Probability and Stochastic Modelling 1

Semester 1, 2022

Dr Alexander Browning

Tarang Janawalkar

Contents

1 Events and Probability

1.1 Events

Definition 1.1 (Event). An event is a set of outcomes in a random experiment commonly denoted by a capital letter. Events can be simple (a single event) or compound (two or more simple events).

Definition 1.2 (Sample space). The set of all possible outcomes of an experiment is known as the sample space for that experiment and is denoted Ω.

Definition 1.3 (Intersection). An intersection between two events A and B describes the set of outcomes that occur in both A and B. The intersection can be represented using the set `AND` operator ($\cap$) — $A \cap B$ (or AB).

Definition 1.4 (Disjoint). Disjoint (mutually exclusive) events are two events that cannot occur simultaneously, or have no common outcomes.

Theorem 1.1.1 (Intersection of disjoint events). *The intersection of disjoint events results in the null set ($\varnothing$).*

Lemma 1.1.1.1. *Disjoint events are **dependent** events as the occurrence of one means the other cannot occur.*

Definition 1.5 (Union). A union of two events A and B describes the set of outcomes in either A or B. The union is represented using the set `OR` operator ($\cup$) — $A \cup B$.

Definition 1.6 (Complement). The complement of an event E is the set of all other outcomes in Ω. The complement of E is denoted $\overline{E}$.

Theorem 1.1.2 (Intersection of complement set).

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

Theorem 1.1.3 (Union of complement set).

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

Definition 1.7 (Subset). A is a (non-strict) subset of B if all elements in A are also in B. This can be denoted as $A \subset B$.

Theorem 1.1.4. *All events E are subsets of Ω.*

Theorem 1.1.5. *Given $A \subset B$*

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

Corollary 1.1.5.1. *Given $\varnothing \subset E$*

$$\varnothing E = \varnothing \qquad \textit{and} \qquad \varnothing \cup E = E$$

Theorem 1.1.6 (Associative Identities).

$$A\left(BC\right) = \left(AB\right)C$$
$$A \cup \left(B \cup C\right) = \left(A \cup B\right) \cup C$$

Theorem 1.1.7 (Distributive Identities).

$$A\left(B \cup C\right) = AB \cup AC$$
$$A \cup BC = \left(A \cup B\right)\left(A \cup C\right)$$

1.2 Probability

Definition 1.8 (Probability)**.** Probability is a measure of the likeliness of an event occurring. The probability of an event E is denoted $\Pr(E)$ (sometimes $\mathrm{P}(E)$).

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

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

Theorem 1.2.1 (Probability of Ω)**.**

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

Theorem 1.2.2 (Complement rule)**.** *The probability of the complement of E is given by*

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

Theorem 1.2.3 (Multiplication rule for independent events)**.** *The probability of the intersection between two independent events A and B is given by*

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

Theorem 1.2.4 (Addition rule for independent events)**.** *The probability of the union between two independent events A and B is given by*

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

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

Corollary 1.2.4.1 (Addition rule for 3 events)**.** *The addition rule for 3 events is as follows*

$$\Pr(A \cup B \cup C) = \Pr(A) + \Pr(B) + \Pr(C) - \Pr(AB) - \Pr(AC) - \Pr(BC) + \Pr(ABC).$$

Proof. If we write $D = A \cup B$ and apply the addition rule twice, we have

$$\begin{aligned}
\Pr(A \cup B \cup C) &= \Pr(D \cup C) \\
&= \Pr(D) + \Pr(C) - \Pr(DC) \\
&= \Pr(A \cup B) + \Pr(C) - \Pr((A \cup B)C) \\
&= \Pr(A) + \Pr(B) - \Pr(AB) + \Pr(C) - \Pr(AC \cup BC) \\
&= \Pr(A) + \Pr(B) - \Pr(AB) + \Pr(C) - (\Pr(AC) + \Pr(BC) - \Pr(ACBC)) \\
&= \Pr(A) + \Pr(B) + \Pr(C) - \Pr(AB) - \Pr(AC) - \Pr(BC) + \Pr(ABC)
\end{aligned}$$

□

Theorem 1.2.5 (De Morgan's laws)**.** *Recall De Morgan's Laws:*

$$\begin{aligned}
\overline{A \cup B} &= \overline{A}\ \overline{B} \\
\overline{AB} &= \overline{A} \cup \overline{B}.
\end{aligned}$$

Taking the negation of both sides and applying the complement rule yields

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

1.3 Circuits

A signal can pass through a circuit if there is a functional path from start to finish.
We can define a circuit where each component i functions with probability p, and is independent of other components.
Then W_i to be the event in which the associated component i functions, we can determine the event S in which the system functions, and probability $\Pr(S)$ that the system functions.
As the probability that any component functions is p, in other words

$$\Pr(W_i) = p,$$

$\Pr(S)$ will be a function of p defined $f : [0,\ 1] \to [0,\ 1]$.

