Events and Probability

Event

Set of outcomes from an experiment.

Sample Space

Set of all possible outcomes Ω.

Intersection

Outcomes occur in both A and B

$$A \cap B \qquad \text{or} \qquad AB$$

Disjoint

No common outcomes

$$AB = \varnothing$$

$$\Pr(AB) = 0 \implies \Pr(\varnothing) = 0$$

$$\Pr(A \mid B) = 0$$

These events are dependent.

Union

Set of outcomes in either A or B

$$A \cup B$$

Complement

Set of all outcomes not in A, but in Ω — $\overline{A} = \Omega \backslash A$.

$$A\overline{A} = \varnothing$$

$$A \cup \overline{A} = \Omega$$

Subset

A is a (non-strict) subset of B if all elements in A are also in B — $A \subset B$.

$$AB = A \qquad \text{and} \qquad A \cup B = B$$

$$\forall A : A \subset \Omega \wedge \varnothing \subset A$$

$$\Pr(A) \leq \Pr(B)$$

$$\Pr(B \mid A) = 1$$

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

Identities

$$A(BC) = (AB)C$$

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A(B \cup C) = AB \cup AC$$

$$A \cup BC = (A \cup B)(A \cup C)$$

Probability

Measure of the likeliness of an event occurring

$$\Pr(A) \qquad \text{or} \qquad \mathrm{P}(A)$$

$$0 \leq \Pr(E) \leq 1$$

where a probability of 0 never happens, and 1 always happens.

$$\Pr(\Omega) = 1$$

$$\Pr\left(\overline{E}\right) = 1 - \Pr(E)$$

Multiplication Rule

For independent events A and B

$$\Pr(AB) = \Pr(A)\Pr(B).$$

For dependent events A and B

$$\Pr(AB) = \Pr(A \mid B)\Pr(B)$$

Addition Rule

For independent A and B

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(AB).$$

If $AB = \varnothing$, then $\Pr(AB) = 0$, so that $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

De Morgan's Laws

$$\overline{A \cup B} = \overline{A}\,\overline{B}$$

$$\overline{AB} = \overline{A} \cup \overline{B}.$$

$$\Pr(A \cup B) = 1 - \Pr\left(\overline{A}\,\overline{B}\right)$$

$$\Pr(AB) = 1 - \Pr\left(\overline{A} \cup \overline{B}\right)$$

Conditional probability

The probability of event A given B has already occurred

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

A and B are independent events if

$$\Pr(A \mid B) = \Pr(A)$$

$$\Pr(B \mid A) = \Pr(B)$$

the following statements are also true

$$\Pr\left(A \mid \overline{B}\right) = \Pr(A)$$

$$\Pr\left(\overline{A} \mid B\right) = \Pr\left(\overline{A}\right)$$

$$\Pr\left(\overline{A} \mid \overline{B}\right) = \Pr\left(\overline{A}\right)$$

Probability Rules with Conditional

All probability rules hold when conditioning on another event C.

$$\Pr\left(\overline{A} \mid C\right) = 1 - \Pr(A \mid C)$$

$$\Pr(A \cup B \mid C) = \Pr(A \mid C) + \Pr(B \mid C) - \Pr(AB \mid C)$$

$$\Pr(AB \mid C) = \Pr(A \mid BC)\Pr(B \mid C)$$

Conditional Independence

Given $\Pr(A \mid B) \neq \Pr(A)$ A and B are conditionally dependent given C if

$$\Pr(A \mid BC) = \Pr(A \mid C).$$

Futhermore

$$\Pr(AB \mid C) = \Pr(A \mid C)\Pr(B \mid C).$$

Conversely

$$\Pr(A \mid B) = \Pr(A)$$

$$\Pr(A \mid BC) \neq \Pr(A \mid C)$$

$$\Pr(AB \mid C) = \Pr(A \mid BC)\Pr(B \mid C)$$

Pairwise independence does not imply mutual independence

$$\begin{cases} \Pr(AB) = \Pr(A)\Pr(B) \\ \Pr(AC) = \Pr(A)\Pr(C) \\ \Pr(BC) = \Pr(B)\Pr(C) \end{cases} \not\Rightarrow$$

$$\Pr(ABC) = \Pr(A)\Pr(B)\Pr(C).$$

Independence should not be assumed unless explicitly stated.

Marginal Probability

The probability of an event irrespective of the outcome of another variable.

