IAE ACADEMY INSTITUTE FOR ENGINEERS

GATE DA

IAE Academy free study material · Updated 27 October 2026

Probability and Statistics Formula Sheet

INSIDE

- 156 formulas in 7 topics
- When to use each one
- Taken from the free topic notes on iaeacademy.in

Online, one page per topic: **iaeacademy.in/study/probability-statistics/formula-sheet**

FREE COUNSELLING

Preparing for GATE DA? Talk to IAE Academy.

Or call **+91 84204 36524**

Every day, 10:30 am to 7:30 pm

Chat on WhatsApp

Free. Not for sale. Pass it on.

Counting, sample spaces and the axioms of probability

PERMUTATIONS OF R FROM N DISTINCT OBJECTS (ORDER MATTERS, NO REPETITION)

$$P(n,r) = \frac{n!}{(n-r)!}, \qquad 0 \le r \le n$$

Selecting and arranging r of n distinct objects. P(n,n) = n! is the number of orderings of all n, and P(n,0) = 1 because 0! = 1.

COMBINATIONS OF R FROM N DISTINCT OBJECTS (ORDER DOES NOT MATTER)

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!}, \qquad \binom{n}{r} = \binom{n}{n-r}, \qquad 0 \le r \le n$$

Selecting r of n distinct objects with no arrangement. Related to permutations by P(n,r) = r! times the binomial coefficient. The binomial coefficient is taken as 0 when r > n.

PASCAL'S IDENTITY AND THE BINOMIAL SUM

$$\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}, \qquad \sum_{r=0}^{n} \binom{n}{r} = 2^n$$

The sum counts all subsets of an n-element set, so there are 2^n - 1 non-empty subsets. Useful for 'at least one' style selections.

BINOMIAL THEOREM

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}, \qquad n = 0, 1, 2, \dots$$

Putting x = y = 1 gives the sum 2^n. Putting x = -1, y = 1 shows that the alternating sum of the binomial coefficients is 0 for n >= 1.

INCLUSION-EXCLUSION FOR COUNTING

$$|A \cup B| = |A| + |B| - |A \cap B|, \qquad |A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

For finite sets. The count of objects with NONE of the properties is the total minus the union. The same pattern for probabilities is in section 4.

MULTINOMIAL COEFFICIENT

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1!\, n_2! \cdots n_k!}, \qquad n_1 + n_2 + \cdots + n_k = n$$

The number of ways to split n distinct objects into LABELLED groups of sizes n_1, ..., n_k, and equally the number of arrangements of n objects with repeated types. If the groups are unlabelled, divide by m! for each set of m groups of the same size. For example, 6 distinct people split into three unlabelled pairs: 6!/(2! 2! 2! 3!) = 15.

STARS AND BARS

$$\#\{(x_1, \dots, x_k) \in \mathbb{Z}_{\ge 0}^k : \; x_1 + \cdots + x_k = n\} = \binom{n+k-1}{k-1}$$

Distributing n identical objects into k distinct boxes with no limit per box. For a lower bound x_i >= a, substitute y_i = x_i - a and reduce n by a. For an upper bound x_i <= b, count all solutions and subtract those with x_i >= b + 1, using inclusion-exclusion when several variables are bounded.

ONTO FUNCTIONS (NO BOX LEFT EMPTY)

$$\#\{\text{onto } f : \{1, \dots, m\} \to \{1, \dots, n\}\} = \sum_{j=0}^{n} (-1)^j \binom{n}{j} (n-j)^m$$

Also the number of ways to place m distinct objects into n distinct boxes with every box non-empty. It comes from inclusion-exclusion over the set of boxes left empty, and is 0 when m < n. Check: m = 4, n = 2 gives 16 - 2 = 14; m = 4, n = 3 gives 81 - 48 + 3 = 36.

EQUALLY LIKELY OUTCOMES (CLASSICAL PROBABILITY)

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{number of outcomes favourable to } A}{\text{total number of outcomes}}$$

Valid only when the sample space is finite and every outcome is equally likely. State this assumption before using it; the wording of the question (a fair die, chosen at random, uniformly) is what justifies it.

AXIOMS OF PROBABILITY

$$(1)\ P(A) \geq 0; \qquad (2)\ P(\Omega) = 1; \qquad (3)\ P\left(\bigcup_i A_i\right) = \sum_i P(A_i) \ \text{ for pairwise mutually exclusive } A_1, A_2, \ldots$$

Every other rule is derived from these three. Axiom 3 applies to any finite or countably infinite sequence of events, but only if they are pairwise mutually exclusive.

COMPLEMENT RULE

$$P(A^c) = 1 - P(A), \qquad P(\varnothing) = 0$$

Use it for 'at least one' events: compute the probability that none occurs and subtract from 1.

INCLUSION-EXCLUSION FOR TWO EVENTS

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

Holds for any two events. The subtraction removes the overlap, which was counted twice.

INCLUSION-EXCLUSION FOR THREE EVENTS

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

Holds for any three events.

INCLUSION-EXCLUSION FOR N EVENTS

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n} (-1)^{k+1} S_k, \qquad S_k = \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k})$$

Holds for any events. The signs alternate: singles plus, pairs minus, triples plus, and so on. S_k has C(n,k) terms.

DIFFERENCE, MONOTONICITY AND THE UNION BOUND

$$P(B \setminus A) = P(B) - P(A \cap B); \qquad A \subseteq B \Rightarrow P(B \setminus A) = P(B) - P(A) \text{ and } P(A) \leq P(B);$$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

The first formula holds for any A and B; the simpler form P(B) - P(A) needs A to be a subset of B. The union bound holds with no condition.

FEASIBLE RANGE OF AN INTERSECTION

$$\max\{0,\ P(A) + P(B) - 1\} \leq P(A \cap B) \leq \min\{P(A), P(B)\}$$

The lower bound follows from P(A union B) <= 1; the upper bound from monotonicity. It gives the possible values of the intersection when only the two marginal probabilities are known.

INDEPENDENCE OF TWO EVENTS

$$A \text{ and } B \text{ are independent} \iff P(A \cap B) = P(A)\,P(B)$$

With the conditional probability P(A | B) = P(A and B)/P(B), defined for P(B) > 0, independence is equivalent to P(A | B) = P(A). If A and B are independent, so are A and B^c, A^c and B, and A^c and B^c.

MUTUAL INDEPENDENCE OF THREE EVENTS

$$P(A \cap B) = P(A)P(B), \quad P(A \cap C) = P(A)P(C), \quad P(B \cap C) = P(B)P(C),$$
$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

All four equations are required; pairwise independence is only the first three. For n events, the product rule must hold for every sub-collection of two or more events, which is 2^n - n - 1 equations.

