STAT | Random Variable

Dr. Mostafa Elhosseini

https://youtube.com/drmelhosseini

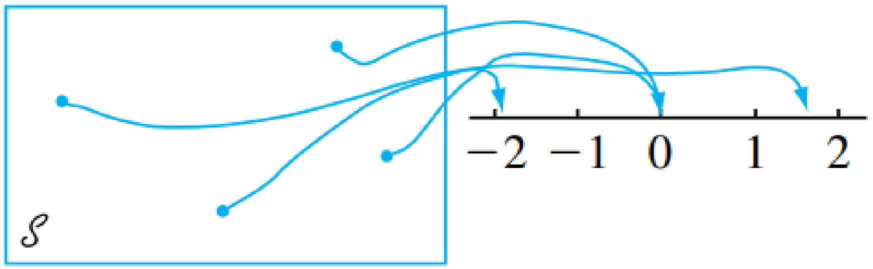
Intro…

- In any experiment, there are numerous characteristics that can be observed or measured,
  - but in most cases an experimenter will focus on some specific aspect or aspects of a sample.
  - For example, in a study of commuting patterns in a metropolitan area, each individual in a sample might be asked about commuting distance and the number of people commuting in the same vehicle, but not about IQ, income, family size, and other such characteristics

Random Variable

- In general, each outcome of an experiment can be associated with a number by specifying a rule of association
  - The number among the sample of ten components that fail to last 1000 hours
  - The total weight of baggage for a sample of 25 airline passengers
- Such a rule of association is called a **random variable**
  - **variable** because different numerical values are possible and
  - **random** because the observed value depends on which of the possible experimental outcomes results

Random Variable



- For a given sample space $\mathcal{S}$ of some experiment, a random variable is any rule that associates a number with each outcome in $\mathcal{S}$.
- In mathematical language, a random variable is a function whose domain is the sample space and whose range is the set of real numbers

Random Variable

- **Random variables** are customarily denoted by uppercase letters, such as $\boldsymbol{X}$ and $\boldsymbol{Y}$, near the end of our alphabet
- we will now use lowercase letters to represent some particular value of the corresponding random variable. The notation $X(s) = x$ means that x is the value associated with the outcome s by the random variable X

Example: 01

- When a student attempts to log on to a computer time-sharing system, either all ports are busy (F), in which case the student will fail to obtain access, or else there is at least one port free (S), in which case the student will be successful in accessing the system. With $\mathcal{S}\{S, F\}$, define random variable X by
  - $X(S) = 1$ $X(F) = 0$
- The random X indicates whether (1) or not (0) the student can log on

Example: 02

- Consider the experiment in which a telephone number in a certain area code is dialed using a random number dialer (such devices are used extensively by polling organizations), and define random variable Y by

$$Y = \begin{cases} 1 & \textit{if the selected number is unlisted} \\ 0 & \textit{if the selected number is listed in the directory} \end{cases}$$

- if 5282966 appears in the telephone directory, then $Y(5282966) = 0$, whereas $Y(7727350) = 1$ tells us that the number 7727350 is unlisted
- Any random variable whose only possible values are 0 and 1 is called a **Bernoulli random variable**

Example 03

- In an experiment in which the number of pumps in use at each of two gas stations was determined. Define random variable X, Y, and U by
  - **X** the total number of pumps in use at the two stations
  - **Y** the difference between the number of pumps in use at station 1 and the number in use at station 2
  - **U** the maximum of the numbers of pumps in use at the two stations
- If this experiment is performed and $s(2, 3)$ results
  - X(s(2,3)) = 5
  - Y(s(2,3)) = -1
  - U(s(2,3)) = 3

Example 04

- Each of the random variables of the above examples can assume only a finite number of possible values. This need not be the case
- the experiment in which batteries were examined until a good one (S) was obtained. The sample space was $\mathcal{S}\ \{S, FS, FFS, \ldots\}$.
  Define random variable X by

  X : the number of batteries examined before the experiment terminates

— X(S) = 1

— X(FS) = 2

— X(FFS) = 3

— X(FFFFFFS) = 7

- Any positive integer is a possible value of X, so the set of possible values is infinite

Example 05

- The experiment of tossing a coin twice. Let X represent number of heads appearing
  - X is a random variable with values 0, 1, 2
  - X(T,T) = 0
  - X(H,T) = 1

Example 05

- Suppose that in some random fashion, a location (latitude and longitude) in the continental United States is selected. <u>Define random Y by</u>
  Y the height above sea level at the selected location
- $Y\left(\left(39°\, 5\grave{0}\, N, 98°\, 3\grave{5}\, W\right)\right) = 1748.26$ ft
- There are an infinite number of numbers in this interval

Two Types of Random Variable

- **discrete**
  - Finite set
  - Infinite "But counted"
- **continuous**
  - Interval
  - $P(X = c) = 0$ for any possible value of c

Probability distributions for discrete Random Variable

- Probabilities assigned to various outcomes in $\mathcal{S}$ in turn determine probabilities associated with the values of any particular random X.
- The probability distribution of X says <u>how the total probability of 1 is distributed among (allocated to) the various possible X values</u>

