Set 8: Random Variables

Stat 260 A01: May 29, 2024

Example 1: Suppose you flip a fair coin 3 times and record the outcomes.

equally likely

(a) Complete the following table of probabilities (called a probability distribution):

outcome	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
probability	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

← 8 outcomes (sample space)

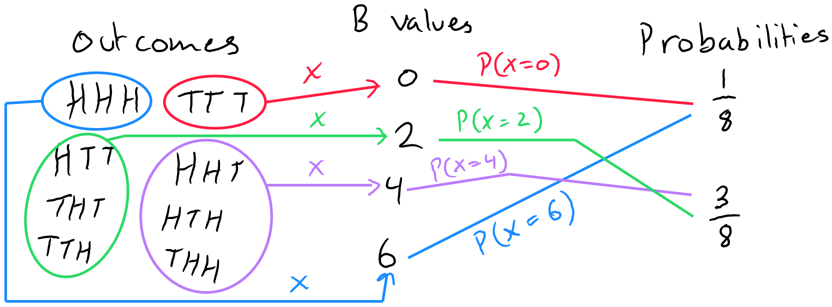
(b) Suppose you win \$2 for each Head (H) you flip. How does this change your experiment?

Experiment remains the same, but how we label the outcome changes.

A ***random variable (r.v.)*** $\underline{X}$ is a *function* that assigns a numeric value to each event in a sample sample S.

↳ Capital letter

Let X = amount of money won in the coin game.



Discrete r.v.: A random variable with a finite or "countability infinite" number of possible values.

e.g. X = amount of students in a classroom
Y = number of cats on a farm } "count"

▷ A ***Bernoulli r.v.*** is a discrete r.v. with exactly 2 outcomes.

Continuous r.v.: A random variable that can take on any value in an interval of real numbers.

X = the weight of an apple
Y = the distance to a cafe } "measure"

Probability distribution: is a table, formula, or graph that describes all possible value of a random variable, and gives the probability that each outcome occurs.

Discrete Random Variables

Probability mass function (pmf): Let X be a discrete random variable, the ***mass function*** $f(x)$ of X assigns a probability to each value of X

$$\underset{\text{pmf}}{f}(x) = P(X = x)$$

Properties of a (discrete) pmf:

(i) $0 \le f(x) \le 1$

(ii) $\sum f(x) = 1$

Example 1 Continued...

(c) Determine the probability distribution for X, the random variable for the winnings from flipping a coin three times.

$f(x) = P(X = x)$

x	0	2	4	6
$f(x)$	1/8	3/8	3/8	1/8

This means:

$f(0) = P(X = 0) = 1/8$

$f(2) = P(X = 2) = 3/8$

$f(4) = P(X = 4) = 3/8$

$f(6) = P(X = 6) = 1/8$

To quickly check: sum probabilities up and see if they're $= 1$

$$\frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = 1$$

Let X = number of offspring of a mule deer

Example 2: Doe mule deer give birth in late May to early June. Although mule deer usually have two offspring, in their first season they often have just one. In a very healthy population, triplets can occur.[1] Suppose that among a certain population the probability of each is as follows.

Number of offspring (x_i)	0	1	2	3
$f(x_i) = P(X = x_i) = p_i$	0.06	0.18	0.73	???

→ $1-(0.06+0.18+0.73)$
$= 0.03$

(a) What is the probability of a random mule deer having 2 offspring?

$f(x) = P(X=2) = 0.73$

(b) What is the probability of a random mule deer having at least 2 offspring?

$P(X \geq 2) = P(X=2) + P(X=3) = f(2) + f(3)$
$= 0.73 + 0.03$
$= 0.76$

(c) What is the probability of a random mule deer having more than 2 offspring?

$P(X>2) = P(X=3) = f(3) = 0.03$

> The ***Cumulative Distribution Function (CDF)*** of a **discrete** random variable X is defined as:
>
> capital → $F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$
>
> cdf

Example 2 (Mule Deer) Continued...

(d) What is the probability of a random mule deer having at most 2 offspring?

$P(X \leq 2) = F(2) = f(0) + f(1) + f(2)$
$= 0.06 + 0.18 + 0.73$
$= 0.97$

(e) Create the CDF for the mule deer birth probabilities.

x	0	1	2	3
$F(x) = P(X \leq x)$	0.06	0.24	0.97	1

Last on the right is always $=1$

$F(1) = f(0) + f(1)$
$= 0.06 + 0.18 = 0.24$

[1] Quadruplets have been observed in the wild, but are so exceedingly rare, we won't include them in this example.

Rules for Discrete CDFs:

(i) $P(X \geq x)$ $= 1 - P(X \leq x-1)$

(ii) $P(X > x)$ $= 1 - P(X \leq x)$

(iii) $P(a \leq X \leq b)$ $= P(X \leq b) - P(X \leq a-1)$

(iv) $P(X = a)$. $= f(a) = P(X \leq a) - P(X \leq a-1)$

Example 2 (Mule Deer) Continued again...

(f) $P(X \geq 1)$ $= f(1) + f(2) + f(3)$

$= 1 - f(0)$

$= 1 - F(0)$

$= 1 - 0.06$

$= 0.94$

(g) $P(X > 1) = 1 - P(X \leq 1)$

$= 1 - F(1) = 1 - 0.24 = 0.76$

want: 2,3
dont want: 0,1

(h) $P(X \geq 2)$. $= P(X > 1) = 0.76$

want: 2,3
dont want: 0,1

(i) $P(1 \leq X \leq 2)$

$= P(X \leq 2) - P(X \leq 0)$

$= F(2) - F(0)$

$= 0.97 - 0.06$

$= 0.91$

want: 1,2
dont want: 0, 3

(j) $P(1 < X \leq 3)$

$= P(X \leq 3) - P(X \leq 1)$

$= F(3) - F(1)$

$= 1 - 0.24$

$= 0.76$

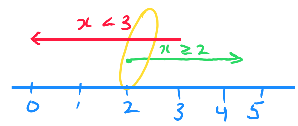
want : 2, 3
dontwant : 0, 1

Midterm → practice building pmf using cdf

Extra Example 1: The Mallampati score is a visual assessment on a scale of 1-4 of the distance from the tongue base to the roof of the mouth. Originally developed as a predictor for the difficulty of endotracheal intubation, it is now used in other medical fields, including as a predictor of obstructive sleep apnea. Suppose that in a certain population, the distribution of Mallampati scores are:

Mallampati Score x_i	1	2	3	4
$f(x) =$ $P(X = x_i)$	0.12	0.27	0.59	???

$\to 0.02 = 1 - (0.12 + 0.27 + 0.59)$



(a) What is the probability that random patient scores at least 3?

$$P(x < 3 \mid x \geq 2) = \frac{P((x<3) \cap (x \geq 2))}{P(x \geq 2)}$$

$$= \frac{P(x=2)}{f(2) + f(3) + f(4)} = \frac{0.27}{0.27 + 0.59 + 0.02} = \frac{0.27}{0.88} = 0.3$$

(b) If a random patient has a Mallampati score of at least 2, what is the probability that they score less than 3?

(c) Two patients are scored at random. What is the probability that they both scored at least 3?

Textbook Readings: Swartz Ch4, 4.1 [EPS 2.1, 2.2]
Practice problems: Swartz 2.1, 2.3, 2.5, 2.9, 2.11, 2.13, 2.17, 2.27