Probability distributions

Overview

A **random variable** X assigns a numerical value to each outcome of an experiment.

Type	Description	Examples
Discrete	Countable values (0, 1, 2, …)	Number of defects, coin flips
Continuous	Any value in an interval	Height, time, temperature

Expected value and variance

For any distribution:

Expected Value: $E(X) = \mu$ Variance: $\text{Var}(X) = \sigma^2$ Standard Deviation: $\sigma = \sqrt{\text{Var}(X)}$

Properties:

- $E(aX + b) = aE(X) + b$
- $\text{Var}(aX + b) = a^2\text{Var}(X)$
- $E(X + Y) = E(X) + E(Y)$
- If X and Y are independent: $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

DISCRETE DISTRIBUTIONS

Binomial distribution

Use when: Counting the number of successes in n independent trials, each with probability p of success.
Notation: $X \sim \text{Binomial}(n, p)$

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } k = 0, 1, 2, \ldots, n$$

$$E(X) = np \qquad \text{Var}(X) = np(1-p)$$

Example: A manufacturer knows 5% of items are defective. In a batch of 20 items, what is the probability that exactly 2 are defective?

Student Learning Support, Teaching and Learning Centre
studentlearning@ontariotechu.ca
ontariotechu.ca/studentlearning

Here $n = 20$, $p = 0.05$, $k = 2$.

$$P(X = 2) = \binom{20}{2}(0.05)^2(0.95)^{18} = 190(0.0025)(0.3972) \approx 0.189$$

Example: What is the probability of **at most 1** defective item?

$$P(X \leq 1) = P(X = 0) + P(X = 1) = \binom{20}{0}(0.05)^0(0.95)^{20} + \binom{20}{1}(0.05)^1(0.95)^{19}$$

$$= (1)(1)(0.3585) + (20)(0.05)(0.3774) = 0.3585 + 0.3774 = 0.736$$

Geometric distribution

Use when: Counting the number of trials until the **first** success.
Notation: $X \sim$ Geometric(p)

$$P(X = k) = (1-p)^{k-1}p \qquad \text{for } k = 1, 2, 3, \ldots$$

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

Example: A basketball player makes 70% of free throws. What is the probability that she makes her first basket on the 3rd attempt?
Here $p = 0.7$, $k = 3$.

$$P(X = 3) = (1 - 0.7)^{3-1}(0.7) = (0.3)^2(0.7) = 0.063$$

Expected attempts until first success: $E(X) = \dfrac{1}{0.7} \approx 1.43$ attempts.

Poisson distribution

Use when: Counting the number of events occurring in a fixed interval of time or space, when events occur independently at a constant average rate λ.
Notation: $X \sim$ Poisson(λ)

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

$$E(X) = \lambda \qquad \text{Var}(X) = \lambda$$

Example: A call center receives an average of 4 calls per hour. What is the probability of receiving exactly 6 calls in an hour?

Here $\lambda = 4$, $k = 6$.

$$P(X = 6) = \frac{4^6 e^{-4}}{6!} = \frac{4096(0.0183)}{720} \approx 0.104$$

Example: What is the probability of receiving **no calls** in 30 minutes?
For 30 minutes, $\lambda = 4 \times 0.5 = 2$.

$$P(X = 0) = \frac{2^0 e^{-2}}{0!} = e^{-2} \approx 0.135$$

Tip: The Poisson distribution can approximate the Binomial when n is large and p is small (use $\lambda = np$).

CONTINUOUS DISTRIBUTIONS

For continuous distributions, probabilities are found by integrating the probability density function (pdf):

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx$$

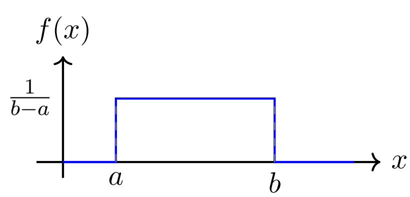
Note: For continuous distributions, $P(X = a) = 0$ for any specific value a. So $P(X \leq a) = P(X < a)$.

Uniform distribution

Use when: All values in an interval $[a, b]$ are equally likely.
Notation: $X \sim$ Uniform(a, b)

$$f(x) = \frac{1}{b-a} \quad \text{for } a \leq x \leq b \qquad P(c \leq X \leq d) = \frac{d-c}{b-a}$$

$$E(X) = \frac{a+b}{2} \qquad \text{Var}(X) = \frac{(b-a)^2}{12}$$



Example: A bus arrives randomly between 8:00 and 8:30 AM. What is the probability it arrives between 8:10 and 8:20?
Here $a = 0$, $b = 30$ (minutes), and we want $P(10 \leq X \leq 20)$.

