Probability and Statistics

Module 2: Random Variables and Distributions

Complete classroom and examination notes

B.Sc. Computer Science • Course Code DASH250011

Based on the NIELIT Deemed to be University 2025 curriculum

Contents

1 Module map 1
2 1. Random variables 1
3 2. Probability mass function (PMF) 1
4 3. Cumulative distribution function (CDF) 1
5 4. Probability density function (PDF) 2
6 5. Expectation and moments 2
7 6. Viance and standard deviation 3
8 7. Binomial distribution 3
9 8. Poisson distribution 4
10 9. Normal distribution 4
10.1 Standardization 4
11 10. Exponential distribution 4
12 11. Comparing the four syllabus distributions 5
13 12. Relationships and useful approximations 5
14 13. Problem-solving checklist 6
15 14. Quick formula sheet 6
16 15. Practice questions with answers 6

1 Module map

The syllabus module contains discrete and continuous random variables; probability mass functions (PMFs); probability density functions (PDFs); cumulative distribution functions (CDFs); expectation; variance; and the Binomial, Poisson, Normal, and Exponential distributions.

> **Learning outcomes.** You should be able to convert an experiment into a numerical random variable, identify its type, construct and validate a PMF/PDF/CDF, calculate probabilities and moments, and select an appropriate standard distribution for a computing or real-world problem.

2 1. Random variables

A **random variable** is a function that maps each outcome ω in a sample space Ω to a real number $X(\omega)$. It is not "random" because the function is unclear; randomness comes from the outcome of the experiment. Examples:

- Number of heads in three coin tosses: $X \in \{0, 1, 2, 3\}$.
- Number of requests arriving at a server in one minute: $X \in \{0, 1, 2, \ldots\}$.
- Lifetime of a battery: $X \in [0, \infty)$.
- Waiting time for a packet: a non-negative continuous quantity.

A random variable is **discrete** if its possible values are finite or countably infinite. It is **continuous** if it can take values over an interval and probabilities are assigned to intervals rather than individual points.

Feature	Discrete	Continuous
Values	Countable values such as $0, 1, 2, \ldots$	Any value in an interval
Model	PMF $p(x)$	PDF $f(x)$
Point probability	$P(X = x)$ may be positive	$P(X = x) = 0$ for every single point
Probability	Sum of bars/values	Area under a curve

3 2. Probability mass function (PMF)

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

$$p_{X(x)} = P(X = x).$$

A valid PMF satisfies:

1. $p_{X(x)} \geq 0$ for every x.
2. $\sum_x p_{X(x)} = 1$, where the sum is over all possible values.
3. For any set A, $P(X \in A) = \sum_{x \in A} p_{X(x)}$.

> **Worked example: number of heads**
>
> For three independent fair tosses, let X be the number of heads. The possible values are $0, 1, 2, 3$ with probabilities $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$. The sum is 1, so this is a valid PMF. $P(X \geq 2) = \frac{3}{8} + \frac{1}{8} = \frac{1}{2}$.

4 3. Cumulative distribution function (CDF)

For any random variable X, the **CDF** is

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

Properties of every CDF:

- $0 \le F_{X(x)} \le 1$.
- $F_{X(x)}$ is non-decreasing.
- $F_{X(x)}$ is right-continuous.
- $\lim_{x\to-\infty} F_{X(x)} = 0$ and $\lim_{x\to\infty} F_{X(x)} = 1$.

For a discrete variable, the CDF is a step function and jumps at values having positive mass. For any $a < b$,

$$P(a < X \le b) = F_{X(b)} - F_{X(a)}.$$

For a discrete variable, $P(X = x) = F_{X(x)} - F_{X(x^-)}$, where $F_{X(x^-)}$ is the left-hand limit.

5 4. Probability density function (PDF)

A continuous random variable is described by a **probability density function** $f_{X(x)}$ satisfying

$$f_{X(x)} \ge 0, \quad \int_{-\infty}^{\infty} f_{X(x)}\,\mathrm{d}x = 1.$$

Probabilities are areas:

$$P(a \le X \le b) = \int_a^b f_{X(x)}\,\mathrm{d}x.$$

The CDF and PDF are connected by

$$F_{X(x)} = \int_{-\infty}^{x} f_{X(t)}\,\mathrm{d}t, \quad f_{X(x)} = F'_{X(x)}$$

where the derivative exists. For a continuous variable, $P(X = x) = 0$; endpoints do not affect interval probabilities.

> **PMF versus PDF.** A PMF value is itself a probability. A PDF value is a density and may exceed 1; only its integrated area over an interval is a probability.

6 5. Expectation and moments

The **expected value** or mean of X is its long-run average.