AT LEAST ONE OF N INDEPENDENT EVENTS

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n} \left(1 - P(A_i)\right)$$

Valid when A_1, ..., A_n are mutually independent: their complements are then mutually independent too, and the probability that none occurs is the product.

DRAWING WITHOUT REPLACEMENT (HYPERGEOMETRIC COUNTING)

$$P(\text{exactly } k \text{ red}) = \frac{\binom{R}{k}\binom{B}{n-k}}{\binom{R+B}{n}}, \qquad 0 \le k \le R, \;\; 0 \le n-k \le B$$

An urn has R red and B blue balls, and n <= R + B balls are drawn at random without replacement. Order is ignored in both the numerator and the denominator.

BIRTHDAY PROBLEM

$$P(\text{all } k \text{ birthdays distinct}) = \prod_{j=0}^{k-1} \left(1 - \frac{j}{N}\right) = \frac{N!}{(N-k)!\, N^k}, \qquad k \le N$$

Assumes N equally likely days and independent birthdays (N = 365, ignoring 29 February). For k > N the probability is 0. The probability that at least two people share a birthday is 1 minus this.

DERANGEMENTS AND THE MATCHING PROBLEM

$$D_n = n! \sum_{k=0}^{n} \frac{(-1)^k}{k!}, \qquad P(\text{no match}) = \frac{D_n}{n!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!}$$

n letters are placed at random in n addressed envelopes, one per envelope. D_n counts the permutations with no fixed point: D_1 = 0, D_2 = 1, D_3 = 2, D_4 = 9, D_5 = 44. Recurrence: D_n = (n-1)(D_{n-1} + D_{n-2}) for n >= 3. As n grows, P(no match) tends to 1/e, about 0.3679.

Conditional, marginal and joint probability, and Bayes' theorem

CONDITIONAL PROBABILITY

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

Defined when $P(B) > 0$. For fixed B, $P(\cdot \mid B)$ is itself a probability measure, so $P(A^c \mid B) = 1 - P(A \mid B)$.

MULTIPLICATION RULE (TWO EVENTS)

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

The first form needs $P(B) > 0$, the second needs $P(A) > 0$. No independence is needed.

CHAIN RULE (MULTIPLICATION RULE FOR N EVENTS)

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap \cdots \cap A_{n-1})$$

Used for sequential experiments such as draws without replacement. Needs $P(A_1 \cap \cdots \cap A_{n-1}) > 0$, which makes every conditioning event positive.

COMPLEMENT AND SPLITTING

$$P(A^c) = 1 - P(A), \qquad P(A) = P(A \cap B) + P(A \cap B^c)$$

Always valid. The second identity splits A by whether B occurs; it is the two-part case of the law of total probability below.

UNION OF TWO EVENTS

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

Always valid. Given any three of $P(A)$, $P(B)$, $P(A \cup B)$ and $P(A \cap B)$, it gives the fourth.

UNION OF THREE EVENTS (INCLUSION-EXCLUSION)

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

Always valid. Add the singles, subtract the pairs, add back the triple.

MARGINAL FROM THE JOINT

$$P(X = x) = \sum_{y} P(X = x, Y = y), \qquad P(Y = y) = \sum_{x} P(X = x, Y = y)$$

Sum over all values of the other variable. All cells of the table sum to 1.

CONDITIONAL FROM THE JOINT

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

Defined when $P(Y = y) > 0$. Fix the column of the conditioning value, then divide each cell in it by that column's total. For each fixed y these conditionals sum to 1 over x.

INDEPENDENCE OF TWO DISCRETE RANDOM VARIABLES

$$P(X = x, Y = y) = P(X = x)\,P(Y = y) \quad \text{for every pair } (x, y)$$

Each cell must equal the product of its row total and column total. One cell that fails is enough to show X and Y are dependent.

MARGINAL AND CONDITIONAL DENSITIES (CONTINUOUS CASE)

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy, \qquad f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

For jointly continuous X, Y with joint density $f_{X,Y}$. Sums become integrals. The conditional density is defined where $f_Y(y) > 0$. Densities are not probabilities: probabilities come from integrating them, and $P(X = x) = 0$ for a continuous X.

LAW OF TOTAL PROBABILITY

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

The B_i must be mutually exclusive and exhaustive with $P(B_i) > 0$. The simplest partition is $\{B, B^c\}$: $P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$.

BAYES' THEOREM (TWO EVENTS)

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

Follows from the multiplication rule written both ways. Needs $P(A) > 0$ and $P(B) > 0$.

BAYES' THEOREM OVER A PARTITION

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

$B_1, \dots, B_n$ is a partition with positive probabilities, and $P(A) > 0$. The denominator is the law of total probability, so the posteriors sum to 1 over j.

BAYES' THEOREM IN ODDS FORM

$$\frac{P(B \mid A)}{P(B^c \mid A)} = \frac{P(A \mid B)}{P(A \mid B^c)} \cdot \frac{P(B)}{P(B^c)}$$

Posterior odds = likelihood ratio × prior odds. It needs $P(A \mid B^c) > 0$, $0 < P(B) < 1$ and $P(A) > 0$. Convert back with probability = odds / (1 + odds).

BAYES WITH TWO CONDITIONALLY INDEPENDENT PIECES OF EVIDENCE

$$P(D \mid E_1 \cap E_2) = \frac{P(E_1 \mid D)\,P(E_2 \mid D)\,P(D)}{P(E_1 \mid D)\,P(E_2 \mid D)\,P(D) + P(E_1 \mid D^c)\,P(E_2 \mid D^c)\,P(D^c)}$$

Valid when E_1 and E_2 are conditionally independent given D and also given D^c. In the example above, two positive tests give $\dfrac{0.81 \times 0.02}{0.81 \times 0.02 + 0.0025 \times 0.98} = \dfrac{0.0162}{0.01865} = \dfrac{324}{373} \approx 0.869$. Equivalently, apply Bayes once and use the first posterior as the prior for the second test.

INDEPENDENCE OF TWO EVENTS

$$P(A \cap B) = P(A)\,P(B) \iff P(A \mid B) = P(A) \quad (P(B) > 0)$$

The product form is the definition and needs no positivity condition. If A and B are independent, so are A and B^c, A^c and B, and A^c and B^c.

MUTUAL INDEPENDENCE OF THREE EVENTS

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

A, B, C are mutually independent when this holds AND every pair is independent: $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, $P(B \cap C) = P(B)P(C)$. For n events, the product rule must hold for every subcollection of two or more.

