DegreeLive

Bachelor of Commerce (B2001) | Level 1 • Part A Core | Monash University

Business and economic statistics

Course Code: ETC1000 • Module 2 Notes

Module 2: Probability Theory, Random Variables, Normal Distribution and Sampling Distributions

2.1 Foundations of Probability, Sample Spaces, and Counting

Business and economic decisions are rarely made under conditions of perfect certainty. Whether an Australian bank prices default risk on a portfolio of residential mortgages, an energy retailer hedges wholesale electricity spot prices in the National Electricity Market, or a Melbourne manufacturer plans warranty reserves, analysts must quantify the likelihood of uncertain future outcomes. Probability theory supplies the mathematical language for modelling uncertainty, bridging the descriptive summaries of Module 1 with the formal statistical inference developed in Module 3.

Random Experiments, Sample Spaces, and Events

A **random experiment** is any well-defined process that generates one outcome from a known set of possible results, where the exact outcome cannot be predicted with certainty prior to execution. The **sample space (S)** is the exhaustive collection of all distinct elementary outcomes (sample points) of the experiment. Any subset $A \subseteq S$ is called an **event**. For instance, when assessing a commercial borrower's quarterly credit outcome, $S = \{\text{Paid in Full},\ \text{30-Day Arrears},\ \text{90-Day Default}\}$, and the event $A = \{\text{30-Day Arrears},\ \text{90-Day Default}\}$ represents credit impairment.

Probability is assigned under three complementary frameworks in commerce and economics:

- **Classical (a priori) probability:** Assumes N equally likely elementary outcomes, assigning $P(A) = n(A)/N$. This applies to symmetric randomisation mechanisms such as lottery draws, audit-selection random number generators, or balanced experimental assignment.
- **Empirical (relative frequency) probability:** Defines $P(A)$ as the long-run proportion of trials in which A occurs under stable conditions, $\lim_{n\to\infty} f_A/n$. Insurers such as IAG or Suncorp rely on empirical claim frequencies across millions of policy-years to price motor and home premiums.
- **Subjective probability:** Quantifies a decision-maker's calibrated degree of belief based on expert judgement, macroeconomic briefings, and incomplete historical precedent—for example, an investment bank's assessed probability that the Reserve Bank of Australia alters the Cash Rate Target at its next policy meeting.

Kolmogorov's Axioms and Set Operations

Regardless of interpretation, every valid probability assignment obeys three foundational axioms:

1. **Non-negativity:** $0 \le P(A) \le 1$ for every event $A \subseteq S$.
2. **Certainty of the sample space:** $P(S) = 1$.
3. **Countable additivity:** If $A_1, A_2, \ldots$ are pairwise **mutually exclusive (disjoint)** events ($A_i \cap A_j = \emptyset$ for $i \neq j$), then $P(A_1 \cup A_2 \cup \ldots) = \sum_i P(A_i)$.

For two events A and B, the **complement** A^c (or $\bar{A}$) consists of all sample points in S not in A; the **union** $A \cup B$ contains outcomes in A, B, or both; and the **intersection** $A \cap B$ contains outcomes belonging simultaneously to both A and B. Two events are **collectively exhaustive** if $A \cup B = S$.

Counting Principles: Permutations and Combinations

When a sample space consists of a large number of equally likely arrangements or selections, direct enumeration becomes impractical. Combinatorial counting rules compute sample-space sizes systematically:

- **Multiplication rule:** If a sequential procedure consists of k steps with $n_1, n_2, \ldots, n_k$ choices at each stage, the total number of distinct sequences is $n_1 \times n_2 \times \cdots \times n_k$.
- **Permutations (P_r^n):** The number of *ordered* arrangements of r distinct items selected without replacement from n distinct items (relevant when sequence matters, such as ranked investment preferences or serial codes):

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

  Where: n is the total number of distinct items available, r is the number of items selected and ordered, and $n! = n(n-1)\cdots(1)$ with $0! = 1$.

- **Combinations ($\binom{n}{r}$):** The number of *unordered* subsets of r distinct items selected without replacement from n distinct items (relevant for audit sampling and equity portfolio selection):

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

  Where: $\binom{n}{r}$ is the binomial coefficient counting unordered subsets of size r chosen from n items.

For example, if an Australian investment committee selects an unordered sub-portfolio of $r = 4$ ASX 50 equities from a shortlist of $n = 12$ candidates, there are $\binom{12}{4} = \frac{12!}{4!8!} = \frac{12\times11\times10\times9}{24} = 495$ distinct equal-weighted portfolios. By contrast, if the committee ranks its top $r = 3$ choices in strict order of priority from $n = 8$ infrastructure projects, sequence matters and the number of ordered rankings is $P_3^8 = \frac{8!}{5!} = 8 \times 7 \times 6 = 336$.

DegreeLive

2.2 Rules of Probability, Conditional Probability, and Independence

Combining probabilities of elementary business risks into composite portfolio or operational probabilities relies on three core algebraic rules derived directly from Kolmogorov's axioms.

Complement and Addition Rules

Because A and A^c partition S, the **complement rule** states $P(A^c) = 1 - P(A)$. This is frequently used to compute "at least one" probabilities (e.g., $P(\text{at least one default}) = 1 - P(\text{zero defaults})$). For any two events A and B, the **general addition rule** avoids double-counting the overlapping intersection $A \cap B$:

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

Where: $P(A \cup B)$ is the probability that event A, event B, or both occur; $P(A)$ and $P(B)$ are marginal probabilities; and $P(A \cap B)$ is the joint probability that both occur simultaneously. When A and B are mutually exclusive ($P(A \cap B) = 0$), the formula simplifies to $P(A \cup B) = P(A) + P(B)$.

