Lecture 4: Discrete Probability Theory

Alfred Rossi, Ph.D.

Let Ω denote a set of "atomic" events. This is known as a sample space. Examples:

- Flipping a coins, $\Omega = \{H, T\}$.
- Rolling a six sided dice once, $\Omega = [6] = \{1, 2, \ldots, 6\}$.
- Flipping 2 coins, $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$.
- Rolling two six sided dice,

$$\begin{aligned}\Omega &= [6] \times [6] = \{(i, j) : 1 \leq i, j \leq 6\} \\ &= \{(1, 1), (1, 2), \ldots (2, 1), (2, 2), \ldots (6, 6)\}.\end{aligned}$$

A (discrete) ***probability space*** $(\Omega, \Pr)$ models a randomized experiment, where Ω is a set containing all possible outcomes, where each $e \in \Omega$ which occurs with probability $\Pr(e)$.

- Intuitively, because $\Pr$ represents a probability, $\Pr : \Omega \to [0, 1]$.
- Because the experiment must give an outcome, contained in Ω,

$$\sum_{e \in \Omega} \Pr(e) = 1.$$

Examples of probability spaces:

- One flip of a fair coin: $\Omega = \{H, T\}$, and $\Pr(H) = \Pr(T) = 1/2$.
- One roll of a fair six-sided dice $\Omega = \{1, 2, 3, 4, 5, 6\}$, and $\Pr(i) = 1/6$ for $1 \leq i \leq 6$.
- k flips of a fair coin, $\Omega = \{H, T\}^k$, $\Pr(e) = 1/2^k$ for all $e \in \Omega$.
- The *uniform distribution*: Ω is finite and for all $e \in \Omega$, $\Pr(e) = 1/|\Omega|$.

An event, E is a subset of Ω. We define $\Pr[E] = \sum_{e \in E} \Pr(e)$.

Let E (O) denote the event that a fair, 6-sided, dice-roll is *even* (*odd*, respectively), then $E = \{2, 4, 6\}$ ($O = \{1, 3, 5\}$).

$$\begin{aligned}
\Pr[E] &= \Pr(2) + \Pr(4) + \Pr(6) \\
&= 1/6 + 1/6 + 1/6 = 1/2. \\
\Pr[O] &= \Pr(1) + \Pr(3) + \Pr(5) \\
&= 1/6 + 1/6 + 1/6 = 1/2.
\end{aligned}$$

What is the probability that a roll of a fair 20-sided dice yields a number less than 8 or more than 18?

- We're talking about events over the probability space where $\Omega = [20]$, and $\Pr(i) = 1/20$ for all $i \in \Omega$.
- The event E is $E = \{1, 2, 3, 4, 5, 6, 7, 19, 20\}$.
- It follows that,

$$
\begin{aligned}
Pr[E] &= \sum_{i \in E} \Pr(i) = 1/20 \sum_{i \in E} 1 = 1/20\,|E| \\
&= 9/20 = 0.45.
\end{aligned}
$$

What is the probability of rolling a 20-sided dice and getting a number less than 8 or more than 18 where the dice biased towards even values such that $\Pr(i) = 1/15$ when i is even and $\Pr(i) = 1/30$ when i is odd?

- We're talking about events over the probability space where $\Omega = [20]$, and $\Pr(i) = 1/15$ when i is even, and $\Pr(i) = 1/30$ when i is odd.
- The event E is $E = \{1, 2, 3, 4, 5, 6, 7, 19, 20\}$.
- It follows that,

$$\Pr[E] = \sum_{i \in E} \Pr(i) = \sum_{i \in E,\, i \text{ is even}} 1/15 + \sum_{i \in E,\, i \text{ is odd}} 1/30$$
$$= 1/15 \cdot 4 + 1/30 \cdot 5 = 13/30 \approx 0.43.$$

