Section 3.1 Random variables

Tuesday, September 6, 2016 1:15 PM

Notes_3_1_
Random_...

3 Discrete Distributions

Complementary reading: Chapter 3 (WMS), except § 3.10 and § 3.11.

3.1 Random variables

PROBABILISTIC DEFINITION: A **random variable** Y is a function whose domain is the sample space S and whose range is the set of real numbers $\mathcal{R} = \{y : -\infty < y < \infty\}$. That is, $Y : S \rightarrow \mathcal{R}$ takes sample points in S and assigns them a real number.

WORKING DEFINITION: In simpler terms, a random variable is a variable whose observed value is determined by chance.

Example 3.1. Suppose that an experiment consists of flipping two fair coins. The sample space is

$$S = \{(H, H), (H, T), (T, H), (T, T)\}. \rightarrow \mathcal{R}$$

Let Y denote the number of heads observed. Before we perform the experiment, we do not know, with certainty, the value of Y. We can, however, list out the possible values of Y corresponding to each sample point:

E_i	$Y(E_i) = y$	E_i	$Y(E_i) = y$
(H, H)	2	(T, H)	1
(H, T)	1	(T, T)	0

For each sample point E_i, Y takes on a numerical value specific to E_i. This is precisely why we can think of Y as a function; i.e.,

$$Y[(H, H)] = 2 \quad Y[(H, T)] = 1 \quad Y[(T, H)] = 1 \quad Y[(T, T)] = 0,$$

(chap 2)

so that

$$\begin{aligned} P(Y = 2) &= P[(H, H)] = 1/4 \\ P(Y = 1) &= P[(H, T)] + P[(T, H)] = 1/4 + 1/4 = 1/2 \\ P(Y = 0) &= P[(T, T)] = 1/4. \end{aligned}$$

NOTE: From these probability calculations; note that we can

- work on the sample space S and compute probabilities from S, or
- work on $\mathcal{R}$ and compute probabilities for events $\{Y \in B\}$, where $B \subset \mathcal{R}$.

X Y Z

NOTATION: We denote a random variable Y using a capital letter. We denote an observed value of Y by y, a lowercase letter. **This is standard notation.** For example, if Y denotes the weight (in ounces) of the next newborn boy in Columbia, SC, then Y is random variable. After the baby is born, we observe that the baby weighs $y = 128$ oz.

3.2 Probability distributions for discrete random variables

TERMINOLOGY: The **support** of a random variable Y is set of all possible values that Y can assume. We will denote the support set by R.

TERMINOLOGY: If the random variable Y has a support set R that is countable (finitely or infinitely), we call Y a **discrete** random variable.

Example 3.2. An experiment consists of rolling an unbiased die. Consider the two random variables:

$$\begin{aligned} X &= \text{face value on the first roll} \\ Y &= \text{number of rolls needed to observe a six.} \end{aligned}$$

The support of X is $R_X = \{1, 2, 3, 4, 5, 6\}$. The support of Y is $R_Y = \{1, 2, 3, ...\}$. R_X is finitely countable and R_Y is infinitely countable; thus, both X and Y are discrete. $\square$

GOAL: For a discrete random variable Y, we would like to find $P(Y = y)$ for any $y \in R$. Mathematically,

$$p_Y(y) \equiv P(Y = y) = \sum P[E_i \in S : Y(E_i) = y],$$

for all $y \in R$.