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, $\Pr(E) = \frac{|E|}{|S|}$.

Practice Problem(s)

Determine which of the following are valid probability spaces (as defined above):

1. The tuple $(\{1, 2, 3, 4\}, \Pr)$ where $\Pr(n) = \frac{1}{2\text{rem}(n,2)+4}$ for all $n \in \{1, 2, 3, 4\}$.
2. The tuple $(\mathbb{N}, \Pr)$ where $\Pr(n) = 2^{-n-1}$ for all $n \in \mathbb{N}$. Note that $1/2+1/4+1/8 \ldots = 1$

Solution(s)

1. This is not a valid probability space. Given the tuple, we have $\Pr(1) = 1/6, \Pr(2) = 1/4, \Pr(3) = 1/6, \Pr(4) = 1/4$, which does not sum to 1.
2. This is a valid probability space.

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 $E_1, ..., E_n$, $\Pr(E_1 \cup ... \cup E_n) \leq \Pr(E_1) + ... + \Pr(E_n)$

Practice Problem(s)

Show that $\Pr(A) = \Pr(A \cap B) + \Pr(A \cap \overline{B})$. (This should be very straightforward from one of the rules.)

Solution(s)

We can split A into two disjoint events: $(A \cap B) \cup (A \cap \overline{B})$. From the sum rule, $\Pr(A) = \Pr((A \cap B) \cup (A \cap \overline{B})) = \Pr(A \cap B) + \Pr(A \cap \overline{B})$.

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!

Practice Problem(s)

1. 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. Let A, B, and C be mutually independent events.

   a) Prove that $A \cap B$ and C are independent.

   b) Prove that $A \cup B$ and C are independent.

2. Suppose that we have two pints of Half Baked^{TM}, each containing a mix of cookie dough ice cream and fudge brownie ice cream. One pint contains three times as much cookie dough as fudge brownie. The other pint contains three times as many fudge brownie as cookie dough. Suppose we choose one of these pints at random. From this pint, we select five spoonfuls at random with replacement (i.e. each selected spoonful is returned to its respective pint, because we're looking for the perfect bite). Each spoonful necessarily has either one brownie or one cookie. The result is that we find 4 spoonfuls of cookie dough and one spoonful of fudge brownie. What is the probability that we were using the pint with mainly cookie dough?

3. Of the high blood pressure patients in a particular clinic, 62 % are treated with medication X, the remainder with medication Y. It is known that 1.4% of the patients using medication X suffer from fainting spells, as do 2.9% of the patients using medication Y. A patient known by the clinic to have high blood pressure suffers a fainting spell but does not remember which medication she is on. Which medication is she more likely to be taking?

4. It is estimated that 50% of emails are spam emails. Some software has been applied to filter these spam emails before they reach your inbox. A certain brand of software claims that it can detect 99% of spam emails, and the probability for a false positive (a non-spam email detected as spam) is 5%. Now if an email is detected as spam, then what is the probability that it is in fact a non-spam email?

Solution(s)

1. a) We have $\Pr((A \cap B) \cap C) = \Pr(A \cap B \cap C)$. Given that A, B, C are mutually independent, $\Pr(A \cap B \cap C) = \Pr(A)\Pr(B)\Pr(C) = (\Pr(A)\Pr(B))\Pr(C)$. So $A \cap B$ and C are independent.

   b) $\Pr((A \cup B) \cap C) = \Pr((A \cap C) \cup (B \cap C)) = \Pr(A \cap C) + \Pr(B \cap C) - \Pr(A \cap B \cap C)$.

   Since A, B, C are mutually independent, $\Pr(A \cap C) = \Pr(A)\Pr(C)$, $\Pr(B \cap C) = \Pr(B)\Pr(C)$, $\Pr(A \cap B \cap C) = \Pr(A)\Pr(B)\Pr(C)$.

   So,

$$\Pr((A \cup B) \cap C) = \Pr(A)\Pr(C) + \Pr(B)\Pr(C) - \Pr(A)\Pr(B)\Pr(C)$$

$$= \Pr(A \cup B)\Pr(C).$$

Thus, $A \cup B$ and C are independent.

