1 Bernoulli distribution

Let X be the random variable denoting the condition of the inspected item. Agree to write $X = 1$ when the item is defective and $X = 0$ when it is not. (This is a convenient notation because, once we inspect n such items, $X_1, X_2, ..., X_n$ denoting their condition, the total number of defectives will be given by $X_1 + X_2 + ... + X_n$.)

Let p denote the probability of observing a defective item. The probability distribution of X, then, is given by

x	0	1
$p(x)$	$q = 1 - p$	p

Such a random variable is said to have a *Bernoulli distribution*. Note that

$$\mathbb{E}(X) = \sum xp(x) = 0 \times p(0) + 1 \times p(1) = 0(q) + 1(p) = p \quad \text{and}$$

$$\mathbb{E}(X^2) = \sum x^2 p(x) = 0(q) + 1(p) = p.$$

Hence, $V(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = p - p^2 = pq$.

2. Binomial distribution

Now, let us inspect n items and count the total number of defectives. This process of repeating an experiment n times is called **Bernoulli trials**. The Bernoulli trials are formally defined by the following properties:

a) The result of each trial is either a success or a failure

b) The probability of success p is constant from trial to trial.

c) The trials are independent

d) The random variable X is defined to be the number of successes in n repeated trials

This situation applies to many random processes with just two possible outcomes: a heads-or-tails coin toss, a made or missed free throw in basketball etc[2]. We arbitrarily call one of these outcomes a "success" and the other a "failure".

Definition : **Binomial RV**

Assume that each Bernoulli trial can result in a success with probability p and a failure with probability $q = 1 - p$. Then the probability distribution of the binomial random variable X, the number of successes in n independent trials, is

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, 1, 2, \ldots, n.$$

The mean and variance of the binomial distribution are

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

We can notice that the mean and variance of the Binomial are n times larger than those of the Bernoulli random variable.

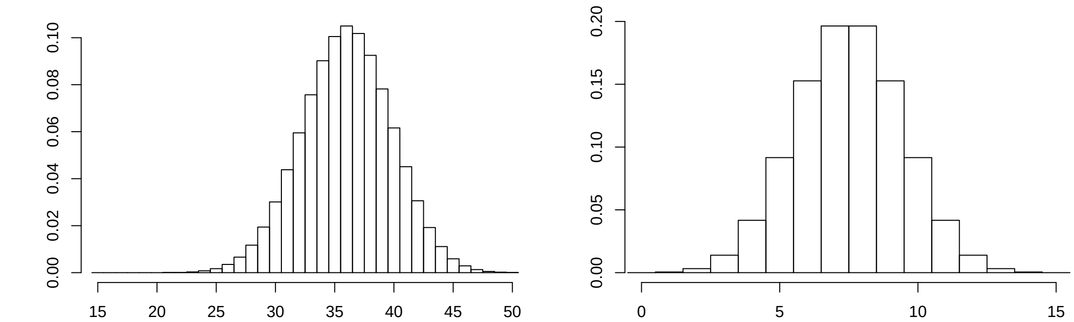

Figure : Binomial PMF: left, with $n = 60$, $p = 0.6$; right, with $n = 15$, $p = 0.5$

Note that Binomial distribution is symmetric when $p = 0.5$. Also, two Binomials with the same n and $p_2 = 1 - p_1$ are mirror images of each other.

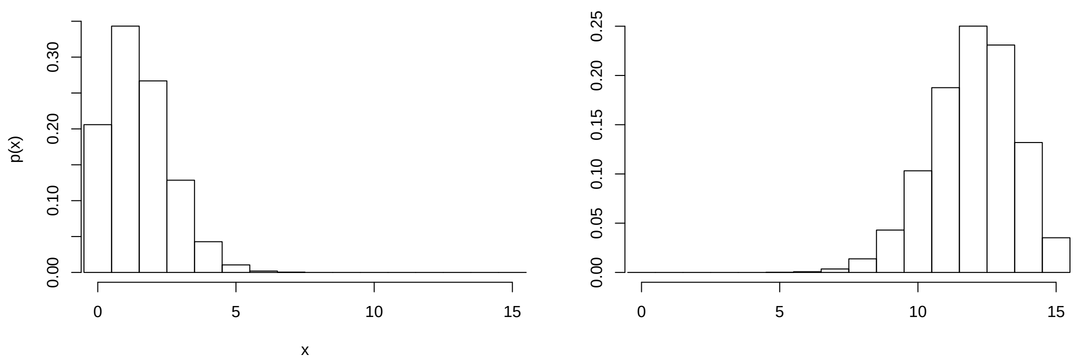

Figure : Binomial PMF: left, with $n = 15$, $p = 0.1$; right, with $n = 15$, $p = 0.8$

Example

The probability that a certain kind of component will survive a shock test is 0.75. Find the probability that

a) exactly 2 of the next 8 components tested survive,

b) at least 2 will survive,

c) at most 6 will survive.

