Then

$$\mathbb{P}(|\bar{X} - \mu| \leq 100) = \mathbb{P}\left(\left|\frac{\bar{X} - \mu}{100}\right| \leq 1\right) = \mathbb{P}(|Z| \leq 1) \approx 68\%.$$

It is also equivalent to say that "The population mean falls within the interval $(\bar{X} - 100, \bar{X} + 100)$ with probability 68%." If we call this 68% as *confidence level*, then $(\bar{X} - 100, \bar{X} + 100)$ is the *confidence interval with confidence level 68%* (or, the 68% confidence interval).

The meaning of a confidence interval is a bit subtle in practice. Suppose now we have this random sample of size 400 at hand, and the sample mean is, after computation, $\bar{X} = \$28,000$. Therefore, we obtain $(27900, 28100)$ as the confidence interval with confidence level 68%.

Question: Can we say that "The population mean falls within the interval (27900, 28100) with probability 68%"?

The answer is "No." Once the confidence interval is calculated for a specific set of data it is is fixed. The population mean is also a *fixed* number – it will either fall within or outside of the interval (27900, 28100).

Confidence interval should be understood as an *random interval*, and the specific "confidence interval" obtained from a specific sample is just an realization. In other words, a confidence interval is an interval computed from sample data by a method that has probability C (the confidence level) of producing an interval containing the population mean (or other parameters of interest).

Example: In order to estimate the population mean, each one of 4 students in the class is asked to (independently) select a random sample of size n from the population, and then construct the 90% confidence interval. What is the probability that all these four confidence intervals will cover the true population mean?

Solution: Let

$$X_i = \begin{cases} 1 & ; \quad \text{if the confidence interval by the } i\text{-th student covers the true population mean} \\ 0 & ; \quad \text{otherwise} \end{cases}$$

Then (X_1, X_2, X_3, X_4) are iid Bernoulli random variables with $\mathbb{P}(X_i = 1) = 0.9$, and $X = X_1 + X_2 + X_3 + X_4 \sim \text{Binomial}(4; 0.9)$. The probability in question is

$$\mathbb{P}(X = 4) = (0.9)^4 \approx .656 \ .$$

General procedure to construct confidence interval for population mean: Suppose that the population standard deviation σ is known, and we are interested in the estimation of population mean μ (which is unknown). Randomly select an sample of size n from the population, and let $\bar{X}$ be the sample mean.

In order to construct the confidence interval of confidence level $0 < C < 1$, we first determine the number z^* such that

$$\mathbb{P}(-z^* \le Z \le z^*) = C, \qquad \text{where } Z \text{ is a standard normal}$$

using the normal table. Then the confidence interval is

$$\left(\bar{X} - z^* \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z^* \frac{\sigma}{\sqrt{n}}\right).$$

Here $z^* \frac{\sigma}{\sqrt{n}}$ is said to be the *margin of error.*

Exercise: Show that

$$\mathbb{P}(z \ge z^*) = \frac{1 - C}{2}.$$

This result will help you find z^* from the table. Also justify from CLT that the interval covers the population mean with probability (approximately) C.

Some observations: The length of the confidence interval is $2z^* \frac{\sigma}{\sqrt{n}}$.

- The length of the confidence interval decreases if the sample size n increases. And if you want to reduce the length by half, you need to quadruple the sample size.
- The length of the confidence interval increases if the confidence level C increases.
- The ideal case would be to make the confidence level as high as possible, while the margin of error as small as possible. But these two things conflict each other (with fixed sample size).

Example: In an area of a large city in which houses are rented, an economist wishes to estimate the average monthly rental correct to within \$20, with probability 95%. If the economist guesses that σ is about \$60, how many houses must be included in his sample?

