Lecture #10: Discrete Probability for Computer Science I
Key Concepts

Much of the material preceding the section on “Random Variables” should be review material that you saw — might not recall — from CPSC 251.

Experiments and Sample Spaces

An ***experiment*** is a procedure (or process) that yields one of a given set of possible ***outcomes***. The set of possible outcomes of the experiment — which we will often name Ω — is called the ***sample space***. In this course we will (almost always) consider experiments where the sample space, Ω, is countable — studying a part of probability theory called ***discrete probability theory***.

Events

An ***event*** is a subset of the experiment's sample space Ω. Events are used to model the kinds of “things that are interested in” that will be studied.

An ***elementary event*** is a set of size one — that is, it is an event that only includes a single outcome. (Some references might use the name “outcome” for these too.)

Probability Distributions

Consider an experiment with sample space Ω. A ***probability distribution*** is a (total) function $\mathsf{P} : \Omega \to \mathbb{R}$ such that $0 \leq \mathsf{P}(x) \leq 1$ for every outcome $x \in \Omega$, and such that

$$\sum_{x \in \Omega} \mathsf{P}(x) = 1.$$

For any set Ω, let $\mathcal{P}(\Omega)$ denote the set of all ***subsets*** of Ω. A probability distribution P (on an experiment with a countable sample space) is "extended" to get a function

$$\mathsf{P} : \mathcal{P}(\Omega) \to \mathbb{R}$$

by setting

$$\mathsf{P}(A) = \sum_{x \in A} \mathsf{P}(x)$$

for every event $A \subseteq \Omega$ (that is, for all $A \in \mathcal{P}(\Omega)$).

Uniform Distributions

If Ω is a finite set then the ***uniform probability distribution*** (for Ω) defines the probability of every outcome to be the same: This is the function $\mathsf{P} : \Omega \to \mathbb{R}$ such that

$$\mathsf{P}(x) = \frac{1}{|\Omega|}$$

for every outcome $x \in \Omega$.

Suppose that $A \subseteq \Omega$ is an event. Then, if P is the uniform distribution for Ω, then

$$\begin{aligned}\mathsf{P}(A) &= \sum_{x \in A} \mathsf{P}(x) \\ &= \sum_{x \in A} \tfrac{1}{|\Omega|} \\ &= \frac{|A|}{|\Omega|}.\end{aligned}$$

Probability of the Complement of an Event

If Ω is a sample space for an experiment and $A \subseteq \Omega$ is an event, then the ***complement***[1] of the event A, $\overline{A}$, is the set of outcomes that *are not* in A.

$$\overline{A} = \{x \in \Omega \mid x \notin A\}.$$

Theorem 1. *Let Ω be a sample space with probability distribution P : $\Omega \to \mathbb{R}$, and let $A \subseteq \Omega$. Then the probability of the complement, $\overline{A}$, of the event A is* $\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A)$.

[1] It is also OK if you use A^C to represent the complement of A, as we did for *languages*. These are both commonly used as the name for this set.

Probability of the Union of Events

Theorem 2. *Let Ω be a sample space with probability distribution P : $\Omega \to \mathbb{R}$. Then, for any events $A, B \subseteq \Omega$,*

$$\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B).$$

Corollary 3. *Let Ω be a sample space with probability distribution P : $\Omega \to \mathbb{R}$. Then, for any events $A, B \subseteq \Omega$,*

$$\mathsf{P}(A \cup B) \leq \mathsf{P}(A) + \mathsf{P}(B).$$

Theorem 4 (Union Bound)**.** *Let Ω be a sample space with probability distribution P : $\Omega \to \mathbb{R}$, let k be a positive integer, and let $E_1, E_2, \ldots, E_k \subseteq \Omega$. Then*

$$\mathsf{P}(E_1 \cup E_2 \cup \cdots \cup E_k) \leq \sum_{i=1}^{k} \mathsf{P}(E_i).$$

Conditional Probability

Let Ω be a sample space and let $A, B \subseteq \Omega$ be events such that $\mathsf{P}(B) > 0$. The ***conditional probability*** of A given B, denoted $\mathsf{P}(A \mid B)$, is

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}.$$

$\mathsf{P}(A \mid B)$ is not defined if $\mathsf{P}(B) = 0$.