Conditional Probability and the Multiplication Rule

Decision-makers constantly update risk assessments when partial information arrives. The **conditional probability** of event A given that event B has occurred (with $P(B) > 0$) restricts the relevant sample space to B:

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

Where: $P(A \mid B)$ is the conditional probability of A given B, $P(A \cap B)$ is the joint probability of A and B, and $P(B) > 0$ is the marginal probability of the conditioning event B. Rearranging yields the **general multiplication rule** for joint probabilities:

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

Where: $P(A \cap B)$ is obtained by multiplying the marginal probability of one event by the conditional probability of the other.

Statistical Independence versus Mutual Exclusivity

Events A and B are **statistically independent** if knowing that B occurred provides zero information that alters the probability of A:

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

Where: $P(A \mid B)$ is the conditional probability and $P(A)P(B)$ is the product of the marginal probabilities. Analysts must never confuse *independence* with *mutual exclusivity*. If A and B have non-zero probabilities ($P(A) > 0$ and $P(B) > 0$) and are mutually exclusive ($P(A \cap B) = 0$), then $P(A \cap B) = 0 \neq P(A)P(B)$. Thus, two positive-probability mutually exclusive events are strongly *dependent*: the occurrence of B guarantees that A cannot occur ($P(A \mid B) = 0$).

In structured credit and operational-risk modelling, analysts also apply **De Morgan's laws** to evaluate joint system failures: $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$. If two independent backup payment gateways have outage probabilities $P(A) = 0.02$ and $P(B) = 0.03$, the probability that *at least one* remains operational is $1 - P(A \cap B) = 1 - (0.02)(0.03) = 0.9994$ (99.94%). However, during systemic macroeconomic stress (such as an RBA rate-tightening shock), default events across borrowers exhibit positive tail dependence, making independence assumptions dangerously optimistic.

Worked Numerical Example 1: Mortgage Arrears Contingency Analysis

An Australian commercial bank audits a random sample of 1,000 residential mortgages across interest-rate structure (F: Fixed-rate; V: Variable-rate) and repayment status (A: 90+ days in arrears; C: Current). The joint count contingency table and resulting joint probability matrix (dividing every cell by 1,000) are shown below:

Rate Structure	Current (C)	90+ Days Arrears (A)	Marginal Total
Fixed-Rate (F)	380 (0.380)	20 (0.020)	400 (0.400)
Variable-Rate (V)	555 (0.555)	45 (0.045)	600 (0.600)
Marginal Total	935 (0.935)	65 (0.065)	1,000 (1.000)

1. **Joint and Marginal Probabilities:** The overall arrears rate is $P(A) = 0.065$ (6.5%), and the joint probability of a mortgage being both variable-rate and in arrears is $P(V \cap A) = 0.045$ (4.5%).
2. **Conditional Probability:** Given a borrower holds a variable-rate loan, what is the probability of arrears?

DegreeLive

$$P(A \mid V) = \frac{P(A \cap V)}{P(V)} = \frac{0.045}{0.600} = 0.0750 \ (7.50\%)$$

Where: $P(A \cap V) = 0.045$ is the joint probability of variable-rate arrears and $P(V) = 0.600$ is the marginal probability of a variable-rate loan. By comparison, $P(A \mid F) = 0.020/0.400 = 0.0500$ (5.00%).

3. **Addition Rule:** What is the probability a randomly selected mortgage is either variable-rate or in arrears?

$$P(V \cup A) = P(V) + P(A) - P(V \cap A) = 0.600 + 0.065 - 0.045 = 0.620 \ (62.0\%)$$

Where: 0.600 is $P(V)$, 0.065 is $P(A)$, and 0.045 removes the double-counted intersection $P(V \cap A)$.

4. **Independence Check:** Since $P(A \mid V) = 0.0750 \neq P(A) = 0.0650$ (equivalently, $P(V)P(A) = 0.600 \times 0.065 = 0.0390 \neq 0.0450$), repayment status and rate structure are *not independent* in this portfolio.

2.3 Law of Total Probability and Bayes' Theorem

In corporate auditing, diagnostic screening, and algorithmic credit scoring, analysts frequently know the conditional reliability of a signal given an underlying hidden state—for example, $P(\text{Alert} \mid \text{Fraud})$—yet need the inverse probability: given an alert has triggered, what is the probability that fraud actually occurred?

Law of Total Probability

Suppose events $B_1, B_2, \ldots, B_k$ form a **partition** of the sample space S (they are mutually exclusive and collectively exhaustive, so $\sum_{i=1}^{k} P(B_i) = 1$). Then any event $A \subseteq S$ can be decomposed across the partition via the **Law of Total Probability**:

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

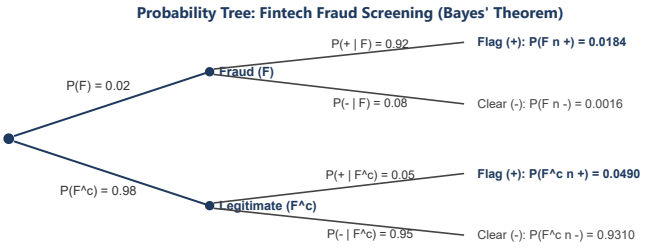
Where: $B_1, \ldots, B_k$ is a partition of S, $P(B_i)$ is the prior probability of state B_i, and $P(A \mid B_i)$ is the conditional likelihood of observing event A when state B_i holds.

Bayes' Theorem

Combining the definition of conditional probability with the Law of Total Probability yields **Bayes' theorem**, which updates initial (*prior*) probabilities $P(B_i)$ into revised (*posterior*) probabilities $P(B_i \mid A)$ after observing evidence A:

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{P(A)} = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{k} P(A \mid B_i)P(B_i)}$$

Where: $P(B_j)$ is the prior probability of state B_j, $P(A \mid B_j)$ is the likelihood of evidence A under state B_j, $P(A)$ is the total marginal probability of observing A across all states, and $P(B_j \mid A)$ is the posterior probability of state B_j given evidence A.