CONDITIONAL INDEPENDENCE

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

Events A and B are conditionally independent given C when this holds, with $P(C) > 0$. In general it neither implies nor is implied by ordinary independence.

Mean, median, mode, standard deviation, covariance and correlation

ARITHMETIC MEAN

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Raw data. For grouped data with frequencies f_i and class midpoints x_i, use $\bar{x} = \sum f_i x_i \big/ \sum f_i$. This treats every value as sitting at its class midpoint, so it is an approximation unless the values really are the midpoints.

MEDIAN OF RAW DATA

$$\text{median} = \begin{cases} x_{((n+1)/2)} & n \text{ odd} \\ \dfrac{x_{(n/2)} + x_{(n/2+1)}}{2} & n \text{ even} \end{cases}$$

Here $x_{(k)}$ is the k-th smallest value; sort first.

MEDIAN OF GROUPED DATA

$$\text{median} = L + \frac{N/2 - F}{f}\,h$$

The median class is the first class whose cumulative frequency is at least $N/2$. L = its lower class boundary, $N = \sum f_i$, F = cumulative frequency of the classes before it, f = its frequency, h = its width. Classes must be continuous: if the class limits leave gaps (10-19, 20-29), use the boundaries (9.5-19.5, 19.5-29.5).

MODE OF GROUPED DATA

$$\text{mode} = L + \frac{f_1 - f_0}{2f_1 - f_0 - f_2}\,h$$

L = lower boundary of the modal class (highest frequency f_1), f_0 and f_2 = frequencies of the classes just before and after it, h = class width. The modal class and its two neighbours must have the same width. If the modal class is the first or last class, take the missing neighbour's frequency as 0.

MEAN, MEDIAN AND MODE OF A RANDOM VARIABLE

$$E[X] = \sum_x x\,p(x) \;\text{ or }\; \int_{-\infty}^{\infty} x\,f(x)\,dx, \qquad F(m) = \tfrac{1}{2}, \qquad \text{mode} = \arg\max_x\, p(x) \;\text{ or }\; \arg\max_x\, f(x)$$

p is the pmf, f the pdf and F the CDF. The median condition $F(m) = \frac{1}{2}$ is for a continuous X (in general a median satisfies $P(X \le m) \ge \frac{1}{2}$ and $P(X \ge m) \ge \frac{1}{2}$). The mean exists only when the sum or integral of $|x|$ converges. Example: for the exponential distribution with rate λ, $1 - e^{-\lambda m} = \frac{1}{2}$ gives median $= \ln 2/\lambda$, while the mean is $1/\lambda$ and the mode is 0.

DEVIATIONS FROM THE MEAN

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0, \qquad \sum_{i=1}^{n}(x_i - c)^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - c)^2$$

For any constant c, the sum of squared deviations is therefore smallest when $c = \bar{x}$. The sum of absolute deviations $\sum |x_i - c|$ is smallest when c is the median (for even n, any c between the two middle values).

POPULATION VARIANCE AND SD (DIVIDE BY N)

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \mu^2, \qquad \sigma = \sqrt{\sigma^2}$$

μ is the population mean. The second form is the shortcut $E[X^2] - (E[X])^2$.

SAMPLE VARIANCE AND SD (DIVIDE BY N - 1)

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2 - n\bar{x}^2\right), \qquad s = \sqrt{s^2}$$

s^2 is an unbiased estimator of the population variance. For the same data it relates to the divide-by-n value by $s^2 = \frac{n}{n-1}\sigma^2$.

VARIANCE OF A RANDOM VARIABLE

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

$\text{SD} = \sqrt{\text{Var}(X)}$. If X picks one of the n data values uniformly at random, this equals the population (divide-by-n) variance of the data.

COEFFICIENT OF VARIATION

$$\text{CV} = \frac{\sigma}{\bar{x}} \quad \left(\text{often quoted as } \tfrac{\sigma}{\bar{x}} \times 100\%\right)$$

Defined for data with a positive mean. It has no units, so it compares the spread of data sets measured on different scales: the one with the larger CV is relatively more variable.

EFFECT OF $Y=AX+B$ ON CENTRE AND SPREAD

$$\bar{y} = a\bar{x} + b, \quad \operatorname{Var}(y) = a^2 \operatorname{Var}(x), \quad \operatorname{SD}(y) = |a| \operatorname{SD}(x)$$

This holds for the population and the sample versions alike. The median and the mode also transform as $a \cdot (\text{value}) + b$.

COMBINED MEAN

$$m = \frac{n_1 m_1 + n_2 m_2}{n_1 + n_2}$$

Weighted by group size. It equals the plain average $(m_1 + m_2)/2$ only when $n_1 = n_2$.

COMBINED VARIANCE (POPULATION FORM, DIVIDE BY N)

$$\sigma^2 = \frac{n_1\left(\sigma_1^2 + d_1^2\right) + n_2\left(\sigma_2^2 + d_2^2\right)}{n_1 + n_2}, \quad d_i = m_i - m$$

Both group variances must be the divide-by-n kind. If sample variances are given, first convert them with $\sigma_i^2 = \frac{n_i - 1}{n_i} s_i^2$.

COVARIANCE (RANDOM VARIABLES)

$$\operatorname{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]\,E[Y]$$

$\operatorname{Cov}(X, X) = \operatorname{Var}(X)$, and $\operatorname{Cov}(X, c) = 0$ for a constant c.

SAMPLE COVARIANCE OF PAIRED DATA

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

For the population version, divide by n instead. State the convention you use.

PEARSON CORRELATION COEFFICIENT

$$\rho_{XY} = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}, \qquad r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Both variances must be positive and finite. The n or $n - 1$ factors cancel in r, so the sample value does not depend on the convention.

LINEAR TRANSFORMATION RULES

$$\operatorname{Cov}(aX + b,\, cY + d) = ac \operatorname{Cov}(X, Y), \qquad \rho_{aX+b,\, cY+d} = \operatorname{sign}(ac)\, \rho_{XY}$$

For $a \neq 0$ and $c \neq 0$. A positive-slope linear change of either variable leaves the correlation unchanged; a negative multiplier on exactly one variable flips its sign.

COVARIANCE IS BILINEAR

$$\operatorname{Cov}(X + Y,\, Z) = \operatorname{Cov}(X, Z) + \operatorname{Cov}(Y, Z), \qquad \operatorname{Var}(X - Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) - 2\operatorname{Cov}(X, Y)$$

Covariance is linear in each argument separately. Note the $+\operatorname{Var}(Y)$ in $\operatorname{Var}(X - Y)$: variances never subtract.