2. Let C be the event that we were using the pint with mainly cookie dough, and let B be the event that we were using the pint with mainly brownie dough. 4C1B represents the event that we got 4 spoonfuls of cookie dough and one spoonful of fudge brownie.

$$\Pr(C \mid \text{4C1B}) = \frac{\Pr(\text{4C1B} \mid C)\Pr(C)}{\Pr(\text{4C1B} \mid C)\Pr(C) + \Pr(\text{4C1B} \mid B)\Pr(B)}$$

$$= \frac{(3/4)^4(1/4)(1/2)}{(3/4)^4(1/4)(1/2) + (1/4)^4(3/4)(1/2)} = \frac{81}{84} = \frac{27}{28}.$$

$\Pr(C \mid \text{4C1B}) = \frac{27}{28} \approx 96.4\%$

3.

$$\Pr(X \mid Faint) = \frac{0.014 \cdot 0.62}{0.014 \cdot 0.62 + 0.029 \cdot 0.38} = \frac{0.00868}{0.00868 + 0.01102} \approx 0.441,$$

$$\Pr(Y \mid Faint) = \frac{0.029 \cdot 0.38}{0.014 \cdot 0.62 + 0.029 \cdot 0.38} = \frac{0.01102}{0.00868 + 0.01102} \approx 0.559.$$

The patient is more likely on medication Y.

4. We want to calculate $\Pr(X|Y)$, where X represents a non-spam email and Y represents an email being detected as spam. Using Bayes' Rule, $\Pr(X|Y) = \Pr(X \cap Y)/\Pr(Y)$, where $\Pr(X \cap Y)$ is the probability of a non-spam email being marked as scam and $\Pr(Y)$ is the probability that an email is detected as spam.

$\Pr(X \cap Y) = (0.5 * 0.05) = 0.025$
Using the law of total probability, $\Pr(Y) = \Pr(X \cap Y) + \Pr(\overline{X} \cap Y) = (0.5 * 0.99) + (0.5 * 0.05) = 0.52$.

Therefore, the answer is ≈ 0.0481.

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.

Practice Problem(s)

1. 100 homeworks are on the table, with two questions to be graded. Andy is in charge of grading question one and Patrick is in charge of grading question two. First, Andy grades 30 homeworks at random (each homework has probability 0.3 of being graded by Andy). Next, Patrick grades 50 homeworks at random (each homework has probability .5 of being graded by Patrick). Assume Andy and Patrick make their choices independently.

   a) Let A be the number of homeworks that Andy graded. What is $\mathbb{E}[A]$?

   b) Let P be the number of homeworks that Patrick graded. What is $\mathbb{E}[P]$?

   c) Let a "fully graded" homework be one graded by both Andy and Patrick. Let B be the number of full graded homeworks . What is $\mathbb{E}[B]$?

   d) Let a "good pair" be a pair of adjacent homeworks such that both questions are graded. Let G be the number of good pairs. What is $\mathbb{E}[G]$?

2. A fair coin is tossed repeatedly until either it lands on heads or a total of five tosses have been made, whichever comes first. Let X be the random variable denoting the number of tosses made. What is the probability distribution for X?

3. A supervisor in a manufacturing plant has three men and three women working for him. He wants to choose two workers for a special job. He decides to select the two workers at random. Let Y denote the number of women in his selection. Find the probability distribution for Y.

Solution(s)

1. a) $\mathbb{E}[A] = 30$ homeworks.

   b) $\mathbb{E}[P] = 50$ homeworks.

   c) A single homework is graded by both with probability $.5 \cdot .3 = .15$ since Andy and Patrick grade independently so we have $\mathbb{E}[B] = .15 \cdot 100 = 15$.

   d) In total, there are 99 adjacent homework pairs $(1-2, 2-3, \text{etc.}, 99-100)$. For a given pair, the probability that Andy grades both is

   $$\frac{30}{100} \cdot \frac{29}{99}$$

   Likewise, the probability that Patrick grades both is

   $$\frac{50}{100} \cdot \frac{49}{99}$$

Since they grade independently, the probability that they both grade both is

$$\frac{30}{100} \cdot \frac{29}{99} \cdot \frac{50}{100} \cdot \frac{49}{99}$$

Thus, by linearity of expectation we get

$$\mathbb{E}[G] = 99 \cdot \frac{30}{100} \cdot \frac{29}{99} \cdot \frac{50}{100} \cdot \frac{49}{99} \approx 2.153$$

2. The possible outcomes are $\{H, TH, TTH, TTTH, TTTTH, TTTTT\}$. Because the probability of flipping heads and tails are both $\frac{1}{2}$, $\Pr(H) = \frac{1}{2}$, $\Pr(TH) = \frac{1}{4}$, $\Pr(TTH) = \frac{1}{8}$, $\Pr(TTTH) = \frac{1}{16}$, $\Pr(TTTTH) = \frac{1}{32}$, and $\Pr(TTTTT) = \frac{1}{32}$. Thus we have the distribution of X is given by

$$f_X(1) = \frac{1}{2}$$
$$f_X(2) = \frac{1}{4}$$
$$f_X(3) = \frac{1}{8}$$
$$f_X(4) = \frac{1}{16}$$
$$f_X(5) = \frac{1}{32} + \frac{1}{32} = \frac{1}{16}$$