Solution: The z^* corresponding to confidence level 95% is determined by

$$\mathbb{P}(-z^* \le Z \le z^*) = 95\% \quad \rightarrow \quad z^* \approx 2.$$

The margin of error is then

$$z^* \frac{\sigma}{\sqrt{n}} = \frac{120}{\sqrt{n}}.$$

The requirement is

$$\frac{120}{\sqrt{n}} \le 20 \quad \Rightarrow \quad n \ge 36,$$

or at least 36 houses should be included in the sample.

7 Hypothesis testing

Let us begin with an example of coin tossing.

Example: Suppose we toss a coin 10 times and observe 9 Heads. Question: is this coin a fair one?

Solution: Assume this is a fair coin. Let X be the total number of Heads in 10 tosses, and it follows that $X \sim \text{Binomial}(10; 0.5)$. We would expect that the sample proportion of Heads is likely to be around 0.5. However, our observed proportion is 0.9. The probability of an observation as extreme as or more extreme than our current one is

$$\mathbb{P}\left(\left|\frac{X}{10} - 0.5\right| \ge \left|\frac{9}{10} - 0.5\right|\right) = \mathbb{P}(X = 10) + \mathbb{P}(X = 9) + \mathbb{P}(X = 1) + \mathbb{P}(X = 0) = 0.02.$$

What does this imply? There are two explanations.

(a) The assumption is right (i.e., the coin is fair) and an extreme event (i.e., low-probability event) is observed.

(b) The assumption is wrong (i.e., the coin is unfair).

Both explanations are may be correct. However, the probability (2%) in (a) is so small[1] that (b) seems more appealing, that is, we are more inclined to claim the coin is *not* fair. In other words, the data is more *compatible* with the conclusion of an unfair coin.

General procedure of hypothesis testing

(1) State the *null hypothesis* H_0 and the *alternative hypothesis* H_a: the null hypothesis is the statement that we try to find evidence *against* (or *reject*). The alternative hypothesis is a statement opposite to the null hypothesis. In the preceding example,

$$H_0: \quad \text{"it's a fair coin"}; \qquad H_a: \quad \text{"it's an unfair coin"}.$$

[1]the definition of "small" depends on our subjective feeling; usually, people regard a probability of no more than 5% as "small". 10% and 1% are occasionally used. Unless specified, we will use the standard 5% (and this is called the *significance level*).

(2) Construct a test statistic – a measurement of the compatibility of the null hypothesis and the data. Usually it will be the same as the statistic we would use in the estimate of the parameter appearing in the null hypothesis.

(3) Calculate the *P-value*: The probability of an outcome as extreme or more extreme than the actually observed outcome *under null hypothesis*. The smaller the P-value, the stronger the evidence is against the null hypothesis. In the preceding example the P-value is 0.02.

(4) Compare the P-value with the pregiven significance level α (usually $\alpha = 0.05$) – if P-value $\leq \alpha$, we say the data is *statistically significant at level* α, and we reject null hypothesis. If P-value $> \alpha$, we do not have enough evidence to reject H_0, and we accept H_0.

Example – testing population mean: Suppose we want to test the hypothesis that the mean μ of a population (suppose the standard deviation σ is known) equals some specific number μ_0. In this case, the null hypothesis and alternative hypothesis are

$$H_0: \quad \mu = \mu_0; \qquad H_a: \quad \mu \neq \mu_0.$$

To this end, a random sample of size n will be picked from the population. Denote the sample mean by $\bar{X}$. It is not difficult to see from the CLT that $\bar{X}$ is (approximately) normally distributed with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. If the null hypothesis is true, the *test-statistic*

$$Z \doteq \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}$$

is a standard normal $N(0, 1)$.

Suppose the actual observation turns out to be $\bar{x}$ with the corresponding $z = \sqrt{n}(\bar{x} - \mu_0)/\sigma$. Then the P-value is $2\mathbb{P}(Z \geq |z|)$ (why?). The null hypothesis H_0 is accepted if the P-value is bigger than α; and rejected otherwise.

