Supplementary 8: Probability Essentials for Modeling and Simulation

Mathematical Modeling Lecture

Mathematical Modeling

Kenneth Sok Kin Cheng

Session Route

Launch	Build	Test
Question Motivating context, key vocabulary, and the first modelling decision.	**Method** Worked example, diagram or table, calculation, and parameter meaning.	**Critique** Checkpoint practice, limitation check, and communication of results.

Columns are used where comparison helps; explanation slides stay clean and spacious.

Learning Goals

By the end of this two-hour supplementary lesson, students should be able to:

- describe a random experiment using a **sample space**, **event**, and probability model;
- compute simple probabilities using counting, complements, and inclusion–exclusion;
- distinguish **conditional probability**, **independence**, and disjointness;
- use **Bayes' theorem** to update probabilities from evidence;
- model uncertainty using **random variables**, PMF/PDF, CDF, expectation, and variance;
- recognise when binomial, Poisson, exponential, normal, or Markov-chain models are appropriate;
- explain how the Law of Large Numbers and Central Limit Theorem justify Monte Carlo simulation.

Why probability belongs in mathematical modeling

Core explanation, worked reasoning, and application

Why probability belongs in mathematical modeling

A deterministic model gives one output when the input is fixed. Real systems often require a different language. Customers do not arrive exactly every five minutes; a sensor reading contains noise; a production line occasionally creates defective products; weather changes from day to day. A modeling report should therefore be able to say not only “what will happen”, but also “how likely”, “with what variability”, and “how confident are we?”

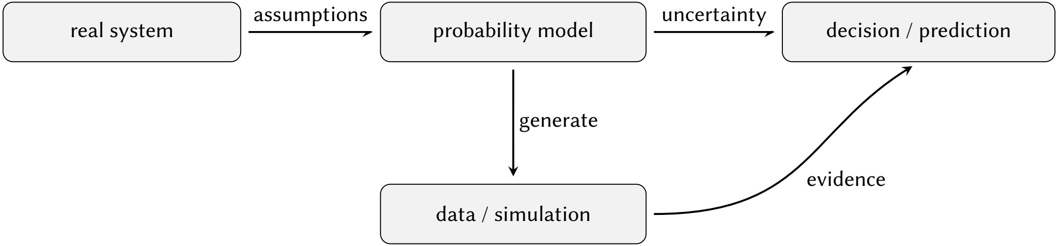
Activity: Launch activity: deterministic or stochastic?

Classify each situation as mainly deterministic, stochastic, or mixed.

1. A robot travels 120 m at a constant speed of 1.5 m/s.
2. Customers enter a canteen during lunch time.
3. A component has a 2% chance of failing during one month.
4. A delivery route is planned, but travel time depends on weather and crowding.

A strong modeling answer usually separates the predictable structure from the random variation.

Illustration: Why probability belongs in mathematical modeling



Probability enters modeling when assumptions must describe uncertainty, variability, and evidence rather than a single fixed outcome.

Sample spaces, events, and counting

Core explanation, worked reasoning, and application

Concept: Sample space and event

A **random experiment** is a procedure whose outcome is uncertain but whose possible outcomes can be described. The **sample space** Ω is the set of all possible outcomes. An **event** is a subset $A \subseteq \Omega$.

Worked Example: Worked example 1: two dice

For rolling two fair dice,

$$\Omega = \{(i, j) : 1 \leq i, j \leq 6\}, \qquad |\Omega| = 36.$$

The event "sum is 7" is

$$\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\},$$

so $\mathbb{P}(\text{sum} = 7) = 6/36 = 1/6$.

Concept: Probability rules

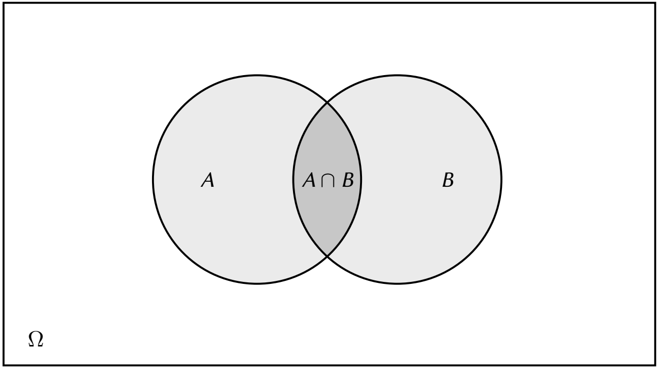
For events A and B,

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A), \qquad \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

If all outcomes are equally likely and Ω is finite, then

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}.$$

Illustration: Sample spaces, events, and counting



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

Inclusion–exclusion subtracts the overlap once because it is counted in both A and B.

