1 Probability

Counting

If an experiment has n outcomes, and another experiment has m outcomes then the two experiments jointly have $n \times m$ outcomes.

Permutations

Define $H = \{h_1, \dots, h_n\}$ to be a set of n different objects. The permutations of H are the different orders in which you can write all of its elements.

$$n!$$

$0! = 1$ (special case)

Permutations with Repetitions

k-Permutations

Let $H = \{h_1, h_2, \dots, h_n\}$ be a set of n different objects. The k-permutationsof H are the different ways in which one can pick and write k of its elements **in order**.

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

k-Permutations with Repetitions

k-Combinations

Let $H = \{h_1, h_2, \dots, h_n\}$ be a set of n different objects. The k-combinations of H are the different ways in which one can pick and write k of its elements **without order**.

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

("n choose k")
Note that this is just the k-permutations divided by $k!$.

Events

A mathematical model for experiments:

- Sample space: Ω (set of all possible outcomes)
- An event is a collection of possible outcomes $E \subseteq \Omega$

We can use sets and subsets and logic to represent events.

Axioms of Probability

The probability P on a sample space Ω assigns numbers to events of Ω in such a way that:

1. The probability of an event is non-negative (i.e. $P(E) \geq 0$)
2. The probability of the entire sample space is 1 (i.e. $P(\Omega) = 1$)
3. For countable many mutually exclusive events $E_1, E_2, \dots$:

$$P\left(\bigcup_i E_i\right) = \sum_i P(E_i)$$

Proposition

For any event: $P(\bar{E}) = 1 - P(E)$

Corollary

We have that $P(\emptyset) = P(\bar{\Omega}) = 1 - P(\Omega) = 0$
For any event, $P(E) = 1 - P(\bar{e}) \leq 1$

Proposition

For any two events:

$$\begin{aligned} &P(E \cup F) \\ &= P(E) + P(F) - P(E \cap F) \end{aligned}$$

Boole's Inequality

For any events $E_1, E_2, \dots, E_n$:

$$P\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} P(E_i)$$

Inclusion-Exclusion Principle

For any events E, F, and G:

$$\begin{aligned} &P(E \cup F \cup G) \\ &= P(E) + P(F) + P(G) - P(E \cap F) \\ &- P(E \cap G) - P(F \cap G) \\ &+ P(E \cap F \cap G) \end{aligned}$$

Proposition

If $E \subseteq F$, then $P(F - E) = P(F) - P(E)$.

Corollary

If $E \subseteq F$, then $P(E) \leq P(F)$.

Equally Likely Outcomes

If all outcomes are equally likely, then the probability of any event is the number of outcomes in the event divided by the number of outcomes in the sample space.

$$P(w) = \frac{1}{\|\Omega\|} \forall w \in \Omega$$

Conditional Probability

Let F be an event with $P(F) > 0$. The conditional probability of E given F is:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Axioms of Conditional Probability

1. Conditional probability is non-negative: $P(E|F) \geq 0$
2. Conditional probability of sample space is one: $P(\Omega|F) = 1$
3. For countably many mutually exclusive events $E_1, E_2, \dots$:

$$P\left(\bigcup_i E_i | F\right) = \sum_i P(E_i|F)$$

Corollary

1. $P(\bar{E}|F) = 1 - P(E|F)$
2. $P(\emptyset|F) = 0$
3. $P(E|F) = 1 - P(\bar{E}|F) \leq 1$
4. $P(E \cup G|F) = P(E|F) + P(G|F) - P(E \cap G|F)$
5. If $E \subseteq G$, then $P(G - E|F) = P(G|F) - P(E|F)$
6. If $E \subseteq G$, then $P(E|F) \leq P(G|F)$

Note: don't change the condition. $P(E|F)$ and $P(E|\bar{F})$ have nothing to do with each other.

