THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics

Mathematical Modelling Project Team
mathmodel@math.cuhk.edu.hk

Exercise (Probability and Statistics)

Last updated: 11/5/2026

Part A: Key Concepts

1 Probability

Probability gives a numerical measure for the degree of uncertainty or the occurrence of an event.
If there are n total possible outcomes in a sample space S, and m of those are favorable for an event A, then the probability of event A is given as

$$P(A) = \frac{\text{number of favorable outcomes}}{\text{total number of possible outcomes}} = \frac{n(A)}{n(S)} = \frac{m}{n}$$

Example. Find the probability of getting a 3 or 5 when throwing a die.
Solution. Sample space $S = \{1, 2, 3, 4, 5, 6\}$ and event $A = \{3, 5\}$.
We have $n(A) = 2$ and $n(S) = 6$. So, $P(A) = \frac{n(A)}{n(S)} = \frac{1}{3}$.

2 Review of set notation

1. **Definition** (Complement) The **complement** of event A is the set of all outcomes in a sample that are not included in the event A. The complement of event A is denoted by A'.

2. **Definition** (Intersection) The event $A \cap B$ is the **intersection** of the events A and B and consists of outcomes that are contained within both events A and B.

3. **Definition** (Mutually Exclusive) Two events are said to be **mutually exclusive** if $A \cap B = \emptyset$.

4. **Definition** (Union) The event $A \cup B$ is the **union** of events A and B and consists of the outcomes that are contained within at least one of the events A and B.

5. **Theorem** (Distributive Laws)
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
and
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

6. **Theorem** (De Morgan's Law)
$$(A \cap B)' = A' \cup B'$$
$$(A \cup B)' = A' \cap B'$$

3 Law of probability

1. **Definition** (Axioms of Probability) For an experiment with sample space $S = \{e_1, e_2, \cdots, e_n\}$, we can assign probability $P(e_1), P(e_2), \cdots, P(e_n)$ provided that

   (a) $0 \leq P(e_i) \leq 1$

   (b) $P(e_1) + P(e_2) + \cdots + P(e_n) = 1$.

2. **Theorem** (Complement Rule)

$$P(A') = 1 - P(A)$$

3. **Theorem** (Addition Law) If A and B are two different events then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

4 Counting rules

1. **Theorem** (Permutations) The number of permutations of n distinct objects taken r at a time is

$${}_nP_r = \frac{n!}{(n-r)!}$$

2. **Theorem** (Combinations) The number of distinct subsets or combinations of size r that can be selected from n distinct objects, $(r \leq n)$, is given by

$${}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

5 Conditional probability and independence

1. **Definition** (Conditional Probability) The **conditional probability** of event A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ for } P(B) > 0$$

2. **Definition** (Independence) The events A and B are called **independent** if

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

3. **Definition** (Partition) Events $B_1, B_2, \cdots, B_k$ are said to be a **partition** of the sample space S if the following two conditions are satisfied.

   (a) $B_i \cap B_j = \emptyset$ for each pair i, j

   (b) $B_1 \cup B_2 \cup \cdots \cup B_k = S$

4. **Theorem** (Bayes Rule) If $B_1, B_2, \cdots, B_k$ form a partition of the sample space S such that $P(B_i) \neq 0$ for $i = 1, 2, \cdots, k$, then for any event A of S,

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$$

   Subsequently,

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{P(A)}$$

6 Discrete probability distributions

1. **Definition** (Random Variable) A **random variable (RV)** is a number associated with each outcome of some random experiment. A random variable X is said to be **discrete** if it can take on only a finite or countable number of possible values x.

   Example. Toss two coins and record the number of heads: $0, 1$ or 2. Then, the following outcomes can be observed.

Outcome	TT	HT	TH	HH
Number of heads	0	1	1	2

   The random variables will be denoted by capital letters $X, Y, Z, \cdots$, and the lowercase x will represent a particular value of X. For the above example, $x = 2$ if heads comes up twice. Now, we want to look at the probabilities of the outcomes. For the probability that the random variable X has the value x, we write $P(X = x)$, or just $p(x)$. For the coin flipping random variable X, we can make the table:

x	0	1	2
$p(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

   The table represents the **probability distribution** of the random variable X.