Worked Numerical Example 2: Automated Payment Fraud Detection

A Sydney fintech processes card-not-present transactions using a machine-learning anomaly filter. Historically, 2% of incoming transactions are genuinely fraudulent ($P(F) = 0.02$, so $P(F^c) = 0.98$). The filter correctly flags 92% of fraudulent transactions ($P(+ \mid F) = 0.92$, sensitivity) but mistakenly flags 5% of legitimate transactions ($P(+ \mid F^c) = 0.05$, false-positive rate). If a transaction triggers a fraud alert ($+$), what is the posterior probability that it is actually fraudulent?

Step 1: Compute the total probability of an alert via the Law of Total Probability.

DegreeLive

$$P(+) = P(+ \mid F)P(F) + P(+ \mid F^c)P(F^c) = (0.92)(0.02) + (0.05)(0.98) = 0.0184 + 0.0490 = 0.0674$$

Where: 0.0184 is the joint probability of a true positive and 0.0490 is the joint probability of a false positive, giving an overall alert rate of 6.74%.

Step 2: Apply Bayes' Theorem to find the posterior probability $P(F \mid +)$.

$$P(F \mid +) = \frac{P(+ \mid F)P(F)}{P(+)} = \frac{0.0184}{0.0674} = 0.2730 \ (27.30\%)$$

Where: $P(F \mid +)$ is the posterior fraud probability given an alert. Even with a 92% true-positive rate, nearly three-quarters (72.70%) of flagged alerts are legitimate transactions because the underlying base rate of fraud (2%) is small relative to the 98% volume of legitimate traffic. This base-rate insight explains why Australian banks send automated SMS verification prompts rather than immediately cancelling cards on an initial flag.

Multi-State Bayesian Updating Worksheet

When a sample space is partitioned into three or more macroeconomic regimes—such as Expansion (E, prior 0.50), Stagnation (M, prior 0.35), and Contraction (R, prior 0.15)—Bayes' theorem is systematically organised in a tabular worksheet. Suppose an economic consultancy evaluates a leading signal S (positive Melbourne CBD commercial leasing absorption), whose conditional likelihoods are $P(S \mid E) = 0.80$, $P(S \mid M) = 0.40$, and $P(S \mid R) = 0.10$:

Economic Regime (B_i)	Prior $P(B_i)$	Likelihood $P(S \mid B_i)$	Joint $P(S \cap B_i) = P(B_i)P(S \mid B_i)$	Posterior $P(B_i \mid S) = P(S \cap B_i)/P(S)$
Expansion (E)	0.50	0.80	0.400	$0.400/0.555 = 0.7207$
Stagnation (M)	0.35	0.40	0.140	$0.140/0.555 = 0.2523$
Contraction (R)	0.15	0.10	0.015	$0.015/0.555 = 0.0270$
Total	**1.00**	—	$P(S) = 0.555$	**1.0000**

Observing positive leasing absorption shifts the posterior probability of Expansion upward from 50.0% to 72.07%, while compressing the probability of Contraction from 15.0% down to 2.70%.

2.4 Discrete Random Variables and Probability Distributions

A **random variable (X)** is a numerical function that assigns a real number to each elementary outcome in a sample space S. Random variables are denoted by uppercase letters (X, Y), while specific realised values are denoted by lowercase letters (x, y). A random variable is **discrete** if it can assume at most a finite or countably infinite list of isolated values.

Probability Mass Function (PMF) and Cumulative Distribution Function (CDF)

The probability distribution of a discrete random variable X is specified by its **probability mass function (PMF)**, $p(x) = P(X = x)$, which satisfies two axioms: $0 \le p(x) \le 1$ for every x, and $\sum_{\text{all } x} p(x) = 1$. Its **cumulative distribution function (CDF)** gives the accumulated probability up to and including x: $F(x) = P(X \le x) = \sum_{t \le x} p(t)$.

Expected Value, Variance, and Linear Combinations

The **expected value ($E(X)$ or μ)** of a discrete random variable is the probability-weighted average of all possible values, representing the long-run theoretical mean over infinitely many independent repetitions. The **variance ($\mathrm{Var}(X)$ or σ^2)** measures expected squared dispersion around μ:

$$E(X) = \mu = \sum_{\text{all } x} x\,p(x) \qquad \text{and} \qquad \mathrm{Var}(X) = \sigma^2 = \sum_{\text{all } x} (x - \mu)^2 p(x) = E(X^2) - [E(X)]^2$$

Where: x ranges over all possible values of X, $p(x) = P(X = x)$ is the PMF, $E(X^2) = \sum x^2 p(x)$, and $\sigma = \sqrt{\mathrm{Var}(X)}$ is the standard deviation.

Under any linear transformation $Y = a + bX$ (for constants a, b), expectation and variance transform as:

$$E(a + bX) = a + bE(X) \qquad \text{and} \qquad \mathrm{Var}(a + bX) = b^2 \mathrm{Var}(X)$$

Where: a is an additive shift constant and b is a multiplicative scaling factor. In portfolio finance, analysts combine two risky assets X and Y with weights w_1 and w_2 into a portfolio return $W = w_1 X + w_2 Y$. While expected return is strictly additive, portfolio variance depends on the **covariance** $\mathrm{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$ or correlation $\rho_{XY} = \mathrm{Cov}(X, Y)/(\sigma_X \sigma_Y)$:

DegreeLive

$$E(w_1X + w_2Y) = w_1E(X) + w_2E(Y)$$

$$\text{Var}(w_1X + w_2Y) = w_1^2\text{Var}(X) + w_2^2\text{Var}(Y) + 2w_1w_2\text{Cov}(X,Y)$$

Where: w_1, w_2 are portfolio weights, $\text{Var}(X)$ and $\text{Var}(Y)$ are individual asset return variances, and $\text{Cov}(X,Y) = \rho_{XY}\sigma_X\sigma_Y$ captures co-movement. Whenever $\rho_{XY} < 1$, diversification reduces portfolio risk below the weighted average of individual standard deviations.