Total Probability

$A = AB \cup A\overline{B}$

$$\Pr(A) = \Pr(AB) + \Pr\left(A\overline{B}\right)$$

$$\Pr(A) = \Pr(A \mid B)\Pr(B) + \Pr\left(A \mid \overline{B}\right)\Pr\left(\overline{B}\right)$$

In general, partition Ω into disjoint events B_1, B_2, …, B_n, such that $\bigcup_{i=1}^{n} B_i = \Omega$

$$\Pr(A) = \sum_{i=1}^{n} \Pr(A \mid B_i)\Pr(B_i)$$

Bayes' Theorem

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

Combinatorics

Number of outcomes

Let $|A|$ denote the number of outcomes in an event A.
For k disjoint events $\{S_1, \ldots, S_k\}$ where the ith event has n_i possible outcomes,

Addition principle

Number of possible samples from any event

$$\left|\bigcup_{i=0}^{k} S_i\right| = \sum_{i=1}^{k} n_i$$

Multiplication principle

Number of possible samples from every event

$$\left|\bigcap_{i=0}^{k} S_i\right| = \prod_{i=1}^{k} n_i$$

Counting probability

If S_i has equally likely outcomes

$$\Pr(S_i) = \frac{|S_i|}{|S|}$$

Ordered Sampling with Replacement

Number of ways to choose k objects from a set with n elements

$$n^k$$

Ordered Sampling without Replacement

Number of ways to arrange k objects from a set of n elements, or the k-permutation of n-elements

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

for $0 \leq k \leq n$.
An n-permutation of n elements is the permutation of those elements.

$${}^{n}P_n = n!$$

Unordered Sampling without Replacement

Number of ways to choose k objects from a set of n elements, or the k-combination of n-elements

$${}^{n}C_k = \frac{{}^{n}P_k}{k!} = \frac{n!}{k!\,(n-k)!}$$

for $0 \leq k \leq n$.

Unordered Sampling with Replacement

Number of ways to choose k objects from a set with n elements

$$\binom{n+k-1}{k}$$

Distribution	Restrictions	PMF	CDF	$\mathrm{E}(X)$	$\mathrm{Var}(X)$
$X \sim \mathrm{Uniform}(a,\ b)$	$x \in \{a, \ldots, b\}$	$\frac{1}{b-a+1}$	$\frac{x-a+1}{b-a+1}$	$\frac{a+b}{2}$	$\frac{(b-a+1)^2-1}{12}$
$X \sim \mathrm{Bernoulli}(p)$	$p \in [0,1], x \in \{0,1\}$	$p^x (1-p)^{1-x}$	$1-p$	p	$p(1-p)$
$X \sim \mathrm{Binomial}(n,\ p)$	$x \in \{0, \ldots, n\}$	$\binom{n}{x} p^x (1-p)^{n-x}$	$\sum_{u=0}^{x} \binom{n}{u} p^u (1-p)^{n-u}$	np	$np(1-p)$
$N \sim \mathrm{Geometric}(p)$	$n \geq 1$	$(1-p)^{n-1} p$	$1-(1-p)^n$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$Y \sim \mathrm{Geometric}(p)$	$y \geq 0$	$(1-p)^y p$	$1-(1-p)^{y+1}$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
$N \sim \mathrm{NB}(k,\ p)$	$n \geq k$	$\binom{n-1}{k-1} (1-p)^{n-k} p^k$	$\sum_{u=k}^{n} \binom{u-1}{k-1} (1-p)^{u-k} p^k$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$
$Y \sim \mathrm{NB}(k,\ p)$	$y \geq 0$	$\binom{y+k-1}{k-1} (1-p)^y p^k$	$\sum_{u=0}^{y} \binom{u+k-1}{k-1} (1-p)^u p^k$	$\frac{k(1-p)}{p}$	$\frac{k(1-p)}{p^2}$
$N \sim \mathrm{Poisson}(\lambda)$	$n \geq 0$	$\frac{\lambda^n e^{-\lambda}}{n!}$	$e^{-\lambda} \sum_{u=0}^{n} \frac{\lambda^u}{u!}$	λ	λ