2. **Definition** (Probability Mass Function) The function $p_X(x)$ or simply $p(x)$ is called **probability mass function (PMF)** of X if it satisfies:

   (a) $P(X = x) = p_X(x) \geq 0$

   (b) $\sum_x P(X = x) = 1$, where the sum is over all possible x

   Example. A shipment of 8 computers contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the PMF for the number of defectives.

   Solution. Let X be a random variable whose values x are the possible numbers of defective computers purchased by the school. Then x must be $0, 1$ or 2. For each case, we have

   $$P(X = 0) = \frac{\binom{3}{0}\binom{5}{2}}{\binom{8}{2}} = \frac{10}{28}$$

   $$P(X = 1) = \frac{\binom{3}{1}\binom{5}{1}}{\binom{8}{2}} = \frac{15}{28}$$

   $$P(X = 2) = \frac{\binom{3}{2}\binom{5}{0}}{\binom{8}{2}} = \frac{3}{28}$$

   Thus, the PMF of X is given by

x	0	1	2
$p(x)$	$\frac{10}{28}$	$\frac{15}{28}$	$\frac{3}{28}$

3. **Definition** (Cumulative Distribution Function) The **cumulative distribution function (CDF)** $F(x)$ for a random variable X is defined as

   $$F(x) = P(X \leq x)$$

   If X is discrete,

   $$F(x) = \sum_{y \leq x} p(y)$$

   where $p(x)$ is the probability mass function.

   Example. Find the CDF of the random variable in the last example.

Solution. The CDF of the random variable X is:
$F(0) = p(0) = \frac{10}{28}$
$F(1) = p(0) + p(1) = \frac{10}{28} + \frac{15}{28} = \frac{25}{28}$
$F(2) = p(0) + p(1) + p(2) = \frac{10}{28} + \frac{15}{28} + \frac{3}{28} = 1$
Hence,

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{10}{28} & \text{for } 0 \leq x < 1 \\ \frac{25}{28} & \text{for } 1 \leq x < 2 \\ 1 & \text{for } x \geq 2 \end{cases}$$

4. **Definition** (Expected Value) The **mean** or **expected value** of a discrete random variable X with probability mass function $p(x)$ is given by

$$\mathbb{E}(X) = \sum_x xp(x)$$

5. **Definition** (Variance) The **variance** of a random variable X with expected value μ is given by

$$V(X) = \sigma^2 = \mathbb{E}(X - \mu)^2 = \mathbb{E}(X^2) - \mu^2,$$

where

$$\mathbb{E}(X^2) = \sum_x x^2 p(x)$$

6. **Definition** (Standard Deviation) The **standard deviation** of a random variable X is the square root of the variance, and is given by

$$\sigma = \sqrt{V(X)} = \sqrt{\mathbb{E}(X - \mu)^2}$$

Remark. The mean describes the *center* of the probability distribution, while the standard deviation describes the *spread*.
Example. The number of fire emergencies at a rural county in a week has the following distribution

x	0	1	2	3	4
$P(X = x)$	0.52	0.28	0.14	0.04	0.02

Find $\mathbb{E}(X)$, $V(X)$ and σ.
Solution. By definition, we see that

$$\mathbb{E}(X) = 0(0.52) + 1(0.28) + 2(0.14) + 3(0.04) + 4(0.02) = 0.76 = \mu$$

$$\mathbb{E}(X^2) = 0^2(0.52) + 1^2(0.28) + 2^2(0.14) + 3^2(0.04) + 4^2(0.02) = 1.52$$

$$V(X) = 1.52 - (0.76)^2 = 0.9424$$

$$\sigma = \sqrt{0.9424} \approx 0.9708$$

7. **Definition** (Bernoulli Distribution) Let X be the random variable denoting the condition of the inspected item. Agree to write $X = 1$ when the item is defective and $X = 0$ when it is not. (This is a convenient notation because, once we inspect n such items, $X_1, X_2, \cdots, X_n$ denoting their condition, the total number of defectives will be given by $X_1 + X_2 + \cdots + X_n$)
Let p denote the probability of observing a defective item. The probability distribution of X, then, is given by

x	0	1
$P(X = x)$	$q = 1 - p$	p

Such a random variable is said to have a **Bernoulli distribution**. Note that

$$\mathbb{E}(X) = 0 \times p(0) + 1 \times p(1) = p$$

$$\mathbb{E}(X^2) = 0^2 \times p(0) + 1^2 \times p(1) = p$$

Hence,

$$V(X) = p - p^2 = pq$$

8. **Definition** (Binomial Distribution) Now, let us inspect n items and count the total number of defectives. This process of repeating an experiment n times is called Bernoulli trials. The Bernoulli trials are formally defined by the following properties:

   (a) The result of each trial is either a success or a failure.

