Basic Probability Rules

Addition Rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Multiplication Rule: $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

Conditional Probability: $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$

Baye's Rule: $P(A_1 \mid B) = \dfrac{P(B \mid A_1)P(A_1)}{\sum P(B \mid A_i)P(A_i)}$

DeMorgan's Laws: $P\left[(A \cup B)'\right] = P(A' \cap B')$

$P\left[(A \cap B)'\right] = P(A' \cup B')$

Law of Total Probability: $P(B) = P(B \cap A) + P(B \cap A')$

***A* and *B* are independent** $(A \perp B)$ **if & only if:**

- $P(A \cap B) = P(A)P(B)$
- $P(A \mid B) = P(A)$
- $P(B \mid A) = P(B)$

Combinatorics

Multiplication Rule: The number of ways to make n choices, having k_i options for choice number i, is equal to $k_1 \cdot k_2 \cdot \ldots \cdot k_n$.

Permutations of *n* objects: $n!$

Permutations of *k* out of *n* objects: $nPk = \dfrac{n!}{(n-k)!}$

Partitions: The number of ways to partition n objects into k non-overlapping groups with sizes $n_1, n_2, \ldots, n_k$ is equal to:

- $\dbinom{n}{n_1 n_2 \cdots n_t} = \dfrac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_t!}$

Combinations: The number of ways to choosing *k* out of *n* objects:

- $nCk = \dbinom{n}{k} = \dfrac{n!}{k! \cdot (n-k)!}$

Distribution and Density Functions

Discrete Distribution Functions

PMF: $f(x) = P[X = x]$, $P[a \le x \le b] = \sum_{x=a}^{b} f(x)$

CDF: $F(x) = P[X \le x] = \sum_{x=a}^{b} f(x)$

Survival Fn: $S(x) = P[X > x] = 1 - F(x)$

Continuous Distribution Functions

PDF: $f(x) \approx \dfrac{1}{2\varepsilon} P[x - \varepsilon < x < x + \varepsilon]$, $P[a \le x \le b] = \int_a^b f(x)\,dx$

CDF: $F(x) = P[X \le x] = \int_{-\infty}^{x} f(t)\,dt$

Survival: $S(x) = P[X > x] = 1 - F(x)$

Derivatives: $F'(x) = -S'(x) = f(x)$

Hazard Rate: $h(x) = \dfrac{f(x)}{1 - F(x)} = -\dfrac{d}{dx} \ln[1 - F(x)]$

Summation & Integration Formulas

The following formulas are useful to know:

- $1 + 2 + 3 + \ldots + n = \dfrac{n(n+1)}{2}$
- $a + ar + ar^2 + \ldots + ar^{n-1} = \dfrac{a - ar^n}{1 - r}$
- $a + ar + ar^2 + ar^3 + \ldots = \dfrac{a}{1 - r}$, $|r| < 1$
- $1 + 2r + 3r^2 + 4r^3 + \ldots = \dfrac{1}{(1-r)^2}$, $|r| < 1$
- $\int_0^{\infty} x^k e^{-ax}\,dx = \dfrac{k!}{a^{k+1}}$

Moments and MGF's

Expected Value (Discrete): $E[X] = \sum x f(x)$

$E[h(X)] = \sum h(x) f(x)$

Expected Value (Continuous): $E[X] = \int_{-\infty}^{\infty} x f(x)\,dx$

$E[h(X)] = \int_{-\infty}^{\infty} h(x) f(x)\,dx$

Darth Vader Rule: If $X \ge 0$, then $E[X] = \int_0^{\infty} S(x)\,dx$

Variance: $\mathrm{Var}[X] = E\left[(X - \mu)^2\right] = E[X^2] - (E[X])^2$

Algebraic Properties: $E[aX + b] = a\,E[X] + b$

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

Moments

- ***n*-th Moment:** $E[X^n]$
- ***n*-th Central Moment:** $E\left[(X - \mu)^n\right]$

Moment Generating Functions

MGF Definition: $M_X(t) = E[e^{tX}]$

MGF Properties:

- $M_X(0) = 1$
- $M'_X(0) = E[X]$ and $M_X^{(n)}(0) = E[X^n]$
- $\dfrac{d^2}{dt^2} \ln[M_X(t)]\big|_{t=0} = \mathrm{Var}[X]$
- If *X* is discrete with $f(x_i) = p_i$, then $M_X(t) = \sum p_i e^{t x_i}$.