Distribution	Restrictions	PMF	CDF	$\mathrm{E}(X)$	$\mathrm{Var}(X)$
$X \sim \mathrm{Uniform}(a,\ b)$	$a < x < b$	$\frac{1}{b-a}$	$\frac{x-a}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$T \sim \mathrm{Exp}(\eta)$	$t > 0$	$\eta e^{-\eta t}$	$1-e^{-\eta t}$	$\frac{1}{\eta}$	$p(1-p)$
$X \sim \mathrm{N}(\mu,\ \sigma^2)$	$x \in \{0, \ldots, n\}$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\frac{1}{2}\left(1+\mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right)$	μ	σ^2

Univariate X	Discrete	Continuous
Valid probabilities	$0 \leq p_x \leq 1$	$f(x) \geq 0$
Cumulative probability	$\sum_{u \leq x} p_u$	$\int_{-\infty}^{x} f(u)\,\mathrm{d}u$
$\mathrm{E}(X)$	$\sum_{\Omega} x p_x$	$\int_{\Omega} x f(x)\,\mathrm{d}x$
$\mathrm{Var}(X)$	$\sum_{\Omega} (x-\mu)^2 p_x$	$\int_{\Omega} (x-\mu)^2 f(x)\,\mathrm{d}x$

Random Variables

Measurable variable whose value holds some uncertainty. An event is when a random variable assumes a certain value or range of values.

Probability distribution

The probability distribution of a random variable X is a function that links all outcomes $x \in \Omega$ to the probability that they will occur $\Pr(X = x)$.

Probability mass function

$$\Pr(X = x) = p_x$$

Probability density function

$$\Pr(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)\,\mathrm{d}x$$

Cumulative distribution function

Computes the probability that the random variable is less than or equal to a particular realisation x.
$F(x)$ is a valid CDF if:

1. F is monotonically increasing and continuous
2. $\lim_{x \to -\infty} F(x) = 0$
3. $\lim_{x \to \infty} F(x) = 1$

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x} \int_{-\infty}^{x} f(u)\,\mathrm{d}u = f(x)$$

Complementary CDF (survival)

$$\Pr(X > x) = 1 - \Pr(X \leq x) = 1 - F(x)$$

p-Quantile

$$F(x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u = p$$

Median

$$\int_{-\infty}^{m} f(u)\,\mathrm{d}u = \int_{m}^{\infty} f(u)\,\mathrm{d}u = \frac{1}{2}$$

Lower and upper quartile

$$\int_{-\infty}^{q_1} f(u)\,\mathrm{d}u = \frac{1}{4}$$

and

$$\int_{-\infty}^{q_2} f(u)\,\mathrm{d}u = \frac{3}{4}$$

Quantile function

$$x = F^{-1}(p) = Q(p)$$

Expectation (mean)

Expected value given an infinite number of observations. For $a < c < b$:

$$\mathrm{E}(X) = -\int_{a}^{c} F(x)\,\mathrm{d}x + \int_{c}^{b} (1 - F(x))\,\mathrm{d}x + c$$

Variance

Measure of spread of the distribution (average squared distance of each value from the mean).

$$\mathrm{Var}(X) = \sigma^2$$

Variance is also denoted as σ^2.

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

Standard deviation

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

Uniform Distribution

Single trial X in a set of equally likely elements.

Bernoulli (binary) Distribution

Boolean-valued outcome X, i.e., success (1) or failure (0). $(1-p)$ is sometimes denoted as q.

Binomial Distribution

Number of successes X for n independent trials with the same probability of success p.

$$X = Y_1 + \cdots + Y_n$$

$$Y_i \overset{\text{iid}}{\sim} \mathrm{Bernoulli}(p) : \forall i \in \{1,\ 2,\ \ldots,\ n\}.$$

Geometric Distribution

Number of trials N up to and including the first success, where each trial is independent and has the same probability of success p.

Alternate Geometric

Number of failures $Y = N - 1$ until a success.

Negative Binomial Distribution

Number of trials N until $k \geq 1$ successes, where each trial is independent and has the same probability of success p.

$$N = Y_1 + Y_2 + \cdots + Y_k$$

$$Y_i \overset{\text{iid}}{\sim} \mathrm{Geom}(p) : \forall i \in \{1,\ 2,\ \ldots,\ k\}.$$

Alternate Negative Binomial

Number of failures $Y = N - k$ until k successes:

Poisson Distribution

Number of events N which occur over a fixed interval of time λ.