Example 01

- Suppose, for example, that a business has just purchased four laser printers, and let X be the number among these that require service during the warranty period.
  Possible X values are then 0, 1, 2, 3, and 4.
- The probability distribution will tell us how the probability of 1 is subdivided among these five possible values
  - how much probability is associated with the X value 0, how much is apportioned to the X value 1, and so on
  - $p(0)$ = the probability of the X value 0 = $p(X = 0)$
  - $p(1)$ = the probability of the X value 1 = $p(X = 1)$
  - in general $p(x)$ will denote the probability assigned to the value x

Example 02

- A certain gas station has six pumps. Let X denote the number of pumps that are in use at a particular time of day. Suppose that the probability distribution of X is as given in the following table; the first row of the table lists the possible X values and the second row gives the probability of each such value.

x	0	1	2	3	4	5	6
$p(x)$	.05	.10	.15	.25	.20	.15	.10

Example 02

x	0	1	2	3	4	5	6
$p(x)$	.05	.10	.15	.25	.20	.15	.10

We can now use elementary probability properties to calculate other probabilities of interest

- the probability that at most 2 pumps are in use is
- $p(x \leq 2) = p(x = 0\ or\ 1\ or\ 2) = p(0) + p(1) + p(2) = 0.3$
- Since the event at least 3 pumps are in use is complementary to at most 2 pumps are in use,
- $p(x \geq 3) = 1 - p(x \leq 2) = 1 - 0.3 = 0.7$
- which can, of course, also be obtained by adding together probabilities for the values, 3, 4, 5, and 6
- The probability that between 2 and 5 pumps inclusive are in use ?
- $P(2 \leq x \leq 5)$

Probability distribution

- The probability distribution or probability mass function (pmf) of a discrete random variable is defined for every number x

$$p(x) = p(X = x) = p(all\ s \in \mathcal{S} : X(s) = x)$$

- In words, for every possible value x of the random variable, the pmf specifies the probability of observing that value when the experiment is performed
- For any PMF
  - $p(x) \geq 0$
  - $\sum_{all\ possible\ x} p(x) = 1$

Example

- Six lots of components are ready to be shipped by a certain supplier. The number of defective components in each lot is as follows:

Lot	1	2	3	4	5	6
Number of defectives	0	2	0	1	2	0

- One of these lots is to be randomly selected for shipment to a particular customer. Let X be the number of defectives in the selected lot. The three possible X values are 0, 1, and 2
- $p(0) = p(X = 0) = p(lot\ 1\ or\ 3\ or\ 6\ is\ sent) = \frac{3}{6} = 0.5$
- $p(1) = p(X = 1) = p(lot\ 4\ is\ sent) = \frac{1}{6} = 0.167$
- $p(2) = p(X = 2) = p(lot\ 2\ or\ 5\ is\ sent) = \frac{2}{6} = 0.333$

Example

- The experiment of tossing a coin twice. Let X represent number of heads appearing
  - X is a random variable with values 0, 1, 2
  - $p(X = 0) = p(TT) = \frac{1}{4}$
  - $p(X = 1) = p(HT, TH) = \frac{2}{4}$
  - $p(X = 2) = p(HH) = \frac{1}{4}$

Example

- Considering the experiment of tossing two dice. The random variable X represents the sum of the two numbers. Find the probability distribution $p(X)$

Example

- Consider a group of five potential blood donors—a, b, c, d, and e—of whom only a and b have type O+ blood. Five blood samples, one from each individual, will be typed in random order until an O+ individual is identified. Let the random variable Y the number of typings necessary to identify an O+ individual.

Example

$$p(1) = P(Y = 1) = P(a \text{ or } b \text{ typed first}) = \frac{2}{5} = .4$$

$$p(2) = P(Y = 2) = P(c,\ d, \text{ or } e \text{ first, and then } a \text{ or } b)$$

$$= P(c,\ d, \text{ or } e \text{ first}) \cdot P(a \text{ or } b \text{ next} \mid c,\ d, \text{ or } e \text{ first}) = \frac{3}{5} \cdot \frac{2}{4} = .3$$

$$p(3) = P(Y = 3) = P(c,\ d, \text{ or } e \text{ first and second, and then } a \text{ or } b)$$

$$= \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{2}{3}\right) = .2$$

$$p(4) = P(Y = 4) = P(c,\ d, \text{ and } e \text{ all done first}) = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{1}{3}\right) = .1$$

$$p(y) = 0 \quad \text{if } y \neq 1, 2, 3, 4$$

Example

y	1	2	3	4
$p(y)$	.4	.3	.2	.1

$$p(1) = P(Y = 1) = P(a \text{ or } b \text{ typed first}) = \frac{2}{5} = .4$$

$$p(2) = P(Y = 2) = P(c, d, \text{ or } e \text{ first, and then } a \text{ or } b)$$

$$= P(c, d, \text{ or } e \text{ first}) \cdot P(a \text{ or } b \text{ next} \mid c, d, \text{ or } e \text{ first}) = \frac{3}{5} \cdot \frac{2}{4} = .3$$

$$p(3) = P(Y = 3) = P(c, d, \text{ or } e \text{ first and second, and then } a \text{ or } b)$$

$$= \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{2}{3}\right) = .2$$

$$p(4) = P(Y = 4) = P(c, d, \text{ and } e \text{ all done first}) = \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{1}{3}\right) = .1$$

$$p(y) = 0 \quad \text{if } y \neq 1, 2, 3, 4$$