

Discrete probability

Sample spaces

For any probabilistic experiment or process, the set Ω of all its possible outcomes is called its **sample space**.

E.g. Consider the following probabilistic (random) experiment:

Flip a fair coin 7 times in a row, and see what happens

The possible outcomes of the example above, are all the sequences of Heads (H) and Tails (T), of length 7. In other words, they are the set of strings $\Omega = \{H, T\}^7$.

The cardinality of this sample space is $|\Omega| = 2^7$, because there are two choices for each of the seven characters of the string.

Sample spaces may be infinite.

E.g. **Flip a fair coin repeatedly until a Heads is flipped.**

$\Omega = \{H, TH, TTH, TTTH, TTTTH, \dots\}$

However, in **discrete probability**, we focus on finite and countable sample spaces.

Probability distributions

A **probability distribution** over a finite or countable set Ω , is a function:

$$P : \Omega \rightarrow [0, 1]$$

such that $\sum_{s \in \Omega} P(s) = 1$.

In other words, each outcome $s \in \Omega$ is assigned a probability $P(s)$ such that $0 \leq P(s) \leq 1$ and the sum of the probabilities of all outcomes sum to 1.

E.g. Suppose a fair coin is tossed 7 times consecutively. This random experiment defines a probability distribution $P : \Omega \rightarrow [0, 1]$ on $\Omega = \{H, T\}^7$, where for all $s \in \Omega$, $P(s) = \frac{1}{2^7}$.

Since $|\Omega| = 2^7$, we have $\sum_{s \in \Omega} P(s) = 2^7 \cdot \frac{1}{2^7} = 1$

E.g. suppose a fair coin is tossed repeatedly until it lands heads. This random experiment defines a probability distribution $P : \Omega \rightarrow [0, 1]$, on $\Omega = T^*H$, such that, $\forall k \geq 0$,

$$P(T^k H) = \frac{1}{2^{k+1}}$$

Note that $\sum_{s \in \Omega} P(s) = P(H) + P(TH) + P(TTH) + \dots = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1$

Events

For a **countable** sample space Ω , and **event** E , is simply a subset $E \subseteq \Omega$ of the set of possible outcomes.

Given a probability distribution $P : \Omega \rightarrow [0, 1]$, we define **the probability of the event** $E \subseteq \Omega$ to be $P(E) = \sum_{s \in E} P(s)$.

E.g. For $\Omega = \{H, T\}^7$, the following are events:

- The third coin toss came up heads

$$E_1 = \{H, T\}^2 H \{H, T\}^4$$

$$P(E_1) = \frac{1}{2}$$

- The fourth and fifth coin tosses did not both come up tails

$$E_2 = \Omega - \{H, T\}^3 TT \{H, T\}^2$$

$$P(E_2) = 1 - \frac{1}{4} = \frac{3}{4}$$

E.g. For $\Omega = T^* H$, the following are events:

- The first time the coin comes up heads is after an even number of coin tosses

$$E_3 = \{T^k H \mid k \text{ is odd}\}$$

$$P(E_3) = \sum_{k=1}^{\infty} \frac{1}{2^{2k}} = \frac{1}{3}$$

Basic facts about probabilities of events

For event $E \subseteq \Omega$, define the **complement event** to be $\bar{E} = \Omega - E$.

Theorem

Suppose E_0, E_1, E_2, \dots are a (finite or countable) sequence of pairwise disjoint events from the sample space Ω . In other words, $E_i \in \Omega$ and $E_i \cap E_j = \emptyset$, $\forall i, j \in \mathbb{N}$. Then

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

Furthermore, for each event $E \subseteq \Omega$, $P(\bar{E}) = 1 - P(E)$.

Proof: Follows easily from definitions:

$$\forall E_i, P(E_i) = \sum_{s \in E_i} P(s), \text{ thus, since the sets } E_i \text{ are disjoint,}$$

$$P\left(\bigcup_i E_i\right) = \sum_{s \in \bigcup_i E_i} P(s) = \sum_i \sum_{s \in E_i} P(s) = \sum_i P(E_i).$$

Likewise, since $P(\Omega) = \sum_{s \in \Omega} P(s) = 1$,

$$P(\bar{E}) = P(\Omega - E) = \sum_{s \in \Omega - E} P(s) = \sum_{s \in \Omega} P(s) - \sum_{s \in E} P(s) = 1 - P(E)$$

Conditional probability

Definition: Let $P : \Omega \rightarrow [0, 1]$ be a probability distribution, and let $E, F \subseteq \Omega$ be two events, such that $P(F) > 0$.

The **conditional probability** of E given F , denoted $P(E|F)$, is defined by:

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

E.g. A fair coin is flipped three times. Suppose we know that the event $F =$ **heads came up exactly once** occurs.