For a discrete variable:

$$E[X] = \sum_x x p_{X(x)}.$$

For a continuous variable:

$$E[X] = \int_{-\infty}^{\infty} x f_{X(x)}\,\mathrm{d}x.$$

More generally, for a function g:

$$E[g(X)] = \sum_x g(x) p_{X(x)} \quad \text{(discrete)}$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_{X(x)}\,\mathrm{d}x \quad \text{(continuous)}$$

The **linearity of expectation** holds even without independence:

$$E[aX + b] = aE[X] + b, \quad E[X + Y] = E[X] + E[Y].$$

The r-th raw moment is $E[X^r]$. The r-th central moment is $E[(X-\mu)^r]$, where $\mu = E[X]$.

7 6. Viance and standard deviation

Viance measures spread around the mean:

$$V(X) = E[(X-\mu)^2].$$

The computational form is

$$V(X) = E[X^2] - (E[X])^2.$$

Standard deviation is $\sigma = \sqrt{V(X)}$ and has the same units as X.

Useful rules:

$$V(aX+b) = a^2 V(X)$$

$$V(X+Y) = V(X) + V(Y) + 2C(X,Y)$$

If X and Y are independent, $C(X,Y) = 0$, so $V(X+Y) = V(X) + V(Y)$.

Worked example: a small PMF

Let $P(X=0) = 0.2$, $P(X=1) = 0.5$, and $P(X=2) = 0.3$. Then $E[X] = 0(0.2) + 1(0.5) + 2(0.3) = 1.1$. Also $E[X^2] = 0 + 0.5 + 4(0.3) = 1.7$, so $V(X) = 1.7 - 1.1^2 = 0.49$ and $\sigma = 0.7$.

8 7. Binomial distribution

The **Binomial** model counts successes in n independent, identical Bernoulli trials, each with success probability p. Let X be the number of successes and $q = 1-p$. Then $X \sim \text{Binomial}(n,p)$ and

$$P(X=k) = \binom{n}{k} p^{k(1-p)^{n-k}}, \quad k = 0, 1, \ldots, n.$$

Conditions: fixed number of trials; two outcomes per trial; constant success probability; independence of trials.

Mean and variance:

$$E[X] = np, \quad V(X) = npq.$$

Question	Use
Exactly k successes	$P(X=k)$
At most k	$P(X \le k) = \sum_{i=0}^{k} P(X=i)$
At least k	$P(X \ge k) = 1 - P(X \le k-1)$
No successes	$P(X=0) = q^n$
At least one	$P(X \ge 1) = 1 - q^n$

Worked example: packet errors

If each of 10 packets is independently lost with probability 0.1, the number lost $X \sim \text{Binomial}(10, 0.1)$. The probability of exactly two losses is $\binom{10}{2}(0.1)^2(0.9)^8 \approx 0.1937$.

9 8. Poisson distribution

The **Poisson** distribution models the number of events in a fixed interval when events occur independently at a constant average rate λ. If $X \sim \text{Poisson}(\lambda)$,

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots$$

Mean and variance:

$$E[X] = \lambda, \quad V(X) = \lambda.$$

For a time interval of length t with rate r per unit time, use $\lambda = rt$. The probability of no events is $e^{-\lambda}$, and the probability of at least one event is $1 - e^{-\lambda}$.

Poisson is often a good approximation to Binomial(n, p) when n is large, p is small, and $\lambda = np$ is moderate.

Worked example: server arrivals

A server receives an average of 3 requests per minute. Model the arrivals in one minute by $X \sim \text{Poisson}(3)$. Then $P(X = 5) = e^{-3}\frac{3^5}{5!} \approx 0.1008$.

10 9. Normal distribution

A continuous variable X has a **Normal** distribution with mean μ and variance σ^2, written $X \sim N(\mu, \sigma^2)$, when its density is

$$f_{X(x)} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty.$$

The curve is bell-shaped and symmetric about μ. Its mean is μ, variance is σ^2, and standard deviation is σ.

10.1 Standardization

Convert X to a standard normal variable Z:

$$Z = \frac{X - \mu}{\sigma}, \quad Z \sim N(0, 1).$$

Then use the standard normal CDF $\Phi(z) = P(Z \le z)$:

$$P(X \le x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

For an interval: $P(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$.

Symmetry: $\Phi(-z) = 1 - \Phi(z)$. The empirical rule is approximately 68% within 1σ, 95% within 2σ, and 99.7% within 3σ of the mean.

Worked example: response time

Suppose response time is approximately $N(200, 30^2)$ milliseconds. The probability that it is at most 230 ms is $P(X \le 230) = \Phi\left(\frac{230-200}{30}\right) = \Phi(1) \approx 0.8413$.