Worked Numerical Example 3: Warranty Claims and Two-Asset ASX Portfolio

Part A (Discrete Random Variable): An Adelaide electronics retailer tracks daily warranty replacement claims X with empirical PMF given in the table below. Each claim costs the firm A$180 in replacement hardware plus A$400 in fixed daily technician standby overhead ($C = 400 + 180X$).

Claims (x)	$p(x)$	$x \cdot p(x)$	$x^2 \cdot p(x)$
0	0.25	0.00	0.00
1	0.35	0.35	0.35
2	0.22	0.44	0.88
3	0.12	0.36	1.08
4	0.06	0.24	0.96
Total	**1.00**	$E(X) = 1.39$	$E(X^2) = 3.27$

Using the shortcut variance formula:

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = 3.27 - (1.39)^2 = 3.27 - 1.9321 = 1.3379 \implies \sigma_X = \sqrt{1.3379} = 1.1567 \text{ claims}$$

Where: $E(X) = 1.39$ is expected daily claims and $\sigma_X = 1.1567$ is the standard deviation. Applying the linear cost transformation $C = 400 + 180X$ gives expected daily cost $E(C) = 400 + 180(1.39) = \text{AUD } 650.20$ and standard deviation $\sigma_C = 180(1.1567) = \text{AUD } 208.21$.

Part B (Two-Asset Portfolio): Suppose an investor splits capital 60% in an Australian Mining ETF (X, $E(X) = 11\%$, $\sigma_X = 18\%$) and 40% in an Australian Infrastructure Bond fund (Y, $E(Y) = 5\%$, $\sigma_Y = 6\%$), with correlation $\rho_{XY} = -0.20$. Then $\text{Cov}(X,Y) = (-0.20)(18)(6) = -21.6$ (in $\%^2$). Portfolio expected return and standard deviation are:

$$E(W) = 0.60(11\%) + 0.40(5\%) = 8.60\%$$

$$\text{Var}(W) = (0.60)^2(18)^2 + (0.40)^2(6)^2 + 2(0.60)(0.40)(-21.6) = 116.64 + 5.76 - 10.368 = 112.032 \implies \sigma_W = \sqrt{112.032} = 10.58\%$$

Where: $w_1 = 0.60, w_2 = 0.40$ are asset weights and $\sigma_W = 10.58\%$ is portfolio return volatility (lower than $0.60(18\%) + 0.40(6\%) = 13.20\%$ because of negative correlation). In general, portfolio volatility depends critically on the correlation coefficient $\rho_{XY} \in [-1, +1]$: if $\rho_{XY} = +1$, $\sigma_W = w_1\sigma_X + w_2\sigma_Y = 13.20\%$ (zero diversification gain); if $\rho_{XY} = -1$, choosing $w_1^* = \sigma_Y/(\sigma_X + \sigma_Y) = 6/(18+6) = 0.25$ and $w_2^* = 0.75$ eliminates portfolio variance entirely ($\sigma_W = |0.25(18\%) - 0.75(6\%)| = 0\%$).

Joint Discrete Probability Distributions and Independence

When modelling two discrete variables (X, Y) simultaneously—such as weekly demand in online (X) and physical (Y) store channels—the **joint PMF** $p(x,y) = P(X = x, Y = y)$ satisfies $p(x,y) \geq 0$ and $\sum_x \sum_y p(x,y) = 1$. Summing across rows or columns yields the **marginal PMFs** $p_X(x) = \sum_y p(x,y)$ and $p_Y(y) = \sum_x p(x,y)$. The conditional PMF is $p_{Y|X}(y \mid x) = p(x,y)/p_X(x)$, and X and Y are statistically independent if and only if $p(x,y) = p_X(x)p_Y(y)$ for every pair (x,y)—in which case $\text{Cov}(X,Y) = 0$ and $\rho_{XY} = 0$.

Consider an Australian omni-channel retailer tracking weekly unit sales in its Online channel ($X \in \{100, 200\}$) and Physical store channel ($Y \in \{50, 100, 150\}$) with joint PMF below:

Online (X) / Physical (Y)	$Y = 50$	$Y = 100$	$Y = 150$	Marginal $p_X(x)$
$X = 100$	0.20	0.15	0.05	0.40
$X = 200$	0.10	0.25	0.25	0.60
Marginal $p_Y(y)$	0.30	0.40	0.30	**1.00**

DegreeLive

From the marginal distributions, $E(X) = 100(0.40) + 200(0.60) = 160$ units and $E(Y) = 50(0.30) + 100(0.40) + 150(0.30) = 100$ units. Computing $E(XY) = \sum_x \sum_y xy\,p(x,y)$ yields 100(50)(0.20) + 100(100)(0.15) + 100(150)(0.05) + 200(50)(0.10) + 200(100)(0.25) + 200(150)(0.25) = 1,000 + 1,500 + 750 + 1,000 + 5,000 + 7,500 = 16,750. Therefore, $\text{Cov}(X,Y) = E(XY) - E(X)E(Y) = 16,750 - (160)(100) = +750 > 0$, confirming positive co-movement (strong promotional weeks boost both channels together) and statistical dependence since $p(100, 50) = 0.20 \neq p_X(100)p_Y(50) = 0.12$.