3. The possible values of Y are $\{0, 1, 2\}$, and we have $\binom{6}{2} = 15$ ways to choose 2 workers. To choose 0 women, we need to choose 2 men, of which there are $\binom{3}{2} = 3$ ways. To choose 1 woman, we need to choose 1 woman and 1 man, so there are $\binom{3}{1} \times \binom{3}{1} = 3 \times 3 = 9$ ways. Lastly, to choose 2 women, we need to choose 2 women, so again there are 3 ways. So, $\Pr(Y = 0) = \frac{3}{15} = 0.2$, $\Pr(Y = 1) = \frac{9}{15} = 0.6$, and $\Pr(Y = 2) = \frac{3}{15} = 0.2$.

$$f_Y(0) = 0.2$$
$$f_Y(1) = 0.6$$
$$f_Y(2) = 0.2$$

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 $a_1, ..., a_n$,

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

Practice Problem(s)

1. A coin is biased so that the probability a head comes up when it is flipped is 0.6. What is the expected number of heads that come up when it is flipped 10 times?

2. The final exam of a discrete mathematics course consists of 50 true/false questions, each worth two points, and 25 multiple-choice questions, each worth four points. The probability that Linda answers a true/false question correctly is 0.9, and the probability that she answers a multiple-choice question correctly is 0.8. What is her expected score on the final?

Solution(s)

1. 6

2. $50 * 0.9 * 2 + 25 * 0.8 * 4 = 170$

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* $\text{Var}[R]$ 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,

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

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

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

Practice Problem(s)

1. Let X be the random variable that denotes the number that comes up when a fair die is rolled. What is the variance of X?

2. What is the variance of the number of heads that come up when a fair coin is flipped 4 times?

3. Suppose that the number of tin cans recycled in a day at a recycling center is a random variable with an expected value of 50,000 and a variance of 10,000.

   a) Use Markov's inequality to find an upper bound on the probability that the center will recycle more than 55,000 cans on a particular day.

   b) Use Chebyshev's inequality to provide a lower bound on the probability that the center will recycle 40,000 to 60,000 cans on a certain day.

Solution(s)

1. We have

$$E[X] = \frac{1+2+3+4+5+6}{6} = 3.5, \quad E[X^2] = \frac{1^2+2^2+3^2+4^2+5^2+6^2}{6} = \frac{91}{6}.$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \tfrac{91}{6} - (3.5)^2 = \tfrac{35}{12}.$$

2. Let X be the number of heads that come up over 4 flips. We have

$$E[X] = 4 * 0.5 = 2, \text{ so } E[X]^2 = 4$$

To calculate $E[X^2]$, we can use the following formula:

$$E[X^2] = \sum_{k=0}^{4} k^2 * \Pr(X = k) \text{ , where } \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Thus, $E[X^2] = 0^2 \binom{4}{0} \frac{1}{2}^4 + 1^2 \binom{4}{1} \frac{1}{2}^4 + 2^2 \binom{4}{2} \frac{1}{2}^4 + 3^2 \binom{4}{3} \frac{1}{2}^4 + 4^2 \binom{4}{4} \frac{1}{2}^4 = 80 * \frac{1}{2}^4 = 5$

Then,

$$\text{Var}(X) = E[X^2] - (E[X])^2 = 5 - 4 = 1$$

3. a) Given $E[X] = 50{,}000$, we want $\Pr(X \geq 55{,}000)$.

   Markov gives:

$$\Pr(X \geq 55{,}000) \leq \frac{E[X]}{55{,}000} = \frac{50{,}000}{55{,}000} = \tfrac{10}{11} \approx 0.909.$$

   b) We want $\Pr(40{,}000 \leq X \leq 60{,}000) = \Pr(|X - 50{,}000| < 10{,}000)$.

   Chebyshev gives:

$$\Pr(|X - \mathbb{E}[X]| \geq a) \leq \frac{\text{Var}(X)}{a^2}.$$

   Here $\mathbb{E}[X] = 50{,}000$, so $a = 10{,}000$.

$$\Pr(|X - 50{,}000| \geq 10{,}000) \leq \frac{10{,}000}{(10{,}000)^2} = \frac{1}{10{,}000} = .0001.$$

   Thus

$$\Pr(40{,}000 \leq X \leq 60{,}000) \geq 1 - .0001.$$