For example, we hope to show that a random number generator produces a different proportion of odd and even numbers with significance level .01 ($\alpha = .01$, since we wish to be quite sure of our conclusion). We observe the parity (1 for odd, 0 for even) of $n = 1000$ numbers. The data model is $\{X_i\}$ iid with

$$P(X_i = 1) = p, \qquad P(X_i = 0) = 1 - p.$$

If the random number generator is working properly then p should be 1/2. Our hypotheses are

$$H_0: \ p = \frac{1}{2}, \qquad H_a: \ p \neq \frac{1}{2}.$$

Under the assumption H_0 we have $\mu = \mathbb{E}(X_i) = 1/2$ and $\sigma^2 = V(X_i) = 1/4$, so

$$\begin{aligned} Z &= \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu) \\ &= \frac{\sqrt{n}}{.5}(\bar{X} - .5) \\ &\sim \text{Normal}(0, 1) \end{aligned}$$

from central limit theorem.

Suppose the sample mean based on the 1000 observations is $\bar{X} = .48$. Substitution gives, in this case,

$$z = \frac{\sqrt{1000}}{.5}(.48 - .5) = -1.26.$$

Therefore the P-value is

$$2\mathbb{P}(Z \geq |-1.26|) \sim .21$$

Since P-value is bigger that α, we cannot reject H_0 — we cannot refute that $p = 1/2$ with the evidence providence by data.

Note that in hypothesis testing, the null hypothesis H_0 is accepted or rejected according to the relative sizes of the P-value and the significance level. There are two possible errors associated with the hypothesis testing:

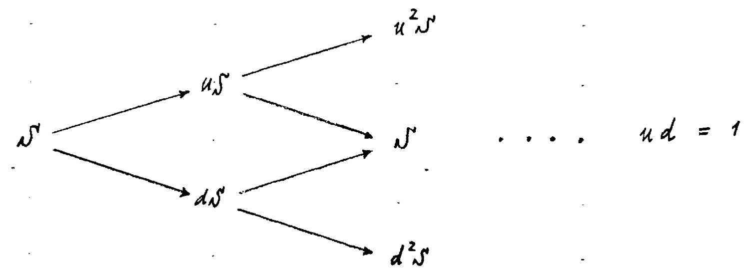
(1) *type*-I *error*. The null hypothesis H_0 is true, but we reject it from the sample.

(2) *type*-II *error*. The null hypothesis H_0 is not true, but we accept the null hypothesis H_0.

Type-I error is always no more than the significance level α (usually they are equal, why?). Type-II error is usually denoted by β, and $1 - \beta$ is said to be the *power* of the test.

8 Application to mathematical finance

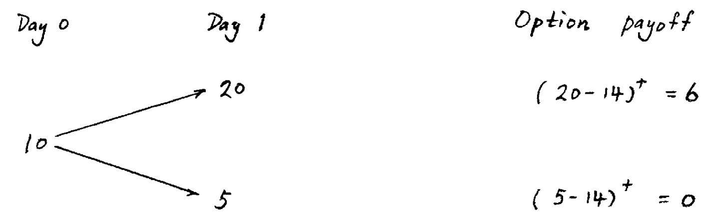
In this section, we will discuss the *Binomial asset pricing model* that has been widely used in financial practice. Stock prices are modeled in discrete time, assuming that at each time step, the stock price will change to one of two possible values. See the following figure for illustration.



Even though the dynamics of the binomial pricing model are simple compared to the real-world stock price movement, it provides a good approximation to continuous-time models with sufficiently many time steps. Besides its advantage of computational tractability, it also help illustrate the idea of "arbitrage pricing" and "risk-neutral pricing".

8.1 A Preliminary Example

Consider the following one-period binomial pricing model. The movement of the stock price is indicated in the following figure.



