Probability & Statistics Formula Sheet

1. Probability Theory

Basic Concepts

- **Sample Space (S)**: Set of all possible outcomes
- **Event Probability**:
$$P(A) = \frac{\text{Number of favorable outcomes}}{\text{Total outcomes}} \quad \text{(Classical)}$$
- **Axioms**:
  - $0 \leq P(A) \leq 1$
  - $P(S) = 1$
  - For mutually exclusive events: $P(\bigcup A_i) = \sum P(A_i)$

Conditional Probability

- **Definition**:
$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$
- **Bayes' Theorem**:
$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

Independence

- **Events A and B are independent** iff:
$$P(A \cap B) = P(A)P(B)$$

2. Random Variables

Discrete Random Variables

- **Probability Mass Function (PMF)**:
$$p_X(x) = P(X = x)$$
- **Cumulative Distribution Function (CDF)**:
$$F_X(x) = P(X \leq x) = \sum_{k \leq x} p_X(k)$$

Continuous Random Variables

- **Probability Density Function (PDF)**:
$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$
- **CDF**:
$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

Expectation and Variance

- **Expected Value**:
$$E[X] = \begin{cases} \sum_x x p_X(x) & \text{(Discrete)} \\ \int_{-\infty}^{\infty} x f_X(x)dx & \text{(Continuous)} \end{cases}$$
- **Variance**:
$$Var(X) = E[X^2] - (E[X])^2$$

3. Special Distributions

Discrete Distributions

- **Bernoulli** ($X \sim \text{Bern}(p)$):
$$P(X = x) = p^x(1-p)^{1-x}, \quad x \in \{0, 1\}$$
- **Binomial** ($X \sim \text{Bin}(n, p)$):
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$
- **Poisson** ($X \sim \text{Pois}(\lambda)$):
$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Continuous Distributions

- **Uniform** ($X \sim U(a, b)$):
$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$
- **Exponential** ($X \sim \text{Exp}(\lambda)$):
$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$
- **Normal** ($X \sim N(\mu, \sigma^2)$):
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

4. Descriptive Statistics

Measures of Central Tendency

- **Mean**:
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$
- **Median**: Middle value of ordered data
- **Mode**: Most frequent value

Measures of Dispersion

- **Variance** (Sample):
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$
- **Standard Deviation**:
$$s = \sqrt{s^2}$$

Skewness and Kurtosis

- **Skewness**:
$$\text{Skew} = \frac{\frac{1}{n}\sum(x_i - \bar{x})^3}{s^3}$$
- **Kurtosis**:
$$\text{Kurt} = \frac{\frac{1}{n}\sum(x_i - \bar{x})^4}{s^4} - 3$$

5. Bivariate Data

Correlation

- **Covariance**:
$$\text{Cov}(X, Y) = \frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})$$
- **Pearson's Correlation Coefficient**:
$$r = \frac{\text{Cov}(X, Y)}{s_X s_Y}$$

Regression

- **Simple Linear Regression Line**:

$$\hat{y} = a + bx$$

where:

$$b = \frac{\mathrm{Cov}(X, Y)}{s_X^2}, \quad a = \bar{y} - b\bar{x}$$