Random Variables

Al Nosedal.
University of Toronto.

Summer 2019

My momma always said: "Life was like a box of chocolates. You never know what you're gonna get."

Forrest Gump.

Definition

A **random variable** is a real-valued function for which the domain is a sample space.

Example

A fair die is thrown twice. The sample points are $(1,1),\ (1,2),\ (1,3), ...,\ (6,6)$. There are 36 sample points. Let us assign the same probability $1/36$ for each of these points. Suppose we are interested in the sum of the numbers of each outcome. Then we can define a random variable $X = i + j$ associated with the outcome (i,j). List the possible values of X and say what the probability of each value is.

Solution

Value of X	Probability
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Definition

The probability that Y takes on the value y, $P(Y = y)$, is defined as the sum of probabilities of all sample points in S that are assigned the value y. We will sometimes denote $P(Y = y)$ by $p(y)$.

Definition

The **probability distribution** for a discrete variable Y can be represented by a formula, a table, or a graph that provides $p(y) = P(Y = y)$ for all y.

Important Result

For any discrete probability distribution, the following must be true:
1. $0 \leq p(y) \leq 1$ for all y.
2. $\sum_y p(y) = 1$, where the summation is over all values of y with nonzero probability.

Example

A couple plans to have three children. There are 8 possible arrangements of girls and boys. For example, GGB means the first two children are girls and the third child is a boy. All 8 arrangements are (approximately) equally likely.
a. Write down all 8 arrangements of the sexes of three children. What is the probability of any one of these arrangements?
b. Let X be the number of girls the couple has. What is the probability that $X = 2$?
c. Starting from your work in (a), find the distribution of X. That is, what values can X take, and what are the probabilities for each value?

Solution

a. Sample space = S = {(G,G,G), (G,G,B), (G,B,G), (G,B,B), (B,G,G), (B,G,B), (B,B,G), (B,B,B)}.
Probability of any of these arrangements = $\frac{1}{8}$.

b. $P(X = 2) = P(G, G, B) + P(G, B, G) + P(B, G, G) = \frac{3}{8}$.

c. $P(X = 0) = \frac{1}{8}$, $P(X = 1) = \frac{3}{8}$,
$P(X = 2) = \frac{3}{8}$, $P(X = 3) = \frac{1}{8}$

The Binomial setting

1. There are a fixed number n of observations.
2. The n observations are all **independent**. That is, knowing the result of one observation tells nothing about the other observations.
3. Each observation falls into one of just two categories, which for convenience we call "success" and "failure".
4. The probability of a success, call it p, is the same for each observation.

Binomial Distribution

The distribution of the count X of successes in the binomial setting is the **Binomial distribution** with parameters n and p. The parameter n is the number of observations, and p is the probability of a success on any one observation. The possible values of X are the whole numbers from 0 to n.

Example. Blood types

Genetics says that children receive genes from their parents independently. Each child of a particular pair of parents has probability 0.25 of having type O blood. If these parents have 5 children, the number who have type O blood is the count X of successes in 5 independent trials with probability 0.25 of a success on each trial. So X has the Binomial distribution with $n = 5$ and $p = 0.25$.

Example. Inheriting blood type

Each child born to a particular set of parents has probability 0.25 of having blood type O. If these parents have 5 children, what is the probability that exactly 2 of them have type O blood?

Solution

The count of children with type O blood is a Binomial random variable X with $n = 5$ tries and probability $p = 0.25$ of a success on each try. We want $P(X = 2)$.

$$P(X = 2) = 10(0.25)^2(0.75)^3 \approx 0.2637.$$

R Code

```
dbinom(2,size=5,prob=0.25)

## [1] 0.2636719

# 1st number, point where you want to evaluate
# Binomial distribution;
# size = n = number of tries;
# prob = p = probability of success;
```

Example

In order to verify the accuracy of their financial accounts, companies use auditors on a regular basis to verify accounting entries. The company's employees make erroneous entries 5% of the time. Suppose that an auditor randomly checks three entries. (Assume independence)

a. Find the probability distribution for Y, the number of errors detected by the auditor.