Modelling Count Data

- Poisson (mean = variance)
- Binomial (underdispersed, mean > variance)
- Geometric/Negative Binomial (overdispersed, mean < variance)

Uniform Distribution

Outcome X within some interval, where the probability of an outcome in one interval is the same as all other intervals of the same length.

$$m = \frac{a+b}{2}$$

Exponential Distribution

Time T between events with rate η.

$$m = \frac{\ln(2)}{\eta}$$

Memoryless Property

For $T \sim \mathrm{Exp}(\lambda)$:

$$\Pr(T > s + t \mid T > t) = \Pr(T > s)$$

For $N \sim \mathrm{Geometric}(p)$:

$$\Pr(N > s + n \mid N > n) = \Pr(N > s)$$

Normal Distribution

Used to represent random situations, i.e., measurements and their errors. Also used to approximate other distributions.

Standard Normal Distribution

Given $X \sim \mathrm{N}(\mu, \sigma^2)$, consider

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

so that $Z \sim \mathrm{N}(0, 1)$.

Central Limit Theorem

The sum of independent and identically distributed random variables, when properly standardised, can be approximated by a normal distribution, as $n \to \infty$.

Let $X_1, \dots, X_n \overset{iid}{\sim} X$, and $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$

Average of Random Variables

If $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$:

$$\mathrm{E}(\overline{X}) = \mu$$

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

By standardising $\overline{X}$, we can define

$$Z = \lim_{n\to\infty} \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

so that $Z \to \mathrm{N}(0, 1)$ as $n \to \infty$.

Sum of Random Variables

If $\overline{Y} = \sum_{i=1}^{n} X_i$:

$$\mathrm{E}(Y) = n\mu$$

$$\mathrm{Var}(Y) = n\sigma^2$$

$$Y \sim \mathrm{N}(n\mu, n\sigma^2), \quad n \to \infty$$

Binomial Approximations

If $X \sim \mathrm{Binomial}(n, p)$:

$$X \approx Y \sim \mathrm{N}(np, np(1-p))$$

Sufficient for $np > 5$ and $n(1-p) > 5$.

If $np < 5$:

$$X \approx Y \sim \mathrm{Pois}(np).$$

If $n(1-p) < 5$, consider the number of failures $W = n - X$:

$$W \approx Y \sim \mathrm{Pois}(n(1-p)).$$

Continuity Correction

$$\Pr(X \le x) = \Pr(X < x + 1)$$

must hold for any x. Therefore

$$\Pr(X \le x) \approx \Pr\left(Y \le x + \frac{1}{2}\right).$$

Poisson Approximation

If $X_i \sim \mathrm{Poisson}(\lambda)$:

Let $X = \sum_{i=1}^{n} X_i$:

$$\mathrm{E}(X) = n\lambda$$

$$\mathrm{Var}(X) = n\lambda$$

$$X \approx Y \sim \mathrm{N}(n\lambda, n\lambda).$$

Sufficient for $n\lambda > 10$, and for accurate approximations, $n\lambda > 20$.

Bivariate Distributions

Bivariate probability mass function

Distribution over the joint space of two discrete random variables X and Y:

$$\Pr(X = x, Y = y) = p_{x,y} \ge 0$$

$$\sum_{y\in\Omega_2}\sum_{x\in\Omega_1} \Pr(X = x, Y = y) = 1$$

for all pairs of $x \in \Omega_1$ and $y \in \Omega_2$. The joint probability mass function can be shown using a table:

	y_1	$\cdots$	y_n
x_1	$p_{1,1}$	$\cdots$	$p_{1,n}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_n	$p_{n,1}$	$\cdots$	$p_{n,n}$

Bivariate probability density function

Distribution over the joint space of two continuous random variables X and Y:

$$\Pr(x_1 \le X \le x_2, y_1 \le Y \le y_2) = \int_{x_1}^{x_2}\int_{y_1}^{y_2} f(x, y)\,\mathrm{d}y\,\mathrm{d}x$$

This function must satisfy

$$f(x, y) \ge 0$$

$$\int_{x\in\Omega_1}\int_{y\in\Omega_2} f(x, y)\,\mathrm{d}y\,\mathrm{d}x = 1.$$

for all pairs of $x \in \Omega_1$ and $y \in \Omega_2$.

Joint Probability

$$\Pr(X = x, Y = y) = \Pr(X = x \mid Y = y)\Pr(Y = y)$$

Marginal Probability

Probability function of each random variable. Must specify the range of values that variable can take.

