Introduction to probability & statistics (MATH1241)

By **Tuan Ho** [github.com/ttuanho|linkedin.com/in/ttuan-hho/]

- *Definitions, Propositions, Corrolary & Theorems are collected from the coursepack unless noted.*
- *All Solutions are written by me.*

Table of contents

Set Theory

Definition A ***set*** is a collection of objects. These objects are called ***elements***.

Definition.

- A set A is a subset of a set B (written $A \subset B$) if and only if each element of A is also an element of B; that is, if $x \in A$, then $x \in B$.
- The power set $P(A)$ of A is set of all subsets of A.
- The universal set S is the set that denotes all objects of given interest.
- The empty set $\emptyset$ (or $\{\}$) is the set with no elements.

Definition A set S is ***countable*** if its elements can be listed as a sequence.

See MATH1081 for more.

Probability

Definition. The ***conditional probability*** of A given B is denoted and defined by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \text{ provided that } P(B) \neq 0.$$

MultiplicationRule.

$$P(A \cap B) = P(A \mid B)P(B) = P(B \mid A)P(A).$$

Total Probability Rule. If $A_1, ..., A_n$ partition S and B is an event, then

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B \mid A_i)P(A_i)(\text{using Multiplication Rule}).$$

Bayes' Rule. If $A_1, ..., A_n$ partition S and B is an event, then

$$P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}.$$

Random Variables

Definition. A ***random variable*** is a real function defined on a sample space.

Definition. For a random variable X on some sample space S, define for all subsets $A \subset S$ and real numbers $r \in R$,

- $\{X \in A\} = \{s \in S : X(s) \in A\}$
- $\{X = r\} = \{s \in S : X(s) = r\}$
- $\{X \leq r\} = \{s \in S : X(s) \leq r\}$
- $\ldots$ and so on.

Definition. The ***cumulative distribution function*** of a random variable X is given by

$$F_X(x) = P(X \leq x) \text{ for } x \in R.$$

Definition. A random variable X is ***discrete*** if its image is countable.

Definition. The ***probability distribution*** of a discrete random variable X is some description of all the probabilities of all events associated with X.

Definition. The expected value (or ***mean***) of a discrete random variable X with probability distribution p_k

$$E(X) = \sum_{allk} x_k p_k.$$

Theorem. Let X be a discrete random variable with probability distribution $p_k = P(X = x_k)$. Then for any real function $g(x)$, the expected value of $Y = g(X)$ is

$$E(Y) = E(g(X)) = \sum_{k} g(x_k) p_k.$$

Definition.

- The ***variance*** of a discrete random variable X is $Var(X) = E\big(X - E(X)\big)$.
- The standard deviation of X is $SD(X) = \sqrt{Var(X)}$.

Theorem. $Var(X) = E(X^2) - (E(X))^2$.

Theorem. If a and b are constants, then

- $E(aX + b) = aE(X) + b$
- $Var(aX + b) = a^2 Var(X)$
- $SD(aX + b) =\mid a \mid SD(X)$

Special distributions

Definition. The ***Binomial distribution*** $B(n, p)$ for $n \in N$ is the function

$$B(n, p, k) = \binom{n}{k} p^k (1 - p)^{n-k} \text{ where } k = 0, 1, \cdots, n.$$

Theorem. If X is the random variable that counts the successes of some Bernoulli process with n trials having success probability p, then X has the binomial distribution $B(n, p)$.

Theorem. If X is a random variable and $X \sim B(n, p)$, then

- $E(X) = np$
- $Var(X) = npq = np(1 - p)$

Definition. Probabilities such as $P(X \geq t), P(X > t)$, and $P(|X - E(X)| > t)$ are each referred to as a ***tail probability***.

Definition. The ***Geometric Distribution*** $G(p)$ is the function

$$G(p,k) = (1-p)^k p = q^k p \text{ where } k = 1, 2, \cdots$$

Theorem. Consider an infinite Bernoulli process of trials each of which has success probability p. If the random variable X is the number of trials conducted until success occurs for the first time, then X has the geometric distribution $G(p)$.

Notation. We write $X \sim G(p)$ to denote that X has this distribution.

Theorem. If $X \sim G(p)$ and n is a positive integer, then

$$P(X > n) = (1-p)^n = p^n.$$

Corollary. If $X \sim G(p)$, then the cumulative distribution function F is given by

$$F(x) = P(X \leq x) = 1 - (1-p)^{\lfloor x \rfloor} = 1 - q^{\lfloor x \rfloor} \text{ for } x \in \mathbb{R}.$$

Theorem. If X is a random variable and $X \sim G(p)$, then

- $E(X) = \frac{1}{p}$
- $Var(X) = \frac{1-p}{p^2}$

Sign Tests

A sample of data consisting of independent observations of some quantity of interest, and it might be of interest to see whether the observed values differ systematically from some fixed and pre-determined value.