   (b) The probability of success p is constant from trial to trial.

   (c) The trials are independent.

   (d) The random variable X is defined to be the number of successes in n repeated trials.

   Such a random variable X is said to have a **Binomial distribution**. Assume that each Bernoulli trial can result in a success with probability p and a failure with probability $q = 1 - p$. Then the probability distribution of the binomial random variable X, the number of successes in n independent trials, is

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \cdots, n$$

   The **mean** and **variance** of binomial distribution are

$$\mathbb{E}(X) = np \quad \text{and} \quad V(X) = npq$$

   We can notice that the mean and variance of the Binomial are n times those of the Bernoulli random variable.

7 Continuous probability distributions

1. **Definition** (Continuous Random Variables) All of the random variables discussed previously were discrete, meaning they can take only a finite (or, at most, countable) number of values. However, many of the random variables seen in practice have more than a countable collection of possible values. For example, the metal content of ore samples may run from 0.10 to 0.80. Such random variables can take any value in an interval of real numbers. Since the random variables of this type have a continuum of possible values, they are called **continuous random variables**.

2. **Definition** (Probability Density Function) The function $f(x)$ is a **probability density function (PDF)** for the continuous random variable X, defined over the set of real numbers, if

   (a) $f(x) \geq 0$, for all x

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

   (c) $P(a \leq x \leq b) = \int_a^b f(x)dx$.

   Remark. What does this actually mean? Since continuous probability functions are defined for an infinite number of points, the probability at a single point is always **zero**! Probabilities are measured over intervals, not single points. That is, the area under the curve between two distinct points defines the probability for that interval. This means that the height of the probability function can in fact be greater than one. The property that the integral must equal one is equivalent to the property for discrete distributions that the sum of all the probabilities must equal one.

3. **Definition** (Cumulative Distribution Function) The **cumulative distribution function (CDF)** $F(x)$ of a continuous random variable X, with density function $f(x)$, is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)dx$$

Remark. As an immediate consequence of the above equation, one can write these two results:

(a) $P(a \le x \le b) = F(b) - F(a)$

(b) $f(x) = F'(x)$, if the derivative exists

4. **Definition** (Expected Values) The **expected value** or **mean** of a continuous random variable X that has a probability density function $f(x)$ is given by

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

5. **Definition** (Variance) Let X be a random variable with probability density function $f(x)$ and mean $\mathbb{E}(X) = \mu$. The **variance** of X is

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

6. **Definition** (Median) The **median** of a probability distribution is defined as solution m to the equation (F is the CDF)

$$F(m) = 0.5$$

	Discrete	Continuous
Probability	$p(x) = P(X = x)$	$f(x) = \frac{d}{dx}P(X \le x) = F'(x)$ and $P(X = x) = 0$ for any x
CDF	has jumps at every values of X	is continuous
Mean	$\sum xp(x)$	$\int xp(x)dx$
Variance	$\sum(x-\mu)^2 p(x)$	$\int (x-\mu)^2 p(x)dx$

7. **Definition** (Uniform Distribution) One of the simplest continuous distributions is the continuous **uniform distribution**. This distribution is characterized by a density function that is flat and thus the probability is uniform in a finite interval, say $[a, b]$. The **PDF** the continuous uniform random variable X on the interval $[a, b]$ is

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a < x < b \\ 0 & \text{elsewhere} \end{cases}$$

