Probability and Statistics

B.Tech all branches · GATE Engineering Mathematics

Uncodemy Handwritten Notes · free for students

1. Basic Statistics

- **Mean** (average) $\bar{x} = \Sigma x / n$. **Median** = middle value after sorting. **Mode** = most frequent value.
- Empirical relation (moderately skewed data): Mode = 3 Median - 2 Mean.
- **Variance** $\sigma^2 = \Sigma(x - \bar{x})^2 / n = (\Sigma x^2 / n) - \bar{x}^2$. **Standard deviation** σ = sqrt(variance).
- **Coefficient of variation** = $(\sigma / \bar{x}) \times 100$. Lower CV = more consistent data.

➤ Effect of change of origin and scale

Change	Mean	Variance
$y = x + c$	$\bar{x} + c$	same
$y = kx$	$k\,\bar{x}$	$k^2\,\sigma^2$
$y = kx + c$	$k\,\bar{x} + c$	$k^2\,\sigma^2$

EXAMPLE Data 2, 4, 4, 4, 5, 5, 7, 9. Mean = 40/8 = 5. $\Sigma(x - 5)^2$ = 9+1+1+1+0+0+4+16 = 32. Variance = 32/8 = 4, $\sigma = 2$.

2. Probability Basics

- **Sample space** S = all outcomes. **Event** = subset of S.
- Classical: $P(A)$ = favourable outcomes / total outcomes. $0 \le P(A) \le 1$.
- $P(A') = 1 - P(A)$.
- **Mutually exclusive**: $A \cap B$ = empty, so $P(A \cup B) = P(A) + P(B)$.
- **Independent**: $P(A \cap B) = P(A)\,P(B)$.

FORMULA Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

FORMULA Conditional: $P(A/B) = P(A \cap B) / P(B)$, $P(B) > 0$.

➤ Counting

- Permutations nPr = n! / (n - r)! (order matters).
- Combinations nCr = n! / (r! (n - r)!) (order does not matter).

> **EXAMPLE** Two dice thrown. P(sum = 7) = 6/36 = 1/6 (pairs (1,6), (2,5), (3,4), (4,3), (5,2), (6,1)).

3. Total Probability and Bayes' Theorem

- If B1, B2, ..., Bn split S (exclusive and cover all):
  - **Total probability**: P(A) = Σ P(Bi) P(A/Bi).
  - **Bayes**: P(Bi/A) = P(Bi) P(A/Bi) / Σ P(Bj) P(A/Bj).

➤ Solved example

- Machine M1 makes 60% of items, M2 makes 40%. Defect rates: M1 2%, M2 5%.
- P(defective) = 0.6 x 0.02 + 0.4 x 0.05 = 0.012 + 0.020 = 0.032.
- An item is defective. P(it came from M2) = 0.020 / 0.032 = 0.625.

> **TIP** In Bayes problems, first draw a small tree: branch probabilities on the left, conditional ones on the right. Multiply along each path.

4. Random Variables

- **Random variable (RV)** = a rule that gives a number to each outcome.
- **Discrete**: takes countable values, has a **PMF** p(x) with Σ p(x) = 1.
- **Continuous**: has a **PDF** f(x) >= 0 with ∫ f(x) dx = 1 over the whole line. P(X = a) = 0.
- **CDF** F(x) = P(X <= x). For continuous, f(x) = F'(x).

> **FORMULA** E(X) = Σ x p(x) or ∫ x f(x) dx. Var(X) = E(X^2) - [E(X)]^2.

- E(aX + b) = a E(X) + b. Var(aX + b) = a^2 Var(X).
- For independent X, Y: Var(X + Y) = Var X + Var Y.

> **EXAMPLE** Fair die: E(X) = (1+2+...+6)/6 = 3.5. E(X^2) = 91/6. Var = 91/6 - 12.25 = 35/12.

5. Standard Distributions

Distribution	P(X = x) or f(x)	Mean	Variance
Bernoulli (p)	p^x (1-p)^(1-x), x = 0, 1	p	p(1 - p)
Binomial (n, p)	nCx p^x q^(n-x)	np	npq
Poisson (λ)	e^(-λ) λ^x / x!	λ	λ
Geometric (p)	q^(x-1) p, x = 1, 2, ...	1/p	q/p^2
Uniform (a, b)	1/(b - a) on [a, b]	(a + b)/2	(b - a)^2 / 12
Exponential (λ)	λ e^(-λx), x >= 0	1/λ	1/λ^2
Normal (μ, σ)	(1/(σ sqrt(2π))) e^(-(x-μ)^2 / 2σ^2)	μ	σ^2

- q = 1 - p. Poisson is the limit of binomial when n is large, p small, $\lambda = np$.
- Exponential is **memoryless**: P(X > s + t / X > s) = P(X > t).

➤ Solved example (Binomial)

- A coin is tossed 4 times. P(exactly 2 heads) = 4C2 (1/2)^2 (1/2)^2 = 6/16 = 3/8.

