Contents

1 Probability

1.1 Definitions

- A countable **sample space** S is a countable nonempty set. Don't worry too much about the countable part. Usually, we'll work with finite sets. If you're curious about when an infinite set is considered "countable," see the end of recitation 2.
- An element $\omega \in S$ is called an **outcome**.
- A **probability function** on S is a function $\Pr : S \to \mathbb{R}$ with the following two properties:
  1. $\Pr(\omega) \geq 0 \ \forall \omega \in S$
  2. $\sum_{\omega \in S} \Pr(\omega) = 1$
- Together, a sample space and probability function are called a **probability space**.
- A subset $E \subseteq S$ is called an **event**. The probability of E is defined as $\Pr(E) = \sum_{\omega \in E} \Pr(\omega)$
- A probability space is **uniform** if all outcomes have equal probability, that is $\forall \omega \in S$, $\Pr(\omega) = \frac{1}{|S|}$. If this is true, for any event E, $\Pr(E) = \frac{|E|}{|S|}$.

1.2 Rules

Here are some rules about the probabilities of events. You should be comfortable working with them. Some of them are very closely related to counting rules!

- **Sum Rule**: If $E_1, ..., E_n$ are disjoint events (that is, there are no outcomes which are members of more than one event) then $\Pr(E_1 \cup ... \cup E_n) = \sum_{i=1}^{n} \Pr(E_i)$
- **Complement Rule**: For any event E, $\Pr(\overline{E}) = 1 - \Pr(E)$
- **Difference Rule**: For events A and B, $\Pr(B - A) = \Pr(B) - \Pr(B \cap A)$
- **Inclusion-Exclusion**: For events A and B, $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
- **Union Bound**: For events $E_1, ..., E_n$, $\Pr(E_1 \cup ... \cup E_n) \leq \Pr(E_1) + ... + \Pr(E_n)$

1.3 Conditional Probability and Independence

The **conditional probability** $\Pr(A|B)$ is the probability that A happened given that we know B did. Essentially, we limit our set of possibilities to the outcomes in B and find how many of those are also in A. It is defined as

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

Bayes' Rule is a useful rearrangement of the definition of conditional probability and tells us

$$\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B)}$$

A is **independent** of B if knowing B occurred does not give us any additional information about whether A did. Mathematically, A is independent of B if $\Pr(A|B) = \Pr(A)$, or if $\Pr(B) = 0$.

A set of events $\{E_1, ..., E_n\}$ is **mutually independent** if for every subset S of the set of events, the probability of the intersection of the events is equal to the product of the probabilities of each event.

For any set, pairwise independence of the events does *not* guarantee mutual independence!

1.4 Random Variables

A **random variable** is a function from outcomes of a probability space. Usually, the codomain of the function is the real numbers or integers.

Some examples are a mapping from a sequence of coin flips to the number of heads that occur in the sequence or mapping from a person to the number of emails in their inbox.

An **indicator random variable** "indicates" whether an event occurs by mapping all outcomes to either 1 or 0. These are also referred to as Bernoulli variables.

1.5 Expected Value

The **expected value** (or just expectation) of a random variable is a probability-weighted average of its values. That is, if one value is far more likely to occur, we weight it higher in the average. The expected value of a random variable R is defined as

$$\mathbb{E}[R] = \sum_{\omega \in S} R(\omega) \Pr(\omega)$$

It can also be useful to think about summing over the output of R rather than the events in S. This is an equivalent definition of expected value:

$$\mathbb{E}[R] = \sum_{x \in \text{range } R} x \cdot \Pr(R = x)$$

The **conditional expectation** of a random variable R given an event A is defined as

$$\mathbb{E}[R|A] = \sum_{x \in \text{range} R} x \cdot \Pr(R = x|A)$$

Perhaps the most important property from this section is **linearity of expectation**. For random variables $R_1, .., R_n$ and real numbers $a_1, ..., a_n$,

$$\mathbb{E}[a_1 R_1 + ... + a_n R_n] = a_1 \mathbb{E}[R_1] + ... + a_n \mathbb{E}[R_n]$$

1.6 Variance

Sometimes measuring the mean (expectation) of a random variable doesn't give us enough information: it can be helpful to know how much we expect the variable to *stray* from its average.

Markov's inequality gives a generally coarse estimate of the probability that a random variable takes a value much larger than its mean.

Theorem (Markov). If R is a nonnegative random variable, then for all $x > 0$,

$$\Pr[R \geq x] \leq \frac{\mathbb{E}[R]}{x}.$$

Expressed differently:

Corollary. If R is a nonnegative random variable, then for all $c \geq 1$,

$$\Pr[R \geq c \cdot \mathbb{E}[R]] \leq \frac{1}{c}.$$

That is: the probability of R being more than c times its mean is at most $1/c$.

A related notion is that of *variance*:

Definition. The *variance* $\mathrm{Var}[R]$ of a random variable R is defined to be $\mathbb{E}[(R - \mathbb{E}[R])^2]$.

Unpacking this from the inside out: $R - \mathbb{E}[R]$ is a random variable measuring the distance between R and its mean at each outcome. Averaging the square of this gives us a sense of, overall, how far R tends to be from its mean.

There is an equivalent way to state this:

Lemma. For any random variable R,

$$\mathrm{Var}[R] = \mathbb{E}[R^2] - (\mathbb{E}[R])^2.$$

This leads us to state Chebyshev's theorem, an application of Markov's inequality:

Theorem (Chebyshev). Let R be a random variable and $x \in \mathbb{R}^+$. Then

$$\Pr\left[|R - \mathbb{E}\left[R\right]| \geq x\right] \leq \frac{\mathrm{Var}[R]}{x^2}.$$