Statistics & Probability Cheat Sheet

n Comprehensive Quick-Reference Guide for Exams

1. Fundamentals of Probability

Axioms of Probability:

- $0 \leq P(A) \leq 1$
- $P(\Omega) = 1$ (Sample space Ω)
- For disjoint events $A_1, A_2, \ldots$: $P(\bigcup A_i) = \sum P(A_i)$

Core Rules:

- **Complement:** $P(A^c) = 1 - P(A)$
- **Addition Rule:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- **Conditional Probability:**

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad (P(B) > 0)$$

- **Independence:** A, B are independent $\iff P(A \cap B) = P(A)P(B)$

Law of Total Probability: Partition $B_1, B_2, \ldots, B_n$ of Ω:

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

Bayes' Theorem:

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

2. Discrete Random Variables

PMF: $p(x) = P(X = x)$, with $\sum p(x) = 1$
Expectation: $\mathbb{E}[X] = \sum x p(x)$
Variance: $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

Common Discrete Distributions:

1. **Bernoulli(p):** $X \in \{0, 1\}$

- $P(X = 1) = p$
- $\mathbb{E}[X] = p$, $\text{Var}(X) = p(1-p)$

2. **Binomial(n, p):** Number of successes in n trials

- $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
- $\mathbb{E}[X] = np$, $\text{Var}(X) = np(1-p)$

3. **Poisson(λ):** Rare events rate λ

- $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \{0, 1, 2, \ldots\}$
- $\mathbb{E}[X] = \lambda$, $\text{Var}(X) = \lambda$

4. **Geometric(p):** First success on trial k

- $P(X = k) = (1-p)^{k-1} p$
- $\mathbb{E}[X] = \frac{1}{p}$, $\text{Var}(X) = \frac{1-p}{p^2}$

3. Continuous Random Variables

PDF: $f(x) \geq 0$, $\int_{-\infty}^{\infty} f(x)dx = 1$
CDF: $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$
Expectation: $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)dx$

Common Continuous Distributions:

1. **Uniform(a, b):**

- $f(x) = \frac{1}{b-a}$ for $x \in [a, b]$
- $\mathbb{E}[X] = \frac{a+b}{2}$, $\text{Var}(X) = \frac{(b-a)^2}{12}$

2. **Exponential(λ):**

- $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$
- $\mathbb{E}[X] = \frac{1}{\lambda}$, $\text{Var}(X) = \frac{1}{\lambda^2}$

3. **Normal(μ, σ^2):**

- $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
- Standard Normal $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$
- $\mathbb{E}[X] = \mu$, $\text{Var}(X) = \sigma^2$

4. Joint Distributions & Covariance

Joint CDF: $F(x,y) = P(X \le x, Y \le y)$
Covariance:

$$\text{Cov}(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Correlation Coefficient:

$$\varrho_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}, \quad -1 \le \varrho_{XY} \le 1$$

Properties of Linear Combinations:

- $\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$
- $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X,Y)$
- If X, Y independent $\implies \text{Cov}(X,Y) = 0$

5. Descriptive Statistics

Given Sample: $x_1, x_2, \ldots, x_n$

- **Sample Mean:** $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- **Sample Variance:** $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- **Sample Std Dev:** $s = \sqrt{s^2}$
- **Median:** Middle value of ordered data
- **IQR:** $Q_3 - Q_1$ (Interquartile Range)

6. Limit Theorems & Sampling

Central Limit Theorem (CLT): Let $X_1, \ldots, X_n$ be i.i.d. with mean μ and variance σ^2. As $n \to \infty$:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{d} \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

Standardized form: $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$

Empirical Rule (68-95-99.7):

$$\longrightarrow z$$

[color=primary,thick,domain=-3:3,samples=50] plot (,2.5*exp(-*/2)); [dashed] (0,0) – (0,2.5) node[above] μ; [dashed] (-1,0) – (-1,1.52); [dashed] (1,0) – (1,1.52); at (0,0.8) $_{68\%}$; at (-1,-0.4) $_{-\sigma}$; at (1,-0.4) $_{+\sigma}$;

7. Hypothesis Testing Framework

1. State Null (H_0) and Alternative (H_1) Hypotheses. 2. Choose Significance Level α (e.g., 0.05). 3. Calculate Test Statistic. 4. Determine p-value or Critical Region. 5. Reject H_0 if p-value $\le \alpha$.

Decision Matrix:

	Fail to Reject H_0	**Reject H_0**
H_0 **True**	Correct $(1-\alpha)$	Type I Error (α)
H_0 **False**	Type II Error (β)	Correct $(1-\beta)$

8. Common Test Statistics

1. One-Sample Z-Test (σ known):

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

2. One-Sample t-Test (σ unknown):

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$$

3. Two-Sample t-Test (Independent):

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

4. Chi-Square Test for Goodness of Fit:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{k-1-p}$$

9. Simple Linear Regression

Model: $Y = \beta_0 + \beta_1 X + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$
Least Squares Estimates:

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Coefficient of Determination (R^2):

$$R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} = (r_{XY})^2$$