Several results, that can be useful when you need to compute probabilities, were introduced:

Claim 5 (Law of Total Probability)**.** *Let Ω be a sample space and let $P : \Omega \to \mathbb{R}$ be a probability distribution. Then, for any events A and B.*

$$\mathsf{P}(A) = \mathsf{P}(A \mid B) \cdot \mathsf{P}(B) + \mathsf{P}(A \mid \overline{B}) \cdot \mathsf{P}(\overline{B}).$$

Claim 6 (Extended Partition Rule)**.** *Let Ω be a sample space, let P : $\Omega \to \mathbb{R}$ be a probability distribution, let k be a positive integer, and let A and $B_1, B_2, \ldots, B_k$ be events satisfying the following properties.*

(a) $B_1, B_2, \ldots, B_k$ *are* pairwise disjoint. *That is, $B_i \cap B_j = \emptyset$ for all integers i and j such that $1 \leq i, j \leq k$ and $i \neq j$.*

(b) $A \subseteq B_1 \cup B_2 \cup \cdots \cup B_k$.

Then

$$\mathsf{P}(A) = \mathsf{P}(A \mid B_1) \cdot \mathsf{P}(B_1) + \mathsf{P}(A \mid B_2) \cdot \mathsf{P}(B_2) + \cdots + \mathsf{P}(A \mid B_k) \cdot \mathsf{P}(B_k).$$

Claim 7 (Baye's Theorem)**.** *Let Ω be a sample space, let $P : \Omega \to \mathbb{R}$, and let A and B be events such that* $\mathsf{P}(A) > 0$ *and* $\mathsf{P}(B) > 0$*. Then*

$$\mathsf{P}(A \,|\, B) = \frac{\mathsf{P}(B \,|\, A) \cdot \mathsf{P}(A)}{\mathsf{P}(B)}.$$

Independence

Events A and B are said to be ***independent*** if

$$\mathsf{P}(A \cap B) = \mathsf{P}(A) \times \mathsf{P}(B).$$

This can be generalized in two — different — ways when three or more random variables are being considered:

Let Ω be a sample space with probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$. Let $A_1, A_2, \ldots, A_k \subseteq \Omega$ be events, for some integer $k \geq 2$. The events $A_1, A_2, \ldots, A_k$ are ***mutually independent*** if

$$\mathsf{P}\left(\bigcap_{i \in S} A_i \right) = \prod_{i \in S} \mathsf{P}(A_i) \tag{1}$$

for *every* S of $\{1, 2, \ldots, k\}$.

Note: The condition at line (1) is guaranteed to hold whenever $|S| \leq 1$, so this condition only needs to be considered when $|S| \geq 2$.

The events $A_1, A_2, \ldots, A_k$ are ***pairwise independent*** if

$$\mathsf{P}(A_i \cap A_j) = \mathsf{P}(A_i) \times \mathsf{P}(A_j) \tag{2}$$

for every pair of integers i and j such that $1 \leq i, j \leq k$ and $i \neq j$.

If events $A_1, A_2, \ldots, A_k$ are ***mutually independent*** then they are also ***pairwise independent***. The converse does not always hold.

Random Variables

Let Ω be a sample space. A ***random variable over*** $\boldsymbol{\Omega}$ is a (total) function $X : \Omega \to \mathbb{R}$.

We will often be interested in random variable whose ranges are particular *subsets* V of $\mathbb{R}$ — so that these functions can also be viewed as functions $X : \Omega \to V$ (as well as functions $X : \Omega \to \mathbb{R}$).

An ***indicator random variable*** over a sample space Ω, is a random variable X such that $X(\sigma) \in \{0, 1\}$ for all $\sigma \in \Omega$.

Once again, let Ω be a sample space and let $X : \Omega \to \mathbb{R}$.

- We will write "$X = r$" as the name of the event
$$\{\sigma \in \Omega \mid X(\sigma) = r\} \subseteq \Omega.$$
- We will write "$X \geq r$" as the name of the event
$$\{\sigma \in \Omega \mid X(\sigma) \geq r\} \subseteq \Omega.$$
- "$X \leq r$", "$X > r$", $X < r$", and $X \neq r$" can be used as the names for (corresponding) events in the same way.

The Expected Value of a Random Variable

Let Ω be a sample space with probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$, and let $X : \Omega \to \mathbb{R}$ be a random variable over Ω.

Suppose that
$$\sum_{\sigma \in \Omega} \mathsf{P}(\sigma) \times |X(\sigma)|$$
is finite — that is, " less than $+\infty$".[2]

Then the ***expected value of X, with respect to probability distribution P***, is the value
$$\mathsf{E}[X] = \sum_{\sigma \in \Omega} \mathsf{P}(\sigma) \times X(\sigma).$$

Recall that if Ω is a sample space, $P : \Omega \to \mathbb{R}$ is a probability distribution, and $B \subseteq \Omega$ is an event such that $\mathsf{P}(B) > 0$, then a ***conditional probability distribution*** $P_B : \Omega \to \mathbb{R}$ can be defined by setting
$$\mathsf{P}_B(\sigma) = \begin{cases} \frac{\mathsf{P}(\sigma)}{\mathsf{P}(B)} & \text{if } \sigma \in B, \\ 0 & \text{if } \sigma \notin B \end{cases}$$
for every outcome $\sigma \in \Omega$.