Conditional Expectations

- $E[X \mid a < X < b] = \dfrac{1}{F(b) - F(a)} \displaystyle\int_a^b x f(x)\,dx$
- $E[X \mid X < k] = \dfrac{1}{F(k)} \displaystyle\int_{-\infty}^{k} x f(x)\,dx$
- $E[X \mid X > k] = \dfrac{1}{S(k)} \displaystyle\int_k^{\infty} x f(x)\,dx$

Miscellaneous Formulas

- **Skewness:** $\dfrac{E\left[(X - \mu)^3\right]}{\sigma^3}$
- **Coefficient of Variation:** $c_v = \sigma / \mu$
- **100p-th Percentile:** $F(\pi_p) \ge p$
- **Chebyshev's Ineq:** $P\left[|X - \mu_X| > r\,\sigma_X\right] \le \dfrac{1}{r^2}$

Joint Distributions

Joint PDF and CDF

Discrete

- PMF: $f(x,y) = P[X = x \text{ and } Y = y]$
- CDF: $F(x,y) = P[X \le x \text{ and } Y \le y] = \sum_{s \le x}\sum_{t \le y} f(s,t)$

Continuous

- PDF: $f(x,y)$ = joint density function
- CDF: $F(x,y) = P[X \le x \text{ and } Y \le y] = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)\,dt\,ds$
- $P[(X,Y) \in R] = \iint_R f(s,t)\,dt\,ds$
- $\frac{\partial^2}{\partial x \partial y} F(x,y) = f(x,y)$

Marginal Distributions

Discrete:

- Marginal PMF: $f_X(x) = P[X = x] = \sum_y f(x,y)$
- Marginal CDF: $F_X(x) = P[X \le x] = \sum_{t \le x} f_X(t)$

Continuous:

- Marginal PDF: $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$
- Marginal CDF: $F_X(x) = P[X \le x] = \int_{-\infty}^{x} f_X(t)\,dt$

Conditional Distributions

Shorthand Notation

- $g(x \mid y) = f_{X|Y}(x \mid Y = y)$
- $h(y \mid x) = f_{Y|X}(y \mid X = x)$

Definition and Basic Properties

- $g(x \mid y) = \frac{f(x,y)}{f_Y(y)}$
- $h(y \mid x) = \frac{f(x,y)}{f_X(x)}$
- $P[a \le X \le b \mid Y = k] = \int_a^b f(x \mid k)\,dx = \frac{1}{f_Y(k)}\int_a^b f(x,k)\,dx$
- $f(x,y) = g(x \mid y) f_Y(y) = h(y \mid x) f_X(x)$

Expected Values and Variance

Expected value of $h(X,Y)$

- Discrete: $E[h(X,Y)] = \sum_x \sum_y h(x,y) \cdot f(x,y)$
- Continuous: $E[h(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y)\,dx\,dy$
- $E[X + Y] = E[X] + E[Y]$

Marginal Expectation

- $E[X] = \sum_x x f_X(x)$
- $E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$

Conditional Expectation

- $E[X \mid Y = k] = \sum_x x\, g(x \mid k)$
- $E[X \mid Y = k] = \int_{-\infty}^{\infty} x\, g(x \mid k)\,dx$
- $E_Y[E_X[X \mid Y]] = E[X]$

Conditional Variance

- $\operatorname{Var}[X \mid Y = k] = E[X^2 \mid k] + (E[X \mid k])^2$

Law of Total Variance

- $\operatorname{Var}[X] = E_Y[\operatorname{Var}[X|Y]] + \operatorname{Var}_Y[E[X|Y]]$

Joint Distributions (Continued)

Covariance

Definition: $\operatorname{Cov}[X,Y] = E[XY] - E[X]E[Y]$

Properties of Covariance

- $\operatorname{Cov}[X,X] = \operatorname{Var}[X]$
- $\operatorname{Var}[X + Y] = \operatorname{Var}[X] + \operatorname{Var}[Y] + 2\operatorname{Cov}[X, Y]$
- $\operatorname{Cov}[aX, bY] = ab\operatorname{Cov}[X, Y]$
- $\operatorname{Cov}[X + a, Y + b] = \operatorname{Cov}[X, Y]$

Correlation Coefficient: $\rho_{X,Y} = \frac{\operatorname{Cov}[X,Y]}{\sigma_X \sigma_Y}$

Independence of Random Variables

If one of the following statements are true, then they all are:

- X and Y are independent $(A \perp B)$.
- $f(x,y) = f_X(x) \cdot f_Y(y)$ and R is a (possibly infinite) rectangle.
- $F(x,y) = F_X(x) \cdot F_Y(y)$
- $g(x \mid y) = f_X(x)$ and $h(y \mid x) = f_Y(y)$

The following statements are true if *X* and *Y* are independent, but do not themselves imply independence:

