Basic Probability

Definitions

Sample Space: Set of all possible outcomes
Event: Subset of sample space
Probability: Numerical measure of chance

Set Operations

$A \cap B$ = Both A and B occur
$A \cup B$ = A or B (or both) occur
A^c = A does not occur

Probability Rules

Intersection: $\Pr(A \cap B) = \Pr(A \mid B)\Pr(B)$
Union: $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
Complement: $\Pr(A^c) = 1 - \Pr(A)$
Conditional: $\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$

Independence

A and B are independent if: $\Pr(A \cap B) = \Pr(A)\Pr(B)$
If independent: $\Pr(A \mid B) = \Pr(A)$

Discrete Random Variables

PMF & CDF

PMF: $f(x) = \Pr(X = x)$
Expected Value: $E[X] = \sum_{x \in S} x \cdot f(x)$
Variance: $\mathrm{Var}[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$
Standard Deviation: $\mathrm{SD}[X] = \sqrt{\mathrm{Var}[X]}$

Linear Transformations

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

Conditional Expectation

$E[X \mid Y = y] = \sum_x x \cdot \Pr(X = x \mid Y = y)$
Law of Iterated Expectations: $E[E[X \mid Y]] = E[X]$

Moment-Generating Functions

Definition

$M_X(t) = E[e^{tX}] = \sum_x e^{tx} f(x)$ (discrete)
$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)dx$ (continuous)

Finding Moments

r-th moment: $\mu_r = E[X^r] = M_X^{(r)}(0)$

Standardized Moments

$\hat{\mu}_k = \frac{E[(X-\mu)^k]}{E[(X-\mu)^2]^{k/2}}$
Skewness: $\hat{\mu}_3$ (3rd standardized moment)
Kurtosis: $\hat{\mu}_4$ (4th standardized moment)

Discrete Distributions

Bernoulli Distribution

Modeling: Single trial with success/failure outcome $X \sim \mathrm{Bernoulli}(p)$
$P(X = 1) = p, \quad P(X = 0) = 1 - p$
$E[X] = p, \quad \mathrm{Var}[X] = p(1 - p)$
Parameter: $0 < p < 1$ (success probability)

Binomial Distribution

Modeling: Number of successes in n independent trials $X \sim \mathrm{Binomial}(n, p)$
$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$
$E[X] = np, \quad \mathrm{Var}[X] = np(1 - p)$
Parameters: $n \geq 1$ (trials), $0 < p < 1$ (success probability)

Geometric Distribution

Modeling: Number of trials until first success $X \sim \mathrm{Geometric}(p)$
$P(X = x) = p(1 - p)^{x-1}$ for $x = 1, 2, 3, \ldots$
$E[X] = \frac{1}{p}, \quad \mathrm{Var}[X] = \frac{1-p}{p^2}$
Parameter: $0 < p < 1$ (success probability)

Negative Binomial Distribution

Modeling: Number of trials until r-th success $X \sim \mathrm{NegBin}(r, p)$
$P(X = x) = \binom{x-1}{r-1} p^r (1 - p)^{x-r}$ for $x = r, r + 1, \ldots$
$E[X] = \frac{r}{p}, \quad \mathrm{Var}[X] = \frac{r(1-p)}{p^2}$
Parameters: $r > 0$ (target successes), $0 < p < 1$ (success probability)

Poisson Distribution

Modeling: Number of events in fixed interval (rare events) $X \sim \mathrm{Poisson}(\lambda)$
$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ for $k = 0, 1, 2, \ldots$
$E[X] = \lambda, \quad \mathrm{Var}[X] = \lambda$
Parameter: $\lambda > 0$ (rate/mean)

Continuous Random Variables

PDF & CDF

PDF: $f(x) \geq 0$ and $\int_S f(x)dx = 1$
CDF: $F(a) = \Pr(X \leq a) = \int_{-\infty}^{a} f(x)dx$
Expected Value: $E[X] = \int_{-\infty}^{\infty} x f(x)dx$
Variance: $\mathrm{Var}[X] = \int_{-\infty}^{\infty} (x - E[X])^2 f(x)dx$