Worked Example: Worked example 2: birthday problem

Assume birthdays are uniformly distributed over 365 days. For $n = 23$ people, compute the probability that at least two people share a birthday.
It is easier to use the complement. The probability all birthdays are distinct is

$$\frac{365 \cdot 364 \cdot 363 \cdots (365 - 22)}{365^{23}} \approx 0.4927.$$

Therefore

$$\mathbb{P}(\text{at least one shared birthday}) \approx 1 - 0.4927 = 0.5073.$$

Only 23 people are enough for a probability slightly above one half.

Checkpoint: Checkpoint

A bag has 5 red balls and 3 blue balls. Two balls are drawn without replacement. Is "first ball red" independent of "second ball red"? Explain using conditional probability.

Conditional probability, independence, and Bayes

Core explanation, worked reasoning, and application

Concept: Conditional probability

For events A and B with $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

This means: restrict the sample space to B, then renormalise.

Concept: Independence

Events A and B are **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B),$$

or equivalently $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ when $\mathbb{P}(B) > 0$. Independence is not the same as disjointness. If two positive-probability events are disjoint, knowing one happened makes the other impossible, so they are dependent.

Concept: Total probability and Bayes' theorem

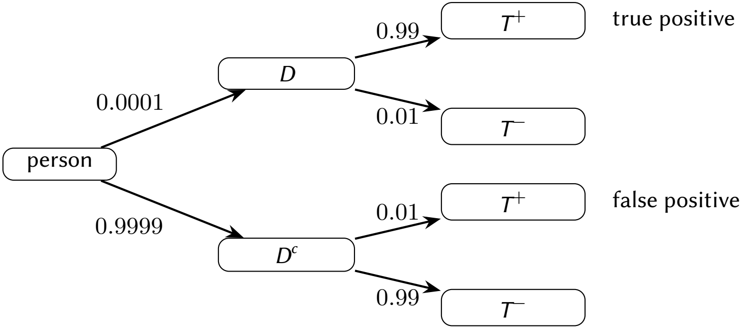
If $B_1, \ldots, B_n$ partition Ω, then

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

Bayes' theorem updates the probability of a cause after observing evidence:

$$\mathbb{P}(B_k \mid A) = \frac{\mathbb{P}(A \mid B_k)\mathbb{P}(B_k)}{\sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)}.$$

Illustration: Conditional probability, independence, and Bayes



A probability tree separates prior probability from test reliability. The false-positive branch can dominate when the base rate is very small.

Worked Example: Worked example 3: medical testing and base rate

A disease affects 1 in 10,000 people. A test has 99% sensitivity and 99% specificity. A randomly selected person tests positive. What is the probability they have the disease?
Let D be "has disease" and T^+ be "positive test". Then

$$\mathbb{P}(D) = 10^{-4}, \quad \mathbb{P}(T^+ \mid D) = 0.99, \quad \mathbb{P}(T^+ \mid D^c) = 0.01.$$

Therefore

$$\mathbb{P}(D \mid T^+) = \frac{0.99(10^{-4})}{0.99(10^{-4}) + 0.01(0.9999)} \approx 0.0098.$$

Even after a positive test, the probability is under 1%. This is not because the test is useless; it is because the disease is rare.

Model Critique: Model criticism: Bayes in reports

When using Bayes' theorem, a report should state the prior probabilities, evidence reliability, and what changes after observing evidence. Many wrong conclusions come from ignoring the base rate.

Random variables, distributions, expectation, and variance

Core explanation, worked reasoning, and application

Concept: Random variable

A **random variable** is a function $X : \Omega \to \mathbb{R}$ that assigns a numerical value to each outcome. It is **discrete** if its possible values are finite or countable, and **continuous** if probabilities are described by intervals and integrals.

