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 independent $\Leftrightarrow \Pr(A \cap B) = \Pr(A)\Pr(B)$
If independent: $\Pr(A \mid B) = \Pr(A)$

Discrete Random Variables

PMF & CDF

PMF: $f(x) = \Pr(X = x)$
CDF: $F(a) = \Pr(X \le a) = \sum_{x \le a} f(x)$
Expected Value: $E[X] = \sum_{x \in S} x f(x)$
Variance: $\mathrm{Var}[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$

Linear Transformations

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

Conditional Expectation

$E[X \mid Y = y] = \sum_x x \Pr(X = x \mid Y = y)$
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)
$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$ (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: $\hat{\mu}_3$
Kurtosis: $\hat{\mu}_4$

Discrete Distributions

Bernoulli

$X \sim \text{Bernoulli}(p)$, $P(X = 1) = p$, $P(X = 0) = 1 - p$
$E[X] = p$, $\mathrm{Var}[X] = p(1 - p)$

Binomial

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

Geometric

$X \sim \text{Geometric}(p)$ counts trials to first success
$P(X = x) = p(1 - p)^{x-1}$, $x = 1, 2, \dots$
$E[X] = \frac{1}{p}$, $\mathrm{Var}[X] = \frac{1-p}{p^2}$

Negative Binomial

$X \sim \text{NegBin}(r, p)$ counts trials to r-th success
$P(X = x) = \binom{x-1}{r-1} p^r (1 - p)^{x-r}$
$E[X] = \frac{r}{p}$, $\mathrm{Var}[X] = \frac{r(1-p)}{p^2}$

Poisson

$X \sim \text{Poisson}(\lambda)$
$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$
$E[X] = \lambda$, $\mathrm{Var}[X] = \lambda$

Continuous Random Variables

PDF & CDF

PDF: $f(x) \ge 0$, $\int_S f(x)dx = 1$
CDF: $F(a) = \Pr(X \le a) = \int_{-\infty}^{a} f(x)dx$
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 satisfies $p = \int_{-\infty}^{\pi_p} f(x)dx$

Continuous Distributions

Uniform $[a, b]$

$f(x) = \frac{1}{b-a}$ for $a < x < b$, else 0
$F(x) = 0$ if $x < a$, $\frac{x-a}{b-a}$ if $a \le x < b$, 1 if $x \ge b$
$E[X] = \frac{a+b}{2}$, $\mathrm{Var}[X] = \frac{(b-a)^2}{12}$

Normal

$Z \sim N(0, 1)$ has $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$
$X \sim N(\mu, \sigma^2)$ has $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
Standardize: $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$
$M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$
Empirical rule: 68/95/99.7% within $\mu \pm 1/2/3\sigma$

Exponential

$W \sim \text{Exp}(\lambda)$ with $f(w) = \lambda e^{-\lambda w}$ for $w \ge 0$
$F(w) = 1 - e^{-\lambda w}$, $E[W] = 1/\lambda$, $\mathrm{Var}[W] = 1/\lambda^2$
Memoryless: $\Pr(W > s + t \mid W > s) = \Pr(W > t)$

Gamma

$X \sim \text{Gamma}(\alpha, \theta)$
$f(x) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$ for $x \ge 0$
$E[X] = \alpha\theta$, $\mathrm{Var}[X] = \alpha\theta^2$
$\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy$, $\Gamma(t) = (t-1)\Gamma(t-1)$

Special Cases & Relationships

Exponential: $\text{Gamma}(1, \theta) = \text{Exp}(\lambda)$, $\theta = 1/\lambda$
Geometric: $\text{NegBin}(1, p) = \text{Geom}(p)$
Binomial $\to$ Poisson: $n \to \infty$, $p \to 0$, $np \to \lambda$

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

Bivariate Distributions

Expectations & Covariance

$E[X+Y] = E[X] + E[Y]$
$\text{Cov}[X,Y] = E[(X - E[X])(Y - E[Y])]$
Interpretation: > 0 positive association, < 0 negative, $= 0$ no linear link
$\text{Cov}(aX + b, cY + d) = ac\,\text{Cov}(X,Y)$
$\text{Cov}(X, Y + Z) = \text{Cov}(X,Y) + \text{Cov}(X,Z)$
Independence $\Rightarrow \text{Cov}(X,Y) = 0$, converse need not hold

Variance of a Sum

$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\,\text{Cov}(X,Y)$
If X and Y independent: $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$

Correlation Coefficient

$\rho = \frac{\text{Cov}(X,Y)}{\text{SD}(X)\,\text{SD}(Y)}$
Scale-free measure, $\rho \in [-1, 1]$, $\rho = 0$ means no linear association