Discrete versus Continuous Random Variables: Structural Comparison

Property	Discrete Random Variable	Continuous Random Variable
Support (Possible Values)	Finite or countably infinite isolated points	Uncountable real interval (a, b) or $(-\infty, \infty)$
Probability Function	Probability Mass Function $p(x) = P(X = x)$	Probability Density Function $f(x) \geq 0$
Point Probability	$P(X = c) \geq 0$ (can be strictly positive)	$P(X = c) = 0$ always; intervals have positive area
Total Probability Axiom	$\sum_{\text{all } x} p(x) = 1$	$\int_{-\infty}^{\infty} f(x)\,dx = 1$
Expected Value & Variance	$\mu = \sum xp(x), \sigma^2 = \sum(x-\mu)^2 p(x)$	$\mu = \int xf(x)\,dx, \sigma^2 = \int (x-\mu)^2 f(x)\,dx$

2.5 The Binomial Distribution (and Poisson Overview)

Certain discrete experiments recur so frequently across quality control, marketing response modelling, and credit default analysis that they are governed by named parametric probability distributions.

Binomial Assumptions and Probability Mass Function

A **Binomial random variable** $X \sim \text{Bin}(n, p)$ counts the total number of "successes" across n repeated trials satisfying four strict conditions: (1) the number of trials n is fixed in advance; (2) each trial produces one of two mutually exclusive outcomes ("success" or "failure", i.e., a Bernoulli trial); (3) the probability of success p remains constant across every trial; and (4) all trials are statistically independent. Under these conditions, the binomial PMF, mean, and variance are:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x \in \{0, 1, \ldots, n\}$$

$$\mu = E(X) = np \qquad \text{and} \qquad \sigma^2 = \text{Var}(X) = np(1-p)$$

Where: n is the fixed number of independent trials, p is the single-trial probability of success, x is the exact count of successes, and $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ counts the distinct ways those x successes can occur among n trials.

When sampling without replacement from a finite population of size N, trials are technically slightly dependent; however, as a standard rule of thumb, the binomial model remains an accurate approximation whenever the sample represents less than 5% of the population ($n/N \leq 0.05$). By contrast, when counting rare events occurring randomly over a continuous interval of time or space at an average rate $\lambda > 0$ (such as customer arrivals per hour at a bank branch or server outages per month), the **Poisson distribution** applies: $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ for $x = 0, 1, 2, \ldots$, with $E(X) = \text{Var}(X) = \lambda$.

Worked Numerical Example 4: Brisbane Manufacturing Quality Control

An automotive components plant in Brisbane produces sensor brackets with a known long-run defect rate of $p = 0.08$ (8%). A quality engineer randomly inspects $n = 10$ independent brackets from the production line. Let X denote the number of defective brackets found, so $X \sim \text{Bin}(10, 0.08)$.

1. **Expected Defects and Standard Deviation:**

$$\mu = np = 10(0.08) = 0.80 \text{ defects}, \qquad \sigma = \sqrt{np(1-p)} = \sqrt{10(0.08)(0.92)} = \sqrt{0.736} = 0.8579 \text{ defects}$$

   Where: $n = 10$ is the batch sample size and $p = 0.08$ is the defect probability.

2. **Probability of Exactly Two Defective Brackets ($P(X = 2)$):**

$$P(X = 2) = \binom{10}{2}(0.08)^2(0.92)^8 = 45 \times 0.0064 \times 0.513219 = 0.1478 \ (14.78\%)$$

   Where: $\binom{10}{2} = \frac{10 \times 9}{2} = 45$ is the number of ways to observe 2 defects among 10 units.

3. **Probability of At Least Two Defective Brackets ($P(X \geq 2)$):** Using the complement rule, $P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)]$:

DegreeLive

$$P(X=0)=\binom{10}{0}(0.08)^0(0.92)^{10}=0.43439,\qquad P(X=1)=\binom{10}{1}(0.08)^1(0.92)^9=0.37773$$

$$P(X\geq 2)=1-(0.43439+0.37773)=1-0.81212=0.1879\ (18.79\%)$$

Where: $P(X\leq 1)=0.81212$ is the cumulative probability of zero or one defect. Nearly 18.8% of inspected batches will exhibit two or more defective brackets.

Poisson Distribution for Arrival Counts and Rare Events

When events occur independently and randomly at a constant average intensity $\lambda>0$ over a continuous interval of time, distance, or area—such as customer arrivals at a Melbourne CBD bank branch per 15-minute window, cyber-intrusion attempts per day, or machinery breakdowns per month—the count $X\sim \text{Pois}(\lambda)$ follows the **Poisson distribution**:

$$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!},\qquad x=0,1,2,\ldots\qquad\text{with}\qquad E(X)=\lambda,\quad \text{Var}(X)=\lambda$$

Where: $\lambda>0$ is the mean arrival rate per unit interval, x is the observed non-negative integer count, and $e\approx 2.71828$. For example, if branch arrivals average $\lambda=4.5$ customers per 15-minute interval, then $P(X=0)=e^{-4.5}=0.0111$, $P(X=1)=4.5e^{-4.5}=0.0500$, $P(X=3)=\frac{e^{-4.5}(4.5)^3}{6}=0.1687$, and the probability of at least two customers arriving is $P(X\geq 2)=1-(0.0111+0.0500)=0.9389$ (93.89%).

Furthermore, when n is large ($n\geq 50$) and p is very small ($np\leq 5$), the Poisson distribution accurately approximates $\text{Bin}(n,p)$ by setting $\lambda=np$. For an Australian insurer holding $n=1,000$ commercial cyber policies each carrying an annual claim probability $p=0.003$, $\lambda=1,000(0.003)=3.0$, so the probability of two or more claims is $P(X\geq 2)\approx 1-e^{-3}(1+3)=1-0.1991=0.8009$ (80.09%).