VARIANCE OF A SUM

$$\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) + 2ab \operatorname{Cov}(X, Y)$$

The cross term vanishes when the covariance is zero (for example, when X and Y are independent).

VARIANCE OF A SUM OF N VARIABLES

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i) + 2 \sum_{i<j} \operatorname{Cov}(X_i, X_j)$$

If the X_i are pairwise uncorrelated, the variance of the sum is the sum of the variances.

VARIANCE OF THE MEAN OF N INDEPENDENT VALUES

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}, \qquad \mathrm{SD}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

For $X_1, \ldots, X_n$ independent (pairwise uncorrelated is enough) with common variance σ^2; here $\bar{X} = \frac{1}{n}\sum X_i$.

CORRELATION AND THE REGRESSION SLOPES

$$b_{yx} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} = r\,\frac{s_y}{s_x}, \qquad b_{xy} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)} = r\,\frac{s_x}{s_y}, \qquad r^2 = b_{yx}\,b_{xy}$$

b_{yx} is the least-squares slope of the line of y on x, and b_{xy} that of x on y. Both lines pass through $(\bar{x}, \bar{y})$. The two slopes and r share one sign, so r is the square root of $b_{yx}b_{xy}$ taken with that sign. Use the same divisor in the covariance and the variances.

SPEARMAN RANK CORRELATION

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

d_i is the difference between the ranks of x_i and y_i. This shortcut is exact only when there are no tied ranks. With ties, compute Pearson's r on the ranks instead.

Random variables, expectation and variance, including conditional expectation

PMF CONDITIONS (DISCRETE)

$$p_X(x) = P(X = x) \geq 0, \qquad \sum_x p_X(x) = 1$$

Use the second condition to find an unknown constant in a given PMF.

PDF CONDITIONS (CONTINUOUS)

$$f_X(x) \geq 0, \qquad \int_{-\infty}^{\infty} f_X(x)\,dx = 1, \qquad P(a \leq X \leq b) = \int_a^b f_X(x)\,dx$$

A density can exceed 1 at a point; only its integral is a probability.

CDF

$$F_X(x) = P(X \leq x) = \begin{cases} \sum_{t \leq x} p_X(t) & \text{discrete} \\ \int_{-\infty}^{x} f_X(t)\,dt & \text{continuous} \end{cases}$$

The CDF is non-decreasing, right-continuous, with limits 0 at $-\infty$ and 1 at $+\infty$. For any random variable, $P(a < X \leq b) = F_X(b) - F_X(a)$.

RECOVERING PMF AND PDF FROM THE CDF

$$f_X(x) = \frac{dF_X(x)}{dx}, \qquad p_X(x) = F_X(x) - F_X(x^-)$$

For a continuous variable differentiate the CDF where it is differentiable; for a discrete variable the PMF is the size of the jump at x.

EXPECTATION

$$E[X] = \sum_x x\,p_X(x) \quad \text{or} \quad E[X] = \int_{-\infty}^{\infty} x\,f_X(x)\,dx$$

Defined when the sum or integral converges absolutely.

EXPECTATION OF A FUNCTION (LOTUS)

$$E[g(X)] = \sum_x g(x)\,p_X(x) \quad \text{or} \quad \int_{-\infty}^{\infty} g(x)\,f_X(x)\,dx$$

No need to find the distribution of $g(X)$. In general $E[g(X)] \neq g(E[X])$; they are equal when g is linear, $g(x) = ax + b$.

LINEARITY OF EXPECTATION

$$E[aX + bY + c] = aE[X] + bE[Y] + c, \qquad E\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n} E[X_i]$$

True whenever the expectations exist, with or without independence.

VARIANCE

$$\operatorname{Var}(X) = E\big[(X - E[X])^2\big] = E[X^2] - (E[X])^2$$

Standard deviation $\mathrm{SD} = \sqrt{\operatorname{Var}(X)}$. Since $\operatorname{Var}(X) \geq 0$, $E[X^2] \geq (E[X])^2$, and $\operatorname{Var}(X) = 0$ only when X is constant with probability 1. Convention: $\operatorname{Var}(X)$ here is the population (distribution) variance. For a list of n data values, the population variance divides by n and the sample variance divides by $n - 1$.

VARIANCE UNDER A LINEAR CHANGE

$$\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X)$$

A shift b does not change the variance; a scale a multiplies it by a^2 (the sign of a is lost).

INDICATOR VARIABLES

$$I_A = \begin{cases} 1 & A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}, \qquad E[I_A] = P(A), \qquad \operatorname{Var}(I_A) = P(A)\big(1 - P(A)\big)$$

To find the expected number of events $A_1, \ldots, A_n$ that occur, write the count as $\sum I_{A_i}$; by linearity its mean is $\sum P(A_i)$, with no independence needed.

MARGINALS

$$p_X(x) = \sum_y p_{X,Y}(x, y), \qquad f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy$$

Sum (or integrate) out the other variable over its whole range. Mind the limits when the support is not a rectangle.

INDEPENDENCE

$$p_{X,Y}(x, y) = p_X(x)\,p_Y(y) \ \text{ for all } x, y \qquad \Big(f_{X,Y}(x, y) = f_X(x)\,f_Y(y)\Big)$$

For a PMF the product rule must hold at every pair (x, y), so one failing pair shows dependence. For a PDF it must hold everywhere apart from sets of zero probability. If the set where the joint PMF or PDF is positive is not a product set $A \times B$ (for example the triangle $0 < x < y < 1$), then X and Y are dependent.

COVARIANCE

$$\operatorname{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

$\operatorname{Cov}(X, X) = \operatorname{Var}(X)$; $\operatorname{Cov}(aX + b, cY + d) = ac \operatorname{Cov}(X, Y)$.

CORRELATION COEFFICIENT

$$\rho_{XY} = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}, \qquad -1 \leq \rho_{XY} \leq 1$$

Defined when both variances are positive. $|\rho| = 1$ exactly when $Y = aX + b$ with probability 1 for some constants $a \neq 0$ and b; then $\rho = 1$ if $a > 0$ and $\rho = -1$ if $a < 0$.

VARIANCE OF A SUM

$$\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) + 2ab \operatorname{Cov}(X, Y)$$

For uncorrelated (in particular independent) variables the covariance term vanishes, so $\operatorname{Var}(X \pm Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$. Note the plus sign, even for the difference.

VARIANCE OF A SUM OF N VARIABLES

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i) + 2 \sum_{i<j} \operatorname{Cov}(X_i, X_j)$$

For independent identically distributed X_i with variance σ^2, the sample mean has $\operatorname{Var}(\bar{X}) = \sigma^2 / n$.

