**CS** **Summer**

# Discrete Probability

Authors: Shaleen Baral

## Contents

# 1. Events and Probability

## 1.1. Events

Random phenonmenon are observed by means of experiments. Each experiment results in an *outcome*. The collection of all possible outcomes $\omega$ is called the *sample space* $\Omega$. Any subset $A \subseteq \Omega$ of the sample space $\Omega$ can be regarded as a representation of some *event*.

**Definition 1.1.1** (Indicator Function): Let $\mathcal{P}$ be a property that an element $x$ of some set $E$ may or may not satisfy. Then the indicator function for $\mathcal{P}$, $\mathbb{1}_{\mathcal{P}} : E \to \{0, 1\}$ is defined by

$$\mathbb{1}_{\mathcal{P}}(x) = \begin{cases} 1 & \text{if } x \text{ satisfies } \mathcal{P}, \\ 0 & \text{otherwise.} \end{cases}$$

*Example*: For any set $A \subseteq \Omega$ we may define an indicator function $\mathbb{1}_A : \Omega \longrightarrow \{0, 1\}$ using the set membership property:

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

We say that an outcome $\omega \in \Omega$ *realizes* an event $A \subseteq \Omega$ if $\omega \in A$. Two event $A, B \subseteq \Omega$ are said to be *incompatible* if $A \cap B = \emptyset$– that is, no outcome can realize both $A$ and $B$.

We refer to $\emptyset$ as the *impossible event*. Conversely, we refer to $\Omega$ as the *certain event*.

For a family of sets $\{A_k\}_{k \in \mathbb{N}}$, we use the notation $\sum_{k=1}^{\infty} A_k$ to denote $\cup_{k=1}^{\infty} A_k$ if the family is pairwise disjoint.

**Definition 1.1.2** (Exhaustive): A family of events is *exhaustive* if any outcome $\omega$ realizes at least one of them.

**Definition 1.1.3** (Mutually Exclusive): A family of events is *mutually exclusive* if any two distinct events among them are incompatible.

**Definition 1.1.4** (Partition): A family of sets $\{A_k\}$ partition $\Omega$ if

$$\sum_{k=1}^{\infty} A_k = \Omega.$$

In other words, we say that the events $\{A_k\}$ are *mutually exclusive* and *exhaustive*. Furthermore, in terms of indicator functions,

$$\sum_{k=1}^{\infty} \mathbb{1}_{A_k} = 1.$$

If $B \subseteq A$, event $B$ is said to *imply* event $A$, because $\omega \in \Omega$ realizes $A$ whenever it realizes $B$. In terms of indicator functions, $\mathbb{1}_B(\omega) \leq \mathbb{1}_A(\omega)$.

## 1.2. Probability

Probability theory assigns to each event a number, the *probability* of said event. We require the collection, $\mathcal{F}$, of events to which a probability is assigned to be a $\sigma$-field.

**Definition 1.2.1** ($\sigma$-field): Let $\mathcal{F}$ be a collection of subsets of $\Omega$ such that

a. the certain event $\Omega$ is in $\mathcal{F}$,
b. if $A \in \mathcal{F}$, then $\overline{A} \in \mathcal{F}$,
c. if $\{A_k\}_{k \in \mathbb{N}}$ is a countable collection of sets in $\mathcal{F}$ then $\cup_{k=1}^{\infty} A_k \in \mathcal{F}$.

Then $\mathcal{F}$ is a *$\sigma$-field* (or *$\sigma$-algebra*) on $\Sigma$. Here, we may also refer to it as the *$\sigma$-field of events*.

*Remark*: Particularly, $\mathcal{F}$ may not be the collection of all subsets of $\Omega$.

The *trivial $\sigma$-field* is the collection of all subsets $2^{\Omega}$ of $\Omega$. The *gross $\sigma$-field* is the $\sigma$-field of two members $\{\emptyset, \Omega\}$.

Usually, if $\Omega$ is countable we take the $\sigma$-field of events to be the trivial $\sigma$-field.