Suppose we buy a share of *European call option* at day 0 with strike price $K = 14$ dollars and expiration time day 1. That is, the payoff from exercising the option is

$$Y = (S_1 - K)^+ = \max\{S_1 - K, 0\}.$$

Remark 1. The holder of the call-option has the *right*, not the obligation (hence the name "option") to buy a share of stock at the strike price . A European put-option gives the holder the right to sell a share of stock with strike price K; that is, the payoff is

$$Y = (K - S_1)^+ = \max\{K - S_1, 0\}.$$

Remark 2. The name "European" means that the option can only be exercised at the "expiration date". On the contrast, "American" options can be exercised at any time before or at the expiration date. The price (value) of the American option is obviously higher than its European counterpart.

We also assume that the interest rate is zero, that is, \$1 today is worth \$1 tomorrow. Now the question is: what should be the price (say, p) of this call-option at day 0?

Arbitrage Free Principle: Suppose at day 0, we construct a portfolio by adding x share of stock besides the call-option. The value of x is yet to be determined. This portfolio is worth $10x+p$ at time day 0. Now at day 1, the portfolio is worth either $20x+6$ (if the stock price goes up to \$20 with \$6=\$20-\$14) or $5x$ (if the stock price comes down to \$5). However, if we pick x so that

$$20x + 6 = 5x \quad \Rightarrow \quad x = -\frac{2}{5} = -0.4,$$

This portfolio yields a riskless payoff of $20x + 6 = 5x = -2$ dollar at day 1. The arbitrage free principle says that

$$10x + p = -2 \quad \Rightarrow \quad -4 + p = -2 \quad \Rightarrow \quad p = 2.$$

That is, the option is worth \$2 at day 0.

Observation: We did not specify real-life probability for the two possible outcomes. In other words, whatever the real probabilities are, the option price is always \$2.

Delta: The quantity $x = -0.4$ is called "delta" in lots of occasions. Its role is to "hedge" away the risk ("delta hedging").

Risk-neutral Probability: The risk-neutral probability is an (artificial) probability measure, under which the expected return of the option payoff equals the option price. In this example, one can find the risk-neutral probability by solving the equations

$$a + b = 1, \quad 6a + 0b = 2 \quad \Rightarrow \quad a = \frac{1}{3}, \; b = \frac{2}{3}.$$

Observation: Under the risk-neutral probability, the expected stock return is

$$\frac{1}{3} \cdot 20 + \frac{2}{3} \cdot 5 = 10 = S_0.$$

This is indeed a general phenomenon (Martingale property).

Observation: No matter how the option payoff changes, we always get the *same* risk-neutral probability. For example, suppose the strike price is now $K = 15$. Constructing portfolio with value $10x + p$ at day 0, with riskless value

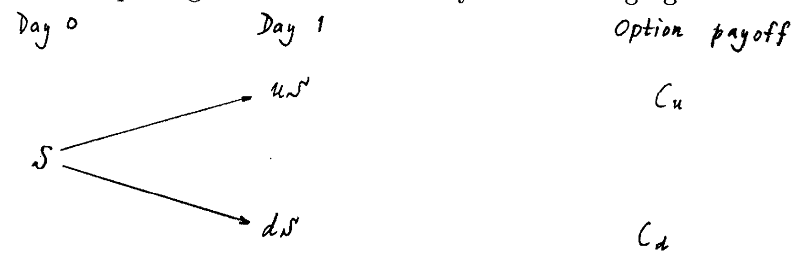
$$20x + 5 = 5x \quad \Rightarrow \quad x = -\frac{1}{3} \quad \Rightarrow \quad 10x + p = -\frac{5}{3} \quad \Rightarrow \quad p = \frac{5}{3},$$

which equals the expected return under probability measure $(\frac{1}{3}, \frac{2}{3})$:

$$\frac{1}{3} \cdot 5 + \frac{2}{3} \cdot 0 = \frac{5}{3}.$$

8.2 General Pricing Formulae

A general one-period binomial pricing model is indicated by the following figure.



Here C_u and C_d are two non-negative constants, representing the option payoff when the stock price goes up or down, respectively.