Concept: PMF, PDF, and CDF

For a discrete random variable, the **probability mass function** is

$$p_X(x) = \mathbb{P}(X = x), \qquad \sum_x p_X(x) = 1.$$

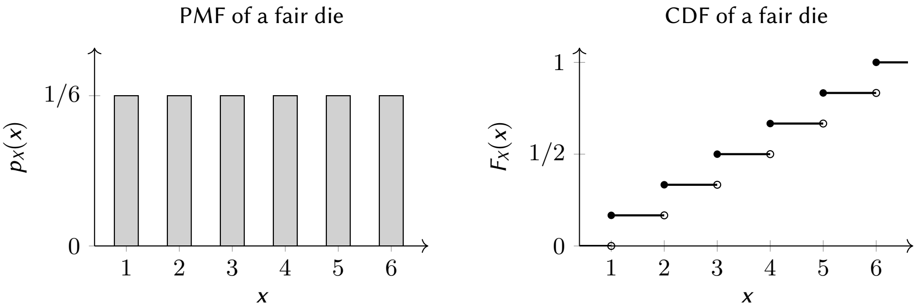
For a continuous random variable, the **probability density function** satisfies

$$\mathbb{P}(a \le X \le b) = \int_a^b f_X(x)\, dx.$$

For any random variable, the **cumulative distribution function** is

$$F_X(x) = \mathbb{P}(X \le x).$$

Illustration: Random variables, distributions, expectation, and variance



A discrete distribution has mass at points. Its CDF increases in jumps.

Concept: Expectation and variance

The **expected value** is the long-run average:

$$\mathbb{E}[X] = \sum_x x p_X(x) \quad \text{(discrete)}, \qquad \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx \quad \text{(continuous)}.$$

The **variance** measures spread:

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

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

Worked Example: Worked example 4: die roll mean and variance

Let X be the result of a fair six-sided die. Then

$$\mathbb{E}[X] = \frac{1+2+3+4+5+6}{6} = 3.5,$$

$$\mathbb{E}[X^2] = \frac{1+4+9+16+25+36}{6} = \frac{91}{6},$$

so

$$\mathrm{Var}(X) = \frac{91}{6} - (3.5)^2 = \frac{35}{12} \approx 2.917.$$

The standard deviation is about 1.708.

Worked Example: Worked example 5: expected profit

An insurer sells a one-year policy for a premium of \$1,500. The claim is \$200,000 if death occurs, and the probability of death in the year is $p = 0.01$. The insurer's expected profit is

$$0.99(1500) + 0.01(1500 - 200000) = 1485 - 1985 = -500.$$

At this premium the policy loses \$500 in expectation. This connects directly to decision theory and risk modeling.

Named distributions for modeling

Core explanation, worked reasoning, and application

Summary: When to use which distribution

Distribution	**Modeling question**	**Key quantities**
Bernoulli(p)	one success/failure trial	mean p, variance $p(1-p)$
Binomial(n, p)	number of successes in n independent trials	mean np, variance $np(1-p)$
Poisson(λ)	number of random arrivals in a fixed interval	mean λ, variance λ
Exponential(λ)	waiting time between Poisson arrivals	mean $1/\lambda$, memoryless
Normal(μ, σ^2)	measurement error, averages, natural variability	mean μ, variance σ^2

Worked Example: Worked example 6: binomial quality control

A factory has defect rate $p = 0.02$. In a batch of $n = 100$ items, let X be the number of defective items. Then $X \sim \mathrm{Bin}(100, 0.02)$ and

$$\mathbb{E}[X] = 100(0.02) = 2.$$

The probability of at most 2 defects is

$$\mathbb{P}(X \leq 2) = \sum_{k=0}^{2} \binom{100}{k} (0.02)^k (0.98)^{100-k} \approx 0.677.$$

Worked Example: Worked example 7: Poisson arrivals and exponential waiting

Customers arrive at a service window at average rate 12 per hour. If arrivals are modeled as Poisson, then the number of arrivals in one hour is $X \sim \mathrm{Pois}(12)$:

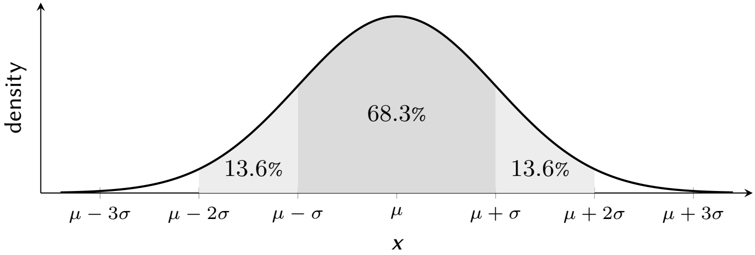
$$\mathbb{P}(X = 10) = e^{-12}\frac{12^{10}}{10!} \approx 0.105.$$

The waiting time until the next customer is $T \sim \mathrm{Exp}(12)$ hours, so

$$\mathbb{E}[T] = \frac{1}{12}\text{ hour} = 5\text{ minutes.}$$

This is the probability foundation of queue simulation.

