Random Variables and Expectations

Robert Y. Lewis

CS 0220 2024

April 10, 2024

Overview

Numerical Values of Outcomes

Sometimes it makes sense to attach a numerical value to each outcome of a probability space.

Example: We ask people to name all the dinosaurs they can.

- Tyrannosaurus, stegosaurus
- Tyrannosaurus, brontosaurus, stegosaurus, velociraptor
- Tyrannosaurus, brontosaurus, velociraptor, apatosaurus, pterodactyl, diplodocus
- stegosaurus

How many outcomes? 2^{700}, according to Google.

If we want to summarize the results, we might assign each outcome a *statistic*, that is, a numerical summary. A natural choice is the number of dinosaurs they named: 2, 4, 6, 1.

Random variable

Definition: A *random variable* R on a probability space is a function whose domain is the sample space.

Example: Let's say the sample space is a deck of cards and R maps a number card to its value and a face card to 10 and ace to 1. So, $R(2\heartsuit) = 2$ and $R(J\clubsuit) = 10$.

Typically, codomain of R is subset of reals. A random variable is used kind of like a variable, but it is "implemented" as a function.

Coin example

We flip 3 fair coins. Let C be the random variable that is the number of coins that come up heads. Let M be a random variable that is 1 if all three coins come up heads or all three coins come up tails and 0 otherwise. They are random variables in that they map all possible outcomes to values, integers in this case.

Example: $C(THH) = 2$. $M(THH) = 0$. $C(TTT) = 0$. $M(TTT) = 1$.

C is *counting* the number of heads, M tells us whether or not all the coins *match*.

In terms of sample space

$$\mathcal{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

$$\begin{aligned} C(HHH) &= 3 & C(THH) &= 2 \\ C(HHT) &= 2 & C(THT) &= 1 \\ C(HTH) &= 2 & C(TTH) &= 1 \\ C(HTT) &= 1 & C(TTT) &= 0 \end{aligned}$$

$$\begin{aligned} M(HHH) &= 1 & M(THH) &= 0 \\ M(HHT) &= 0 & M(THT) &= 0 \\ M(HTH) &= 0 & M(TTH) &= 0 \\ M(HTT) &= 0 & M(TTT) &= 1 \end{aligned}$$

Definition

Definition: An *indicator random variable* is a random variable that maps every outcome to either 0 or 1. Indicator random variables are also called *Bernoulli variables*.

Example: The random variable M. It "indicates" whether the three coins match.

Connection between indicator random variables and events. Recall, an event is a subset of the sample space—a set of outcomes. An indicator random variable can be interpreted as a set, since it maps each outcome to whether it is *in* the set (1) or *out* of the set (0).

If E is an event, we can define the corresponding indicator random variable I_E, where $I_E(\omega) = 1$ if $\omega \in E$ and 0 otherwise.

Example: If we take E to be the event where all 3 coins match, $M = I_E$.

Partitioning sample space

An indicator random variable partitions sample space:

$$\underbrace{HHT\ HTH\ HTT\ THH\ THT\ TTH}_{M=0}\ \underbrace{HHH\ TTT}_{M=1}$$

So does any other random variable:

$$\underbrace{TTT}_{C=0}\ \underbrace{HTT\ THT\ TTH}_{C=1}\ \underbrace{HHT\ HTH\ THH}_{C=2}\ \underbrace{HHH}_{C=3}$$

Statements about random variables

Each block is a subset of the sample space and therefore an event.

The assertion that $C = 2$ defines an event: $\{THH, HTH, HHT\}$.

$\Pr[C = 2] = 3/8$.

$\Pr[M = 1] = 1/4$.

Statements about random variables can also be viewed as events.

$\Pr[C \leq 1] = 1/2$.

$\Pr[M \cdot C \text{ is odd}] = 1/8$.

This last statement is a funny way of saying “all heads”. Why?

Concept

The *expected value* (often *expectation*) of a random variable is its mean or probability weighted average.

Example: Define a random variable R to be the alphabetic position of the first letter of the outcome of a coin flip, $R(H) = 8$, $R(T) = 20$. The expected value of R is 14. It is $1/2 \times 8 + 1/2 \times 20$.

We write $\mathbb{E}[R] = 14$. (Book uses "Ex", but I can't pretend that's ever used.)

Suppose we select a student uniformly at random from the class, and let R be the student's homework 2 score. Then, $\mathbb{E}[R]$ is just the class average. The expected value is a useful thing to know.

Definition

Definition: If R is a random variable defined on a sample space $\mathcal{S}$, then the expectation of R is

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

Example: $\mathbb{E}[C] = \frac{0+1+1+1+2+2+2+3}{8} = 3/2.$

Example: $\mathbb{E}[M] = \frac{1+0+0+0+0+0+0+1}{8} = 1/4.$

Exercise for the reader: If E is an event, $\Pr[E] = \mathbb{E}[I_E]$.

Fair die

Let R be the random variable corresponding to a fair die. Here, the outcomes are numbers, so we'll just define $R(\omega) = \omega$.

$\mathbb{E}[R] = \frac{1+2+3+4+5+6}{6} = 7/2$ or 3.5.

Does that mean we expect the die to come up 3.5? No, it will *never* come up 3.5. Maybe "expected value" was a bad choice of name.

In general, if R is a random variable with a uniform distribution over $[1, n]$, $\mathbb{E}[R] = \sum_{i=1}^{n} i \cdot \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} i = \frac{1}{n} n(n+1)/2 = (n+1)/2$.

Nonuniform distribution

A cherry tomato plant produces about 100 tomatoes a season. Plum tomato plant: 40. Heirloom tomato plant: 20.

60% of the plants in my garden are cherry tomato plants. 25% plum. 15% heirloom.

What's the expected yield of a plant from my garden?

$.6 \cdot 100 + .25 \cdot 40 + .15 \cdot 20 = 73$ tomatoes: a *weighted* average.

One and die

Let R again be the random variable corresponding to a fair die.

$1 + \mathbb{E}[R] = 1 + 3.5 = 4.5.$

$\mathbb{E}[1 + R] = \frac{2+3+4+5+6+7}{6} = \frac{27}{6} = \frac{9}{2}$ or 4.5.

Sometimes the expectation of a function matches the function of the expectation.

One over die

Let R again be the random variable corresponding to a fair die.

$\frac{1}{\mathbb{E}[R]} = \frac{1}{3.5} = \frac{2}{7}$ (.29 ish).

$\mathbb{E}[\frac{1}{R}] = \frac{1+1/2+1/3+1/4+1/5+1/6}{6} = \frac{49}{120}$ (.41 ish).

Sometimes the expectation of a function matches the function of the expectation.

Sometimes not.