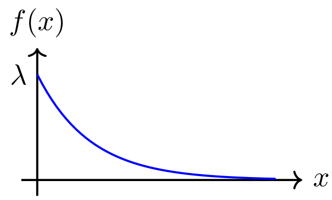
$$P(10 \leq X \leq 20) = \frac{20-10}{30-0} = \frac{10}{30} = \frac{1}{3} \approx 0.333$$

Exponential distribution

Use when: Modeling the time between events in a Poisson process (e.g., time until next failure, time between arrivals).
Notation: $X \sim \text{Exponential}(\lambda)$ where λ is the rate parameter.

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0 \qquad P(X \leq x) = 1 - e^{-\lambda x}$$

$$E(X) = \frac{1}{\lambda} \qquad \text{Var}(X) = \frac{1}{\lambda^2}$$



Example: Light bulbs fail at a rate of $\lambda = 0.01$ per hour (i.e., average lifetime is 100 hours). What is the probability a bulb lasts more than 150 hours?

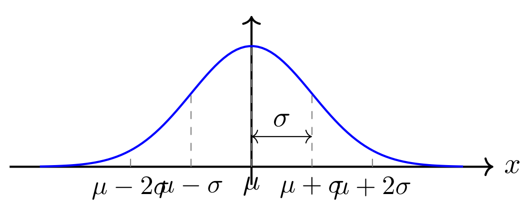
$$P(X > 150) = 1 - P(X \leq 150) = 1 - (1 - e^{-0.01 \times 150}) = e^{-1.5} \approx 0.223$$

Memoryless property: $P(X > s + t \mid X > s) = P(X > t)$
The probability of surviving another t hours doesn't depend on how long you've already survived.

Normal distribution

Use when: Data is symmetric and bell-shaped; many natural phenomena follow this distribution.
Notation: $X \sim N(\mu, \sigma^2)$ where μ is the mean and σ^2 is the variance.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Empirical Rule (68-95-99.7):

- 68% of data falls within $\mu \pm \sigma$
- 95% of data falls within $\mu \pm 2\sigma$
- 99.7% of data falls within $\mu \pm 3\sigma$

Student Learning Support, Teaching and Learning Centre
studentlearning@ontariotechu.ca
ontariotechu.ca/studentlearning

Standard normal and z-scores

The **standard normal distribution** has $\mu = 0$ and $\sigma = 1$: $Z \sim N(0,1)$.
To convert any normal variable to standard normal, use the **z-score**:

$$z = \frac{x - \mu}{\sigma}$$

The z-score tells you how many standard deviations x is from the mean.
Example: Test scores are normally distributed with $\mu = 72$ and $\sigma = 8$. What is the probability a student scores above 84?
First, convert to a z-score:

$$z = \frac{84 - 72}{8} = \frac{12}{8} = 1.5$$

From the z-table: $P(Z < 1.5) = 0.9332$
Therefore: $P(X > 84) = P(Z > 1.5) = 1 - 0.9332 = 0.0668$
About 6.7% of students score above 84.
Example: What score corresponds to the 90th percentile?
From the z-table, the 90th percentile corresponds to $z \approx 1.28$.
Convert back: $x = \mu + z\sigma = 72 + 1.28(8) = 72 + 10.24 = 82.24$

Distribution selection guide

Scenario	Distribution
Fixed number of trials, counting successes	Binomial
Trials until first success	Geometric
Events in fixed time/space at constant rate	Poisson
All values equally likely in an interval	Uniform
Time between Poisson events	Exponential
Symmetric, bell-shaped data	Normal

Summary of formulas

Distribution	Parameters	$E(X)$	Var(X)
Binomial(n,p)	n trials, prob p	np	$np(1-p)$
Geometric(p)	prob p	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson(λ)	rate λ	λ	λ
Uniform(a,b)	interval $[a,b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential(λ)	rate λ	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal(μ,σ^2)	mean μ, var σ^2	μ	σ^2

Student Learning Support, Teaching and Learning Centre
studentlearning@ontariotechu.ca
ontariotechu.ca/studentlearning