The **CDF** of a uniformly distributed X is given by

$$F(x) = \int_{a}^{x} \frac{1}{b-a} dx = \frac{x-a}{b-a}, \quad a \le x \le b$$

The **mean** and **variance** of the uniform distribution are

$$\mu = \frac{a+b}{2} \quad \text{and} \quad \sigma^2 = \frac{(b-a)^2}{12}$$

8. **Definition** (Exponential Distribution) The continuous random variable X has an **exponential distribution**, with parameter β, if its **PDF** is given by

$$f(x) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

The **mean** and **variance** of the exponential distribution are

$$\mu = \beta \quad \text{and} \quad \sigma^2 = \beta^2$$

CDF for the exponential distribution has the simple form:

$$F(x) = P(X \le x) = \int_{0}^{t} \frac{1}{\beta} e^{-\frac{x}{\beta}} = 1 - e^{-\frac{x}{\beta}} \quad \text{for } x \ge 0$$

9. **Definition** (Normal Distribution) The most widely used of all the continuous probability distributions is the **normal distribution** (also known as **Gaussian**). It serves as a popular model for measurement errors, particle displacements under Brownian motion, stock market fluctuations, human intelligence and many other things. It is also used as an approximation for Binomial (for large n).

   The PDF of normal distribution follows the well-known symmetric bell-shaped curve. The curve is centered at the mean value μ and its spread is, of course, measured by the standard deviation σ. These two parameters, μ and σ^2, completely determine the shape and center of the normal density function.

   The normal random variable X has the **PDF**

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \quad \text{for} -\infty < x < \infty$$

   It will be denoted as $X \sim \mathcal{N}(\mu, \sigma^2)$.

Part B: Basic Questions

1. The probability that John passes a Math exam is $\frac{4}{5}$ and that he passes a Chemistry exam is $\frac{5}{6}$. If the probability that he passes both exams is $\frac{3}{4}$, find the probability that he will pass at least one exam.
2. A package of 6 light bulbs contains 2 defective bulbs. If 3 bulbs are selected for use, find the probability that none of the three is defective.
3. Three bits (0 or 1 digits) are transmitted over a noisy channel, so they will be flipped independently with probability 0.1 each. What is the probability that
   (a) At least one bit is flipped?
   (b) Exactly one bit is flipped?

Part C: Advanced Questions

4. Let X be a random variable having a probability mass function given in the following.

x	0	1	2	3	4
$P(X = x)$	0.52	0.28	0.14	0.04	0.02

Calculate the mean and variance of the random variable $Y = 4X + 3$.

5. The probability that a certain kind of component will survive a shock test is 0.75. Find the probability that

   (a) exactly 2 of the next 8 components tested survive,

   (b) at least 2 will survive,

   (c) at most 6 will survive.

6. Suppose that the error in the reaction temperature, in $°C$, for a controlled laboratory experiment is a continuous random variable X having the density

$$f(x) = \begin{cases} \dfrac{x^2}{3} & \text{for } -1 \le x \le 2 \\ 0 & \text{elsewhere} \end{cases}$$

   (a) Verify that the probability indeed adds up to 1

   (b) Find $P(0 < X < 1)$.

7. The life length of batteries X (in hundreds of hours) has the density

$$f(x) = \begin{cases} \frac{1}{2}e^{-\frac{x}{2}} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

   (a) Find the probability that the life of a battery of this type is less than 200 or greater than 400 hours.

   (b) Find the probability that a battery of this type lasts more than 300 hours, given that it has already been in use for more than 200 hours.

8. The failure of a circuit board interrupts the work of a computing system until a new board is delivered. The delivery time X is uniformly distributed over an interval of at least one and at most five days. The cost C of this failure and interruption consists of a fixed cost C_0 for the new part and a cost that increases proportionally to X^2, so that

$$C = C_0 + C_1X^2$$

   (a) Find the probability that the delivery time is two or more days.

   (b) Find the expected cost of a single failure in terms of C_0 and C_1.

9. A downtime due to equipment failure is estimated to have an Exponential distribution with the mean $\beta = 6$ hours. What is the probability that the next downtime will last between 5 and 10 hours?