2 Independence

Definition 2.1 (Conditional probability)**.** When discussing multiple events, it is possible that the occurrence of one event changes the probability that another will occur. This can be denoted using a vertical bar, and is read as "the probability of event A given B":

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

Definition 2.2 (Multiplication rule)**.** For events A and B, the general multiplication rule states that

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

Theorem 2.0.1 (Independent events)**.** *If A and B are independent events then*

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

Theorem 2.0.2 (Complement of independent events)**.** *If A and B are independent, all complement pairs are also independent. Given*

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

the following statements are also true

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

2.1 Probability Rules with Conditional

ALl probability rules hold when conditioning on some event C.

Theorem 2.1.1 (Complement rule with condition)**.**

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

Theorem 2.1.2 (Addition rule with condition)**.**

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

Theorem 2.1.3 (Multiplication rule with condition)**.**

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

In the above examples, all probabilities are conditional on the sample space, hence we are effectively changing the sample space.

2.2 Conditional Independence

Definition 2.3 (Conditional independence)**.** Suppose events A and B are not independent, i.e.,

$$\Pr\left(A \mid B\right) \neq \Pr\left(A\right)$$

but they become independent when conditioned with another event C, i.e.,

$$\Pr\left(A \mid BC\right) = \Pr\left(A \mid C\right)$$

Here we say that A and B are **conditionally independent** given C. Furthermore

$$\Pr\left(AB \mid C\right) = \Pr\left(A \mid C\right) \Pr\left(B \mid C\right)$$

Conversely, events A and B may be conditionally dependent but unconditionally independent, i.e.,

$$\begin{aligned} \Pr\left(A \mid B\right) &= \Pr\left(A\right) \\ \Pr\left(A \mid BC\right) &\neq \Pr\left(A \mid C\right) \\ \Pr\left(AB \mid C\right) &= \Pr\left(A \mid BC\right) \Pr\left(B \mid C\right) \end{aligned}$$

Theorem 2.2.1. *Given events A, B, and C. Pairwise independence does not imply mutual independence. I.e.,*

$$\begin{cases} \Pr\left(AB\right) = \Pr\left(A\right) \Pr\left(B\right) \\ \Pr\left(AC\right) = \Pr\left(A\right) \Pr\left(C\right) \\ \Pr\left(BC\right) = \Pr\left(B\right) \Pr\left(C\right) \end{cases}$$

does not imply

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

In summary, independence should not be assumed unless explicitly stated.

2.3 Disjoint Events

Theorem 2.3.1 (Probability of disjoint events)**.** *The probability of disjoint events A and B is given by*

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

Disjoint events are highly dependent events, since the occurrence of one means the other cannot occur. This implies

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

2.4 Subsets

Theorem 2.4.1 (Probability of subsets)**.** *If $A \subset B$ then $\Pr(A) \leq \Pr(B)$. We also know that $\Pr(AB) = \Pr(A)$ and $\Pr(A \cup B) = \Pr(B)$.*
Here, if A happens, then B definitely happens.

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

Given $\Pr(AB) = \Pr(A)$

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

These events are also highly dependent.

3 Total Probability

Definition 3.1 (Marginal probability)**.** Marginal probability is the probability of an event irrespective of the outcome of another variable.

Theorem 3.0.1 (Total probability for complements)**.** *By writing the event A as $AB \cup A\overline{B}$, and noting that AB and $A\overline{B}$ are disjoint, the marginal probability of A is given by*

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

By applying the multiplication rule to each joint probability:

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

Theorem 3.0.2 (Law of total probability)**.** *The previous theorem partitioned Ω into disjoint events B and $\overline{B}$.*
By partitioning Ω into a collection of disjoint events $B_1, B_2, \ldots, B_n$, such that $\bigcup_{i=1}^{n} B_i = \Omega$, we have

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

Theorem 3.0.3 (Bayes' Theorem)**.** *Given the probability for A given B, the probability of the reverse direction is given by*

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

4 Combinatorics

Definition 4.1 (Number of outcomes)**.** Let $|A|$ denote the number of outcomes in an event A.

Theorem 4.0.1 (Addition principle)**.** *Given a sample space S with k disjoint events $\{S_1, \dots, S_k\}$, where the ith event has n_i possible outcomes, the number of possible samples from any event is given by*

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

Theorem 4.0.2 (Multiplication principle)**.** *Given a sample space S with k events $\{S_1, \dots, S_k\}$, where the ith event has n_i possible outcomes, the number of possible samples from every event is given by*

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

Theorem 4.0.3 (Counting probability)**.** *Given a sample space S with equally likely outcomes, the probability of an event $S_i \subset S$ is given by*

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

4.1 Ordered Sampling with Replacement

When ordering is important and repetition is allowed, the total number of ways to choose k objects from a set with n elements is

$$n^k$$

4.2 Ordered Sampling without Replacement

When ordering is important and repetition is not allowed, the total number of ways to arrange k objects from a set of n elements is known as a k-permutation of n-elements denoted ${}^nP_k$

$$\begin{aligned} {}^nP_k &= n \times (n-1) \times \cdots \times (n-k+1) \\ &= \frac{n!}{(n-k)!} \end{aligned}$$

for $0 \leq k \leq n$.

