Probability Cheat Sheet

Algebra of Sets

$S \cup T = T \cup S$
$S \cap (T \cup U) = (S \cap T) \cup (S \cap U)$
$S \cup (T \cap U) = (S \cup T) \cap (S \cup U)$
$S \cup (T \cap U) = (S \cup T) \cap (S \cup U)$

De Morgan's laws

everything not in at least one set
$\left(\bigcup_n S_n\right)^c = \bigcap_n S_n^c$
everything not in all sets
$\left(\bigcap_n S_n\right)^c = \bigcup_n S_n^c$

Probability Axioms

`Non-negativity`	$P(A) \geq 0$
`Additivity`	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
`Normalization`	$P(\Omega) = 1$

for sample space Ω

Conditional Probability

Definition of Conditional Probability

$P(A|B) = \frac{P(A \cap B)}{P(B)}$

Multiplication Rule

$P\left(\cap_{i=1}^{n} A_i\right) =$
$P(A_1) P(A_2|A_1) P(A_3|A_1 \cap A_2) \dots P\left(A_n | \cap_{i=1}^{n-1} A_i\right)$

Total Probability Theorem

Let $A_1, \dots, A_n$ be disjoint events that partition the sample space. The for any event B:
$P(B) = P(A_1) P(B|A_1) + \dots + P(A_n) P(B|A_n)$

Bayes' Rule

Let $A_1, A_2, \dots, A_n$ be **disjoint events** that partition the sample space. Then for any event B:

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i) P(B|A_i)}{P(B)} \\ &= \frac{P(A_i) P(B|A_i)}{P(A_1) P(B|A_1) + \dots + P(A_n) P(B|A_n)} \end{aligned}$$

Independence of Events

A and B are independent if
$P(A \cap B) = P(A) P(B)$
A and B are conditionally independent of C if
$P(A \cap B|C) = P(A|C) P(B|C)$

Independence of Several Events

Events $A_1, A_2, \dots, A_n$ are independent if
$P\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} P(A_i)$
for every subset S of $\{1, 2, \dots, n\}$

Counting

The Counting Principle

Consider a process that consists of r stages. Suppose: For any sequence of of stages $i-1$, there are n_i possible results as stage i. Then, the total number of possible results at stage r is: $n_1 n_2 \dots n_r$

Counting Formulas for an n-Element Set

`Number of subsets`	2^n
`Permutations (order matters)`	$\frac{n!}{(n-k)!}$
`Combinations (order does not matter)`	$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

`Partitions (r disjoint subsets)`
$\binom{n}{n_1, n_2, \dots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$

Discrete Random Variables

A discrete random variable, X, is described by a *probability mass function* (PMF) that assigns the probability, $p(x)$, that X takes the value x:

$$p_X(x) = \boldsymbol{P}(\{X = x\})$$

Expectation

The expected value (or mean) of a RV with PMF p_X:

$$\boldsymbol{E}[X] = \sum_x x p_X(x)$$

well defined only if the sum converges absolutely

Expected Value Rule

Let X be a RV with PMF p_X. The function $g(X)$ is a RV with expectation:

$$\boldsymbol{E}[g(X)] = \sum_x g(x) p_X(x)$$

Variance

The variance of a RV with PMF p_X:

$$\begin{aligned} var(X) &= \boldsymbol{E}\left[(X - \boldsymbol{E}[X])^2\right] \\ &= \sum_x (x - \boldsymbol{E}[X])^2 p_X(x) \\ &= \boldsymbol{E}\left[X^2\right] - (\boldsymbol{E}[X])^2 \end{aligned}$$

Mean & Variance Linear RV*

For a RV X let

$$Y = aX + B$$

where a and b are scalars. Then
$\boldsymbol{E}[Y] = a\boldsymbol{E}[X] + b$ and $var(Y) = a^2 var(X)$
*also holds for continuous random variables

Joint Distributions

For RVs X and Y with joint PMF $p_{X,Y}(x, y)$:

Marginal PMFs

$$p_X(x) \sum_y p_{X,Y}(x, y), \quad p_Y(y) \sum_x p_{X,Y}(x, y)$$

Expectation

$$\boldsymbol{E}[g(X, Y)] = \sum_x \sum_y g(x, y) p_{X,Y}(x, y)$$

If g is linear, of the form $aX + bY + c$:

$$\boldsymbol{E}[aX + bY + c] = a\boldsymbol{E}[X] + b\boldsymbol{E}[Y] + c$$

Conditioning

The joint PMF, $p_{X,Y}(x, y)$, is related to the conditional PMFs:

$$\begin{aligned} p_{X,Y}(x, y) &= p_Y(y) p_{X|Y}(x|y) \\ &= p_X(x) p_{Y|X}(y|x) \end{aligned}$$