➤ Solved example (Poisson)

- Average 2 calls per minute. P(no call in a minute) = e^(-2) ≈ 0.135.
- P(at least 1 call) = 1 - e^(-2) ≈ 0.865.

6. Normal Distribution

- Bell shaped, symmetric about μ. Mean = median = mode.
- **Standard normal**: Z = (X - μ) / σ, has mean 0 and variance 1.

➤ Area rule

Range	Area (approx)
μ ± 1σ	68.27%
μ ± 2σ	95.45%
μ ± 3σ	99.73%

EXAMPLE Marks ~ N(60, 10^2). P(X > 80) = P(Z > 2) = 0.5 - 0.4772 = 0.0228 (about 2.3%).

REMEMBER Normal table gives area from 0 to z. Area from 0 to 1 = 0.3413, 0 to 2 = 0.4772, 0 to 3 = 0.4987.

7. Joint Distribution, Correlation and Regression

- **Covariance**: $Cov(X, Y) = E(XY) - E(X)E(Y)$. Independent $\Rightarrow$ Cov = 0 (not the other way round).
- **Correlation** $r = Cov(X, Y) / (\sigma x\, \sigma y)$. Always $-1 <= r <= 1$.
  - $r = 1$ perfect positive, $r = -1$ perfect negative, $r = 0$ no linear relation.

➤ Regression lines

- Y on X: $y - \bar{y} = byx\,(x - \bar{x})$, with $byx = r\,\sigma y / \sigma x$.
- X on Y: $x - \bar{x} = bxy\,(y - \bar{y})$, with $bxy = r\,\sigma x / \sigma y$.

FORMULA $r^2 = byx \times bxy$. r has the same sign as the regression coefficients. Both lines pass through $(\bar{x}, \bar{y})$.

EXAMPLE $byx = 0.8$, $bxy = 0.45$. $r^2 = 0.36 \Rightarrow r = 0.6$ (positive since both are positive).

➤ Spearman rank correlation

FORMULA $R = 1 - 6 \Sigma d^2 / (n(n^2 - 1))$, d = difference of ranks.

8. Sampling and Hypothesis Testing

- **Population** = whole group. **Sample** = part of it. **Parameter** (μ, σ) vs **statistic** ($\bar{x}$, s).
- **Central Limit Theorem**: for large n, $\bar{x}$ is approximately normal with mean μ and SD $\sigma / \sqrt{n}$ (standard error).
- **Null hypothesis H0**: no difference. **Alternative H1**: there is a difference.

Error	Meaning
Type I (α)	Reject H0 when it is true
Type II (β)	Accept H0 when it is false

- Level of significance α is usually 5% or 1%. Critical z (two-tailed): 1.96 at 5%, 2.58 at 1%.

➤ Which test?

Test	Use
z-test	Large sample ($n > 30$), or σ known
t-test	Small sample, σ unknown, $df = n - 1$
Chi-square	Goodness of fit, independence of attributes
F-test	Compare two variances (ANOVA)

FORMULA $z = (\bar{x} - \mu) / (\sigma / \sqrt{n})$. Chi-square = $\Sigma (O - E)^2 / E$.

➤ Solved example (z-test)

- Claim: mean life of bulbs $\mu = 1000$ h, $\sigma = 100$ h. A sample of 64 bulbs gives $\bar{x} = 980$ h. Test at 5%.
- Standard error = $100 / \sqrt{64} = 12.5$.
- $z = (980 - 1000) / 12.5 = -1.6$. $|z| = 1.6 < 1.96$.
- Accept H0: the difference is not significant at 5%.

➤ Solved example (chi-square)

- A die is thrown 60 times: faces 1–6 come 8, 12, 9, 11, 10, 10 times. Expected E = 10 each.
- $\Sigma (O - E)^2 / E = (4 + 4 + 1 + 1 + 0 + 0)/10 = 1.0$.
- $df = 6 - 1 = 5$. Table value at 5% = 11.07. $1.0 < 11.07 \Rightarrow$ the die is fair (no evidence against it).

REMEMBER Degrees of freedom: chi-square goodness of fit = (classes − 1 − parameters estimated); contingency table = $(r - 1)(c - 1)$.

9. Most Asked Questions (practice)

1. Find mean, median, mode and SD of a grouped frequency table.
2. State and prove Bayes' theorem. Solve a three-machine problem.
3. Find k so that $f(x) = kx^2$ on $[0, 3]$ is a PDF. Then find $P(1 < X < 2)$.
4. Derive mean and variance of the binomial distribution.
5. Fit a Poisson distribution to given data.

6. If X ~ N(50, 5^2), find P(45 < X < 60).

7. Find the correlation coefficient and both regression lines for given data.

8. Explain Type I and Type II errors with an example.

9. Apply the chi-square test for goodness of fit (die thrown 120 times).

10. State the Central Limit Theorem and its use.

TIP GATE loves Bayes' theorem, expectation/variance of standard distributions, and the Poisson and normal probabilities. Learn the table in Unit 5 by heart.