PRODUCT OF INDEPENDENT VARIABLES

$$X, Y \text{ independent } \Rightarrow\ E[XY] = E[X]E[Y],\ \operatorname{Cov}(X, Y) = 0$$

The converse is false: zero covariance (uncorrelated) does not imply independence.

CONDITIONAL PMF AND PDF

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \qquad f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Requires $p_Y(y) > 0$ (or $f_Y(y) > 0$). If X, Y are independent the conditional equals the marginal.

CONDITIONAL EXPECTATION

$$E[X \mid Y = y] = \sum_x x\, p_{X|Y}(x \mid y) \quad \text{or} \quad \int x\, f_{X|Y}(x \mid y)\, dx$$

For each fixed y this is a number; $E[X \mid Y]$ is the random variable obtained by replacing y with Y.

CONDITIONAL VARIANCE

$$\operatorname{Var}(X \mid Y) = E\big[(X - E[X \mid Y])^2 \mid Y\big] = E[X^2 \mid Y] - (E[X \mid Y])^2$$

Also a random variable, a function of Y.

LAW OF TOTAL EXPECTATION

$$E[X] = E\big[E[X \mid Y]\big] = \sum_y E[X \mid Y = y]\, P(Y = y)$$

Replace the sum by an integral against f_Y when Y is continuous. For events $A_1, A_2, \ldots$ that partition the sample space, $E[X] = \sum_i E[X \mid A_i]\, P(A_i)$.

LAW OF TOTAL VARIANCE

$$\operatorname{Var}(X) = E\big[\operatorname{Var}(X \mid Y)\big] + \operatorname{Var}\big(E[X \mid Y]\big)$$

Average within-group variance plus variance of the group means.

USEFUL PROPERTIES

$E[aX + bZ \mid Y] = aE[X \mid Y] + bE[Z \mid Y], \qquad E[g(Y)X \mid Y] = g(Y)\, E[X \mid Y], \qquad E[X \mid Y] = E[X]$ if X, Y independent

Once Y is given, any function of Y behaves like a constant. The last property does not work in reverse: $E[X \mid Y] = E[X]$ does not imply independence.

RANDOM SUM

$$S = \sum_{i=1}^{N} X_i: \qquad E[S] = E[N]\, \mu, \qquad \operatorname{Var}(S) = E[N]\, \sigma^2 + \operatorname{Var}(N)\, \mu^2$$

Here $X_1, X_2, \ldots$ are independent and identically distributed with mean μ and variance σ^2, and N is a non-negative integer random variable independent of the X_i. It follows from the two total laws with $E[S \mid N] = N\mu$ and $\operatorname{Var}(S \mid N) = N\sigma^2$.

MARKOV'S INEQUALITY

$$P(X \geq a) \leq \frac{E[X]}{a}$$

For $X \geq 0$ and $a > 0$. Needs only the mean.

CHEBYSHEV'S INEQUALITY

$$P\big(|X - \mu| \geq k\sigma\big) \leq \frac{1}{k^2}, \qquad \text{equivalently } P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

Holds for any distribution with mean μ and finite variance σ^2; $k, t > 0$. Complementary form: $P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$.

Discrete distributions: uniform, Bernoulli and binomial

DISCRETE UNIFORM PMF

$$P(X = x) = \frac{1}{n}, \quad x = a, a+1, \ldots, b, \quad n = b - a + 1$$

Applies when each of the n consecutive integers is equally likely; zero elsewhere.

MEAN AND VARIANCE (UNIFORM ON A TO B)

$$E[X] = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{n^2 - 1}{12}$$

Here n = b - a + 1 and the values are consecutive integers. The variance depends only on n, not on where the block starts.

UNIFORM ON 1 TO N

$$E[X] = \frac{n+1}{2}, \qquad \mathrm{Var}(X) = \frac{n^2 - 1}{12}$$

Special case a = 1, b = n. For a fair die (n = 6): mean 3.5, variance 35/12.

BERNOULLI PMF

$$P(X = x) = p^x (1-p)^{1-x}, \quad x \in \{0, 1\}$$

One trial, success probability p, 0 <= p <= 1.

BERNOULLI MEAN AND VARIANCE

$$E[X] = p, \qquad \mathrm{Var}(X) = p(1-p)$$

Because X^2 = X for a 0/1 variable, E[X^2] = p, so Var(X) = p - p^2. The variance is largest at p = 1/2, where it equals 1/4.

BINOMIAL PMF

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

Requires fixed n, independent trials and the same p on every trial. The probabilities sum to 1 by the binomial theorem.

BINOMIAL MEAN AND VARIANCE

$$E[X] = np, \qquad \mathrm{Var}(X) = np(1-p)$$

Follows by adding n independent Bernoulli(p) variables: means add, and variances add because of independence. When 0 < p < 1 the variance is always less than the mean.

SECOND MOMENTS OF THE BINOMIAL

$$E[X(X-1)] = n(n-1)p^2, \qquad E[X^2] = np(1-p) + (np)^2$$

E[X^2] = Var(X) + (E[X])^2. Useful when a question asks for E[X^2] or the expectation of a quadratic in X.

RATIO OF CONSECUTIVE TERMS

$$\frac{P(X=k+1)}{P(X=k)} = \frac{(n-k)\,p}{(k+1)(1-p)}, \quad 0 < p < 1$$

Use it to compute successive probabilities quickly and to locate the mode.

MODE OF THE BINOMIAL

$$0 < p < 1: \ \text{mode} = \lfloor (n+1)p \rfloor; \ \text{if } (n+1)p \text{ is an integer, both } (n+1)p \text{ and } (n+1)p - 1 \text{ are modes}$$

Valid for 0 < p < 1. At the end points the distribution is degenerate: the mode is 0 when p = 0 and n when p = 1.

SUM OF INDEPENDENT BINOMIALS, SAME P

$$X \sim \text{Bin}(n_1, p),\ Y \sim \text{Bin}(n_2, p),\ X, Y \text{ independent } \Rightarrow\ X + Y \sim \text{Bin}(n_1 + n_2, p)$$

Both independence and the common p are essential. With different success probabilities the sum is in general not binomial, although means still add, and variances still add because of independence.

GEOMETRIC PMF (TRIALS TO FIRST SUCCESS)

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

Independent trials, constant p with 0 < p <= 1. X counts the trial on which the first success occurs, so X >= 1. Mean 1/p and variance (1-p)/p^2. If instead the count is the number of failures before the first success, the PMF is (1-p)^k p for k = 0, 1, 2, ..., the mean is (1-p)/p and the variance is still (1-p)/p^2.