Multiplication Rule

$$\begin{aligned} &P(E_1 \cup \dots \cup E_n) \\ &= P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2) \\ &\dots P(E_n|E_1 \cap \dots \cap E_{n-1}) \end{aligned}$$

Bayes' Theorem

Partition Theorem

$$\begin{aligned} P(E) &= P(E|F)P(F) \\ &+ P(E|\bar{F})P(\bar{F}) \end{aligned}$$

Bayes' Theorem

$$\begin{aligned} &P(F|E) \\ &= \frac{P(E|F)P(F))}{P(E|F)P(F)) + P(E|\bar{F}P(\bar{F}))} \end{aligned}$$

Independence

Two events E and F are independent if:

$$P(E \cap F) = P(E)P(F)$$
$$P(E|F) = P(E)$$
$$P(F|E) = P(F)$$

Three events E, F, and G are independent if:

$$P(E \cup F) = P(E) + P(F)$$
$$P(E \cup G) = P(E) + P(G)$$
$$P(F \cup G) = P(F) + P(G)$$
$$P(E \cup F \cup G) = P(E) + P(F) + P(G)$$

Proposition

If E and F are independent events, then E and $\bar{F}$ are also independent.

$$\text{independent} \neq \text{mutually exclusive}$$

2 Discrete Probability

Random Variables

A random variable is a function from the sample space Ω to the real numbers $\mathbb{R}$.
A random variable X is discrete if it takes on a finite or countable number of values.

Probability Mass Function

The probability mass function (PMF) or distribution of a discrete random variable X gives the probabilities of its possible values.

$$P(X = x) \geq 0$$

(all probabilities are non-negative)

$$\sum_i P(X = x_i) = 1$$

(the probabilities sum to 1)

Cumulative Distribution Function

The cumulative distribution function (CDF) of a discrete random variable X gives the probability that X is less than or equal to x.

$$F : \mathbb{R} \to [0, 1]$$
$$F(x) = P(X \leq x)$$

Similar to PDF graph except it adds them up (cumulative)

$$\begin{aligned} &P(a < x \leq b) \\ &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a) \end{aligned}$$

A cumulative distribution function F:

- is non-decreasing: $F(x) \leq F(y)$ for all $x \leq y$
- has limit 0: $F(-\infty) = 0$ on the left
- has limit 1: $F(\infty) = 1$ on the right

Expected Value

The expected value of a discrete random variable X is the average value of X.

$$E(X) = \sum_i x_i P(X = x_i)$$

(provided the sum exists)
Note: Expected value need not be a possible value of X.
For $g : \mathbb{R} \to \mathbb{R}$:

$$\begin{aligned} &E(g(X)) \\ &= \sum_i g(x_i) P(X = x_i) \end{aligned}$$

(provided the sum exists)
Expectation is linear, so $E(aX+b) = aE(X)+b$ for any constants a and b.

Variance

The variance tells us how surprised we should be if we observe a value of X.

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

Standard Deviation

The standard deviation is the square root of the variance.

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

Bernoulli and Binomial Distributions

$$X \sim \text{Binom}(n, p)$$

X has the Binomial distribution with parameters n and p if, for n independent trials, each succeeding with probability p, the random variable X counts the number of successes within the n trials.
Special case $n = 1$ is called the Bernoulli distribution with parameter p. In this case, X is 1 if the trial succeeds and 0 if it fails (indication variable).

PMF

Let $X \sim \text{Binom}(n, p)$ and $X = 0, 1, \ldots, n$. Then:

$$\begin{aligned} &P(X = k) \\ &= \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

The Bernouilli(p) distribution can take on values 0 or 1 with properties:

$$P(X = 0) = 1 - p$$
$$P(X = 1) = p$$

Newton's Binomial Theorem:

$$\begin{aligned} &\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} \\ &= (a + b)^n \end{aligned}$$

Expected Value

$$E(X) = np$$

Variance

$$\text{Var}(X) = np(1-p)$$

Poisson Distribution

$$X \sim \text{Poisson}(\lambda)$$

The random variable X is Poisson distributed with parameter λ if λ is non-negative integer valued and its mass function is:

$$P(X = k) = e^{-\lambda} \times \frac{\lambda^i}{i!}$$

Poisson Approximation to Binomial

Take $Y \sim \text{Binom}(n, p)$ with large n and small p, such that $np \approx \lambda$. Then Y is approximately Poisson(λ) distributed.

Expected Value and Variance

$$E(X) = \text{Var}(X) = \lambda$$

... since Binomial expectation and variance are np and $np(1-p)$ which both converge to λ for large n.