Quantiles

p-th quantile π_p: $p = \int_{-\infty}^{\pi_p} f(x)dx$

Uniform Distribution

Modeling: Equal probability over interval [a,b] $X \sim \mathrm{Uniform}(a, b)$

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } x \geq b \end{cases}$$

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

Normal Distribution

Modeling: Bell-shaped distribution, many natural phenomena

Standard Normal

$Z \sim N(0,1)$
$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$

General Normal

$X \sim N(\mu, \sigma^2)$
$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
$E[X] = \mu, \quad \mathrm{Var}[X] = \sigma^2$

Standardization

If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

MGF of Normal

$M(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$

Empirical Rule

68% within $\mu \pm \sigma$
95% within $\mu \pm 2\sigma$
99.7% within $\mu \pm 3\sigma$

Linear Combinations

If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ are independent:
$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
$a + bZ \sim N(a, b^2)$ where $Z \sim N(0,1)$

Exponential Distribution

Modeling: Time between events in Poisson process
Parameter: $\lambda > 0$ (rate), $\theta = 1/\lambda$ (scale parameter)
$W \sim \mathrm{Exponential}(\lambda)$

$$f(w) = \begin{cases} \lambda e^{-\lambda w} & \text{if } w \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$$F(w) = \begin{cases} 0 & \text{if } w < 0 \\ 1 - e^{-\lambda w} & \text{if } w \geq 0 \end{cases}$$

$E[W] = \frac{1}{\lambda} = \theta, \quad \mathrm{Var}[W] = \frac{1}{\lambda^2} = \theta^2$

Gamma Distribution

Modeling: Time until α-th event in Poisson process
Parameters: $\alpha > 0$ (shape), $\theta > 0$ (scale), $\theta = 1/\lambda$
$X \sim \mathrm{Gamma}(\alpha, \theta)$
$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$ for $x \geq 0$
$E[X] = \alpha\theta, \quad \mathrm{Var}[X] = \alpha\theta^2$
Gamma Function: $\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy$
Recurrence: $\Gamma(t) = (t-1)\Gamma(t-1)$, $\Gamma(1) = 1$
Special Values: $\Gamma(n) = (n-1)!$ for integer n, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Special Cases & Relationships

Special Cases

Exponential: $\mathrm{Gamma}(1, \theta) = \mathrm{Exponential}(\lambda)$ where $\theta = 1/\lambda$
Geometric: $\mathrm{NegBin}(1, p) = \mathrm{Geometric}(p)$

Limiting Relationships

Binomial to Poisson: $\mathrm{Binomial}(n, p) \to \mathrm{Poisson}(\lambda)$ as $n \to \infty$, $p \to 0$, $np \to \lambda$

Memoryless Property

Exponential: $\Pr(X > s + t \mid X > s) = \Pr(X > t)$
Geometric: $\Pr(X > k + j \mid X > k) = \Pr(X > j)$

Useful Formulas

Variance Identity

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

Combinatorics

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

DeMorgan's Laws

$(A \cup B)^c = A^c \cap B^c$
$(A \cap B)^c = A^c \cup B^c$

Set Laws

Commutativity: $A \cup B = B \cup A$, $A \cap B = B \cap A$
Associativity: $(A \cup B) \cup C = A \cup (B \cup C)$
Distributive: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Key Concepts

Monty Hall Problem

Always switch! Switching gives $\frac{2}{3}$ probability vs $\frac{1}{3}$ for staying.

Bayes' Theorem

$\Pr(A \mid B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$

Law of Total Probability

$\Pr(A) = \sum_i \Pr(A \mid B_i)\Pr(B_i)$ where $\{B_i\}$ partition the sample space.