Continuous random variables and the named distributions: exponential, Poisson, normal, t and chi-squared

CONDITIONS ON A PDF

$$f(x) \geq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f(x)\,dx = 1$$

Use the total area equal to 1 to find an unknown constant in a given density.

CUMULATIVE DISTRIBUTION FUNCTION (CDF)

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt, \qquad f(x) = \frac{dF(x)}{dx}$$

F is non-decreasing and, for a continuous X, continuous. Its limits are 0 as x goes to -infinity and 1 as x goes to +infinity. The PDF is the derivative of the CDF wherever the derivative exists.

INTERVAL PROBABILITY

$$P(a < X \leq b) = \int_{a}^{b} f(x)\,dx = F(b) - F(a)$$

Valid for any continuous X. Also P(X > a) = 1 - F(a).

PERCENTILE AND MEDIAN

$$F(x_p) = p, \qquad \text{median } m: \ F(m) = \tfrac{1}{2}$$

Solve F(x) = p for x. The p-th percentile (p as a fraction) is x_p; the median is the 50th percentile.

MEAN AND VARIANCE

$$E[X] = \int_{-\infty}^{\infty} x f(x)\,dx, \qquad \mathrm{Var}(X) = E[X^2] - (E[X])^2$$

E[X] exists only when the integral of |x| f(x) converges. Some distributions have no mean, for example t with 1 degree of freedom. Convention used throughout: E[X^2] = integral of x^2 f(x) dx, and SD = sqrt(Var(X)).

LINEAR TRANSFORMATION

$$E[aX + b] = aE[X] + b, \qquad \mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$$

Holds for any constants a and b. Adding a constant shifts the mean and leaves the variance unchanged. SD(aX + b) = |a| SD(X).

CONDITIONAL PDF GIVEN X > A

$$f(x \mid X > a) = \frac{f(x)}{P(X > a)} = \frac{f(x)}{1 - F(a)}, \quad x > a$$

Requires P(X > a) > 0, and the conditional PDF is zero for x <= a. In general, conditioning on an event A with P(X in A) > 0 gives f(x)/P(X in A) for x in A. The result integrates to 1 over the conditioning region.

CONDITIONAL PDF FROM A JOINT DENSITY

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

Defined only where f_Y(y) > 0. For a fixed y it is a density in x and integrates to 1. If X and Y are independent, f_{X|Y}(x | y) = f_X(x).

UNIFORM(A, B): PDF AND CDF

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

Zero density outside [a, b]; F is 0 below a and 1 above b. For a sub-interval inside [a, b], probability = length / (b - a).

UNIFORM(A, B): MEAN AND VARIANCE

$$E[X] = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

The median equals the mean for the uniform distribution.

EXPONENTIAL(RATE LAMBDA): PDF AND CDF

$$f(x) = \lambda e^{-\lambda x},\ F(x) = 1 - e^{-\lambda x},\ P(X > x) = e^{-\lambda x} \quad (x \ge 0)$$

Support is x >= 0; lambda is the rate. P(X > x) is the survival function.

EXPONENTIAL: MEAN, VARIANCE, MEDIAN, PERCENTILE

$$E[X] = \frac{1}{\lambda},\ \mathrm{Var}(X) = \frac{1}{\lambda^2},\ \mathrm{median} = \frac{\ln 2}{\lambda},\ x_p = \frac{-\ln(1-p)}{\lambda}$$

Mean and SD are equal. The median (about 0.693/lambda) is smaller than the mean 1/lambda, and the distribution is right-skewed. ln is the natural logarithm.

MEMORYLESS PROPERTY

$$P(X > s + t \mid X > s) = P(X > t) = e^{-\lambda t}$$

Holds for all s, t >= 0. The exponential is the only continuous distribution with this property. Time already waited does not change the distribution of the remaining wait.

MINIMUM OF TWO INDEPENDENT EXPONENTIALS (RATES LAMBDA1, LAMBDA2)

$$\min(X_1, X_2) \sim \text{Exponential}(\text{rate } \lambda_1 + \lambda_2)$$
$$P(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

Requires X1 and X2 independent. Reason: P(min > x) = P(X1 > x) P(X2 > x) = e^{-(lambda1 + lambda2) x}. The same holds for more than two: the rates add.

POISSON PMF

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

lambda > 0 is the average number of events in the interval considered. P(X >= 1) = 1 - e^{-lambda}.

POISSON MEAN AND VARIANCE

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

Mean equals variance. If X1 ~ Poisson(l1) and X2 ~ Poisson(l2) are independent, then X1 + X2 ~ Poisson(l1 + l2).

POISSON PROCESS WITH RATE LAMBDA PER UNIT TIME

$$N(t) \sim \text{Poisson}(\lambda t), \qquad T_{\text{inter-arrival}} \sim \text{Exponential}(\text{rate } \lambda)$$

Events occur independently at a constant average rate. The count in a window of length t has mean lambda*t, and the gaps between successive events are exponential with mean 1/lambda.

POISSON APPROXIMATION TO THE BINOMIAL

$$\text{Binomial}(n, p) \approx \text{Poisson}(\lambda = np)$$

Use when n is large and p is small, so that np is moderate. The binomial's variance np(1-p) is close to np only when p is small.

NORMAL PDF

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

Mean = median = mode = mu; variance sigma^2; symmetric about mu.

STANDARDISATION

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1), \qquad P(a < X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Divide by the standard deviation sigma, not the variance. P(X > a) = 1 - Phi((a - mu)/sigma).

SYMMETRY OF THE STANDARD NORMAL

$$\Phi(-z) = 1 - \Phi(z), \qquad P(|Z| \le z) = 2\Phi(z) - 1$$

Phi(0) = 0.5. Use this to turn negative z into positive z when reading a table.

THE 68-95-99.7 RULE

$$P(|X - \mu| \le \sigma) \approx 0.6827,\ P(|X - \mu| \le 2\sigma) \approx 0.9545,\ P(|X - \mu| \le 3\sigma) \approx 0.9973$$

Applies to every normal distribution, whatever mu and sigma are.

LINEAR COMBINATION OF INDEPENDENT NORMALS

$$aX + bY + c \sim N\left(a\mu_X + b\mu_Y + c,\ a^2\sigma_X^2 + b^2\sigma_Y^2\right)$$

Requires X and Y independent and each normal, with a and b not both zero. Variances add with squared coefficients, even for a difference: Var(X - Y) = sigma_X^2 + sigma_Y^2.