Geometric Distribution

When is the first success?

$$X \sim \text{Geom}(p)$$

Suppose that independent trials, each succeeding with probability p, are repeated until the first success. The total number X of trials made has the Goemetric(p) distribution.
X can take on positive integers, with probabilities:

$$P(X = i) = (1-p)^{i-1} p$$

The Goemetric random variable is (discrete) memoryless:

$$\begin{aligned} &P(X > n + k | X > n) \\ &= P(X > k) \end{aligned}$$

... for every $k \geq 1$, $n \geq 0$.

Expectation and Variance

$$E(X) = \frac{1}{p}$$

$$\text{Var}(X) = \frac{1-p}{p^2}$$

3 Conditional Probability

Continuous Random Variables

A continuous random variable is one that takes values over a continuous range.
A continuous random variable X must have the property that $P(X = x) = 0 \forall x \in \mathbb{R}$.
(This only applies to individual values, ranges may have non-zero probabilities).

Probability Density Function

The probability density function (PDF) of a continuous random variable X is a function $f(x)$ such that for any two numbers $a \leq b$ we have the following:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

For any PDF we know that $f(x) \geq 0$ for all values of x and the total area under the whole graph is 1:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

Uniform Distribution

A continuous random variable X has uniform distribution on the interval $[a, b]$ for values $a \leq b$ if the PDF is given by:

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

We write this as $X \sim \text{Unif}(a, b)$.

Cumulative Distribution Function

For a continuous random variable X with PDF $f(x)$, the cumulative distribution function (CDF) is given by:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

For any number x, $F(x)$ is the probability that the observed value of X will be no more than x.
If X is a continuous random variable with PDF $f(x)$ and CDF $F(x)$ then at every x where the derivative $F'(x)$ is defined we have:

$$F'(x) = f(x)$$

For any value a we have:

$$P(X \leq a) = F(a)$$

$$P(X > a) = 1 - F(a)$$

... and for any two values $a < b$ we have:

$$P(a \leq X \leq b) = F(b) - F(a)$$

Conversion between PDF and CDF gives different ways to calculate the probabilities involved.

Percentiles of Continuous Distributions

Let X be a continuous random variable with PDF $f(x)$ and CDF $F(x)$ and p any real value between 0 and 1.
The $(100p)$th percentile of X is the value η_p such that $P(X \leq \eta_p) = p$.
So we have:

$$p = \int_{-\infty}^{\eta_p} f(x)dx = F(\eta_p)$$

and

$$\eta_p = F^{-1}(p)$$

Expected Value

Let X be a continuous random variable with PDF $f(x)$. The expected value $E(x)$ is calculated as a weighted integral:

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

This is also known as the mean of the distribution and written as μ_X or simply μ.

Proposition

Let X be a continuous random variable with PDF $f(x)$. If $h(x)$ is any real-valued functio nof X then we can calculate an expected value for that, too:

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

Note: $E(h(x))$ does not necessarily equal $h(E(x))$.

Variance and Standard Deviation

Let X be a continuous random variable with PDF $f(x)$ and mean μ. Its variance $\text{Var}(X)$ is the expected value of the squared distance to the mean.

$$\begin{aligned} &\text{Var}(X) \\ &= E((X-\mu)^2) \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \end{aligned}$$

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

Properties

Variance Shortcut:

$$\begin{aligned} &\text{Var}(X) \\ &= E(X^2) - \mu^2 \\ &= \int_{-\infty}^{\infty} x^2 f(x)dx - \left(\int_{-\infty}^{\infty} x f(x)dx\right)^2 \end{aligned}$$

Chebyshev's Inequality:
For any constant value $k \geq 1$, the probability that X is more than k standard deviations away from the mean is no more than $\frac{1}{k^2}$.

$$P(|X - \mu| \geq k\text{SD}(X)) \leq \frac{1}{k^2}$$

Linearity of Expectation:
For any functions $h_1(x)$ and $h_2(x)$ and constants a_1, a_2 and b, the expected values of these in linear combinations is the linear combination of the expected values.

$$\begin{aligned} &E(a_1 h_1(X) + a_2 h_2(X) + b) \\ &= a_1 E(h_1(X)) + a_2 E(h_2(X)) + b \end{aligned}$$

Rescaling:
For any constants a and b, the mean, variance and standard deviation of $(aX + b)$ can be calculated from the corresponding values for X:

$$E(aX + b) = aE(X) + b$$

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

$$\text{SD}(aX + b) = |a|\text{SD}(X)$$