Definition 4.2 (Permutation of n elements)**.** An n-permutation of n elements is the permutation of those elements. In this case, $k = n$, so that

$$\begin{aligned} {}^nP_n &= n \times (n-1) \times \cdots \times (n-n+1) \\ &= n! \end{aligned}$$

4.3 Unordered Sampling without Replacement

When ordering is not important and repetition is not allowed, the total number of ways to choose k objects from a set of n elements is known as a k-combination of n-elements denoted ${}^nC_k$ or $\binom{n}{k}$

$$\begin{aligned} {}^nC_k &= \frac{{}^nP_k}{k!} \\ &= \frac{n!}{k!\,(n-k)!} \end{aligned}$$

for $0 \leq k \leq n$. We divide by $k!$ because any k-element subset of n-elements can be ordered in $k!$ ways.

4.4 Unordered Sampling with Replacement

When ordering is not important and repetition is allowed, the total number of ways to choose k objects from a set with n elements is

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

5 Random Variables and Distributions

Definition 5.1 (Discrete random variables)**.** A discrete random variable has countably many outcomes.

Definition 5.2 (Continuous random variables)**.** A continuous random variable can take an infinite number of individual outcomes. Note that the probability of a continuous random variable being equal to a specific value is always zero.

5.1 Probability distributions

Definition 5.3 (Probability mass function)**.** For discrete random variables, the distribution is described with a Probability Mass Function (PMF)

$$p(n) = \Pr(N = n)$$

For this to be a valid PMF,

$$\forall n,\ \Pr(N = n) \geq 0$$

and

$$\sum_n \Pr(N = n) = 1$$

Definition 5.4 (Probability density function)**.** For continuous variables, the distribution is described with a Probability Density Function (PDF) and the associated Cumulative Distribution Function (CDF).

Definition 5.5 (Cumulative distribution function)**.** Here, probabilities are represented by areas under the PDF:

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

and the CDF is defined as

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u.$$

Note that by the fundamental theorem of calculus we can recover the PDF given the CDF by taking the derivative.
$f(x)$ is a valid PDF provided

$$f(x) \geq 0 : \forall x \quad \text{and} \quad \int_{-\infty}^{\infty} f(u)\,\mathrm{d}u = 1$$

while $F(x)$ is a valid CDF if:

1. F is a non-decreasing right continuous function
2. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$

Note that integrals are replaced with sums for the discrete equivalents.

5.2 Quantiles

Definition 5.6 (Median)**.** For a continuous random variable, the median is defined as m such that

$$\int_{-\infty}^{m} f(x)\,\mathrm{d}x = \int_{m}^{\infty} f(x)\,\mathrm{d}x = 0.5$$

Definition 5.7 (p-Quantile)**.** For a continuous random variable, the p-quantile is defined as a such that

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

Definition 5.8 (Inverse quantile)**.** The inverse quantile or inverse CDF is the inverse function of the CDF and can be used to find the a that a certain p provides. I.e.

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

Note: there will not necessarily be an analyticial form of this function, regardless of if the CDF has one.

5.3 Expected Value and Variance

Definition 5.9 (Expectation)**.** The expected value $\mathrm{E}(X)$, sometimes written $\mathbb{E}$, of a random variable is the average outcome that could be expected from an infinite number of observations of that variable. This is also known as the mean of the variable, denoted μ.

$$\mathrm{E}(X) = \begin{cases} \sum_{\Omega} x \cdot p(x) & \text{for discrete variables} \\ \int_{\Omega} x \cdot f(x)\,\mathrm{d}x & \text{for continuous variables} \end{cases}$$

Another identity for the expected value is

$$\mathrm{E}\left(X\right) = -\int_{x<0} F(x)\,\mathrm{d}x + \int_{x>0} \left(1 - F(x)\right)\mathrm{d}x$$

Definition 5.10 (Variance)**.** The variance $\mathrm{Var}\left(X\right)$, of a random variable is a measure of spread of the distribution (defined as the average squared distance of each value from the mean). $\mathrm{Var}\left(X\right)$ is also denoted as σ^2.

$$\begin{aligned}\mathrm{Var}\left(X\right) &= \begin{cases} \sum_\Omega \left(x-\mu\right)^2 \cdot p(x) & \text{for discrete variables} \\ \int_\Omega \left(x-\mu\right)^2 \cdot p(x) \cdot f(x)\,\mathrm{d}x & \text{for continuous variables} \end{cases} \\ &= \mathrm{E}\left(X^2\right) - \mathrm{E}\left(X\right)^2\end{aligned}$$

Definition 5.11 (Standard deviation)**.** The standard deviation is defined as

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

5.3.1 Transformations

For a simple linear function of a random variable

$$\begin{aligned}\mathrm{E}\left(aX \pm b\right) &= a\,\mathrm{E}\left(X\right) \pm b \\ \mathrm{Var}\left(aX \pm b\right) &= a^2\,\mathrm{Var}\left(X\right)\end{aligned}$$