Suppose the interest rate is r. That is, 1 dollar at day 0 is worth $R = 1 + r$ dollars at day 1. We also assume the bounds

$$d < R < u.$$

Still let p denote the price of the option. Then we can construct a portfolio with x stocks such that it has a riskless return, or,

$$uS \cdot x + C_u = dS \cdot x + C_d \quad \Rightarrow \quad x = -\frac{C_u - C_d}{uS - dS}.$$

The price of the option is

$$(S \cdot x + p)R = uS \cdot x + C_u \quad \Rightarrow \quad p = \frac{1}{R}\left(\frac{R-d}{u-d} \cdot C_u + \frac{u-R}{u-d} \cdot C_d\right)$$

The option price can also be written as

$$p = \frac{1}{R}\mathbf{E}^{\mathbb{P}^*}[C],$$

where $\mathbb{P}^*$ is the *risk-neutral probability*

$$\left(\frac{R-d}{u-d}, \frac{u-R}{u-d}\right),$$

Note that the risk-neutral probability is independent of the option payoff, and the option price is the discounted expected payoff under risk neutral probability.

Furthermore, the discounted expected value of the stock at day 1 is

$$\frac{1}{R}\left(\frac{R-d}{u-d} \cdot uS + \frac{u-R}{u-d} \cdot dS\right) = S,$$

or

$$\mathbf{E}^{\mathbb{P}^*}\left[R^{-1}S_1 \,|S_0\right] = S_0.$$

This indeed says that the discounted stock price is a martingale under risk-neutral measure.

Replication: Let p be the price for the option with payoff C. Then with initial wealth p, one can construct a portfolio so that its value at day 1 completely replicates the option payoff. Indeed, put

$$x = \frac{C_u - C_d}{u - d}$$

amount of money into the stock and the rest $p - x$ amount of money into the bank. Such a portfolio will generate

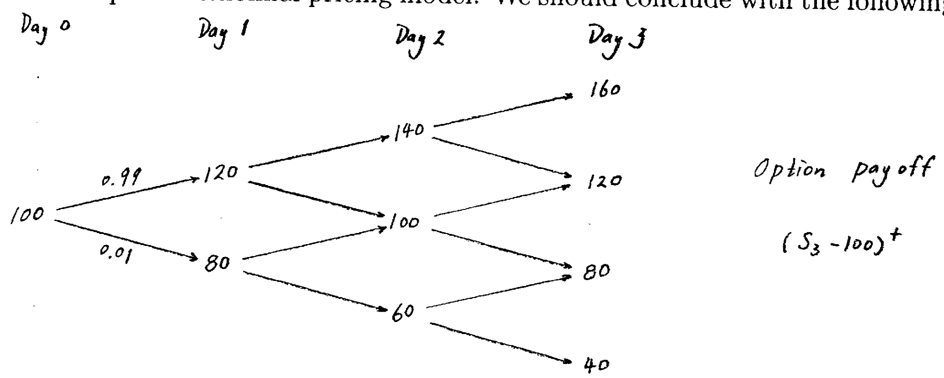
$$uS \cdot \frac{x}{S} + (p - x)R = (u - R)x + pR = \frac{u - R}{u - d}(C_u - C_d) + \left(\frac{R - d}{u - d} \cdot C_u + \frac{u - R}{u - d} \cdot C_d\right) = C_u$$

when the stock price is uS at day 1, and generate

$$dS \cdot \frac{x}{S} + (p - x)R = (d - R)x + pR = \frac{d - R}{u - d}(C_u - C_d) + \left(\frac{R - d}{u - d} \cdot C_u + \frac{u - R}{u - d} \cdot C_d\right) = C_d$$

when the stock price becomes dS at day 1.

In general, we have multi-period binomial pricing model. We should conclude with the following concrete example.



Assuming the interest rate is zero, the risk-neutral probability $(q, 1 - q)$ is such that

$$S_{\text{now}} = qS_{\text{up}} + (1 - q)S_{\text{down}} \quad \Rightarrow \quad q = \frac{1}{2}.$$

We have the following tree.