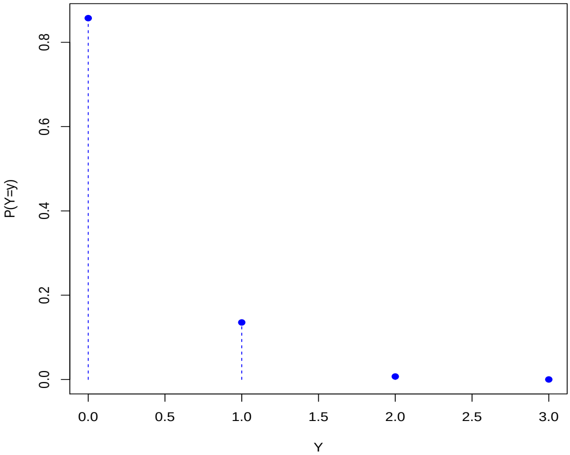
b. Construct a graph for $p(y)$.

c. Find the probability that the auditor will detect more than one error.

Solution

Let E denote an error on a single entry and let N denote no error. There are 8 sample points: {(E,E,E), (E,E,N), (E,N,E), (E,N,N), (N,E,E), (N,E,N), (N,N,E), (N,N,N)}.
We also have that $P(E) = 0.05$ and $P(N) = 0.95$. Then
$P(Y = 0) = (0.95)^3 = 0.857375$
$P(Y = 1) = 3(0.05)(0.95)^2 = 0.135375$
$P(Y = 2) = 3(0.05)^2(0.95) = 0.007125$
$P(Y = 3) = (0.05)^3 = 0.000125.$

Solution b)



Solution c)

$$P(Y > 1) = P(Y = 2) + P(Y = 3)$$

$$= 0.007125 + 0.000125 = 0.00725.$$

Example

The probability that parents with a certain type of blue-brown eyes will have a child with blue eyes is $\frac{1}{4}$. If there are six children in the family, what is the probability that at least half of them will have blue eyes?

Solution

To solve this problem the six children in the family will be treated as six independent trials of an experiment for which the probability of success in a single trial is $\frac{1}{4}$. Thus Y (number of blue-eyed children in the family) has a Binomial distribution with $n = 6$ and $p = 1/4$. It is necessary to calculate $P(Y = 3)$, $P(Y = 4)$, $P(Y = 5)$, $P(Y = 6)$ and sum.

Solution

$$P(Y=3) = \frac{6!}{3!3!}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^3 = \frac{540}{4096}$$
$$P(Y=4) = \frac{6!}{4!2!}\left(\frac{1}{4}\right)^4\left(\frac{3}{4}\right)^2 = \frac{135}{4096}$$
$$P(Y=5) = \frac{6!}{5!1!}\left(\frac{1}{4}\right)^5\left(\frac{3}{4}\right)^1 = \frac{18}{4096}$$
$$P(Y=6) = \frac{6!}{6!0!}\left(\frac{1}{4}\right)^6\left(\frac{3}{4}\right)^0 = \frac{1}{4096}$$

Solution

$$P(Y \geq 3) = P(Y = 3) + P(Y = 4) + P(Y = 5) + P(Y = 6)$$
$$= \frac{540}{4096} + \frac{135}{4096} + \frac{18}{4096} + \frac{1}{4096} = \frac{694}{4096} \approx 0.169$$

Binomial Probability

If X has the Binomial distribution with n observations and probability p of success on each observation, the possible values of X are $0, 1, 2, ..., n,$. If k is any one of these values,

$$P(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

Example. Inheriting blood type

If the parents in our example have 5 children, the number who have type O blood is a random variable X that has the Binomial distribution with $n = 5$ and $p = 0.25$.

a) What are the possible values of X?

b) Find the probability of each value of X. Make a graph to display this distribution.

Solution a)

Possible values of X are: 0, 1, 2, 3, 4, and 5.