- $E[XY] = E[X] \cdot E[Y]$
- $E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)]$
- $E[X \mid Y = k] = E[X]$ and $E[Y \mid X = k] = E[Y]$
- $\operatorname{Cov}[X,Y] = 0$
- $\rho_{X,Y} = 0$

Joint Moment Generating Functions

- $M_{X,Y}(s,t) = E[e^{sX + tY}]$
- $E[X] = \frac{\partial}{\partial s} M_{X,Y}(s,t)\Big|_{s=t=0}$
- $E[Y] = \frac{\partial}{\partial t} M_{X,Y}(s,t)\Big|_{s=t=0}$
- $E[X^n Y^m] = \frac{\partial^{n+m}}{\partial^n s\, \partial^m t} M_{X,Y}(s,t)\Big|_{s=t=0}$
- $M_{X,Y}(t,t) = M_{X+Y}(t)$

Bivariate Normal Distribution

If *X* and *Y* have a bivariate normal distribution, then:

- *X* and *Y* are both normally distributed.
- The conditional variables $X \mid (Y = k)$ and $Y \mid (X = k)$ are normally distributed.
- $E[X \mid Y = y] = \mu_X + \rho_{XY}\frac{\sigma_X}{\sigma_Y}(y - \mu_Y) = \mu_X + \frac{\operatorname{Cov}[X,Y]}{\operatorname{Var}[Y]}(y - \mu_Y)$
- $\operatorname{Var}[X \mid Y = y] = \sigma_X^2(1 - \rho_{XY}^2)$

Mixtures of Distributions

Assume $p_1 + p_2 = 1$ and X_1 and X_2 are random variables. Let X be defined as follows: $P[X = x_1] = p_1$ and $P[X = x_2] = p_2$. Then:

- $f(x) = p_1 f_1(x) + p_2 f_2(x)$
- $E[X] = p_1 E[X_1] + p_2 E[X_2]$
- $E[X^2] = p_1 E[X_1^2] + p_2 E[X_2^2]$
- $M_X(t) = p_1 M_{X_1}(t) + p_2 M_{X_2}(t)$

Note: $\operatorname{Var}[X] \ne p_1 \operatorname{Var}[X_1] + p_2 \operatorname{Var}[X_2]$.

Instead, use $\operatorname{Var}[X] = E[X^2] - (E[X])^2$

Transformations

Single Variable

Suppose that *X* is a continuous random variable with density $f_X(x)$. Assume $Y = u(X)$ is a one-to-one trans. with inverse $X = v(Y)$.

- $f_Y(y) = f_X(v(y)) \cdot |v'(y)|$
- If $v(y)$ is increasing, then $F_Y(y) = F_X(v(y))$

Multiple Variable

Suppose *X* and *Y* have joint density $f(x,y)$ and that that *U* and *V* are functions of *X* and *Y*. Let $x(u,v)$ and $y(u,v)$ refer to expressions for *x* and *y*, written in terms of *u* and *v*. The joint pdf of *U* and *V* is given by:

- $g(u,v) = f(x(u,v),\ y(u,v))\,|J|$
- Note that $J = \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix}$

Min/Max and Order Statistics

Minimum and Maximum

Suppose $X_1, \dots, X_n$ are independent random variables with CDF's and survival functions given by $F_1(x), \dots, F_n(x)$ and $S_1(x), \dots, S_n(x)$.

- $F_{max}(x) = F_1(x) \cdot F_2(x) \cdot \dots \cdot F_n(x)$
- $S_{min}(x) = S_1(x) \cdot S_2(x) \cdot \dots \cdot S_n(x)$
- $F_{min}(x) = 1 - [1 - F_1(x)] \cdot \dots \cdot [1 - F_n(x)]$

Order Statistics

Suppose $X_1, \dots, X_n$ are independent observations of a variable X, and $Y_1, \dots, Y_n$ are the associated order statistics. Let g be the joint pdf of the order statistics and let g_k be the marginal pdf of Y_k.

- $g(y_1, \dots, y_n) = n!\, f(y_1) \cdot f(y_2) \cdot \dots \cdot f(y_n)$, where $y_1 \le y_2 \le \dots \le y_n$
- $g_k(t) = k \binom{n}{k} [F(t)]^{k-1} [S(t)]^{n-k} f(t)$

Insurance and Risk Management

Notation

- Let *X* = Loss associated with a claim.
- *Let Y* = Amount paid by insurer.