Part D: Solutions

1. Let M = John passes the math exam, and C = John passes the chemistry exam.
   $P(\text{John passes at least one exam}) = P(M \cup C) = P(M) + P(C) - P(M \cap C) = \frac{4}{5} + \frac{5}{6} - \frac{3}{4} = \frac{53}{60}$

2. $P(\text{none are defective}) =$
   $$= \frac{\text{number of ways 3 non defectives can be chosen}}{\text{total number of ways a sample of 3 can be chosen}} = \frac{\binom{4}{3}}{\binom{6}{3}} = \frac{1}{5}$$

3. (a) Using the complement rule, $P(\text{at least one}) = 1 - P(\text{none})$. If we denote F_k the event that the kth bit is flipped, due to independence, we have
   $$\begin{aligned} &P(\text{at least one}) \\ &= 1 - P(\text{no bits are flipped}) \\ &= 1 - P(F_1' \cap F_2' \cap F_3') \\ &= 1 - (1 - 0.1)^3 \\ &= 0.271 \end{aligned}$$
   (b) Flipping exactly one bit can be accomplished in 3 ways:
   $$P(\text{exactly one}) = P(F_1 \cap F_2' \cap F_3') + P(F_1' \cap F_2 \cap F_3') + P(F_1' \cap F_2' \cap F_3) = 0.243$$

4. In an example, we have found that $\mathbb{E}(X) = 0.76 = \mu$ and $V(X) = 1.52 - (0.76)^2 = 0.9424$
   Then,
   $$\mathbb{E}(Y) = \mathbb{E}(4X+3) = \sum(4x+3)p(x) = 4\sum xp(x) + 3\sum p(x) = 4\mathbb{E}(X) + 3 = 6.04$$
   $$V(Y) = V(4X+3) = V(4X) = 4^2V(X) = 15.08$$

5. (a) Assuming that the tests are independent and $p = 0.75$ for each of the 8 tests, we get
   $$P(X=2) = \binom{8}{2}(0.75)^2(0.25)^{8-2} = 0.003843$$
   (b)
   $$\begin{aligned} &P(X \geq 2) = 1 - P(X=0) - P(X=1) \\ &= 1 - [8(0.75)(0.25)^7 + (0.25)^8] \approx 0.9996 \end{aligned}$$
   (c)
   $$\begin{aligned} &P(X \leq 6) = 1 - P(X=7) - P(X=8) \\ &= 1 - [8(0.75)^7(0.25) + (0.75)^8] \approx 0.633 \end{aligned}$$

6. (a)
   $$\int_{\infty}^{\infty} f(x)dx = \int_{-1}^{2} \frac{x^2}{3}dx = \frac{8}{9} + \frac{1}{9} = 1$$
   (b)
   $$P(0 < X < 1) = \int_0^1 \frac{x^2}{3}dx = \frac{1}{9}$$

7. (a) Let A denote the event that X is less than 2, and let B denote the event that X is greater than 4. Then
   $$\begin{aligned} &P(A \cup B) \\ &= P(A) + P(B) \quad \text{think about why?} \\ &= \int_0^2 \frac{1}{2}e^{-\frac{x}{2}}dx + \int_4^{\infty} \frac{1}{2}e^{-\frac{x}{2}}dx \\ &= 1 - e^{-1} + e^{-2} \\ &\approx 0.767 \end{aligned}$$

(b) We are interested in $P(X > 3|X > 2)$ and by the definition of conditional probability,

$$P(X > 3|X > 2) = \frac{P(X > 3)}{P(X > 2)} = \frac{1 - \int_0^3 \frac{1}{2} e^{-\frac{x}{2}} dx}{1 - \int_0^2 \frac{1}{2} e^{-\frac{x}{2}} dx} = e^{-\frac{1}{2}} \approx 0.606$$

8. (a)

$$f(x) = \begin{cases} \frac{1}{4} & \text{for } 1 \leq x \leq 5 \\ 0 & \text{elsewhere} \end{cases}$$

Thus,

$$P(X \geq 2) = \int_2^5 \frac{1}{4} dx = \frac{3}{4}$$

(b) We know that

$$\mathbb{E}(C) = C_0 + C_1 \mathbb{E}(X^2)$$

so it remains for us to find $\mathbb{E}(X^2)$. This value could be found directly from the definition or by using the variance and the fact that $\mathbb{E}(X^2) = V(X) + \mu$. Using the latter approach, we find

$$\mathbb{E}(X^2) = \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2 = \frac{31}{3}$$

Then, $\mathbb{E}(C) = \frac{31}{3} C_1 + C_0$.

9. $P(5 < X < 10) = F(10) - F(5) = 1 - \exp(-\frac{10}{6}) - \left(1 - \exp(-\frac{5}{6})\right) \approx 0.2457$.