A variety of corn currently used yields N bushels per acre. We want to know whether the new variety improves on the existing one – that is, are the above values centred around a true value for yield of N, or are they systematically different from the value N?

To answer this question, one may use a "sign test" approach as follows:

1. Count the number of observations that are strictly greater than the target value ("+").
2. Count the total number of observations that are either strictly greater ("+") or strictly smaller ("–") than the target value.
3. Calculate the tail probability that measures how often one would expect to observe as many increases ("+") as were observed, if there were equal probability of "+" and "–".

Continuous random variables

Definition. Random variable X is ***continuous*** if and only if $F_X(x)$ is continuous.

Definition. The ***probability density function*** $f(x)$ of a continuous random variable X is defined by

$$f(x) = f_X(x) = \frac{d}{dx}F(x), x \in \mathbb{R}.$$

if $F(x)$ is differentiable, and $\lim_{x\to a^-} \frac{d}{dx}F(x)$ if $F(x)$ is not differentiable at $x = a$.

Definition. The ***expected value*** (or mean) of a continuous random variable X with probability density function $f(x)$ is defined to be

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

Theorem. If X is a continuous random variable with density function $f(x)$, and $g(x)$ is a real function, then the expected value of $Y = g(X)$ is

$$E(Y) = E(g(x)) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

Definition.

The ***variance*** of a continuous random variable X is

$$Var(X) = E((X - E(X)^2)) = E(X^2) - (E(X))^2.$$

The ***standard deviation*** of X is $\sigma = SD(X) = \sqrt{Var(X)}$.

Theorem. If a and b are constants, then

- $E(aX + b) = aE(X) + b$
- $Var(aX + b) = a^2Var(X)$
- $SD(aX + b) =\mid a \mid SD(X)$.

Theorem. If $E(X) = \mu$ and $Var(X) = \sigma^2$ and $Z = \frac{X-\mu}{\sigma}$, then $E(Z) = 0$ and $Var(Z) = 1$.

Notation. $Z = \frac{X-\mu}{\sigma}$ is referred to as the ***standardised random variable*** obtained from X.

Special Continuous distribution

Definition. The continuous random variable X has ***normal distribution*** $N(\mu, \sigma^2)$ if it has probability density

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \text{ where } -\infty < x < \infty.$$

Theorem. If X is a continuous random variable and $X \sim N(\mu, \sigma^2)$, then

- $E(X) = \mu$
- $Var(X) = \sigma^2$

Theorem. If $X \sim N(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0, 1)$.

Definition. A continuous random variable T has ***exponential distribution*** $Exp(\lambda)$ if it has probability density

$$f(t) = \begin{cases} \lambda e^{-\lambda t} \text{ if } t \geq 0 \\ 0 \text{ if } t < 0 \end{cases}$$

Definition. The cumulative distribution function $F_T(t)$ of an exponentially distributed random variable $T \sim Exp(\lambda)$ is easily expressed:

$$F_T(x) = P(T \leq t) = \int_{-\infty}^{t} f(x)dx = \begin{cases} 1 - e^{-\lambda t} & , \quad t \geq 0 \\ 0 & , \quad t < 0 \end{cases}$$

Theorem. If T is a continuous random variable and $T \sim Exp(\lambda)$, then

- $E(T) = \frac{1}{\lambda}$
- $Var(T) = \frac{1}{\lambda^2}$

Practice

Coursepack exercises

Problem 9.4.27. Let X be a random variable with the Poisson probability distribution $p_k = \frac{e^{-\lambda}\lambda^k}{k!}$. Find $E((1 + X)^{-1})$.

Problem 9.4.29. A coin is tossed 50 times. What is the probability of it coming down heads exactly 25 times?

$$B(50, 0.5, 25) = \binom{50}{25} 0.5^{25} \times 0.5^{50-25}$$

Problem 9.4.33. For the $B(n, p)$ distribution, by considering $\frac{p_k}{p_k - 1}$, show that pk is largest when $k = \lfloor (n+1)p \rfloor$. This k is called the "mode" of the distribution.

Problem 9.6.57. If X is a normal random variable with mean μ and variance σ^2, show that

$$E(\mid X - \mu \mid) = \sqrt{\frac{2}{\pi}}\sigma.$$

$$E(\mid X - \mu \mid) =$$

Problem 9.6.58. Find $E(X)$ and $Var(X)$ for the random variable X with probability density function proportional to

$$e^{-x^2+x}.$$

$E(x) = \int_{-\infty}^{\infty} xf(x)dx$

So $E(x) = Var(x) = \frac{1}{2}$.