Marginal probability mass function

$$p_x = \sum_{y\in\Omega_2} \Pr(X = x, Y = y)$$

$$p_y = \sum_{x\in\Omega_1} \Pr(X = x, Y = y)$$

Marginal probability density function

$$f(x) = \int_{y_1}^{y_2} f(x, y)\,\mathrm{d}y$$

$$f(y) = \int_{x_1}^{x_2} f(x, y)\,\mathrm{d}x$$

Conditional probability mass function

$$\Pr(X = x \mid Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)}$$

$$\sum_{x\in\Omega_1} \Pr(X = x \mid Y = y) = 1$$

Conditional probability density function

$$f(x \mid y) = \frac{f(x, y)}{f(y)}$$

$$\int_{x_1}^{x_2} f(x \mid y)\,\mathrm{d}x = 1$$

Independence

Two discrete random variables X and Y are independent if

$$\Pr(X = x \mid Y = y) = \Pr(X = x)$$

for all pairs of x and y.

Two continuous random variables X and Y are independent if

$$f(x, y) \propto g(x)\,h(y)$$

so that

$$f(x \mid y) = f(x).$$

Conditional Expectation

Given the conditional distribution $X \mid Y = y$ for discrete random variables:

$$\mathrm{E}(X \mid Y = y) = \sum_{x\in\Omega_1} x p_{x\mid y}$$

For continuous random variables:

$$\mathrm{E}(X \mid Y = y) = \int_{x_1}^{x_2} x f(x \mid y)\,\mathrm{d}x$$

Conditional Variance

$$\mathrm{Var}(X \mid Y = y) = \mathrm{E}(X^2 \mid Y = y) - \mathrm{E}(X \mid Y = y)^2$$

Law of Total Expectation

By treating $\mathrm{E}(X \mid Y)$ as a random variable of Y:

$$\mathrm{E}(X) = \mathrm{E}(\mathrm{E}(X \mid Y))$$

Joint expectation

$$\mathrm{E}(XY) = \sum_{x\in\Omega_1}\sum_{y\in\Omega_2} xy p_{x,y}$$

$$\mathrm{E}(XY) = \int_{x_1}^{x_2}\int_{x_1}^{x_2} xy f(x, y)\,\mathrm{d}y\,\mathrm{d}x.$$

Transformation rules

$$\mathrm{E}(aX \pm b) = a\,\mathrm{E}(X) \pm b$$

$$\mathrm{E}(X \pm Y) = \mathrm{E}(X) \pm \mathrm{E}(Y)$$

$$\mathrm{Var}(aX \pm b) = a^2\,\mathrm{Var}(X)$$

$$\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\,\mathrm{Cov}(X, Y)$$

$$\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y)$$

If X and Y are independent:

$$\mathrm{E}(X \mid Y = y) = \mathrm{E}(X)$$

$$\mathrm{Var}(X \mid Y = y) = \mathrm{Var}(X)$$

$$\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$$

$$\mathrm{Var}(XY) = \mathrm{Var}(X)\,\mathrm{Var}(Y) + \mathrm{E}(X)^2\,\mathrm{Var}(Y) + \mathrm{E}(Y)^2\,\mathrm{Var}(X)$$

for constants a, b, c, and d.

Covariance

Measure of the dependence between two random variables

$$\begin{aligned}\mathrm{Cov}(X,\,Y) &= \mathrm{E}\left((X - \mathrm{E}(X))(Y - \mathrm{E}(Y))\right) \\ &= \mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)\end{aligned}$$

Describes the direction of a relationship, but does not quantify the strength of such a relationship.

The covariance of X and Y is:

Positive if an increase in one variable is more likely to result in an increase in the other variable.

Negative if an increase in one variable is more likely to result in a decrease in the other variable.

Zero if X and Y are independent. Note that the converse is not true.

Correlation

Explains both the direction and strength of a linear relationship between two random variables.

$$\rho(X,\,Y) = \frac{\mathrm{Cov}(X,\,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

where $-1 \leq \rho(X,\,Y) \leq 1$.

These value can be interpretted as follows:

- $\rho(X,\,Y) > 0$ iff X and Y have a positive linear relationship.
- $\rho(X,\,Y) < 0$ iff X and Y have a negative linear relationship.
- $\rho(X,\,Y) = 0$ if X and Y are independent. Note that the converse is not true.
- $\rho(X,\,Y) = 1$ iff X and Y have a perfect linear relationship with positive slope.
- $\rho(X,\,Y) = -1$ iff X and Y have a perfect linear relationship with negative slope.

The slope of a perfect linear relationship cannot be obtained from the correlation.