SAMPLE MEAN OF A NORMAL SAMPLE

$$X_1, \ldots, X_n \text{ i.i.d. } N(\mu, \sigma^2) \;\Rightarrow\; \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Exact for a normal population. The standard deviation of X-bar (its standard error) is sigma/sqrt(n).

CENTRAL LIMIT THEOREM

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{\text{approx.}}{\sim} N(0,1) \quad (\text{large } n)$$

For X_1, ..., X_n i.i.d. from any population with mean mu and finite variance sigma^2. One consequence: Binomial(n, p) is approximately N(np, np(1-p)) when np and n(1-p) are both large. For integer counts, a continuity correction of 0.5 improves the approximation.

CHI-SQUARED DISTRIBUTION: DEFINITION

$$Z_1, \ldots, Z_k \text{ i.i.d. } N(0,1) \;\Rightarrow\; \chi^2_k = Z_1^2 + \cdots + Z_k^2$$

Takes only values >= 0. It is right-skewed and becomes more symmetric as k grows. The sum of independent chi-squared variables with k1 and k2 df is chi-squared with k1 + k2 df.

CHI-SQUARED: MEAN AND VARIANCE

$$E[\chi^2_k] = k, \qquad \mathrm{Var}(\chi^2_k) = 2k$$

Follows from E[Z^2] = 1 and Var(Z^2) = 2 for one standard normal, together with the independence of the Z_i.

CHI-SQUARED FOR THE SAMPLE VARIANCE

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

For a random sample of size n from N(mu, sigma^2). Here S^2 is the sample variance, which divides by n - 1. One df is lost because the sample mean is estimated from the data.

T DISTRIBUTION: DEFINITION

$$T = \frac{Z}{\sqrt{V/k}} \sim t_k, \quad Z \sim N(0,1),\ V \sim \chi^2_k \text{ independent}$$

Symmetric about 0 and bell-shaped, with heavier tails than N(0,1). As k grows, t_k approaches N(0,1).

T: MEAN AND VARIANCE

$$E[T] = 0 \ (k > 1), \qquad \mathrm{Var}(T) = \frac{k}{k-2} \ (k > 2)$$

For k = 1 the mean does not exist; for k = 2 the variance is infinite. For k > 2 the variance exceeds 1, reflecting the heavier tails, and it tends to 1 as k grows.

T STATISTIC FOR THE SAMPLE MEAN

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

For a random sample of size n from a normal population with unknown sigma, where S is the sample standard deviation (divisor n - 1). If sigma is known, use (X-bar - mu)/(sigma/sqrt(n)) ~ N(0,1) instead.

Central limit theorem, confidence intervals and hypothesis tests

MEAN AND STANDARD ERROR OF THE SAMPLE MEAN

$$E[\bar{X}] = \mu, \qquad \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}, \qquad \mathrm{SE}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

Needs independent observations with a common mean μ and finite variance σ^2; no normality is needed. SE (standard error) is the standard deviation of $\bar{X}$. When σ is unknown it is estimated by $s/\sqrt{n}$, where s is the sample standard deviation (divisor $n-1$, below).

SAMPLE VARIANCE (DIVIDES BY N – 1)

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \qquad E[S^2] = \sigma^2$$

Dividing by $n-1$ makes S^2 an unbiased estimator of σ^2. Dividing by n instead gives an estimator whose expectation is $\frac{n-1}{n}\sigma^2$. Throughout these notes s^2 is the $n-1$ version.

CENTRAL LIMIT THEOREM

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{\;d\;} N(0,1) \quad (n \to \infty)$$

For i.i.d. observations with finite mean and variance, $\bar{X}$ is approximately $N(\mu, \sigma^2/n)$ for large n, whatever the shape of the population. Equivalently, the sum $S_n = X_1 + \cdots + X_n$ is approximately $N(n\mu,\, n\sigma^2)$. If the population is itself normal, $\bar{X}$ is exactly normal for every n.

NORMAL APPROXIMATION TO THE BINOMIAL WITH CONTINUITY CORRECTION

$$P(X \le k) \approx \Phi\left(\frac{k + 0.5 - np}{\sqrt{np(1-p)}}\right), \quad P(X \ge k) \approx 1 - \Phi\left(\frac{k - 0.5 - np}{\sqrt{np(1-p)}}\right), \quad P(X = k) \approx \Phi\left(\frac{k + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

For $X \sim \mathrm{Binomial}(n, p)$, which has mean np and variance $np(1-p)$. A common rule of thumb is $np \ge 5$ and $n(1-p) \ge 5$ (some use 10). Φ is the standard normal CDF. The ± 0.5 moves each integer boundary to the middle of the gap between integers.

CONFIDENCE INTERVAL FOR A MEAN, SIGMA KNOWN (Z)

$$\bar{x} \pm z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}$$

Independent observations; population normal, or n large (CLT). Commonly used values: $z_{0.05} = 1.645$ (90%), $z_{0.025} = 1.96$ (95%), $z_{0.005} = 2.576$ (99%).

CONFIDENCE INTERVAL FOR A MEAN, SIGMA UNKNOWN (T)

$$\bar{x} \pm t_{\alpha/2,\,n-1}\,\frac{s}{\sqrt{n}}, \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Population normal (for large n the interval is approximately valid without normality, and $t \approx z$). Here s^2 is the sample variance with divisor $n-1$. The t distribution has $n-1$ degrees of freedom and heavier tails than the normal, so $t_{\alpha/2,n-1} > z_{\alpha/2}$; it approaches $z_{\alpha/2}$ as n grows.

CONFIDENCE INTERVAL FOR A PROPORTION

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Large-sample (normal) interval, valid when $n\hat{p}$ and $n(1-\hat{p})$ are both large. $\hat{p} = x/n$ is the sample proportion.

WIDTH AND SAMPLE SIZE FOR A MEAN (SIGMA KNOWN)

$$\text{width} = 2\,z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \qquad n \ge \left(\frac{z_{\alpha/2}\,\sigma}{E}\right)^2$$

E is the required margin of error (half-width). Round n up to the next integer.

ERROR PROBABILITIES AND POWER

$$\alpha = P(\text{reject } H_0 \mid H_0 \text{ true}), \quad \beta = P(\text{fail to reject } H_0 \mid H_0 \text{ false}), \quad \text{power} = 1 - \beta$$

β depends on which alternative value is true. For a composite H_0 (such as $\mu \le \mu_0$), α is the largest Type I error probability over H_0. For a fixed sample size and a fixed alternative, reducing α increases β. Increasing n can reduce both.