Discrete Model	Parameters	PMF Formula	Mean & Variance	Business Context
Bernoulli	$p\in(0,1)$	$p^x(1-p)^{1-x},\ x\in\{0,1\}$	$\mu=p,\ \sigma^2=p(1-p)$	Single loan default or conversion
Binomial	$n\in\mathbb{N},\ p$	$\binom{n}{x}p^x(1-p)^{n-x}$	$\mu=np,\ \sigma^2=np(1-p)$	Defective items in a fixed batch
Poisson	$\lambda>0$	$e^{-\lambda}\lambda^x/x!$	$\mu=\lambda,\ \sigma^2=\lambda$	Customer arrivals or claims per period

2.6 Continuous Random Variables and the Normal Distribution

A **continuous random variable** takes values over an unbroken real interval. Because there are uncountably many points on an interval, the probability that X equals any exact single point is zero ($P(X=c)=0$). Instead, probabilities are defined over intervals as areas beneath a non-negative **probability density function (PDF)**, $f(x)\geq 0$, whose total area integrates to unity ($\int_{-\infty}^{\infty}f(x)\,dx=1$):

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

Where: $f(x)$ is the probability density function, and $a<b$ are interval endpoints. Because single points carry zero probability, $P(a\leq X\leq b)=P(a<X<b)$. Two foundational continuous benchmarks precede the Normal curve: the **Uniform distribution** $U(a,b)$, where density is constant $f(x)=1/(b-a)$ across $[a,b]$, mean is $(a+b)/2$, and variance is $(b-a)^2/12$ (e.g., if a Melbourne warehouse shuttle departs every 12 minutes, waiting time $W\sim U(0,12)$ has $E(W)=6$ min, $\text{SD}(W)=\sqrt{144/12}=3.46$ min, and $P(3\leq W\leq 9)=(9-3)/12=0.50$); and the **Exponential distribution** $\text{Exp}(\lambda)$ for waiting times between Poisson arrivals, with $f(x)=\lambda e^{-\lambda x}$ $(x\geq 0)$, $P(X\leq x)=1-e^{-\lambda x}$, mean $1/\lambda$, and standard deviation $1/\lambda$.

The Normal Distribution and Standardisation

The **Normal (Gaussian) distribution**, denoted $X\sim N(\mu,\sigma^2)$, is the cornerstone distribution of business statistics and econometrics. Its symmetric, bell-shaped PDF is completely governed by two parameters—mean μ (location) and variance σ^2 (spread):

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

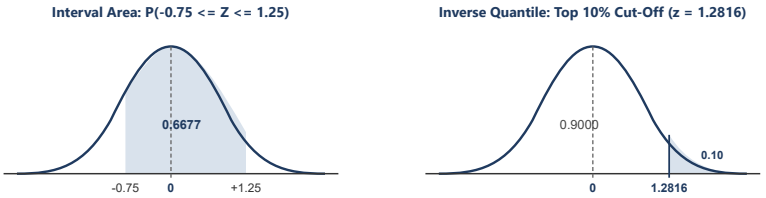
Where: μ is the population mean (equal to median and mode), $\sigma>0$ is the population standard deviation, $\pi\approx 3.14159$, and $e\approx 2.71828$. Inflection points occur at $\mu\pm\sigma$.

Degree Live

Because the normal PDF has no closed-form elementary antiderivative, probabilities are evaluated by transforming any $X \sim N(\mu, \sigma^2)$ into the **Standard Normal random variable** $Z \sim N(0,1)$ via z-standardisation and consulting cumulative standard normal tables ($\Phi(z) = P(Z \leq z)$):

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1) \qquad \text{and (inversely)} \qquad x = \mu + z\sigma$$

Where: Z is the standard normal variate with mean 0 and standard deviation 1, X is the original normal variate, μ is its mean, and σ is its standard deviation.



Worked Numerical Example 5: Forward and Inverse Normal Probabilities

Suppose weekly gross earnings of full-time employees in a Victorian service sector follow $X \sim N(\mu = 1,900,\ \sigma^2 = 200^2)$ in Australian dollars ($\mu = \text{AUD } 1,900, \sigma = \text{AUD } 200$).

1. **Interval Probability:** Find the proportion of employees earning between A$1,750 and A$2,150 per week. Standardise both endpoints:

$$z_1 = \frac{1,750 - 1,900}{200} = -0.75, \qquad z_2 = \frac{2,150 - 1,900}{200} = +1.25$$

From the cumulative standard normal table, $\Phi(1.25) = P(Z \leq 1.25) = 0.8944$ and $\Phi(-0.75) = P(Z \leq -0.75) = 0.2266$. Subtracting gives:

$$P(1,750 \leq X \leq 2,150) = P(-0.75 \leq Z \leq 1.25) = 0.8944 - 0.2266 = 0.6678\ (66.78\%)$$

Where: $\Phi(z)$ denotes the cumulative standard normal distribution function up to z.

2. **Upper-Tail Probability:** What percentage of employees earn more than A$2,250?

$$z = \frac{2,250 - 1,900}{200} = +1.75 \implies P(X > 2,250) = 1 - \Phi(1.75) = 1 - 0.9599 = 0.0401\ (4.01\%)$$

Where: $z = 1.75$ is the standardised upper threshold and $1 - \Phi(1.75)$ is the right-tail probability.

3. **Inverse Quantile Calculation:** What minimum weekly salary places an employee in the top 10% of earners? The top 10% corresponds to cumulative left-tail area 0.9000. From the z-table, $\Phi(z) = 0.9000$ at $z_{0.10} = 1.2816$. Unstandardising:

$$x_{90} = \mu + z_{0.10}\sigma = 1,900 + 1.2816(200) = 1,900 + 256.32 = \text{AUD } 2,156.32$$

Where: $z_{0.10} = 1.2816$ is the 90th percentile of $N(0,1)$ and x_{90} is the 90th percentile salary cut-off.