Notation

- A set of sets is called a ***set-system***.
- We will sometimes write $\{E_i\}_{i=1}^n$ to denote a set-system comprised of sets E_1, E_2, $\ldots$, E_n.
- Two sets A, B are ***disjoint*** if $A \cap B = \varnothing$.
- A set-system $\mathcal{S}$ is said to be ***pairwise disjoint*** if, $A, B \in \mathcal{S}$, are disjoint whenever $A \neq B$.
- Let Ω be a set. We say a set-system, $\mathcal{S}$, is a set-system over Ω if every set in $\mathcal{S}$ is a subset of Ω.
- Let Ω be a set. We say that a set-system $\mathcal{S}$ ***partitions*** Ω provided that $\mathcal{S}$ is pair-wise disjoint, and $\bigcup_{E \in \mathcal{S}} E = \Omega$.

Let $(\Omega, \Pr)$ be a (discrete) probability space, and let $E_1, E_2 \subset \Omega$ be ***disjoint*** events, then

$$\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2]$$

Proof.

Note that for any $e \in E_1 \cup E_2$, it must be that either $e \in E_1$ or $e \in E_2$ but not both. Therefore,

$$\begin{aligned}\Pr[E_1 \cup E_2] &= \sum_{e \in E_1 \cup E_2} \Pr(e) = \sum_{e \in E_1} \Pr(e) + \sum_{e \in E_2} \Pr(e) \\ &= \Pr[E_1] + \Pr[E_2].\end{aligned}$$

$\square$

Let $(\Omega, \Pr)$ be a (discrete) probability space, and let $E_1, E_2 \subset \Omega$ be events, then

$$\Pr[E_1 - E_2] = \Pr[E_1] - \Pr[E_1 \cap E_2]$$

Proof.

Note that E_1 is the union of disjoint sets $(E_1 - E_2)$ and $E_1 \cap E_2$. Therefore,

$$\begin{aligned} \Pr[E_1] &= \Pr[(E_1 - E_2) \cup (E_1 \cap E_2)] \\ &= \Pr[E_1 - E_2] + \Pr[E_1 \cap E_2]. \end{aligned}$$

It follows immediately that

$$\Pr[E_1 - E_2] = \Pr[E_1] - \Pr[E_1 \cap E_2].$$

□

Let $(\Omega, \Pr)$ be a (discrete) probability space, and let $E_1, E_2 \subset \Omega$ be events, then

$$\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2] - \Pr[E_1 \cap E_2].$$

Proof.

The set $E_1 \cup E_2$ can be written as the pairwise disjoint union of sets $E_1 - E_2$, $E_1 \cap E_2$, and $E_2 - E_1$. Therefore,

$$\begin{aligned}
\Pr[E_1 \cup E_2] &= \Pr[(E_1 - E_2) \cup (E_1 \cap E_2) \cup (E_2 - E_1)] \\
&= \Pr[E_1 - E_2] + \Pr[E_1 \cap E_2] + \Pr[E_2 - E_1] \\
&= \Pr[E_1] - \Pr[E_1 \cap E_2] + \Pr[E_1 \cap E_2] \\
&\quad + \Pr[E_2] - \Pr[E_2 \cap E_1] \\
&= \Pr[E_1] + \Pr[E_2] - \Pr[E_1 \cap E_2].
\end{aligned}$$

$\square$

Let $(\Omega, \Pr)$ be a (discrete) probability space, and $E \subset \Omega$ be an event. Then $E^c = (\Omega - E)$ is an event with $\Pr[E^c] = 1 - \Pr[E]$.

Proof.

Note that $\Omega = E \cup E^c$, and that E and E^c are disjoint. Therefore,

$$\begin{aligned} 1 &= \sum_{e \in \Omega} \Pr(e) \\ &= \sum_{e \in E} \Pr(e) + \sum_{e \in E^c} \Pr(e) \\ &= \Pr[E] + \Pr[E^c]. \end{aligned}$$

It follows immediately that $\Pr[E^c] = 1 - \Pr[E]$. $\square$

A biased coin is heads with probability p. What is the probability of seeing at least two heads if it is flipped 3 times?