What is the probability that the event $E =$ **the first coin flip came up heads** occurs?

For the example above, there are 8 flip sequences, $\{H, T\}^3$, all with probability $\frac{1}{8}$. The event that **heads came up exactly once** is $F = \{HTT, THT, TTH\}$. The event $E \cap F = \{HTT\}$. So,

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{\frac{1}{8}}{\frac{3}{8}} = \frac{1}{3}$$

Independence of two events

Intuitively, two events are independent if knowing whether one occurred does not alter the probability of the other. Formally:

Definition: Events A and B are called **independent** if $P(A \cap B) = P(A)P(B)$.

Note that if $P(B) > 0$ then A and B are independent if and only if

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

Thus, the probability of A is not altered by knowing B occurs.

E.g. A fair coin is flipped three times. Are the events
 $A =$ **the first coin toss came up heads** and
 $B =$ **an even number of coin tosses came up head**, independent?

Yes, because $P(A \cap B) = \frac{1}{4}$, $P(A) = \frac{1}{2}$, and $P(B) = \frac{1}{2}$, so $P(A \cap B) = P(A)P(B)$.

Pairwise and mutual independence

Definition: Events E_1, E_2, \dots, E_n are called **pairwise independent** if for every pair $i, j \in \{1, \dots, n\}, i \neq j, E_i$ and E_j are independent. (i.e. $P(E_i \cap E_j) = P(E_i)P(E_j)$)

Events E_1, E_2, \dots, E_n are called **mutually independent**, if for every subset $J \subseteq \{1, \dots, n\}$, $P(\bigcap_{j \in J} E_j) = \prod_{j \in J} P(E_j)$.

Clearly, mutual independence implies pairwise independence. However, **pairwise independence does not imply mutual independence**.

Biased coins and Bernoulli trials

A **Bernoulli trial** is a probabilistic experiment that has two outcomes: **success** or **failure** (e.g. heads or tails).

We suppose that p is the probability of success, and $q = 1 - p$ is the probability of failure.

We can of course have repeated Bernoulli trials. We typically assume the different trials are **mutually independent**.

E.g. A *biased coin*, which comes up heads with probability $p = \frac{2}{3}$, is flipped 7 times consecutively. What is the probability that it comes up heads exactly 4 times?

The binomial distribution

Theorem

The probability of exactly k successes in n mutually independent Bernoulli trials, with probability p of success and $q = 1 - p$ of failure in each trial, is:

$$\binom{n}{k} p^k q^{n-k}$$

Proof

We can associate n Bernoulli trials with outcomes $\Omega = \{H, T\}^n$. Each sequence $s = (s_1, \dots, s_n)$ with exactly k heads and $n - k$ tails occurs with probability $p^k q^{n-k}$. There are $\binom{n}{k}$ such sequences with exactly k heads.

Definition

The **binomial distribution** with parameters n and p , denoted $b(k; n, p)$, defines a probability distribution on $k \in \{0, \dots, n\}$, given by:

$$b(k; n, p) = \binom{n}{k} \cdot p^k q^{n-k}$$

Pairwise and mutual independence

A finite set of events $\{E_i\}_{i=1}^n$ is **pairwise independent** if every pair of events is independent - that is, if and only if for all distinct pairs of indices m, k :

$$P(E_m \cap E_k) = P(E_m)P(E_k)$$

A finite set of events is **mutually independent** if every event is independent of any intersection of the other events - that is, if and only if for every k -element subset of $\{E_i\}_{i=1}^n$:

$$P\left(\bigcap_{i=1}^k E_i\right) = \prod_{i=1}^k P(E_i)$$

Bayes' theorem

Let A and B be events such that $0 < P(A) < 1$ and $P(B) > 0$.

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

Random variables

Definition: A **random variable**, is a function $X : \Omega \rightarrow \mathbb{R}$, that assigns a real value to each outcome in a sample space Ω .

E.g. Suppose a biased coin is flipped n times. The sample space is $\Omega = \{H, T\}^n$. The function $X : \Omega \rightarrow \mathbb{N}$ that assigns to each outcome $s \in \Omega$ the number $X(s) \in \mathbb{N}$ of coin tosses that came up heads, is one random variable.

For a random variable $X : \Omega \rightarrow \mathbb{R}$ we write $P(X = r)$ as shorthand for the probability $P(\{s \in \Omega \mid X(s) = r\})$. The **distribution** of a random variable X is given by the set of pairs $\{(r, P(X = r)) \mid r \text{ is in the range of } X\}$.

Note: These definitions of a random variable and its distribution are only adequate in the context of **discrete** probability distributions.