Deductible = *d*

- $Y = \begin{cases} 0 & \text{if } X \le d \\ x - d & \text{if } X > d \end{cases}$
- $E[Y] = \int_d^\infty (x-d) f_X(x)\,dx = \int_d^\infty S_X(x)\,dx$

Policy Limit = *u*

- $Y = \begin{cases} X & \text{if } X \le u \\ u & \text{if } X > u \end{cases}$
- $E[Y] = \int_0^u x f_X(x)\,dx + u S_X(u) = \int_0^u S_X(x)\,dx$

Deductible = *d* and Policy Limit = *u*

- $Y = \begin{cases} 0 & \text{if } X \le d \\ X - d & \text{if } d < X < d + u \\ u & \text{if } X > d + u \end{cases}$
- $E[Y] = \int_d^{d+u} (x-d) f_X(x)\,dx + u S_X(d+u) = \int_d^{d+u} S_X(x)\,dx$

Sum of Random Variables

Expected Value and Variance

Assume $Y = \sum_{i=1}^{n} X_i$. Then:

- $E[Y] = E[X_1] + E[X_2] + \dots + E[X_n]$
- $\text{Var}[Y] = \sum \text{Var}[X_i] + 2 \sum \sum \text{Cov}[X_i, X_j]$
- Note: $(X \perp Y) \Rightarrow \text{Cov}[X, Y] = 0$

Covariance

Assume $X = \sum_{i=1}^{n} X_i$ and $Y = \sum_{i=1}^{m} Y_i$. Then:

- $\text{Cov}[X, Y] = \sum_n \sum_m \text{Cov}[X_i, Y_j]$

Convolution Method (Discrete)

Let $Y = X_1 + X_2$, where $X_1, X_2 \ge 0$. Then $f_Y(y)$ is given by:

- $f_Y(y) = \sum_{x_1=0}^{y} f(x_1, y - x_1)$
- $(\boldsymbol{X \perp Y}) \Rightarrow f_Y(y) = \sum_{x_1=0}^{y} f_1(x_1) f_2(y - x_1)$

Convolution Method (Continuous)

Let $Y = X_1 + X_2$. Then $f_Y(y)$ is given by:

- $f_Y(y) = \int_{-\infty}^{\infty} f(x_1, y - x_1)\, d_{x_1}$
- $(\boldsymbol{X \perp Y}) \Rightarrow f_Y(y) = \int_{-\infty}^{\infty} f_1(x_1) f_2(y - x_1)\, d_{x_1}$
- $(\boldsymbol{X_1, X_2 \ge 0}) \Rightarrow f_Y(y) = \int_0^y f(x_1, y - x_1)\, d_{x_1}$

Moment Generating Functions

If $Y = \sum_{i=1}^{n} X_i$ and the X_i 's are pairwise independent, then:

- $M_Y(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot \dots \cdot M_{X_n}(t)$

Central Limit Theorem

- Assume $X_1, \dots, X_n$ are independent and identically distributed (IID) with mean μ and variance σ^2, and let $Y = \sum_{i=1}^{n} X_i$.
- Then $E[Y] = n\mu$ and $\text{Var}[Y] = n\sigma^2$.
- If *n* is large, then $Y \overset{\text{approx}}{\sim} N(n\mu, n\sigma^2)$.

Sums of Specific Distributions

Assume $X_1, \dots, X_k$ are independent random var's and $Y = \sum_{i=1}^{k} X_i$.

Distribution of X_i	Distribution of Y
Bernoulli, $BIN(1, p)$	Binomial, $BIN(k, p)$
Binomial, $BIN(n_i, p)$	Binomial, $BIN(\sum n_i, p)$
Poisson, mean λ_i	Poisson, mean $\sum \lambda_i$
Geometric, p	Neg Binom, k, p
Neg Binom, r_i, p	Neg Binom, $\sum r_i, p$
Normal, $N(\mu_i, \sigma_i^2)$	Normal, $N(\sum \mu_i, \sum \sigma_i^2)$
Exp, mean μ	Gamma, $\alpha = k, \beta = 1/\mu$
Gamma, α_i, β	Gamma, $\sum \alpha = \alpha_i, \beta = \beta$
Chi-Square, k_i df	Chi-Square, $\sum k_i$ df