The *probability* of an event measures the likeliness of its occurence.

**Definition 1.2.2** (Probability Measure): A *probability* on $(\Omega, \mathcal{F})$ is a mapping $P : \mathcal{F} \longrightarrow \mathbb{R}$ such that

a. (*non-negativity*) $0 \leq P(A) \leq 1$,
b. (*unit measure*) )$P(\Omega) = 1$,
c. (*sigma additivity*) $P\left(\sum_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$.

The triple $(\Omega, \mathcal{F}, P)$ is called a *probability space* or *probability model*.

*Remark*: Notationally, we may replace intersection with commas i.e. $P(A, B) = P(A \cap B)$.

*Remark*: One may think of the axioms above as being motivatived by the heuristic interpretation of probability as *emperical frequency*.

Some basic consequences of the axioms above are as follows.

**Proposition 1.2.1**: For any event $A \in \mathcal{F}$

$$P\left(\overline{A}\right) = 1 - P(A)$$

and in particular, $P(\emptyset) = 0$.

*Proof*: Using additivity,

$$1 = P(\Omega) = P\left(A + \overline{A}\right) = P(A) + P\left(\overline{A}\right).$$

By the fact that $P(\Omega) = 1$, $P(\emptyset) = 0$ immediately follows.

□

**Proposition 1.2.2**: Probability is *monotone*, that is to say

$$A \subseteq B \Longrightarrow P(A) \leq P(B).$$

*Proof*: Using subadditivity,

$$P(B) = P(A \cup (B - A)) = P(A) + P(B - A) \geq P(A).$$

□

**Proposition 1.2.3**: Probability is sub-$\sigma$-additive, that is to say

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k).$$

*Proof*: Using subadditvity and monotonicity,

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = P\left(\bigcup_{k=1}^{\infty}\left(A_k - \bigcup_{\ell=1}^{k-1} A_k\right)\right) = \sum_{k=1}^{\infty} P\left(A_k - \bigcup_{\ell=1}^{k-1} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k).$$

□

**Definition 1.2.3** (Negligible): A set $N \subset \Omega$ is called $P$*-negligible* if it is contained in an event $A \in \mathcal{F}$ of probability $P(A) = 0$.

**Proposition 1.2.4**: A countable union of negligible sets is a negligible set.

*Proof*: Let $N_k$, $k \geq 1$, be $P$-negligble sets. By definition, there exists a sequence $A_k$, $k \geq 1$ of events of null probability such that $N_k \subseteq A_k$, $k \geq 1$. We then have,

$$N := \cup_{k\geq 1} N_k \subset, = A := \cup_{k \geq 1} A_k.$$

Note that $P(A) = 0$. So, by $\sigma$-subadditivity and non-negativity of $P$, $P(N) = 0$.

□

**Definition 1.2.4** (Almost Surely): A property $\mathcal{P}$ relative to the samples $\omega \in \Sigma$ is said to hold $P$*-almost-surely* ("$P$-a.s.") if

$$P(\{\omega; \omega \text{ verifies property } \mathcal{P}\}) = 1.$$

*Remark*: If there is no ambiguity, we may abbreviate $P$-almost-surely to just almost-surely.

**Definition 1.2.5** (Non-Decreasing Sets): A sequence of events $\{A_n\}_{n\geq 1}$ is *non-decreasing* if $A_n \subseteq A_{n+1}$ for all $n \geq 1$.

**Definition 1.2.6** (Non-Increasing Sets): A sequence of events $\{A_n\}_{n\geq 1}$ is *non-decreasing* if $A_{n+1} \subseteq A_{n+1}$ for all $n \geq 1$.