Consequently, the marginal PMF of X is related to the condition PMF:

$$p_X(x) = \sum_y p_Y(y) p_{X|Y}(x|y)$$

Conditional expectation of X given $Y = y$:

$$\boldsymbol{E}[X|Y = y] = \sum_x x p_{X|Y}(x|y)$$

Total Expectation Theorem

If $A_1, \dots, A_n$ are disjoint and form a partition of Ω with $\boldsymbol{P}(A_i) > 0$:

$$\boldsymbol{E}[X] = \sum_{i=1}^{n} P(A_i) \boldsymbol{E}[X|A_i]$$

For any event B with $P(A_i \cap B) > 0$:

$$\boldsymbol{E}[X|B] = \sum_{i=1}^{n} P(A_i|B) \boldsymbol{E}[X|A_i \cap B]$$

Finally,

$$\boldsymbol{E}[X] = \sum_y p_Y(y) \boldsymbol{E}[X|Y = y]$$

Independence of Random Variables

Random variables X and Y are independent if the joint probability mass function satisfies

$$p_{X,Y}(x, y) = p_X(x) p_Y(y) \quad \forall x, y$$

and conditionally independent of event A if

$$p_{X,Y}(x, y|A) = p_X(x|A) p_Y(y|A) \quad \forall x, y$$

If X and Y are independent

$$\boldsymbol{E}[XY] = \boldsymbol{E}[X] \boldsymbol{E}[Y]$$

$$var(X + Y) = var(X) + var(Y)$$

The above hold for multiple independent random variables.

Continuous Random Variables

A continuous random variable, X, is described by a *probability density function* (PDF) that assigns the probability density, $f(x)$, associated with the value x so that the probability that X takes on a value in the real interval $a \leq X \leq b$ is:

$$\boldsymbol{P}(a \leq X \leq b) = \int_a^b f_X(x)\, dx$$

Expectation

The expected value (or mean) of a RV with PDF f_X:

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

well defined only if $\int_{-\infty}^{\infty} |x| f_X(x)\,dx < \infty$

Expected Value Rule

Let X be a RV with PDF f_X. The function $g(X)$ is a RV with expectation:

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

Variance

The variance of a RV with PDF f_X:

$$\begin{aligned} var(X) &= \boldsymbol{E}\left[(X - \boldsymbol{E}[X])^2\right] \\ &= \int_{-\infty}^{\infty} (x - \boldsymbol{E}[X])^2 f_X(x)\,dx \\ &= \boldsymbol{E}\left[X^2\right] - (\boldsymbol{E}[X])^2 \end{aligned}$$

Cumulative Distribution Function

The cumulative distribution function (CDF) of a random variable, X, is denoted by F_X and represents the probability $\boldsymbol{P}(X \le x)$:

$$F_X(k) = \boldsymbol{P}(X \le x) = \begin{cases} \sum\limits_{k \le x} p_X(k) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{x} f_X(t)\,dt & \text{if } X \text{ is continuous} \end{cases}$$

The PDF, f_X, can be obtained from the CDF as:

$$f_X(x) = \frac{dF_X}{dx}(x)$$

for all x where F_x is continuous if X is continuous, and the PMF as:

$$p_x(k) = \boldsymbol{P}(X \le k) - \boldsymbol{P}(X \le k-1) = F_X(k) - F_X(k-1)$$

for all integers k if X is discrete.

Joint Distributions

For RVs X and Y with joint PDF $f_{X,Y}(x,y)$:

$$\boldsymbol{P}(a \le X \le b, c \le Y \le d) = \int_c^d \int_a^b f_{X,Y}(x,y)\,dxdy$$

Marginal PDFs

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

Joint CDF

$$F_{X,Y}(x,y) = \boldsymbol{P}(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s,t)\,dtds$$

Joint PDF recoverd from CDF as:

$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y)$$

Expectations

$$\boldsymbol{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dxdy$$

If g is linear, $aX + bY + c$

$$\boldsymbol{E}[aX + bY + c] = a\boldsymbol{E}[X] + b\boldsymbol{E}[Y] + c$$

Conditional PDF Given an Event

Let X be a continuous RV, given event A representing a subset of the real line, and conditional PDF, $f_{X|A}$ with $\boldsymbol{P}(X \in A) > 0$ then

$$f_{X|\{X \in A\}}(x) = \begin{cases} \frac{f_X(x)}{\boldsymbol{P}(X \in A)} & \text{if } X \in A \\ 0 & \text{otherwise} \end{cases}$$

Let $A_1, A_2, \ldots, A_n$ be a disjoint partition of the sample space, then

$$f_X(x) = \sum_{i=1}^{n} \boldsymbol{P}(A_i) f_{X|A_i}(x)$$

Conditional PDF Given a Random Variable

For jointly continuous RVs X and Y with PDF $f_{X,Y}$, the conditional PDF of X given $Y = y$ is:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \quad f(y) > 0$$

The marginal PDF of X is then

$$f_X(x) = \int_{-\infty}^{\infty} f_Y(y) f_{X|Y}(x|y)\,dy$$

Conditional Expectations

For jointly continuous RVs X and Y:
Conditional expectation of X given $Y = y$:

$$\boldsymbol{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx$$

Expected value rule

$$\boldsymbol{E}[g(X)|Y = y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)\,dx$$