- $\Omega = \{H, T\}^3$, $\Pr(e) = p^{n_H}(1-p)^{3-n_H}$, where n_H denotes the number of heads in the triple (example: $\Pr((H, H, T)) = p^2(1-p)$).
- Let E_i denote the event where the triple contains exactly i heads for $0 \le i \le 3$. Note that E_i partition Ω. That is, Ω is their union but $E_i \cap E_j = \varnothing$ for all $i \ne j$, $0 \le i, j \le 3$.
- We are uninterested in E_0 or E_1. The event we are after is $E = E_2 \cup E_3$. Since they are disjoint, $\Pr[E] = \Pr[E_2] + \Pr[E_3]$.
- Note $E_2 = \{(H, H, T), (H, T, H), (T, H, H)\}$, Note $\Pr[E_2] = 3p^2(1-p)$. Similarly $E_3 = \{(H, H, H)\}$, $\Pr[E_3] = p^3$. Therefore,

$$\Pr[E] = \Pr[E_2] + \Pr[E_3] = 3p^2(1-p) + p^3.$$

A biased coin is heads with probability p. What is the probability of seeing at least two heads if it is flipped n times?

Hard way:

- $\Omega = \{H, T\}^n$, $\Pr(e) = p^{n_H}(1-p)^{n-n_H}$, where n_H denotes the number of heads in the event n-tuple.
- Let E_i denote the event where the event contains exactly i heads for $0 \leq i \leq n$. Note that E_i partition Ω.
- We are uninterested in E_0 or E_1. The event we are after is $E = E_2 \cup E_3 \cup \cdots \cup E_n$. Since they are disjoint, $\Pr[E] = \Pr[E_2] + \Pr[E_3] + \cdots + \Pr[E_n]$.
- Note: for $e \in E_i$, $\Pr(e) = p^i(1-p)^{n-i}$. Therefore,

$$\Pr[E_i] = \sum_{e \in E_i} \Pr(e) = p^i(1-p)^{n-i}|E_i|.$$

- Recall, $\Pr[E_i] = p^i(1-p)^{n-i}|E_i|$.
- To figure out $|E_i|$, note that this is equal to the number of ways to choose i items to be heads out out n. There are n different positions to choose for the first, $n-1$ for the second, ..., $n-(i-1) = n-i+1$ ways for the i-th. So there are $n \cdot (n-1) \cdot (n-2) \cdots (n-i+1) = n!/(n-i)!$, but this logic is order-dependent, and so we count the same set of picked positions $i!$ times. Therefore,

$$|E_i| = \frac{n!}{i!(n-i)!} = \binom{n}{i}.$$

- So,

$$\Pr[E_i] = \binom{n}{i} p^i (1-p)^{n-i}$$

$$\Pr[E] = \sum_{i=2}^{n} \Pr[E_i] = \sum_{i=2}^{n} \binom{n}{i} p^i (1-p)^{n-i}.$$

Recall, from some past math class, the binomial theorem:

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}.$$

Picking $x = p$, $y = 1 - p$ gets us pretty close with

$$1 = (p + 1 - p)^n = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i},$$

except the RHS has two extra terms $\binom{n}{0} p^0 (1-p)^n = (1-p)^n$, and $\binom{n}{1} p^1 (1-p)^{n-1} = np(1-p)^{n-1}$.

Subtracting these terms from both sides,

$$\begin{aligned}\Pr[E] &= \sum_{i=2}^{n} \binom{n}{i} p^i (1-p)^{n-i} \\ &= 1 - (1-p)^n - np(1-p)^{n-1}.\end{aligned}$$

The easy way. Use the event algebra.

- Recall $\{E_i\}_{i=0}^n$ partition Ω. So E_i's are pairwise disjoint and $\Omega = \bigcup_{i=0}^n E_i$.
- Recall, $E = E_2 \cup E_3 \cup \cdots \cup E_n$.
- Recall, $\Pr[E] = 1 - \Pr[E^c]$
- Recall, $E^c = \Omega - E$. It follows that $E^c = E_0 \cup E_1$.
- Recall, E_0 is the event consisting of n-tuples with no heads. There is just one of those $(T, T, \ldots T)$, and it occurs with probability $(1-p)^n$. So $\Pr[E_0] = (1-p)^n$.
- Recall, E_1 is the event consisting of n-tuples with exactly 1 heads. There are n of these, one for when the H occurs in each possible position. Each occur with probability $p(1-p)^{n-1}$, so $\Pr[E_1] = np(1-p)^{n-1}$.