Discrete Distributions

Distribution	Parameters	$f(x)$	$E[x]$	$\mathrm{Var}[x]$	$M_x(t)$	Description
Uniform $X \sim UNIF(N)$	$N > 0$ N an integer	$\frac{1}{N}$ $x = 1, 2, \ldots, N$	$\frac{N+1}{2}$	$\frac{N^2-1}{12}$	$\frac{e^t(e^{Nt}-1)}{N(e^t-1)}$	Each outcome $x = 1, 2, \ldots, N$ is equally likely.
Bernoulli $X \sim BIN(1, p)$	$0 < p < 1$	$\begin{cases} q & \text{if } x=0 \\ p & \text{if } x=1 \end{cases}$ $x = 0, 1$	p	pq	$q + pe^t$	$X = 0$ indications "failure" $X = 1$ indicates "success"
Binomial $X \sim BIN(n, p)$	$n > 0$ n an integer $0 < p < 1$	$\binom{n}{x} p^x q^{n-x}$ $x = 0, 1, \ldots, n$	np	npq	$(q + pe^t)^n$	X = number of successes in n trials
Poisson $X \sim POI(\lambda)$	$\lambda > 0$	$\frac{e^{-\lambda}\lambda^x}{x!}$ $x = 0, 1, 2, \ldots$	λ	λ	$e^{\lambda(e^t-1)}$	X = number of times an event occurs in a unit of time or space
Geometric $X \sim GEO(p)$	$0 < p < 1$	$q^{x-1}p$ $x = 1, 2, 3, \ldots$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^t}{1-qe^t}$	X = number of trials required to get first success.
Negative Binomial $X \sim NB(r, p)$	$r > 0$ $0 < p < 1$	$\binom{r+x-1}{x} p^r q^x$ $x = 0, 1, 2, \ldots$	$\frac{r}{p}$	$\frac{rq}{p^2}$	$\left[\frac{pe^t}{1-qe^t}\right]^r$	X = number of trials required to get r successes.
Hyper-geometric $X \sim HYP(N, r, n)$	$N > 0$ $0 \le r \le N$ $1 \le n \le N$ All are integers	$\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$ $x \le \min[n, K]$	$n\left(\frac{r}{N}\right)$	$n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$		r objects of desired type T objects total n = sample size X = number of desired objects in sample

Continuous Distributions

Distribution	Parameters	$f(x)$	$E[x]$	$\mathrm{Var}[x]$	$M_x(t)$	Comments
Uniform $X \sim UNIF(a, b)$	$a < b$	$\frac{1}{b-a}$ $a < x < b$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{(b-a)\cdot t}$	
Pareto $X \sim PAR(a, n)$	$a > 0$ $n > 1$	$\frac{na^n}{x^{n+1}}$	$\frac{na}{n-1}$	$\frac{na^2}{(n-1)^2(n-2)}$ for $n > 2$	Not a simple function.	
Normal $X \sim N(\mu, \sigma^2)$	$\mu \in \mathbb{R}$ $\sigma^2 > 0$	$\frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty < x < \infty$	μ	σ^2	$\exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right]$	
Exponential $X \sim EXP(\lambda)$	$\lambda > 0$	$\lambda e^{-\lambda x}$ $x > 0$ $F(x) = 1 - e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - t}$	No memory property: $P[X > x+y \mid X > x] = P[X > y]$ Used to model time between events.
Gamma $X \sim GAM(\alpha, \beta)$	$\alpha > 0$ $\beta > 0$	$\frac{x^{\alpha-1} \cdot e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}$ $x > 0$	$\alpha\beta$	$\alpha\beta^2$	$(1-\beta t)^{-\alpha}$	If $X_i \sim EXP(\lambda)$ and $Y = X_1 + \ldots + X_n$ then $Y \sim GAM(n, 1/\lambda)$. It follows that $GAM(1, 1/\lambda) \sim EXP(\lambda)$.
Chi-Square $X \sim \chi^2(n)$	$n = 1, 2, \ldots$	$\frac{x^{n/2-1} e^{-x/2}}{2^{n/2}\Gamma(n/2)}$ $x > 0$	v	$2v$	$(1-2t)^{-n/2}$	$X \sim \chi^2(n) \Leftrightarrow X \sim GAM\left(\frac{n}{2}, 2\right)$

Additional comments:

- **Relationship between Poisson and Exponential:** Assume X = the time between successive events, and has an exponential distribution with mean $1/\lambda$. Let N = the number of events occurring in one unit of time. Then N has a Poisson distribution with mean λ.
- **Gamma CDF:** Assume $X \sim \mathrm{GAM}(\alpha, \beta)$, where $\alpha \in \mathbb{Z}^+$. Let $k > 0$, $\lambda = \beta k$, and $Y \sim POI(\lambda)$. Then $F_X(k) = 1 - F_Y(\alpha - 1)$.
- **MIN of Exponential Variables:** Assume $Y_1, Y_2, \ldots, Y_n$ have exponential distributions with means $1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_n$. Let $Y = \min\{Y_1, Y_2, \ldots, Y_n\}$. Then Y has an exponential distribution with mean $1/(\lambda_1 + \lambda_2 + \ldots + \lambda_n)$.