Total expectation theorem

$$\boldsymbol{E}[X] = \int_{-\infty}^{\infty} \boldsymbol{E}[X|Y = y] f_Y(y)\,dy$$

Independence of Continuous RVs

Continuous RVs X and Y are independent if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) \quad \forall x, y$$

Independence implies CDFs are "independent" i.e.

$$F_{X,Y}(x,y) = F_X(x) F_Y(y)$$

and for expected values and variances:

$$\boldsymbol{E}[XY] = \boldsymbol{E}[X]\,\boldsymbol{E}[Y]$$
$$\boldsymbol{E}[g(X)h(Y)] = \boldsymbol{E}[g(X)]\,\boldsymbol{E}[h(Y)]$$
$$var(X+Y) = var(X) + var(Y)$$

Baye's Rule for Continuous RVs

Let Y be a continuous RV.
If X is a continuous RV:

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{f_Y(y)} = \frac{f_X(x) f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(t) f_{X|Y}(y|t)\,dt}$$

If N is a discrete RV:

$$\boldsymbol{P}(N = n|Y = y) = \frac{p_N(n) f_{Y|N}(y|n)}{f_Y(y)} = \frac{p_N(n) f_{Y|N}(y|n)}{\sum\limits_i p_N(i) f_{Y|N}(y|i)}$$

$$\begin{aligned} f_{Y|N}(y|n) &= \frac{f_Y(y)\,\boldsymbol{P}(N = n|Y = y)}{p_N(n)} \\ &= \frac{f_Y(y)\,\boldsymbol{P}(N = n|Y = y)}{\int_{\infty}^{\infty} f_Y(t)\,\boldsymbol{P}(N = n|Y = t)\,dt} \end{aligned}$$

Common Discrete RVs

In all cases, $k = 0, 1, 2, \ldots$

Bernoulli RV

Models probability of binary outcome experiment

$$p_X(k) = \begin{cases} p & k = 1 \\ 1-p & k = 0 \end{cases}$$
$$\boldsymbol{E}[X] = p$$
$$var(X) = p(1-p)$$

Binomial RV

Models probability of k successes in n independent binary outcome experiments with success probability p

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$
$$\boldsymbol{E}[X] = np$$
$$var(X) = np(1-p)$$

Geometric RV

Models probability of first success occurring on trial k for an infinite set of independent binary outcome experiments with success probability p

$$p_X(k) = p(1-p)^{k-1}$$
$$\boldsymbol{E}[X] = \frac{1}{p}$$
$$var(X) = \frac{1-p}{p^2}$$
$$F_X(k) = \sum_{k=1}^{n} p(1-p)^{k-1} = 1 - (1-p)^n$$

Poisson RV

An approximation to the binomial RV for large n and small p and $\lambda = np$

$$p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$
$$\boldsymbol{E}[X] = \lambda$$
$$var(X) = \lambda$$

Discrete Uniform RV

Models probability of equally likely events

$$p_X(k) = \begin{cases} \frac{1}{b-a+1} & k = a, a+1, \ldots, b \\ 0 & otherwise \end{cases}$$
$$\boldsymbol{E}[X] = \frac{a+b}{2}$$
$$var(X) = \frac{(b-a)(b-a+2)}{12}$$

Common Continuous RVs

Continuous Uniform RV

Models probability of equally likely real values

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & otherwise \end{cases}$$
$$\boldsymbol{E}[X] = \frac{a+b}{2}$$
$$var(X) = \frac{(b-a)^2}{12}$$

Exponential RV

Models the waiting time until an event of interest

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & otherwise \end{cases}$$
$$\boldsymbol{P}(X \ge a) = \int_a^\infty \lambda e^{-\lambda x} dx = e^{-\lambda a}$$
$$\boldsymbol{E}[X] = \frac{1}{\lambda}$$
$$var(X) = \frac{2}{\lambda^2}$$
$$F_X(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$$

Normal (aka Gaussian) RV

Models data symmetrically distributed around the expected value (i.e. the mean)

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$
$$\boldsymbol{E}[X] = \mu$$
$$var(X) = \sigma^2$$

Normality is Invariant under Linear Transformation

If X is a normal RV with mean μ and variance σ^2, and if $a \neq 0$, b are scalars, then the RV

$$Y = aX + b$$

is normal with mean $a\mu + b$ and variance $a^2\sigma^2$

Standard Normal RV

A normal RV, Y with $mu = 0$ and $\sigma^2 = 1$ is said to be **standard normal** with CDF, Φ:

$$\Phi(y) = \boldsymbol{P}(Y < y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} \exp\left(-t^2/2\right) dt$$

Note: $\Phi(-y) = 1 - \Phi(y)$

Standardizing a Normal RV

Any normal variable, X with mean μ and variance σ^2 can be "standardized" by defining

$$Y = \frac{X-\mu}{\sigma}$$

where Y has zero mean and unit variances. Then

$$\boldsymbol{P}(X \le x) = \boldsymbol{P}\left(Y \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$