Standard Normal Reference Quantiles and Continuity-Corrected Binomial Approximation

Business risk modelling and statistical inference rely repeatedly on benchmark Standard Normal critical values, summarised below:

Central Probability / Confidence	Two-Tailed Area (α)	One-Tailed Area ($\alpha/2$)	Standard Normal Cut-Off ($z_{\alpha/2}$)
80.0%	0.20	0.100	1.2816
90.0%	0.10	0.050	1.6449
95.0%	0.05	0.025	1.9600
98.0%	0.02	0.010	2.3263
99.0%	0.01	0.005	2.5758

DegreeLive

Additionally, when n is large and both $np \geq 5$ and $n(1-p) \geq 5$, a discrete Binomial variable $X \sim \text{Bin}(n,p)$ is accurately approximated by $N(\mu = np,\ \sigma^2 = np(1-p))$. Applying a **continuity correction** (± 0.5) aligns discrete integer bars with the continuous normal curve: for example, if $n = 200$ and $p = 0.40$ ($\mu = 80$, $\sigma = \sqrt{48} = 6.9282$), then $P(X \geq 90) \approx P(X_{\text{cont}} \geq 89.5) = P(Z \geq (89.5 - 80)/6.9282) = P(Z \geq 1.3712) = 0.0851$.

Exponential Waiting-Time Example and Selected Cumulative z-Values

To illustrate continuous waiting-time modelling in Australian banking operations, suppose customer arrivals at a Sydney branch follow a Poisson process at rate $\lambda = 12$ customers per hour (0.20 customers per minute). The inter-arrival time T (in minutes) follows $\text{Exp}(0.20)$ with mean $1/0.20 = 5.0$ minutes. The probability that the next customer arrives within 3 minutes is $P(T \leq 3) = 1 - e^{-0.20(3)} = 1 - e^{-0.6} = 0.4512$ (45.12%), while the probability of waiting longer than 10 minutes is $P(T > 10) = e^{-0.20(10)} = e^{-2} = 0.1353$ (13.53%).

z	-2.50	-2.00	-1.50	-1.00	-0.50	0.00	+0.50	+1.00	+1.50	+2.00	+2.50
$\Phi(z)$	0.0062	0.0228	0.0668	0.1587	0.3085	0.5000	0.6915	0.8413	0.9332	0.9772	0.9938

2.7 Sampling Distributions, the Central Limit Theorem, and Standard Error

In practice, analysts observe only a single random sample of size n rather than the entire population. Because a different random sample would yield a slightly different sample mean $\bar{x}$ or sample proportion $\hat{p}$, every sample statistic is itself a random variable possessing a probability distribution across all possible samples of size n, termed its **sampling distribution**.

Sampling Error versus Non-Sampling Error

The unavoidable random discrepancy between a sample statistic and its target population parameter ($|\bar{X} - \mu|$) caused solely by observing a subset rather than a census is called **sampling error**; it diminishes predictably as n increases. By contrast, **non-sampling errors** cannot be cured merely by enlarging n and must be controlled through rigorous survey design (as practised by the Australian Bureau of Statistics). Common commercial sources of non-sampling bias include:

Bias Type	Mechanism	Australian Business Example
Coverage Error	Sampling frame excludes or under-represents target sub-populations.	Surveying only urban postcodes to estimate national SME broadband reliability.
Non-Response Bias	Non-respondents differ systematically from survey completers.	Only dissatisfied customers completing post-call telecommunications feedback surveys.
Survivorship Bias	Analysing only surviving entities while ignoring failed or delisted units.	Computing 10-year ASX managed-fund average returns excluding funds closed after losses.
Measurement Error	Leading wording or recording inaccuracies distort responses.	Asking ambiguous expenditure recall questions without receipt verification.

Sampling Distribution of the Sample Mean ($\bar{X}$) and the Central Limit Theorem

Consider drawing a simple random sample $X_1, X_2, \ldots, X_n$ (independent and identically distributed, i.i.d.) from a population with finite mean μ and finite standard deviation σ. The expected value and standard deviation—called the **standard error of the mean (**$\text{SE}(\bar{X})$**)**—are exact for any population shape:

$$E(\bar{X}) = \mu_{\bar{X}} = \mu \qquad \text{and} \qquad \text{SE}(\bar{X}) = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Where: $E(\bar{X})$ is the mean of the sampling distribution of $\bar{X}$, σ is the population standard deviation, and n is the sample size. (When sampling without replacement from a finite population of size N where the sampling fraction $n/N > 0.05$, the standard error is multiplied by the finite population correction factor $\sqrt{(N-n)/(N-1)}$.)

Notice the fundamental square-root law: reducing the standard error of $\bar{X}$ by half requires quadrupling the sample size n. Furthermore, what is the *shape* of the distribution of $\bar{X}$?

- If the underlying population is itself normal, $X \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N(\mu, \sigma^2/n)$ **exactly for every sample size** $n \geq 1$.
- If the underlying population is *non-normal* (such as right-skewed retail transaction values or insurance claim sizes), the **Central Limit Theorem (CLT)** guarantees that as n grows large, the sampling distribution of $\bar{X}$ converges to a normal distribution:

$$\bar{X} \xrightarrow{d} N\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{so that} \qquad Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1)$$

Where: $\bar{X}$ is the sample mean, μ is the population mean, $\sigma/\sqrt{n}$ is the standard error of the mean, and Z is the standardised sample mean. As a practical benchmark, $n \geq 30$ suffices for approximate normality across moderately skewed business datasets, while heavily skewed distributions may require $n \geq 50$ or 100.