Solution b)

$$P(X=0) = 1(0.25)^0(0.75)^5 \approx 0.2373$$
$$P(X=1) = 5(0.25)^1(0.75)^4 \approx 0.3955$$
$$P(X=2) = 10(0.25)^2(0.75)^3 \approx 0.26367$$
$$P(X=3) = 10(0.25)^3(0.75)^2 \approx 0.08789$$
$$P(X=4) = 5(0.25)^4(0.75)^1 \approx 0.014648$$
$$P(X=5) = 1(0.25)^5(0.75)^0 \approx 0.00097$$

R Code

```
dbinom(0:2,size=5,prob=0.25);

## [1] 0.2373047 0.3955078 0.2636719

dbinom(3:5,size=5,prob=0.25);

## [1] 0.0878906250 0.0146484375 0.0009765625
```

R Code

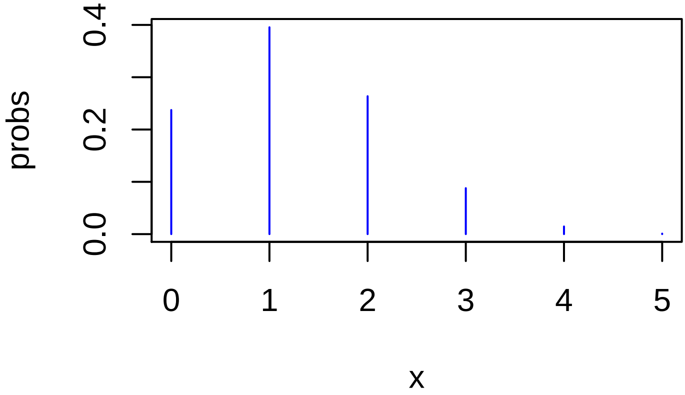
```r
x=0:5;

probs=dbinom(x,size=5,prob=0.25);

plot(x,probs,type="h",col="blue");

# type=h tells R to draw "needles";
```

R Code



Definition

Let Y be a discrete random variable with the probability function $p(y)$. Then the **expected value** of Y, $E(Y)$, is defined to be

$$E(Y) = \sum_{\text{all values of y}} y p(y).$$

Exercise

A single fair die is tossed once. Let Y be the number facing up. Find
a) the expected value of Y,
b) the expected value of $W = 3Y$,

Solution a)

By definition,

$E(Y) = \sum_{y=1}^{6} yP(Y = y)$

$$= 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right)$$

$E(Y) = \left(\frac{1}{6}\right)(1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6} = 3.5$

Solution b)

By definition,

$E(W) = \sum_{all\ ws} wP(W = w)$

$$= 3\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) + 9\left(\frac{1}{6}\right) + 12\left(\frac{1}{6}\right) + 15\left(\frac{1}{6}\right) + 18\left(\frac{1}{6}\right)$$

$$= 3[1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right)]$$

$E(W) = 3[3.5] = 3E(Y)$

Example

The probability distribution for a random variable Y is given below. Find the expected value of Y and $W = Y^2$.

y	p(y)
-2	1/5
-1	1/5
0	1/5
1	1/5
2	1/5

Solution

By definition

$$E(Y) = (-2)p(-2) + (-1)p(-1) + (0)p(0) + (1)p(1) + 2p(2)$$

$$= \left(\frac{1}{5}\right)(-2 - 1 + 0 + 1 + 2) = 0$$

Solution

Note that the probability distribution of W is given by:

w	p(w)
0	$1/5$
1	$1/5 + 1/5 = 2/5$
4	$1/5 + 1/5 = 2/5$

Solution

By definition

$$E(W) = (0)P(W = 0) + (1)P(W = 1) + (4)P(W = 4)$$

$$= 0 + 2/5 + 8/5 = 10/5 = 2$$

Solution

Now, let us rewrite this in terms of Y

$$E(W) = (0)[P(Y=0)] + (1)[P(Y=-1) + P(Y=1)] \\ + (4)[P(Y=-2) + P(Y=2)]$$

$$E(W) = (0)P(Y=0) + (1)P(Y=-1) + (1)P(Y=1) \\ + (4)P(Y=-2) + (4)P(Y=2)$$

$$E(W) = (0)P(Y=0) + (-1)^2P(Y=-1) + (1)^2P(Y=1) \\ + (-2)^2P(Y=-2) + (2)^2P(Y=2)$$

$$E(W) = \sum_{y=-2}^{2} y^2 P(Y=y)$$

$$E(Y^2) = \sum_{y=-2}^{2} y^2 P(Y=y)$$

Definition

If Y is a random variable with mean $E(Y) = \mu$, the **variance** of a random variable Y is defined to be the expected value of $(Y - \mu)^2$. That is,

$$V(Y) = E[(Y - \mu)^2].$$

The **standard deviation** of Y is the positive square root of $V(Y)$.