Therefore,

$$\begin{aligned}
\Pr[E] &= 1 - \Pr[E^c] && \text{(by the event algebra)} \\
&= 1 - \Pr[E_0 \cup E_1] \\
&= 1 - \Pr[E_0] - \Pr[E_1] && \text{(since } E_0 \cap E_1 = \varnothing\text{)} \\
&= 1 - (1-p)^n - np(1-p)^{n-1}.
\end{aligned}$$

Note! Even though we essentially did the same thing, we did not have to know anything about the binomial theorem or $\binom{n}{i}$.

A ***random variable*** is a function $x : \Omega \to \mathbb{R}$.

The ***expectation*** of x, denoted $\mathbb{E}[x]$, is given by

$$\mathbb{E}[x] = \sum_{e \in \Omega} x(e) \Pr(e).$$

Let $c \in \mathbb{R}$. We use the notation $\{x = c\}$ to denote the event $x = c$. That is, if

$$\{x = c\} = \{e : e \in \Omega \text{ and } x(e) = c\}.$$

It is sometimes useful to express the expectation of x as

$$\mathbb{E}[x] = \sum_{c \in \mathrm{Range}(x)} c \Pr[\{x = c\}].$$

This works because the events $E = \{\{x = c\} : c \in \mathrm{Range}(x)\}$ partition Ω, so

$$\begin{aligned}
\mathbb{E}[x] &= \sum_{e \in \Omega} x(e)\,\Pr(e) \\
&= \sum_{c \in \mathrm{Range}(x)} \sum_{e \in \{x=c\}} x(e)\,\Pr(e) \\
&= \sum_{c \in \mathrm{Range}(x)} \sum_{e \in \{x=c\}} c\,\Pr(e) \\
&= \sum_{c \in \mathrm{Range}(x)} c \sum_{e \in \{x=c\}} \Pr(e) \\
&= \sum_{c \in \mathrm{Range}(x)} c\,\Pr[\{x = c\}]
\end{aligned}$$

Expressions of random variables are random variables. E.g. Let X, Y be random variables. Then $S = X + Y$ is a random variable. That is $S : \Omega \to \mathbb{R}$ such that $S(e) = X(e) + Y(e)$.

Indicator Variables

Let S be a set. A ***predicate*** on S is a function $p : S \to \{\mathsf{T}rue, \mathsf{F}alse\}$.
Let p be a ***predicate*** on the sample space, Ω. An ***indicator*** variable for a predicate p, denoted $\mathbb{1}_p$, is the random variable

$$\mathbb{1}_p(e) = \begin{cases} 1 & \text{if } p(e) = \mathsf{T}rue, \\ 0 & \text{otherwise.} \end{cases}$$

Cool Trick!

The expectation of an indicator variable gives the probability that the predicate holds for any $e \in \Omega$. To see this, observe

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_p] &= \sum_{a \in \mathrm{Range}(\mathbb{1}_p)} a \Pr\left[\{\mathbb{1}_p = a\}\right] \\
&= \sum_{a \in \{0,1\}} a \Pr\left[\{\mathbb{1}_p = a\}\right] \\
&= 0 \cdot \Pr\left[\{\mathbb{1}_p = 0\}\right] + 1 \cdot \Pr\left[\{\mathbb{1}_p = 1\}\right] \\
&= \Pr\left[\{\mathbb{1}_p = 1\}\right] \\
&= \Pr\left[\{e : p(e) = \mathsf{T}rue\}\right].
\end{aligned}$$

This may look like a lot of nothing but it will prove to save us work!