DegreeLive

Sampling Distribution of the Sample Proportion ($\hat{p}$)

For binary attributes (e.g., whether a customer renews a subscription or a voter supports a policy), let $X \sim \text{Bin}(n, p)$ count successes and define the sample proportion $\hat{p} = X/n$. Its sampling mean and standard error are:

$$E(\hat{p}) = p \qquad \text{and} \qquad \text{SE}(\hat{p}) = \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Where: p is the true population proportion and n is the sample size. Whenever both expected counts of successes and failures are sufficiently large—conventionally $np \geq 10$ and $n(1-p) \geq 10$ (or at minimum ≥ 5)—the Central Limit Theorem implies that $\hat{p}$ is approximately normally distributed:

$$\hat{p} \approx N\left(p,\ \frac{p(1-p)}{n}\right) \qquad \text{so that} \qquad Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx N(0,1)$$

Where: $\hat{p}$ is the sample proportion, p is the population proportion, and $\sqrt{p(1-p)/n}$ is the standard error of $\hat{p}$.

Worked Numerical Example 6: Applying the Central Limit Theorem

Part A (Sample Mean of SME Invoices): Business-to-business invoice values issued by Australian wholesale suppliers are positively skewed with population mean $\mu = \text{AUD } 2,400$ and standard deviation $\sigma = \text{AUD } 350$. An auditor draws a random sample of $n = 49$ invoices. Find the probability that the sample mean invoice exceeds A2,500.

Because $n = 49 \geq 30$, the Central Limit Theorem applies despite the underlying skewness. Compute the standard error and z-statistic:

$$\text{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{350}{\sqrt{49}} = \frac{350}{7} = \text{AUD } 50.00$$

$$z = \frac{2,500 - 2,400}{50} = +2.00 \implies P(\bar{X} > 2,500) = P(Z > 2.00) = 1 - \Phi(2.00) = 1 - 0.9772 = 0.0228 \ (2.28\%)$$

Where: $\text{SE}(\bar{X}) = 50$ is the standard error of the mean across $n = 49$ invoices and $z = 2.00$ is the standardised distance above $\mu = 2,400$.

Part B (Sample Proportion of Consumer Preferences): Suppose 30% ($p = 0.30$) of Australian households subscribe to a bundled green-energy tariff. In a random survey of $n = 200$ households, what is the probability that at least 35% ($\hat{p} \geq 0.35$) subscribe? First verify normality conditions: $np = 200(0.30) = 60 \geq 10$ and $n(1-p) = 200(0.70) = 140 \geq 10$. Next compute the standard error and z-score:

$$\text{SE}(\hat{p}) = \sqrt{\frac{0.30 \times 0.70}{200}} = \sqrt{0.00105} = 0.03240$$

$$z = \frac{0.35 - 0.30}{0.03240} = +1.5430 \implies P(\hat{p} \geq 0.35) \approx 1 - \Phi(1.54) = 1 - 0.9382 = 0.0618 \ (6.18\%)$$

Where: $\text{SE}(\hat{p}) = 0.03240$ is the standard error of the sample proportion for $n = 200$ and $z = 1.5430$ measures how many standard errors 0.35 sits above the true proportion 0.30.

Part C (Finite Population Correction in Internal Auditing): Suppose a corporate auditor samples $n = 64$ contracts without replacement from a finite ledger of $N = 400$ annual supplier contracts ($\sigma = \text{AUD } 120$). Because the sampling fraction $n/N = 64/400 = 0.16 > 0.05$, applying the finite population correction shrinks the standard error:

$$\text{SE}_{\text{FPC}}(\bar{X}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{120}{8}\sqrt{\frac{400-64}{399}} = 15.00 \times 0.9177 = \text{AUD } 13.77$$

Where: $\sqrt{(N-n)/(N-1)} = 0.9177$ is the finite population correction factor, reducing the standard error from A$15.00 to A$13.77 because a substantial fraction (16%) of the finite ledger has been observed.

DegreeLive

Summary of Key Sampling Distributions

Sample Statistic	Target Parameter	Sampling Mean	Standard Error (Infinite Pop.)	Normality Condition
Sample Mean ($\bar{X}$)	Population Mean (μ)	μ	$\sigma/\sqrt{n}$	Population normal, or $n \geq 30$ (CLT)
Sample Sum ($T = \sum X_i$)	Aggregate Total ($n\mu$)	$n\mu$	$\sqrt{n}\,\sigma$	Population normal, or $n \geq 30$ (CLT)
Sample Proportion ($\hat{p}$)	Population Proportion (p)	p	$\sqrt{p(1-p)/n}$	$np \geq 10$ and $n(1-p) \geq 10$

CLT Extension: Sampling Distribution of Aggregate Portfolio Sums ($T = \sum X_i$)

Insurance actuaries, superannuation treasurers, and retail CFOs frequently evaluate the total sum $T = X_1 + X_2 + \cdots + X_n = n\bar{X}$ rather than the average $\bar{X}$. Because multiplying $\bar{X}$ by n scales both mean and standard deviation by n, the CLT implies that for large n, $T \approx N(n\mu,\ n\sigma^2)$:

$$E(T) = n\mu \qquad \text{and} \qquad \text{SD}(T) = \sqrt{n}\,\sigma$$

Where: T is the aggregate sum across n independent units, μ is the per-unit mean, and σ is the per-unit standard deviation. Notice the crucial risk-pooling insight: while aggregate standard deviation grows as $\sqrt{n}$, relative risk ($\text{CV}(T) = \frac{n\sigma}{n\mu} = \frac{\sigma}{\mu\sqrt{n}}$) shrinks at rate $1/\sqrt{n}$. For example, if $n = 100$ independent motor insurance claims have mean $\mu = \text{AUD } 3,200$ and $\sigma = \text{AUD } 1,500$, then total claims have expected value $E(T) = 100(3,200) = \text{AUD } 320,000$ and standard deviation $\text{SD}(T) = \sqrt{100}(1,500) = \text{AUD } 15,000$. Consequently, the probability that total payout exceeds A350,000 is $P(T > 350,000) = P\left(Z > \frac{350,000-320,000}{15,000}\right) = P(Z > 2.00) = 0.0228$ (2.28%), providing the quantitative basis for actuarial capital reserves.

DegreeLive