Important Result

Let Y be a discrete random variable with probability function $p(y)$ and mean $E(Y) = \mu$; then

$$V(Y) = \sigma^2 = E[(Y - \mu)^2] = E(Y^2) - \mu^2.$$

or

$$V(Y) = E(Y^2) - [E(Y)]^2.$$

Example

The probability distribution for a random variable Y is given below. Find $V(Y)$.

y	p(y)
-2	1/5
-1	1/5
0	1/5
1	1/5
2	1/5

Exercise

A potential customer for an $85,000 fire insurance policy possesses a home in an area that, according to experience, may sustain a total loss in a given year with probability of 0.001 and a 50% loss with probability 0.01. Ignoring all other partial losses, what premium should the insurance company charge for a yearly policy in order to break even on all $85,000 policies in this area?

Solution

Let Y = Payout of an individual policy. First, we will find the probability distribution of Y.

$P(Y = 85,000) = 0.001$
$P(Y = 42,500) = 0.01$
$P(Y = 0) = 1 - (0.001 + 0.01) = 0.989$

Solution (cont.)

By definition,

$$E(Y) = 85,000P(Y = 85,000) + 42,500P(Y = 42,500) + 0P(Y = 0)$$

$$E(Y) = 85,000(0.001) + 42,500(0.01) + 0(0.989) = 85 + 425 = 510.$$

Definition

A **binomial experiment** possesses the following properties:

- The experiment consists of a fixed number, n, of identical trials.
- Each trial results in one of two outcomes: success, S, or failure, F.
- The probability of success on a single trial is equal to some value p and remains the same from trial to trial. The probability of a failure is equal to $q = (1 - p)$.
- The trials are **independent**.
- The random variable of interest is Y, the number of successes observed during the n trials.

Definition

A random variable Y is said to have a **binomial distribution** based on n trials with success probability p if and only if

$$p(y) = \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}, \;\; y = 0, 1, 2, ..., n \text{ and } 0 \leq p \leq 1.$$

Important Result

Let Y be a binomial random variable based on n trials and success probability p. Then

$$\mu = E(Y) = np \text{ and } \sigma^2 = V(Y) = npq.$$

Inheriting blood type (cont.)

What are the mean and standard deviation of the number of children with type O blood in our example.

Solution

The number of children with type O blood is a random variable X that has a Binomial distribution with $n = 5$ and $p = 0.25$.

$$\text{mean} \; = \mu = np = 5(0.25) = 1.25$$

$$\text{standard deviation} = \sqrt{np(1-p)} = \sqrt{5(0.25)(0.75)} \approx 0.9682$$

Another Example

A fire-detection device utilizes three temperature-sensitive cells acting independently of each other in such a manner that any one or more may activate the alarm. Each cell possesses a probability of $p = 0.8$ of activating the alarm when the temperature reaches 100 degrees Celsius or more. Let Y equal the number of cells activating the alarm when the temperature reaches 100 degrees.

a. Find the probability distribution for Y.

b. Find the probability that the alarm will function when the temperature reaches 100 degrees.

Solution

a) Let Y = number of cells activating the alarm. Y has a Binomial distribution with parameters $n = 3$ and $p = 0.8$

b)
$$\begin{aligned} P[\text{alarm will function}] &= P[Y \geq 1] \\ &= P[Y \geq 1] \\ &= 1 - P[Y = 0] \\ &= 1 - 0.008 = 0.992 \end{aligned}$$

R Code

```
1 - dbinom(0,size=3,prob=0.8);

## [1] 0.992
```