**Proposition 1.2.5**: Let $\{A_n\}_{n\geq 1}$ be a non-decreasing sequence of events. Then,

$$P(\cup_{k=1}^{\infty} A_n) = \lim_{n\uparrow\infty} P(A_n).$$

*Proof*: Note that

$$A_n = A_1 + (A_2 - A_1) + \dots + (A_n - A_{n-1})$$
$$\bigcup_{k=1}^{\infty} A_k = A_1 + (A_2 - A_1) + \dots + (A_n - A_{n-1})$$

Thus,

$$\begin{aligned} P\left(\bigcup_{k=1}^{\infty} A_k\right) &= P(A_1) + \sum_{j=2}^{\infty} P\left(A_j - A_{j-1}\right) \\ &= \lim_{n\uparrow\infty} \left\{ P(A_1) + \sum_{j=2}^{n} P\left(A_j - A_{j-1}\right) \right\} = \lim_{n\uparrow\infty} P(A_n). \end{aligned}$$

□

**Corollary 1.2.5.1**: Let $\{B_n\}_{n\geq 1}$ be a non-increasing sequence of events. Then,

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\uparrow\infty} P(B_n).$$

*Proof*: Note that $\left\{\overline{B_n}\right\}$ is a non-decreasing sequence of events. Thus, by Proposition 1.2.5 and De-Morgan's law we have,

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = 1 - P\left(\bigcup_{n=1}^{\infty} \overline{B_n}\right) = 1 - \lim_{n\uparrow\infty} P\left(\overline{B_n}\right) = \lim_{n\uparrow\infty} \left(1 - P\left(\overline{B_n}\right)\right) = \lim_{n\uparrow\infty} P(B_n)$$

□

Next, we introduce the lim inf and lim sup for sets.

**Definition 1.2.7** (Liminf I): The lim inf of a sequence $\{A_n\}_{n\in\mathbb{N}}$ of sets is defined to be

$$\liminf_{n\uparrow\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{j\geq n} A_j.$$

**Definition 1.2.8** (Limsup I): The lim sup of a sequence $\{A_n\}_{n\in\mathbb{N}}$ of sets is defined to be

$$\limsup(n \uparrow \infty) A_n = \bigcap_{n=1}^{\infty} \bigcup_{j\geq n} A_j.$$

The following lemma give useful interpretations for these.

**Lemma 1.2.1**: $x \in \liminf_{n\uparrow\infty} A_n$ iff $x \notin A_n$ for finitely many $n$.

*Proof*: $x \in \liminf_{n\uparrow\infty} A_n$ iff $x \in \bigcap_{j\geq n}^{\infty} A_j$ for some $n \in \mathbb{N}$. By definition, the latter is true iff $x \in A_j$ for every $j \geq n$ i.e. $x \notin A_j$ implies $j \leq n$. Thus, $x \notin A_n$ for finitely many $n$.

□

*Remark*: For this reason, a shorthand for the limit infimum is "$x$ is in $A_n$ all but finitely often", also expressed as $A_n$ a.b.f.o.

**Lemma 1.2.2**: $x \in \limsup_{n\uparrow\infty} A_n$ iff for every $n \in \mathbb{N}$, there exists $m \geq n$ such that $x \in A_m$.

*Proof*: $x \in \limsup_{n\uparrow\infty} A_n$ iff $x \in \bigcup_{j\geq n} A_j$ for every $n \in N$. The latter is true iff $x \in A_j$ for some $j \geq n$, for every $n \in \mathbb{N}$.

$\square$

*Remark*: For this reason, a shorthand for the limit supremum is "$x$ is in $A_n$ infinitely often", also expressed as $A_n$ i.o.

We can also define lim sup and lim inf in terms of indicator functions.

**Lemma 1.2.3** (Liminf II): The lim inf of a sequence $\{A_n\}_{n\in\mathbb{N}}$ of sets is given by

$$\liminf_{n\uparrow\infty} A_n = \left\{x \in \Omega \mid \liminf_{n\uparrow\infty} \mathbb{1}_{A_n}(x) = 1\right\}$$

*Proof*: Note that the indicator function only acquires values in $\{1, 0\}$. For any $x \in \Omega$, $\liminf_{n\uparrow\infty} \mathbb{1}_{A_n}(x) = 1$ iff $\mathbb{1}_{A_n}(x) < 1$ for finitely many $n \in \mathbb{N}$. Thus, $\mathbb{1}_{A_n}(x) = 0$ for finitely many $n \in \mathbb{N}$ and $x \notin A_n$ for finitely many $n \in \mathbb{N}$ too.