Solution. (a) Assuming that the tests are independent and $p = 0.75$ for each of the 8 tests, we get

$$P(X = 2) = \binom{8}{2} (0.75)^2 \, (0.25)^{8-2} = \frac{8!}{2! \, (8-2)!} \, 0.75^2 \, 0.25^6 =$$

$$= \frac{40320}{2 \times 720} \, (0.5625)(0.000244) = 0.003843$$

(b)

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - [P(X = 1) + P(X = 0)]$$

$$= 1 - [8(0.75)(0.000061) + 0.000002] = 1 - 0.000386 \approx 0.9996$$

(c)
$$P(X \le 6) = 1 - P(X \ge 7) = 1 - [P(X = 7) + P(X = 8)]$$
$$= 1 - [0.2669 + 0.1001] = 1 - 0.367 = 0.633$$

□

Example

It has been claimed that in 60% of all solar heating installations the utility bill is reduced by at least one-third. Accordingly, what are the probabilities that the utility bill will be reduced by at least one-third in

(a) four of five installations;
(b) at least four of five installations?

Solution.

$$\text{(a)} \quad P(X = 4) = \binom{5}{4}(0.60)^4\,(0.4)^{5-4} = 5(0.1296)(0.4) = 0.2592$$

$$\text{(b)} \quad P(X = 5) = \binom{5}{5}(0.60)^5\,(0.40)^{5-5} = 0.60^5 = 0.0777$$

Hence, $P(\text{reduction for at least four}) = P(X \ge 4) = 0.2592 + 0.0777 = 0.3369$ □

Exercises

1

There's 50% chance that a mutual fund return on any given year will beat the industry's average. What proportion of funds will beat the industry average for at least 4 out of 5 last years?

2

Biologists would like to catch Costa Rican glass frogs for breeding. There is 70% probability that a glass frog they catch is male. If 10 glass frogs of a certain species are caught, what are the chances that they will have at least 2 male and 2 female frogs? What is the expected value of the number of female frogs caught?

3

A 5-member group are testing a new game console. Suppose that there's 50% chance that any given group member approves of the new console, and their opinions are independent of each other.

a) Calculate and fill out the probability distribution for X = number of group members who approve of the new console.

b) Calculate $P(X \ge 3)$.

c) How does your answer in part (b) change when there's 70% chance that any group member approves of the new console?

4

Suppose that the four engines of a commercial airplane were arranged to operate independently and that the probability of in-flight failure of a single engine is 0.01. Find:

a) Probability of no failures on a given flight.

b) Probability of at most one failure on a given flight.

c) The mean and variance for the number of failures on a given flight.

5

Suppose a television contains 60 transistors, 2 of which are defectives. Five transistors are selected at random, removed and inspected. Approximate

a) probability of selecting no defectives,

b) probability of selecting at least one defective.

c) The mean and variance for the number of defectives selected.

6

A train is made up of 50 railroad cars. Each car may need service with probability 0.05. Let X be the total number of cars in the train that need service.

a) Find the mean and standard deviation of X.

b) Find the probability that no cars need service.

c) Find the probability that at least two cars need service.

7

Show that mean and variance of the binomial random variable X are np and npq respectively.

8

If a thumb-tack is flipped, then the probability that it will land point-up is 0.3. If this thumb-tack is flipped 6 times, find:

a) the probability that it lands point-up on exactly 2 flips,

b) at least 2 flips,

c) at most 4 flips.

9

The proportion of people with type A blood in a certain city is reported to be 0.20. Suppose a random group of 20 people is taken and their blood types are to be checked. What is the probability that there are at least 4 people who have type A blood in the sample? What is the probability that at most 5 people in the group have type A blood¿

10

A die and a coin are tossed together. Let us define success as the event that the die shows an odd number and the coin shows a head (assume independence of the tosses). We repeat the experiment 5 times. What is the probability of exactly 3 successes?

3 Geometric distribution

In the case of Binomial distribution, the number of trials was a fixed number n, and the variable of interest was the number of successes. It is sometimes of interest to count instead how many trials are required to achieve a specified number of successes.

The number of trials Y required to obtain the first success is called a *Geometric random variable* with parameter p.

Theorem Geometric RV

The probability mass function for a Geometric random variable is

$$g(y;p) := P(Y = y) = (1-p)^{y-1}p, \quad y = 1, 2, 3, \ldots$$

Its CDF is

$$F(y) = 1 - q^y, \quad y = 1, 2, 3, \ldots, \quad q = 1 - p$$

Its mean and variance are

$$\mu = \frac{1}{p} \quad \text{and} \quad \sigma^2 = \frac{1-p}{p^2}$$

Proof. To achieve the first success on yth trial means to have the first $y-1$ trials to result in failures, and the last yth one a success, and then by independence of trials,

$$P(FF...FS) = q^{y-1}p$$

Now the CDF

$$F(y) = P(Y \leq y) = 1 - P(Y > y)$$

The latter means that all the trials up to and including the yth one, resulted in failures, which equals $P(y \text{ failures in a row}) = q^y$ and we get the CDF subtracting this from 1.