P-VALUE FOR A TWO-SIDED Z-TEST

$$\text{two-sided } z\text{-test: } p = 2\left[1 - \Phi(|z|)\right]$$

For a one-sided test use $1 - \Phi(z)$ when $H_1 : \mu > \mu_0$ (upper tail) and $\Phi(z)$ when $H_1 : \mu < \mu_0$ (lower tail).

ONE-SAMPLE Z-TEST STATISTIC

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Use when σ is known and the population is normal or n is large. Under $H_0 : \mu = \mu_0$, $z \sim N(0,1)$ (exactly for normal data, approximately otherwise).

ONE-SAMPLE T-TEST STATISTIC

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \qquad \text{df} = n - 1$$

Use when σ is unknown and the population is normal; s uses divisor $n - 1$. Under H_0, t follows a t distribution with $n - 1$ degrees of freedom.

THE T DISTRIBUTION WITH NU DEGREES OF FREEDOM

$$T = \frac{Z}{\sqrt{V/\nu}}, \quad Z \sim N(0,1),\ V \sim \chi^2_\nu \text{ independent}; \qquad E[T] = 0\ (\nu > 1), \quad \text{Var}(T) = \frac{\nu}{\nu - 2}\ (\nu > 2)$$

Symmetric about 0 with heavier tails than $N(0,1)$; it tends to $N(0,1)$ as $\nu \to \infty$. For normal data, $\frac{\bar{X} - \mu}{S/\sqrt{n}}$ has this distribution with $\nu = n - 1$.

TWO-SAMPLE POOLED T-TEST (EQUAL VARIANCES)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}, \quad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \quad \text{df} = n_1 + n_2 - 2$$

Independent samples from normal populations assumed to share the same variance; tests $H_0 : \mu_1 = \mu_2$. s_p^2 is the pooled sample variance.

PAIRED T-TEST

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}, \qquad \text{df} = n - 1$$

For n matched pairs (for example before/after on the same subjects), take differences $d_i = x_i - y_i$. $\bar{d}$ and s_d are their mean and sample SD (divisor $n - 1$). The differences are assumed normal; tests $H_0 : \mu_d = 0$. Do not use the two-sample formula on paired data.

TWO-SAMPLE Z-TEST STATISTIC

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Independent samples, variances σ_1^2, σ_2^2 known, populations normal or both samples large; tests $H_0 : \mu_1 = \mu_2$.

ONE-SAMPLE TEST FOR A PROPORTION

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

Large n. The standard error uses p_0 (the value under H_0), whereas the confidence interval uses $\hat{p}$.

TWO-SAMPLE TEST FOR PROPORTIONS

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}, \qquad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Independent large samples; tests $H_0 : p_1 = p_2$. Under H_0 the common proportion is estimated by the pooled $\hat{p}$.

THE CHI-SQUARED DISTRIBUTION WITH K DEGREES OF FREEDOM

$$Z_1, \ldots, Z_k \text{ i.i.d. } N(0,1) \;\Rightarrow\; \sum_{i=1}^{k} Z_i^2 \sim \chi^2_k, \qquad E = k, \quad \mathrm{Var} = 2k$$

Takes only non-negative values and is right-skewed. The sum of independent chi-squared variables is chi-squared, with the degrees of freedom added.

CHI-SQUARED STATISTIC (GOODNESS OF FIT AND INDEPENDENCE)

$$\chi^2 = \sum_{\text{cells}} \frac{(O-E)^2}{E} = \sum_{\text{cells}} \frac{O^2}{E} - N$$

O = observed count, E = expected count under H_0, N = total count. Large values mean poor agreement, so these two tests are upper-tailed: reject when χ^2 exceeds $\chi^2_{\alpha,\mathrm{df}}$. The chi-squared approximation needs expected counts that are not tiny (a common rule is $E \geq 5$).

DEGREES OF FREEDOM, GOODNESS OF FIT

$$\mathrm{df} = k - 1 - m$$

k categories, m parameters estimated from the data. If H_0 fully specifies the probabilities, $m = 0$ and $E_i = n\,p_i$.

TEST OF INDEPENDENCE IN AN R BY C TABLE

$$E_{ij} = \frac{R_i\, C_j}{N}, \qquad \mathrm{df} = (r-1)(c-1)$$

R_i = row total, C_j = column total, N = grand total. H_0: the two classifications are independent.

CHI-SQUARED TEST FOR A SINGLE VARIANCE

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}, \qquad \mathrm{df} = n - 1$$

Tests $H_0 : \sigma^2 = \sigma_0^2$ for a **normal** population (s^2 with divisor $n-1$); the result is sensitive to non-normality. Unlike the count tests above, this test can be two-sided: for $H_1 : \sigma^2 \neq \sigma_0^2$ reject if $\chi^2 > \chi^2_{\alpha/2,n-1}$ or $\chi^2 < \chi^2_{1-\alpha/2,n-1}$. For $H_1 : \sigma^2 > \sigma_0^2$ reject if $\chi^2 > \chi^2_{\alpha,n-1}$.

CONFIDENCE INTERVAL FOR A VARIANCE (NORMAL POPULATION)

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,\,n-1}},\ \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,\,n-1}} \right)$$

With the upper-tail convention, $\chi^2_{\alpha/2,n-1}$ is the larger critical value, so it gives the lower limit. The interval is not symmetric about s^2.

FROM IAE ACADEMY

Preparing for GATE DA?

IAE Academy teaches GATE DA in Ballygunge, Kolkata, and online. Counselling is free: ask us about batches, fees or your preparation.

FREE COUNSELLING

Chat on WhatsApp

Call +91 84204 36524

Open every day, 10:30 am to 7:30 pm.

Scan to chat on WhatsApp

THE GATE DA COURSE

Classroom batches in Ballygunge and online batches. Fees and details are on the course page.

Scan: course page

See the course

FREE ON IAEACADEMY.IN

Notes and solved previous-year questions for Probability and Statistics, topic by topic.

Scan: the notes

Open the notes

Found this useful? Send it to a friend preparing for GATE.

Send to a friend on WhatsApp

IAE Academy free study material. The formulas come from the free Probability and Statistics notes on iaeacademy.in. Prepared and published by IAE Academy, Kolkata. Free to read and free to share, unaltered and with attribution.

IAE ACADEMY INSTITUTE FOR ENGINEERS

GATE coaching in Kolkata since 2012 · EC, EE, IN, CS and IT, DA, BT and CE · in the classroom and online

227A Rash Behari Ave, Ballygunge, Kolkata 700019

+91 84204 36524 · iaeacademy.in · Updated 27 October 2026

Get directions