11 10. Exponential distribution

The **Exponential** distribution models a waiting time until the next event in a Poisson process. If $X \sim \text{Exponential}(\lambda)$ with rate $\lambda > 0$,

$$f_{X(x)} = \lambda e^{-\lambda x}, \quad x \geq 0.$$

CDF and survival function:

$$F_{X(x)} = P(X \leq x) = 1 - e^{-\lambda x}, \quad S(x) = P(X > x) = e^{-\lambda x}.$$

Mean and variance:

$$E[X] = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2}.$$

The exponential distribution is **memoryless**:

$$P(X > s + t \texttt{ mid } X > s) = P(X > t).$$

This means that after already waiting s units, the additional waiting time behaves like a fresh exponential waiting time.

Worked example: time to next request

If requests arrive at rate 4 per minute, the waiting time $X \sim \text{Exponential}(4)$. The probability of waiting more than 30 seconds (0.5 minute) is $P(X > 0.5) = e^{-4(0.5)} = e^{-2} \approx 0.1353$.

12 11. Comparing the four syllabus distributions

Model	Type	Parameters	Typical use
Binomial	Discrete	n, p	Count successes in a fixed number of trials.
Poisson	Discrete	λ	Count events in a fixed interval at a constant rate.
Normal	Continuous	μ, σ^2	Measurements influenced by many small effects; symmetric data.
Exponential	Continuous	λ	Waiting time between Poisson events; non-negative and memoryless.

Model-selection shortcut. Ask: "Am I counting successes in a fixed number of trials?" → Binomial. "Am I counting events in an interval?" → Poisson. "Am I measuring a roughly symmetric continuous quantity?" → Normal. "Am I measuring time until the next event?" → Exponential.

13 12. Relationships and useful approximations

- A sum of independent Bernoulli(p) variables is Binomial(n, p).
- A Binomial(n, p) variable can be approximated by Poisson(np) for rare events.
- A Poisson process has Poisson counts and Exponential inter-arrival times.
- Sums of many small independent effects are often approximately Normal; this idea is formalized further by the Central Limit Theorem in Module 5.
- For a continuous variable, probabilities are areas, not heights.

14 13. Problem-solving checklist

1. Define X in one sentence and list its possible values or support.
2. Decide whether X is discrete or continuous.
3. Identify the model and verify its assumptions.
4. Write the PMF/PDF/CDF before substituting numbers.
5. Use a complement for "at least," "more than," or "not."
6. Standardize a Normal variable with $z = \frac{x-\mu}{\sigma}$.
7. Keep units consistent, especially for rates and time.
8. Check normalization, range, and whether the answer is reasonable.

15 14. Quick formula sheet

Concept	Formula
Discrete expectation	$E[X] = \sum_x xp(x)$
Continuous expectation	$E[X] = \int xf(x)\,\mathrm{d}x$
Viance	$V(X) = E[X^2] - (E[X])^2$
Binomial PMF	$\binom{n}{k}p^{k(1-p)^{n-k}}$
Binomial moments	$E[X] = np, V(X) = np(1-p)$
Poisson PMF	$e^{-\lambda}\frac{\lambda^k}{k!}$
Poisson moments	$E[X] = V(X) = \lambda$
Normal standardization	$z = \frac{x-\mu}{\sigma}$
Exponential survival	$P(X > x) = e^{-\lambda x}$
Exponential moments	$E[X] = \frac{1}{\lambda}, V(X) = \frac{1}{\lambda^2}$

16 15. Practice questions with answers

1. A fair coin is tossed 8 times. Find the probability of exactly 5 heads. **Answer:** $\binom{8}{5}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^3 = \frac{56}{256} = 0.21875$.
2. A web service receives on average 6 requests per second. Find the probability of no requests in the next half-second. **Answer:** $e^{\{-6(0.5)\}} = e^{\{-3\}} \approx 0.0498$.
3. Let $X \sim N(50, 10^2)$. Find $P(40 \le X \le 60)$. **Answer:** $P(-1 \le Z \le 1) \approx 0.6827$.
4. Let $X \sim \text{Exponential}(0.2)$ per minute. Find $P(X \le 5)$. **Answer:** $1 - e^{\{-0.2(5)\}} = 1 - e^{\{-1\}} \approx 0.6321$.
5. A valid PMF has values $0.1, 0.2, 0.4, 0.3$. Find its mean if the corresponding x values are $1, 2, 3, 4$. **Answer:** $1(0.1) + 2(0.2) + 3(0.4) + 4(0.3) = 2.9$.
6. Explain the difference between $P(X = x)$ for a discrete and continuous variable. **Answer:** It can be positive for a discrete variable; it is exactly 0 for a continuous variable, where interval probabilities are obtained by integration.

> **Exam reminder.** State the support, parameters, and assumptions of every distribution you use. A distribution name alone is not a justification: explain why the experiment satisfies its conditions.