Problem 9.6.59. Let T be a continuous random variable and $T \sim Exp(\lambda)$; that is T has an exponential distribution with parameter λ. Prove that

$$Var(T) = \frac{1}{\lambda^2}.$$

Since $T \sim Exp(\lambda)$, $f_T(t) = \lambda e^{-\lambda t}$.

$$\begin{aligned}
\mu = E(T) &= \int_0^\infty tf_T(t)dt \\
&= \lambda \int_0^\infty te^{-\lambda t}dt \\
&= -te^{-\lambda t}\big|_0^\infty + \int_0^\infty e^{-\lambda t}dt \\
&= 0 + \frac{e^{-\lambda t}}{-\lambda}\big|_0^\infty \\
&= \frac{1}{\lambda}
\end{aligned}$$

$$
\begin{aligned}
Var(T) &= E(T^2) - \mu^2 \\
&= \int_0^\infty t^2 f_T(t^2) dt - \frac{1}{\lambda^2} \\
&= -t^2 e^{-\lambda t}|_0^\infty + 2\int_0^\infty te^{-\lambda t} dt - \frac{1}{\lambda^2} \\
&= -t^2 e^{-\lambda t}|_0^\infty - \frac{2te^{-\lambda t}}{\lambda}|_0^\infty - \frac{2}{\lambda^2} e^{-\lambda t}|_0^\infty - \frac{1}{\lambda^2} \\
&= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\
&= \frac{1}{\lambda^2}
\end{aligned}
$$

Problem 9.6.60. Suppose a continuous random variable T has an exponential distribution with parameter λ and $P(T \leq t) = p$ for $0 < p < 1$.

- Find t in terms of p.
- Hence find the median m. In other words, find m such that $P(T \leq m) = 0.5$.

Problem 9.6.61. (Memoryless property) A continuous random variable T has an exponential distribution and $T \sim Exp(\lambda)$. Prove that

$$P(T \geq t + t_0 | T \geq t_0)$$

is independent of t_0 for positive t and t_0.

Problem 9.6.62. Suppose the time, in minutes, required by any particular student to complete a certain two-hour examination has the exponential distribution for which the mean is 90 minutes. The examination starts at 10:00 am.

- a) What is the probability that a student completes the examination before 11:00 am?
- b) What is the probability that a student completes the examination between 11:00 am and 11:30 am?
- c) What is the probability that a student could not complete the examination within 2 hours?
- d) What is the median time for completing the examination?

1. 0.487
2. 0.146
3. 0.264

Problem 9.6.63. During a whale watch season in Sydney, the time, measured in hours from the moment that a cruise enters the whale watch area, to spot the first whale can be modelled by the exponential distribution with parameter $\lambda = 0.4$. A person joins a Sydney whale watch tour during the season. After entering the area,

- what is the probability that no whale can be spotted in the first hour?
- what is the probability that the time to spot the first whale exceeds the mean by one standard deviation?
- The organiser claims a 90 % success rate of finding whales in a trip. How long should the cruise stay in the whale watching area to achieve that?

Let T be the variable of the hours from the moment that a cruise enters the whale watch area, to spot the first whale.

Since $T \sim Exp(\lambda)$, then

the probability of seeing no whale until t^{th} hour is $F_T(t) = 1 - e^{-\lambda t}$ for $t > 0$,

$f(t) = \lambda e^{-\lambda t}$,

$\mu = E(T) = \frac{1}{\lambda} = 2.5$, and

$\sigma^2 = Var(T) = \frac{1}{\lambda^2} = 6.25$.

- the probability that no whale can be spotted in the first hour($t = 1$) $1 - F_T(1) = 1 - (1 - e^{-\lambda}) \approx 0.6703$.

- the probability that the time to spot the first whale exceeds the mean by one standard deviation is

$$1 - F_T(\mu + \sigma) = 1 - (1 - e^{-\lambda(\mu+\sigma)}) \approx 0.135.$$

- To claim a 90 % success rate of finding whales in a trip, $F_T(t) = 1 - e^{-\lambda t} \geq 0.9 \Leftrightarrow t \geq \frac{ln(1-0.9)}{-\lambda} \approx 5.756$. Hence, the cruise stay in the whale watching area more than 5.756 hours to achieve 90% success rate.