Conditional Distributions

- Discrete: $g(x \mid y) = \frac{f(x,y)}{f_Y(y)}$ for $f_Y(y) > 0$
- Independence $\Rightarrow g(x \mid y) = f_X(x)$
- Continuous joint pdf $f(x,y)$: $f(x,y) \geq 0$, $\iint f(x,y)dxdy = 1$
- Continuous marginals: $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$, $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$
- $X \perp Y$ iff $f(x,y) = f_X(x)f_Y(y)$

Bivariate Normal

PDF: $f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{q(x,y)}{2}\right)$
$q(x,y) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]$
Marginals are normal with means μ_X, μ_Y and variances σ_X^2, σ_Y^2
$X + Y$ is normal when (X,Y) jointly normal
Conditional $Y|X = x$ normal with $E[Y|x] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$
$\text{Var}(Y|x) = \sigma_Y^2(1-\rho^2)$
$\sigma_{XY} = \rho\sigma_X\sigma_Y$ so $E[Y|x] = \mu_Y + \frac{\sigma_{XY}}{\sigma_X^2}(x - \mu_X)$

Functions of Random Variables

Distribution Function Technique

If $Y = u(X)$, $G(y) = \Pr(Y \leq y) = \Pr(u(X) \leq y)$, density $g(y) = G'(y)$
Example: $X \sim N(0,1)$, $Y = e^X$ (log-normal) gives $G(y) = \Phi(\ln y)$, $g(y) = \phi(\ln y)\frac{1}{y}$ for $y > 0$

Change-of-Variable (1 Variable)

If $Y = u(X)$, define $v()$ such that $X = v(Y)$.
Suppose $c_1 < x < c_2$ and $d_1 = u(c_1) < y < d_2 = u(c_2)$ ($v()$ is increasing),

$$G(y) = \int_{c_1}^{v(y)} f(x)dx, \qquad d_1 < y < d_2$$

$$G'(y) = g(y) = f(v(y)) \cdot v'(y), \qquad d_1 < y < d_2$$

Log-normal example:
$u(x) = \exp(x)$, $v(y) = \ln(y)$, $v'(y) = 1/y$

$$g(y) = \phi(\ln y) \cdot \frac{1}{y} \qquad \text{when } y > 0$$

Suppose $c_1 < x < c_2$ and $d_1 = u(c_1) > y > d_2 = u(c_2)$ ($v()$ is decreasing),

$$G(y) = \int_{v(y)}^{c_2} f(x)dx, \qquad d_2 < y < d_1$$

$$G'(y) = g(y) = f(v(y)) \cdot (-v'(y)), \qquad d_2 < y < d_1$$

Note: The negative sign appears because $v'(y) < 0$ when $v()$ is decreasing.
In general, we have:

$$g(y) = f(v(y)) \cdot |v'(y)|, \qquad y \in S_Y$$

Cauchy Distribution

$W \sim U(-\pi/2, \pi/2)$, $X = \tan W$ gives standard Cauchy
$f_X(x) = \frac{1}{\pi(1+x^2)}$, heavy tails, symmetric about 0
Median = mode = 0; mean and variance do not exist
Equivalent views: ratio of two independent standard normals; t with 1 degree of freedom
Graph of pdf known as the witch of Agnesi

Multivariate Transformations

For (X_1, X_2) with joint pdf f, let $Y_1 = u_1(X_1, X_2)$, $Y_2 = u_2(X_1, X_2)$ invertible to $X_i = v_i(Y_1, Y_2)$
Joint pdf $g(y_1, y_2) = \begin{vmatrix} \frac{\partial v_1}{\partial y_1} & \frac{\partial v_1}{\partial y_2} \\ \frac{\partial v_2}{\partial y_1} & \frac{\partial v_2}{\partial y_2} \end{vmatrix} f(v_1(y_1, y_2), v_2(y_1, y_2))$
Jacobian determinant captures area/volume distortion from the transformation

Limit Theorems & Inequalities

Sample Mean

$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ for iid X_i with mean μ, variance σ^2
$E[\overline{X}] = \mu$, $\text{Var}[\overline{X}] = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$

Law of Large Numbers

Sample mean converges in probability to μ as n grows

Central Limit Theorem

Standardized mean $W = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$ for large n
Justifies normal approximations even when data not normal, provided finite mean/variance

Inequalities

Markov: For nonnegative X, $\Pr(X \geq a) \leq \frac{E[X]}{a}$
Chebyshev: $\Pr(|X - E[X]| \geq r) \leq \frac{\text{Var}[X]}{r^2}$

MGF Method Insight

If mgfs $M_n(t)$ converge to $M(t)$ near 0, then $X_n \Rightarrow X$ with mgf $M(t)$; used to establish the CLT limit

Key Concepts

Monty Hall Problem: Switching doors wins with probability $\frac{2}{3}$
Bayes' Theorem: $\Pr(A \mid B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$
Total Probability: $\Pr(A) = \sum_i \Pr(A \mid B_i)\Pr(B_i)$ for partition $\{B_i\}$