The mean $\mathbb{E}(Y)$ can be found by differentiating a geometric series:

$$\mathbb{E}(Y) = \sum_{y=1}^{\infty} yp(y) = \sum_{y=1}^{\infty} yp(1-p)^{y-1} = p\sum_{y=1}^{\infty} y(1-p)^{y-1} =$$

$$= p\sum_{y=1}^{\infty} \frac{d}{dq}q^y = p\,\frac{d}{dq}\sum_{y=1}^{\infty} q^y = p\left[\frac{d}{dq}(1 + q + q^2 + q^3 + \cdots - 1)\right] =$$

$$= p\left\{\frac{d}{dq}\left[(1-q)^{-1}\right] - \frac{d}{dq}(1)\right\} = \frac{p}{(1-q)^2} = \frac{1}{p}.$$

The variance can be calculated by differentiating a geometric series twice:

$$\mathbb{E}\{Y(Y-1)\} = \sum_{y=2}^{\infty} y(y-1)p\,q^{y-1} = pq\sum_{y=2}^{\infty} \frac{d^2}{dq^2}(q^y) =$$

$$= pq\frac{d^2}{dq^2}\left[\sum_{y=0}^{\infty} q^y\right] = pq\,\frac{d^2}{dq^2}(1-q)^{-1} = pq\,\frac{2}{(1-q)^3} = \frac{2q}{p^2}$$

$$\text{Hence} \quad \mathbb{E}(Y^2) = \frac{2q}{p^2} + \frac{1}{p} \quad \text{and} \quad V(Y) = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}$$

□

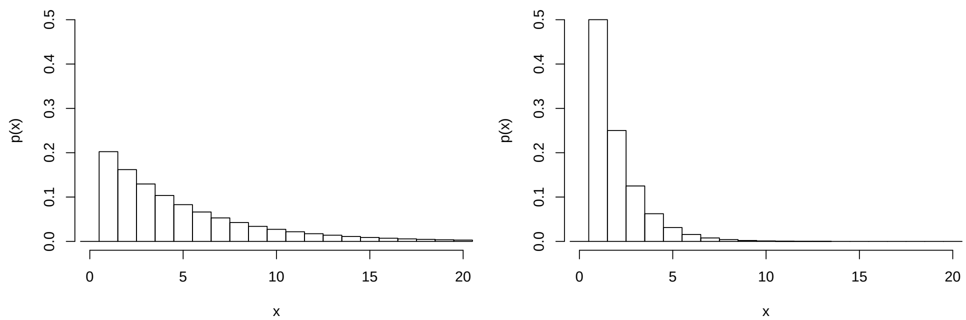

Figure : Geometric PMF: left, with $p = 0.2$; right, with $p = 0.5$

Example

For a certain manufacturing process it is known that, on the average, 1 in every 100 items is defective. What is the probability that the first defective item found is the fifth item inspected? What is the average number of items that should be sampled before the first defective is found?

Solution. Using the geometric distribution with $x = 5$ and $p = 0.01$, we have $g(5; 0.01) = (0.01)(0.99)^4 = 0.0096$.

Mean number of items needed is $\mu = 1/p = 100$. □

Example

If the probability is 0.20 that a burglar will get caught on any given job, what is the probability that he will get caught no later than on his fourth job?

Solution. Substituting $y = 4$ and $p = 0.20$ into the geometric CDF, we get $P(Y \leq 4) = 1 - 0.8^4 = 0.5904$ □

Exercises

1.

The probability to be caught while running a red light is estimated as 0.1. What is the probability that a person is first caught on his 10th attempt to run a red light? What is the probability that a person runs a red light at least 10 times without being caught?

2.

A computing center is interviewing people until they find a qualified person to fill a vacant position. The probability that any single applicant is qualified is 0.15.

a) Find the expected number of people to interview.

b) Find the probability the center will need to interview between 4 and 8 people (inclusive).

3.

From past experience it is known that 3% of accounts in a large accounting population are in error. What is the probability that the first account in error is found on the 5th try? What is the probability that the first account in error occurs in the first five accounts audited?

4.

A rat must choose between five doors, one of which contains chocolate. If the rat chooses the wrong door, it is returned to the starting point and chooses again (randomly), and continues until it gets the chocolate. What is the probability of the rat getting chocolate on the second attempt? Also, find the expected number of tries it takes to get the chocolate.

5.

If the probability of a success is 0.01, how many trials are necessary so that probability of at least one success is greater than 0.5?

6.

For the geometric distribution with $p = 0.02$, find (approximately) the *median* of the distribution, that is, m such that $F(m) \approx 0.5$. Compare to the mean. Find (approximately) the probability for this random variable to be less than its mean.