Problem 9.6.64. A system consists of 3 independent components connected in series. Hence the system fails when at least one of the components fails. Suppose the lengths of life of the components, measured in hours, have the exponential distribution $Exp(0.001)$, $Exp(0.003)$, $Exp(0.004)$. Find the probability that the system can last at least 100 hours.

Since $T \sim Exp(\lambda)$, the culmulative ditribution function of failing after t hours since production is $F_T(t) = 1 - e^{-\lambda t}$.

Hence, the culmulative probability of the component c lasting at least n hours is $P_c = P(t \geq n) = 1 - F_T(t < n) = 1 - (1 - e^{-\lambda n}) = e^{-\lambda n}$.

So the probability that the system can last at least 100 hours is

$$
\begin{aligned}
P_C(t \geq 100) &= \sum_{i=1,2,3} P_{c_i}(100) \\
&= e^{-\lambda_1 100} \times e^{-\lambda_2 100} \times e^{-\lambda_3 100} \\
&= 0.4493.
\end{aligned}
$$

Problem 9.6.65. A system consists of n independent components connected in series. The lengths of life of the components, in hours, have the exponential distributions $Exp(\lambda_i)$, $1 \leq i \leq n$. Let T be the random variable that gives the time until the system fails.

- a) Find $P(T \leq t)$ and hence write down the cumulative distribution function of the random variable T.
- b) Find the probability density function of T .
- c) Name the probability distribution of T, and find the expected value and variance of T.

Exams

2019T2

Question 2e)

At each turn in the game "Rock, Paper, Scissors", the player makes a choice of rock or paper or scissors. When playing the game, Daenerys follows a fairly predictable strategy.

- If she chooses rock one turn, she will choose paper or scissors next turn, with probabilities $\frac{1}{3}$ and $\frac{2}{3}$ respectively.
- If she chooses paper one turn, she will choose any one of the three options next turn, with equal probabilities.
- If she chooses scissors, she will choose rock or paper next turn, with probabilities $\frac{1}{3}$ and $\frac{2}{3}$ respectively.

Let r_n, p_n, s_n be the probabilities that Daenerys chooses rock, paper, scissors on the n^{th} turn in the game, and consider the vector

$$
\vec{v_n} = \begin{pmatrix} r_n \\ p_n \\ s_n \end{pmatrix}
$$

- i) Find a matrix A such that $\vec{v_{n+1}} = A\vec{v_n}$ for all n.
- ii) Determine the long-term probabilities for each of the three choices if Daenerys continues to follow the above strategy.

Question 3a) In a biological experiment, bacteria and phages (organisms that eat bacteria) grow in a container together. Initially, there are bo bacteria and po phages in the container, and after t hours, the number of bacteria is $b(t)$ and the number of phages is $p(t)$. The bacteria and phages each reproduce at a rate of 10% per hour. Each phage eats on average two bacteria per hour. Nothing eats the phages, and in the course of the experiment there are no further limitations on the growth of the populations.

- i) Write down a pair of differential equations which express the growth of the populations and their interactions.
- ii) Solve the equations to find expressions for p and b as functions of time t.
- iii) Under what conditions on bo and po do the phages eventually eat all the bacteria, and if so, when does this occur?

2018S2

Question 3i) An urn contains 200 red balls and 400 blue balls. A player performs the following experiment 180 times: choose a ball from the urn at random, record its colour and put it back in the urn.

Let X be the number of times a red ball was selected.

- Write down, but do not evaluate, an expression for the probability that $X > 100$.
- Calculate the mean and variance of X.
- Use a normal approximation to estimate the probability in part a) .
- For every 10 red balls drawn the player wins \$1. Leftover balls do not count: for example, 57 red balls give a payout of \$5. Write an expression for the average payout from 180 draws (do not evaluate your expression). You may use the notation $p_k = P(X = k)$.

2017S2

Question 2iii)

The probability density function f of a continuous random variable X is given by

$$f(x) = \begin{cases} kx^2 & \text{for } 0 \leq x \leq 3 \\ 0 & otherwise, \end{cases}$$

where k is a constant.

- Find the value of k.
- Evaluate $E(X)$ and $Var(X)$.

2016S2

Question 1iii)

The height of male students in a university is normally distributed with mean 172 cm and standard deviation 5 cm. Calculate the probability that a randomly chosen male student from the university is taller than 180 cm.

2014S2

Question 3iv)

Two (not necessarily distinct) vectors $\vec{v_1}, \vec{v_2}$ are chosen at random from the (finite) set of vectors

$$S = \{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R} : x, y \in \{-1, 0, 1\} \}$$

Define A to be the event that $\vec{v_1}$ and $\vec{v_2}$ are linearly independent.