Linearity of Expectation

Consider $\mathbb{E}[X+Y]$. We have,

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_{e\in\Omega} [X(e)+Y(e)]\Pr(e) \\
&= \sum_{e\in\Omega} X(e)\Pr(e) + \sum_{e\in\Omega} Y(e)\Pr(e) \\
&= \mathbb{E}[X] + \mathbb{E}[Y]
\end{aligned}$$

In other words, the ***expectation of a sum is the sum of expectations***! (May not seem that exciting but this will save us work.)

Linearity of Expectation

Also, consider multiplication of a random variable, X, by a constant, $c \in \mathbb{R}$. The result, cX, is a random variable. Observe,

$$\mathbb{E}[cX] = \sum_{e \in \Omega} cX(e)\,\Pr(e) = c \sum_{e \in \Omega} X(e)\,\Pr(e) = c\mathbb{E}[X]$$

Note, however, that

$$\mathbb{E}[XY] \neq \mathbb{E}[X] \cdot \mathbb{E}[Y].$$

To see why, observe

$$
\begin{aligned}
\mathbb{E}[X] \cdot \mathbb{E}[Y] &= \sum_{x \in \mathrm{Range}(X)} x \Pr[\{X = x\}] \sum_{y \in \mathrm{Range}(Y)} y \Pr[\{Y = y\}] \\
&= \sum_{x \in \mathrm{Range}(X)} \sum_{y \in \mathrm{Range}(Y)} xy \Pr[\{X = x\}] \Pr[\{Y = y\}].
\end{aligned}
$$

While,

$$
\begin{aligned}
\mathbb{E}[XY] &= \sum_{z \in \mathrm{Range}(XY)} z \Pr[\{XY = z\}] \\
&= \sum_{x \in \mathrm{Range}(X)} \sum_{y \in \mathrm{Range}(Y)} xy \Pr[\{XY = xy\}].
\end{aligned}
$$

It follows that these are equal if X, Y are independent. That is,

$$
\Pr[\{XY = xy\}] = \Pr[\{X = x\}] \Pr[\{Y = y\}].
$$

Consider n flips of a biased coin that is heads with probability p. What is X, the expected number of heads? (Long solution.)

$$\mathbb{E}[X] = \sum_{i=0}^{n} i \Pr[\{x = i\}] = \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i}$$

This kind of looks like a binomial expansion of $(p + (1-p))^n$ but there is an i in the way.

Note that,

$$i\binom{n}{i} = n\binom{n-1}{i-1}.$$

To see why, consider two ways to count number of committees formed of i people selected from n people where one of the committee members serves as president:

- On one hand, we have $\binom{n}{i}$ choices of committee members and, for each, i choices for president. Thus, there are $i\binom{n}{i}$ ways to pick this committee.
- On the other, if we pick the president first we have n choices of president. We then must choose the other $i-1$ committee members from the remaining $n-1$ people. That is, we have $n\binom{n-1}{i-1}$ ways to pick this committee.

Consider n flips of a biased coin that is heads with probability p. What is X, the expected number of heads? (Long solution.)

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{n} i \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^{n} i \binom{n}{i} p^i (1-p)^{n-i} \\
&= n \sum_{i=1}^{n} \binom{n-1}{i-1} p^i (1-p)^{n-i} \\
&= n \sum_{i=0}^{n-1} \binom{n-1}{i} p^{i+1} (1-p)^{n-(i+1)} \\
&= np \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{(n-1)-i} \\
&= np(p + (1-p))^{n-1} = np,
\end{aligned}$$

where the second-to-the-last equality follows from the binomial theorem.

Consider n flips of a biased coin that is heads with probability p. What is X, the expected number of heads? (Short solution.)

Let X_i be the indicator variable that the i-th flip is heads. That is,

$$X_i((f_1, f_2, \ldots, f_n)) = \begin{cases} 1 & \text{if } f_i = H, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $X = X_1 + X_2 + \ldots + X_n$, and that $\mathbb{E}[X_i] = p$. Observe,

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=0}^{n} p = np.$$

Moral of the Story: Combining ***indicator variables*** with ***linearity of expectation*** can ***save us work***!