Biased coins and the geometric distribution

Suppose a biased coin comes up heads with probability $p : 0 < p < 1$ each time it is tossed. Suppose we repeatedly flip this coin until it comes up heads.

What is the probability that we flip the coin k times, for $k \geq 1$?

Answer: The sample space is $\Omega = \{H, TH, TTH, \dots\}$. Assuming mutual independence of coin flips, the probability of $T^{k-1}H$ is $(1 - p)^{k-1}p$. Note that this does define a probability distribution on $k \geq 1$ because:

$$\sum_{k=1}^{\infty} (1 - p)^{k-1} p = p \sum_{k=0}^{\infty} (1 - p)^k = p \left(\frac{1}{1 - p} \right) = 1$$

A random variable $X : \Omega \rightarrow \mathbb{N}$, is said to have a **geometric distribution** with parameter $p : 0 \leq p \leq 1$, if for all positive integers $k \geq 1$, $P(X = k) = (1 - p)^{k-1}p$.

Expectation of a random variable

Recall: A **random variable** is a function $X : \Omega \rightarrow \mathbb{R}$, that assigns a real value to each outcome in a sample space Ω .

The **expected value** or **expectation** or **mean**, of a random variable $X : \Omega \rightarrow \mathbb{R}$, denoted by $E(X)$, is defined by:

$$E(X) = \sum_{s \in \Omega} P(s)X(s)$$

Here $P : \Omega \rightarrow [0, 1]$ is the underlying probability distribution on Ω .

Question: Let X be the random variable outputting the number that comes up when a **fair die** is rolled. What is the expected value, $E(X)$, of X ?

$$E(X) = \sum_{i=1}^6 \frac{1}{6} \cdot i = \frac{21}{6} = \frac{7}{2}$$

However, sometimes this way can be inefficient for calculating $E(X)$. Take the sample space of 11 coin flips, for example: $\Omega = \{H, T\}^{11}$. Our summation will have $2^{11} = 2048$ terms!

Better expression for the expectation

Recall $P(X = r)$ denotes the probability $P(\{s \in \Omega \mid X(s) = r\})$.

Recall that for a function $X : \Omega \rightarrow \mathbb{R}$,

$$\text{range}(X) = \{r \in \mathbb{R} \mid \exists s \in \Omega : X(s) = r\}$$

Theorem

For a random variable $X : \Omega \rightarrow \mathbb{R}$,

$$E(X) = \sum_{r \in \text{range}(X)} P(X = r) \cdot r$$

Proof

$E(X) = \sum_{s \in \Omega} P(s)X(s)$, but for each $r \in \text{range}(X)$, if we sum all terms $P(s)X(s)$ such that $X(s) = r$, we get $P(X = r) \cdot r$ as their sum. So, summing over all $r \in \text{range}(X)$ we get $E(X) = \sum_{r \in \text{range}(X)} P(X = r) \cdot r$.

So, if $|\text{range}(X)|$ is small, and if we can compute $P(X = r)$, then we need to sum a lot fewer terms to calculate $E(X)$.

Expected number of successes in n Bernoulli trials

Theorem

The expected number of successes in n (independent) Bernoulli trials, with probability p of success in each, is np .

Linearity of expectation (VERY IMPORTANT)

Theorem

For any random variables X, X_1, \dots, X_n on Ω ,

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + \dots + E(X_n)$$

Furthermore, for any $a, b \in \mathbb{R}$,

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

In other words, the expectation function is a **linear function**.

Variance

The **variance** and **standard deviation** of a random variable X give us ways to measure (roughly) on average, how far off the value of the random variable is from its expectation.

Definition

For a random variable X on a sample space Ω , the **variance** of X , denoted by $V(X)$, is defined by:

$$V(X) = E((X - E(X))^2) = \sum_{s \in \Omega} (X(s) - E(X))^2 P(s)$$

The **standard deviation** of X , denoted by $\sigma(X)$, is defined by

$$\sigma(X) = \sqrt{V(X)}$$

E.g. Consider the random variable X , s.t. $P(X = 0) = 1$, and the random variable Y , s.t. $P(Y = -10) = P(Y = 10) = \frac{1}{2}$.

Then $E(X) = E(Y) = 0$, but $V(X) = 0 = \sigma(X)$, whereas $V(Y) = 100$ and $\sigma(Y) = 10$.

Theorem

For any random variable X ,

$$V(X) = E(X^2) - E(X)^2$$

Proof

$$V(X) = E((X - E(X))^2)$$

$$V(X) = E(X^2 - 2XE(X) + E(X)^2)$$

$$V(X) = E(X^2) - 2E(X)E(X) + E(X)^2$$

$$V(X) = E(X^2) - E(X)^2$$