- a) Show that $P(A) = \frac{16}{27}$.
- b) Define the discrete random variable $X = dim(span \vec{v_1}, \vec{v_2})$. Copy and complete the following table for the probability distribution $p_k = P(X = k)$.
- c) Calculate $E(X)$.
- d) If the vector b is chosen randomly from S, state the value of P(B|A), where B is the event that $b \in span\{\vec{v_1}, \vec{v_2}\}$.

2013S2

MATH1231 - Question 2v)

The probability density function f of a continuous random variable X is given by

$$f(x) = \begin{cases} \frac{1}{2}(2-x) & \text{for } 0 \leq x \leq 2 \\ 0 & otherwise. \end{cases}$$

- Evaluate $E(X)$ and $Var(X)$.
- The **median** of a distribution is defined to be the real number m such that $P(X \leq m) = 2$. Find the median of the above distribution.

MATH1231 - Question 3iv)

A manufacturer claims that only 20 % of the one-litre size soft drinks produced by his factory contain less than 1.05 litres. Consumer Watchdog examined a random sample of 100 bottles of one-litre soft drinks produced by this manufacturer and found that 30 % of the sample contain less than 1.05 litres. To test whether the manufacturer's claim is true, we shall use a binomial distribution B(n,p) to model the number, X, of bottles containing less than 1.05 litres in a random sample of 100.

- a) State the values of n and p.
- b) Find an expression for $P(X \geq 30)$ in terms of binomial coefficients.
- c) Use the normal approximation to the binomial distribution to estimate the probability in the previous part.
- d) Based on this probability, is there evidence that the manufacturer has made a false claim? (Give a reason.)

MATH1241 - Question 3i)

Suppose that a continuous random variable T has an exponential distribution with parameter $\lambda, \lambda > 0$.

- a) Find $Var(8 - 2T)$.
- b) Find $P(\lambda T \geq 1)$.

2012S2

Question 2v)

Suppose that a continuous random variable T has an exponential distribution with parameter λ.

- a) Find $P(4 - 2T \geq 0)$.
- b) Find $E(T^2)$.

Question 3i)

A six-sided die, with faces numbered 1 to 6, is suspected of being unfair so that the number 6 will occur more frequently than should happen by chance. During 300 test rolls of the die, the number 6 occurred 68 times.

- a) Write down an expression for a tail probability that measures the chance of getting a 6 at least 68 times.
- b) Use the normal approximation to the binomial to estimate this probability.
- c) Is this evidence that the die is unfair?

Question 3ii) We roll a die successively, until we roll six consecutive sixes, in which case we stop rolling the die.

- a) What is the probability that we stop after the thirteenth throw?
- b) What is the probability that we roll the die at least ten times?

2011S2

Question 4ii).

A discrete random variable X has the probability distribution given by $pk = \frac{c}{2^k}$ for $k = 0, 1, 2, 3, 4$, where c is a constant.

- a) Find the value of the constant c.
- b) Calculate $P(X = 2)$.
- c) Calculate $P((X - 2)^2 < 4)$.

Question 2iv).

One semester, 65% of the students studying MATH1231/41 attended all their tutorials, and of these students, 45% were awarded credit (CR) or higher in the final exam. In contrast, only 30% of the students who did not attend all their tutorials were awarded credit (CR) or higher in the final exam.

- a) What percentage of the students studying MATH1231/41 were not awarded credit (CR) or higher in the final exam?
- b) Of the students who were awarded credit (CR) or higher in the final exam, what percentage, correct to 1 decimal place, attended all their tutorials ?

2010S2

Question 2v).

On the basis of the health records of a particular group of people, an insurance company accepted 60% of the group for a 10 year life insurance policy and rejected the others. Ten years later the company examined the survival rates for the whole group and found that 80% of those accepted for the policy had survived the 10 years, while 40% of those rejected had survived the 10 years. If a person from the original group did survive 10 years, what is the probability that they had been refused a life insurance policy?

Question 3ii).

According to the Bureau of Meteorology, Sydney experienced rain on 103 out of the 365 days in 2009. Let us make the simple assumptions that

- the probability p of rain on any given day in Sydney is constant; (1)
- the weather on any given day is independent of the weather on other days. (2)
- a) Under assumptions (1) and (2), use the stated data to estimate the probability p that a given day in Sydney is rainy.
- b) Under assumptions (1) and (2), estimate the probability that precisely 5 out of 10 given days are rainy in Sydney.

- c) Suppose that Sydney will experience rain on 1 January, 2011. Using the assumptions (1) and (2), estimate the probability that the 6th of January 2011 is the first dry day of the new year.