

Computer Science 351

Random Variables and Expectation

Instructor: Wayne Eberly

Department of Computer Science
University of Calgary

Lecture #20

Learning Goals

Learning Goals:

- Introduce ***random variable*** and their ***expected values*** as a way to consider numerical information that can be considered as part of an experiment (and may be the main reason why you want to consider the experiment, at all).

Random Variables

When we consider an experiment there is often a numerical value that we wish to **count** or **bound**.

Examples:

- When tossing a sequence of coins, how many coins are tossed before a **head** is tossed for the first time?
- When tossing a sequence of n coins, **how many times** is a “head” tossed? at position 0 of the table?

Random Variables

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

- Any **probability distribution** $P : \Omega \rightarrow \mathbb{R}$ is an example of a “random variable over Ω ”.

Random Variables

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 $X : \Omega \rightarrow V$ (as well as functions $X : \Omega \rightarrow \mathbb{R}$).

- For example, an ***integer-valued random variable*** is a random variable $X : \Omega \rightarrow \mathbb{R}$ such that $X(\sigma) \in \mathbb{Z}$ for all $\sigma \in \Omega$ — so that, in effect, $X : \Omega \rightarrow \mathbb{Z}$.
- The (even more special) case that $X : \Omega \rightarrow \mathbb{N}$ will often be of interest too.

Random Variables

So will be the (even more special) case that $X(\sigma) \in \{0, 1\}$ for all $\sigma \in \Omega$, so that $X : \Omega \rightarrow \{0, 1\}$.

- This kind of random variable is often called an ***indicator random variable*** because it “indicates” an event, namely the event

$$\{\sigma \in \Omega \mid X(\sigma) = 1\} \subseteq \Omega.$$

Random Variables

Example: Consider the experiment of tossing a sequence of three coins — so that

$$\Omega = \{(H, H, H), (H, H, T), (H, T, H), (H, T, T), \\ (T, H, H), (T, H, T), (T, T, H), (T, T, T)\}.$$

The **random variable** “number of heads tossed” is the function $X : \Omega \rightarrow \mathbb{N}$ such that

- $X((H, H, H)) = 3.$
- $X((H, H, T)) = X((H, T, H)) = X((T, H, H)) = 2.$
- $X((H, T, T)) = X((T, H, T)) = X((T, T, H)) = 1.$
- $X((T, T, T)) = 0.$

Random Variables

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

- 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.

Expectation

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

Suppose that

$$\sum_{\sigma \in \Omega} P(\sigma) \times |X(\sigma)|$$

is finite — that is, “less than $+\infty$ ”.¹

Then the ***expected value of X , with respect to probability distribution P*** , is the value

$$E[X] = \sum_{\sigma \in \Omega} P(\sigma) \times X(\sigma).$$

¹This is a “technical restriction” that you will not need to worry about whenever Ω is a finite set.

Expectation

- The phrase “with respect to probability distribution P ” will be dropped when it is clear, from context, which probability distribution is being used.
- This value has other names in the literature including
 - the ***mean*** of X ,
 - the ***expectation*** of X , and
 - the ***first moment*** of X .

Conditional Expectation

Recall, from the initial set of notes about discrete probability theory, that if Ω is a sample space, $P : \Omega \rightarrow \mathbb{R}$ is a probability distribution, and $B \subseteq \Omega$ is an event such that $P(B) > 0$, then a **conditional probability distribution** $P_B : \Omega \rightarrow \mathbb{R}$ can be defined by setting

$$P_B(\sigma) = \begin{cases} \frac{P(\sigma)}{P(B)} & \text{if } \sigma \in B, \\ 0 & \text{if } \sigma \notin B \end{cases}$$

for every outcome $\sigma \in \Omega$.

Definition: 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 :

$$E[X | B] = \sum_{\sigma \in \Omega} P_B(\sigma) \times X(\sigma).$$

Conditional Expectation

Exercise: Prove that Ω is a sample space with probability distribution $P : \Omega \rightarrow \mathbb{R}$, $B \subseteq \Omega$ is an event such that $P(B) > 0$, and X is a random variable, then

$$E[X | B] = \frac{1}{P(B)} \times \sum_{\sigma \in B} (P(\sigma) \times X(\sigma)).$$

Conditional Expectation

Recall that — since P_B is also a probability distribution for the sample space Ω whenever B is an event such that $P(B) > 0$, properties of *probability distributions* could be applied to P_B , in order to establish corresponding results for *conditional probabilities*.