$\square$

**Lemma 1.2.4** (Limsup II): The lim sup of a sequence $\{A_n\}_{n\in\mathbb{N}}$ of sets is given by

$$\limsup_{n\uparrow\infty} A_n = \left\{x \in \Omega \mid \limsup_{n\uparrow\infty} \mathbb{1}_{A_n}(x) = 1\right\}.$$

*Proof*: Note that the indicator function only acquires values in $\{0, 1\}$. For any $x \in \Omega$, $\limsup_{n\uparrow\infty} \mathbb{1}_{A_n}(x) = 1$ iff for every $\sup_{j\geq n} \mathbb{1}_{A_j}(x) = 1$ for every $n \in \mathbb{N}$. The later is true iff there exists $j \geq n$ for every $n \in \mathbb{N}$ such that $\mathbb{1}_{A_j}(x) = 1$.

$\square$

**Proposition 1.2.6** (Borel-Cantelli Lemma): For any sequence of events $\{A_n\}_{n\geq 1}$,

$$\sum_{k=1}^{\infty} P(A_n) < \infty \Longrightarrow P(A_n \text{ i.o.}) = 0$$

*Proof*: The set $\bigcup_{k\geq n}^{\infty} A_k$ decreases with $n$. Thus, by sequential continuity of probability,

$$P(A_n \text{ i.o.}) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k\right) = \lim_{n\uparrow\infty} P\left(\bigcup_{k\geq n} A_k\right).$$

Then, by $\sigma$-subadditivity,

$$P\left(\bigcup_{k\geq n} A_k\right) \leq \sum_{k\geq n} P(A_k).$$

By the fact that $\sum_{k=1}^{\infty} P(A_n) < \infty$, the right hand-side of the inequality goes to 0 as $n \uparrow \infty$.

$\square$

### Counting Models

A common setting is one where the set $\Omega$ of all possible outcomes is finite and (for some reason) we are led to believe that all the outcomes $\omega$ have the same probability. As the probabilities must sum up to one, each outcome has probability $\frac{1}{|\Omega|}$. Since the probability of an event $A$ is the sum of the probabilities of all outcomes $\omega \in A$, we have

$$P(A) = \frac{|A|}{|\Omega|}.$$

In this setting, computing $P(A)$ merely requires *counting* the elements in the sets $A$ and $\Omega$. Now, we review some useful tools that can help with counting.

**Proposition 1.2.7** (Stirling's Formula I):

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

**Proposition 1.2.8** (Stirling's Formula II):

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \leq n! \leq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}.$$

*Remark*: For a simpler, cruder bound, $e^{\frac{1}{12n}} \leq 1.09 < 2$ for all $n \in \mathbb{N}$.

We can apply this to estimate the binomial coefficients.

**Proposition 1.2.9**:

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$$

*Proof*:

$$\begin{aligned}\binom{n}{k} &= \frac{n(n-1)\cdots(n-k+1)}{k!} \\ &\leq \frac{n^k}{k!} \leq \frac{n^k}{\sqrt{2\pi k}\left(\frac{k}{e}\right)^k} \leq \frac{1}{\sqrt{2\pi k}} \cdot \left(\frac{en}{k}\right)^k \leq \left(\frac{en}{k}\right)^k.\end{aligned}$$

□

## 1.3. Independence and Conditioning

# 2. Random Variables

## 2.1. Probability Distribution and Expectation

## 2.2. Generating Function

## 2.3. Conditional Expectation

# 3. Bounds and Inequalities

## 3.1. The Three Basic Inequalities

## 3.2. Frequently Used Bounds