Illustration: Named distributions for modeling



The normal distribution is used for measurement error, natural variation, and sampling distributions. About 68.3% of values lie within one standard deviation of the mean.

Checkpoint: Checkpoint

Which distribution would you use for each case?

1. Number of students arriving in the next 10 minutes.
2. Whether a randomly selected component is defective.
3. Waiting time until the next customer.
4. Measurement error of repeated sensor readings.

Monte Carlo, LLN, and CLT

Core explanation, worked reasoning, and application

Monte Carlo, LLN, and CLT

The link between probability and simulation is simple: if we cannot compute a quantity exactly, we can sometimes estimate it by repeated random trials. The theory tells us why the estimate improves.

Concept: Law of Large Numbers

Let $X_1, X_2, \ldots$ be independent and identically distributed with mean μ. The sample mean

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

converges to μ in probability. In modeling language: repeated simulation averages approach the expected value.

Concept: Central Limit Theorem

If $X_1, \ldots, X_n$ are i.i.d. with mean μ and variance σ^2, then for large n,

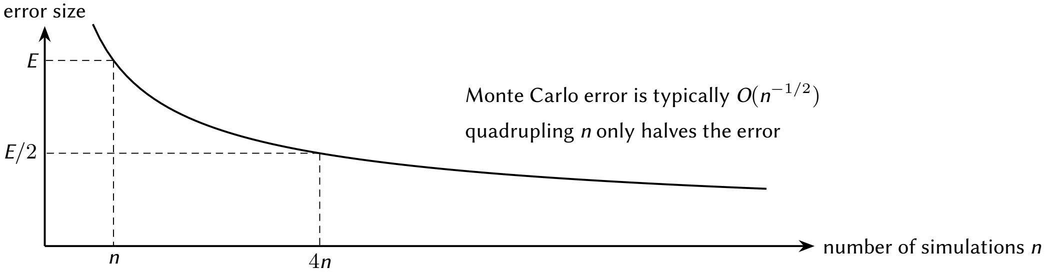
$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \approx Z \sim \mathcal{N}(0, 1).$$

Therefore an approximate 95% confidence interval for μ is

$$\bar{X}_n \pm 1.96 \frac{s}{\sqrt{n}},$$

where s is the sample standard deviation.

Illustration: Monte Carlo, LLN, and CLT



The $1/\sqrt{n}$ convergence rate explains why Monte Carlo is flexible but often slow.

Worked Example: Worked example 8: sample size for Monte Carlo

Suppose a Monte Carlo estimator is an average of Bernoulli indicators, so its standard deviation is at most $1/2$. How large should n be for a 95% confidence interval half-width at most 0.005?

$$1.96\frac{1/2}{\sqrt{n}} \leq 0.005 \quad \Longrightarrow \quad \sqrt{n} \geq 196 \quad \Longrightarrow \quad n \geq 38416.$$

This explains why Monte Carlo may need many samples in one dimension, even though it becomes valuable in high dimensions.

Model Critique: Reporting simulation uncertainty

A competition report should not only give a simulated estimate. It should also report the number of replications, the standard error, a confidence interval, and whether the conclusion changes under reasonable parameter perturbations.

A first Markov-chain model

Core explanation, worked reasoning, and application

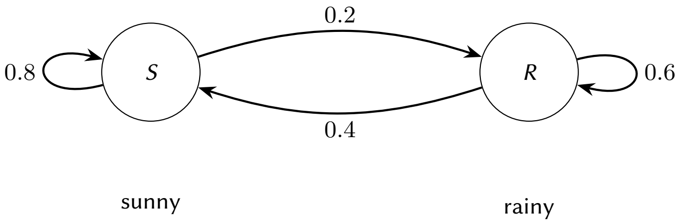
Concept: Discrete-time Markov chain

A sequence $X_0, X_1, X_2, \ldots$ on a finite state space is a **Markov chain** if the next state depends on the present state but not on the full past:

$$\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1}, \ldots, X_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i) = P_{ij}.$$

The matrix $P = (P_{ij})$ is the **transition matrix**; each row sums to 1.

Illustration: A first Markov-chain model



A two-state Markov chain for weather. The probabilities leaving each state add to 1.