- Since a *conditional expectation* is just the expected value of a random variable, defined using the conditional probability distribution P_B , properties of expectations can be used to establish corresponding properties of conditional expectations, in essentially the same way.

Conditional Expectation

The following result is useful just as the “Law of Total Probability” (Theorem #7 from the first set of lecture notes on discrete probability theory) is: It describes a way to compute an expectation by considering cases (corresponding to whether, or not, an event has occurred). Recall that if $B \subseteq \Omega$, then

$$B^C = \{\sigma \in \Omega \mid \sigma \notin B\}.$$

Theorem #1: Let Ω be a sample space with probability distribution $P : \Omega \rightarrow \mathbb{R}$, let $B \subseteq \Omega$ be an event such that $P(B) > 0$ and $P(B^C) > 0$, and let X be a random variable. Then

$$E[X] = E[X \mid B] \times P(B) + E[X \mid B^C] \times P(B^C).$$

Proof: Exercise.

Linearity of Expectation

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

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

for each outcome $\sigma \in \Omega$. Similarly, if $X : \Omega \rightarrow \mathbb{R}$ 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.

Linearity of Expectation

Theorem #2 (Linearity of Expectation): Let Ω be a sample space with probability distribution $P : \Omega \rightarrow \mathbb{R}$.

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

$$E[X_1 + X_2 + \dots + X_n] = E[X_1] + E[X_2] + \dots + E[X_n].$$

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

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

Proof: Another **exercise**.

Independent Random Variables

It follows by Theorem #2 that if X_1 and X_2 are random variables over Ω , then $E[X + Y] = E[X] + E[Y]$.

- It might be reasonable to *wonder whether*

$$E[X \times Y] = E[X] \times E[Y]$$

as well?

- It turns out that ***this is not generally true.***

Independent Random Variables

- Consider a sample space $\Omega = \{H, T\}$, and let $P : \Omega \rightarrow \mathbb{R}$ be the **uniform probability distribution** — so that $P(H) = P(T) = \frac{1}{2}$.
- Consider indicator random variables $X, Y : \Omega \rightarrow \{0, 1\}$ where, for $\sigma \in \Omega$,

$$X(\sigma) = \begin{cases} 1 & \text{if } \sigma = H \\ 0 & \text{if } \sigma = T \end{cases} \quad \text{and} \quad Y(\sigma) = \begin{cases} 0 & \text{if } \sigma = H \\ 1 & \text{if } \sigma = T \end{cases}$$

- $E[X] = E[Y] = \frac{1}{2}$, so that

$$E[X] \times E[Y] = \frac{1}{4}.$$

- However, $(X \times Y)(\sigma) = 0$ for all $\sigma \in \Omega$ — so that

$$E[X \times Y] = 0.$$

Independent Random Variables

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

$$"X = a" = \{\sigma \in \Omega \mid X(\sigma) = a\}$$

and

$$"Y = b" = \{\tau \in \Omega \mid Y(\tau) = b\},$$

so that

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

Independent Random Variables

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

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

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

In other words, X and Y are “independent” if *all* events corresponding to choices of values for X and for Y , respectively, are independent.

Independent Random Variables

Theorem #3: Let Ω be a sample space with probability distribution $P : \Omega \rightarrow \mathbb{R}$, and let $X, Y : \Omega \rightarrow \mathbb{R}$. If X and Y are independent random variables then

$$E[X \times Y] = E[X] \times E[Y].$$

A proof of this claim will be discussed in the lecture presentation.

Pairwise Independent Random Variables

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

$$X_1 : \Omega \rightarrow V_1, X_2 : \Omega \rightarrow V_2, \dots, X_k : \Omega \rightarrow V_k$$

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

Definition: The random variables X_1, X_2, \dots, X_k are **pairwise independent** if the random variables 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 $P : \Omega \rightarrow \mathbb{R}$. Let k be a positive integer and let

$$X_1 : \Omega \rightarrow V_1, X_2 : \Omega \rightarrow V_2, \dots, X_k : \Omega \rightarrow V_k$$

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

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

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

With a bit of work one can show that this is *equivalent* to the definition of the independence of a pair of random variables, X_1 and X_2 , when $k = 2$.

Mutually Independent Random Variables

- If $k \geq 3$ and X_1, X_2, \dots, X_k are random variables, as above, and X_1, X_2, \dots, X_k are ***mutually independent***, then X_1, X_2, \dots, X_k must be ***pairwise independent*** as well.
- However, it is possible that X_1, X_2, \dots, X_k are *pairwise independent*, but ***not mutually independent***.