If X is a random variable then the ***conditional expectation of X given B*** is the expected value of X with the respect to the conditional probability P_B:
$$\mathsf{E}[X \mid B] = \sum_{\sigma \in \Omega} \mathsf{P}_B(\sigma) \times X(\sigma).$$

[2]This is a "technical restriction" that you will not need to worry about whenever Ω is a finite set.

Linearity of Expectation

Let Ω be a sample space with a probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$, and let $X_1, X_2, \dots, X_n : \Omega \to \mathbb{R}$ be random variables, for some positive integer n. "$X_1 + X_2 + \cdots + X_n$" denotes a random variable such that

$$(X_1 + X_2 + \cdots + X_n)(\sigma) = X_1(\sigma) + X_2(\sigma) + \cdots + X_n(\sigma)$$

for each outcome $\sigma \in \Omega$. Similarly, if $X : \Omega \to \mathbb{R}$ and $a, b \in \mathbb{R}$ then "$a \cdot X + b$" denotes a random variable such that

$$(aX + b)(\sigma) = a \cdot X(\sigma) + b$$

for every outcome $\sigma \in \Omega$, as well.

Claim 8 (Linearity of Expectation)**.** *Let Ω be a sample space with probability distribution P : $\Omega \to \mathbb{R}$.*

(a) If $X_1, X_2, \dots, X_n : \Omega \to \mathbb{R}$ are random variables over Ω, for a positive integer n, then

$$\mathsf{E}[X_1 + X_2 + \cdots + X_n] = \mathsf{E}[X_1] + \mathsf{E}[X_2] + \cdots + \mathsf{E}[X_n].$$

(b) If $X : \Omega \to \mathbb{R}$ is a random variable over Ω and $a, b \in \mathbb{R}$ then

$$\mathsf{E}[a \cdot X + b] = a \cdot \mathsf{E}[X] + b.$$

Independent Random Variables

Once again, suppose that Ω is a sample space with probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$. Consider random variables $X : \Omega \to V_X$ and $Y : \Omega \to V_Y$, where $V_X, V_Y \subseteq \mathbb{R}$. Recall that, for $a \in V_X$ and $b \in V_Y$,

$$\text{“}X = a\text{”} = \{\sigma \in \Omega \mid X(\sigma) = a\}$$

and

$$\text{“}Y = b\text{”} = \{\tau \in \Omega \mid Y(\tau) = b\},$$

so that

$$\text{“}X = a \wedge Y = b\text{”} = \{\mu \in \Omega \mid X(\mu) = a \text{ and } Y(\mu) = b\}.$$

The above random variables X and Y are ***independent*** if

$$\mathsf{P}(X = a \wedge Y = b) = \mathsf{P}(X = a) \times \mathsf{P}(Y = b)$$

for all values $a \in V_X$ and $b \in V_Y$.

Claim 9. *Let Ω be a sample space with probability distribution $P : \Omega \to \mathbb{R}$, and let $X, Y : \Omega \to \mathbb{R}$. If X and Y are independent random variables then*

$$\mathsf{E}[X \times Y] = \mathsf{E}[X] \times \mathsf{E}[Y].$$

Pairwise Independent Random Variables

Once again, let Ω be a sample space with probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$. Let k be a positive integer and let

$$X_1 : \Omega \to V_1, X_2 : \Omega \to V_2, \ldots, X_k : \Omega \to V_k$$

be random variables over Ω (so that $V_1, V_2, \ldots, V_k \subseteq \mathbb{R}$).

The random variables $X_1, X_2, \ldots, X_k$ are ***pairwise independent*** if X_i and X_j are independent, for every pair of numbers i and j such that $1 \leq i < j \leq k$.

Mutually Independent Random Variables

Finally, let Ω be a sample space with probability distribution $\mathsf{P} : \Omega \to \mathbb{R}$. Let k be a positive integer and let

$$X_1 : \Omega \to V_1, X_2 : \Omega \to V_2, \ldots, X_k : \Omega \to V_k$$

be random variables over Ω (so that $V_1, V_2, \ldots, V_k \subseteq \mathbb{R}$).

The random variables $X_1, X_2, \ldots, X_k$ are ***mutually independent*** if the following condition is satisfied: ***For every*** subset $S \subseteq \{1, 2, \ldots, k\}$ and ***for all*** combinations of $a_i \in V_i$, for $i \in S$,

$$\mathsf{P}\left(\bigwedge_{i \in S} (X_i = a_i)\right) = \prod_{i \in S} \mathsf{P}(X_i = a_i).$$