Worked Example: Worked example 9: two-day weather prediction

States are S = sunny and R = rainy. Suppose

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{pmatrix},$$

where rows are today's weather and columns are tomorrow's weather. If today is sunny, the distribution after two days is

$$(1,0)P^2 = (1,0)\begin{pmatrix} 0.72 & 0.28 \\ 0.56 & 0.44 \end{pmatrix} = (0.72, 0.28).$$

Thus the probability of rain two days from now is 0.28.

Worked Example: Worked example 10: stationary distribution

A stationary distribution $\pi = (\pi_S, \pi_R)$ satisfies $\pi P = \pi$ and $\pi_S + \pi_R = 1$. From the weather matrix,

$$\pi_S = 0.8\pi_S + 0.4\pi_R \quad \Longrightarrow \quad 0.2\pi_S = 0.4\pi_R \quad \Longrightarrow \quad \pi_S = 2\pi_R.$$

Together with $\pi_S + \pi_R = 1$, we get

$$\pi = (2/3, 1/3).$$

In the long run, about two-thirds of days are sunny.

Differentiated modeling studio

Core explanation, worked reasoning, and application

Activity: Studio task: canteen queue uncertainty

A canteen receives students at an average rate of 18 per 10 minutes. During peak time, arrivals are modeled as Poisson.

1. Compute the probability of exactly 20 arrivals in the next 10 minutes.
2. Compute or approximate the probability of more than 24 arrivals.
3. If one server can handle 20 students per 10 minutes, what risk does this create?
4. Write one paragraph explaining whether the Poisson assumption is reasonable.

Summary: Competition-style reporting paragraph

We model arrivals in each 10-minute period as $X \sim \text{Pois}(18)$, assuming independent arrivals at an approximately constant rate during the short peak interval. This gives $\mathbb{E}[X] = 18$ and $\text{Var}(X) = 18$, so the typical fluctuation is about $\sqrt{18} \approx 4.24$ students. Since the server capacity is 20 students per 10 minutes, the system may become overloaded in a non-negligible fraction of intervals. We therefore recommend evaluating service capacity using not only the mean arrival rate, but also the upper tail probability $\mathbb{P}(X > 20)$ and sensitivity to the estimated rate.

Model Critique: Model criticism

Probability models are assumptions, not facts. A Poisson arrival model assumes independent arrivals and a roughly constant rate. If students arrive in groups after a bell rings, the independence assumption is weak. If the rate changes sharply during lunch, the constant-rate assumption is weak. Validation should compare both the mean and the variability of observed arrivals against the model.

Core practice and extension problems

Core explanation, worked reasoning, and application

Exercises: Core practice

1. Two fair dice are rolled. Find $\mathbb{P}(\text{sum} = 8)$ and $\mathbb{P}(\text{sum} \geq 10)$.
2. A bag contains 5 red and 3 blue balls. Two are drawn without replacement. Find the probability both are red.
3. A factory has two machines. Machine A produces 70% of items with defect rate 1%; machine B produces 30% with defect rate 4%. A randomly chosen item is defective. Find the probability it came from B.
4. Let X be the number of heads in 4 tosses of a fair coin. Write its distribution and compute $\mathbb{E}[X]$.
5. If $X \sim \text{Pois}(5)$, compute $\mathbb{P}(X = 3)$ and $\mathbb{P}(X = 0)$.
6. If $T \sim \text{Exp}(6)$ hours, find $\mathbb{E}[T]$ and $\mathbb{P}(T > 0.5)$.

Exercises: Modeling challenge

A small help desk receives requests according to a Poisson process with rate $\lambda = 10$ per hour. Service time is approximately exponential with mean 5 minutes.

1. Convert the mean service time into a service rate μ per hour.
2. Compute the utilisation ratio $\rho = \lambda/\mu$.
3. Explain what it means if ρ is close to 1.
4. Suggest one data source that could validate the arrival-rate assumption and one source that could validate the service-time assumption.

Exercises: Extension practice

1. Prove that if A and B are independent, then A and B^c are independent.
2. Let $X \sim U(0,1)$ and $Y = -\ln X$. Derive the CDF of Y and identify its distribution.
3. Let $X_1, \ldots, X_n$ be independent Bernoulli(p). Use linearity of expectation to show $\mathbb{E}[\sum X_i] = np$.
4. In the weather Markov chain, if today is rainy, compute the probability of a sunny day three days from now.
5. Give an example where $\mathrm{Cov}(X, Y) = 0$ but X and Y are not independent.