Introduction to Statistics (MAT 283)

Dipti Dubey

Department of Mathematics
Shiv Nadar University

Table of Contents

SOME SPECIAL DISCRETE DISTRIBUTIONS

If a random variable can take on k different values with equal probability, we say that it has a discrete uniform distribution.

1. The Discrete Uniform Distribution: A random variable X has a discrete uniform distribution and it is referred to as a discrete uniform random variable if and only if its probability mass function is given by

$$f(x) = \frac{1}{k}, \; x = x_1, \; x_2, \ldots, x_k$$

where $x_i \neq x_j$ for $i \neq j$.

EXAMPLES:

1. Tossing a fair coin. 2. Rolling a fair die.

EXAMPLES:

1. Tossing a fair coin. 2. Rolling a fair die.

MEAN AND VARIANCE:

$$E(X) = \frac{k+1}{2}$$

$$Var(X) = \frac{k^2-1}{12}$$

If an experiment has two possible outcomes, “success” and “failure”, and their probabilities are, respectively, p and $1 - p$, then the number of successes, 0 or 1, has a Bernoulli distribution; symbolically, we have the following definition:

If an experiment has two possible outcomes, "success" and "failure", and their probabilities are, respectively, p and $1 - p$, then the number of successes, 0 or 1, has a Bernoulli distribution; symbolically, we have the following definition:

2. The Bernoulli Distribution:

A random variable X has a Bernoulli distribution and it is referred to as a Bernoulli random variable if and only if its probability mass function is given by

$$f(x; p) = p^x (1 - p)^{1-x}$$

for $x = 0,\ 1.$

• Observe that we used the notation $f(x;p)$ to indicate explicitly that the Bernoulli distribution has one parameter p.

• We denote the Bernoulli random variable by writing $X \sim \text{BER}(p)$.

• We refer to an experiment to which the Bernoulli distribution applies as a Bernoulli trial, or simply a trial, and to sequences of such experiments as repeated trials.

• Observe that we used the notation $f(x;p)$ to indicate explicitly that the Bernoulli distribution has one parameter p.

• We denote the Bernoulli random variable by writing $X \sim \mathrm{BER}(p)$.

• We refer to an experiment to which the Bernoulli distribution applies as a Bernoulli trial, or simply a trial, and to sequences of such experiments as repeated trials.

EXAMPLES:

1. Tossing a fair coin.
2. What is the probability of getting a score of not less than 5 in a throw of a six-sided die?

THEOREM: If X is a Bernoulli random variable with parameter p, then the mean, variance and moment generating functions are respectively given by

$$E(X) = \mu_X = p$$
$$Var(X) = \sigma_X^2 = p(1-p)$$
$$M_X(t) = (1-p) + pe^t.$$

THEOREM: If X is a Bernoulli random variable with parameter p, then the mean, variance and moment generating functions are respectively given by

$$E(X) = \mu_X = p$$
$$Var(X) = \sigma_X^2 = p(1-p)$$
$$M_X(t) = (1-p) + pe^t.$$

- Note that for the Bernoulli distribution all its moments about zero are same and equal to p.

Consider a fixed number n of mutually independent Bernoulli trails. Suppose these trials have same probability of success, say p.

A random variable X is called a binomial random variable if it represents the total number of successes in n independent Bernoulli trials.

Consider a fixed number n of mutually independent Bernoulli trails. Suppose these trials have same probability of success, say p.

A random variable X is called a binomial random variable if it represents the total number of successes in n independent Bernoulli trials.

3. Binomial Distribution: The random variable X is called

the binomial random variable with parameters p and n if its probability mass function is of the form

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \ x = 0, 1, \ldots, n.$$

where $0 < p < 1$ is the probability of success.

We will denote a binomial random variable with parameters p and n as $X \sim \mathrm{BIN}(n, p)$.

Examples. 1. Tossing a fair coin twice and X denotes the number of heads.
2. On a five-question multiple-choice test there are five possible answers, of which one is correct. If a student guesses randomly and independently, what is the probability that she is correct only on two questions?

Excercise. On a five-question multiple-choice test there are five possible answers, of which one is correct. If a student guesses randomly and independently, what is the probability that she is correct only on questions 1 and 4?

Theorem: If X is a random variable with mean $E(X)$ and variance $Var(X)$, then

$$E(X) = np$$
$$Var(X) = \sigma_X^2 = np(1-p)$$

$$M_X(t) = [(1-p) + pe^t]^n.$$

The geometric distribution is also constructed from independent Bernoulli trials, but from an infinite sequence. Let X denote the trial number on which the first success occurs.

The geometric distribution is also constructed from independent Bernoulli trials, but from an infinite sequence. Let X denote the trial number on which the first success occurs.

4. Geometric Distribution:

The random variable X is called the geometric random variable with parameters p if its probability mass function is of the form

$$f(x) = (1-p)^{x-1}p, \; x = 0, 1, \ldots, \infty,$$

where $0 < p < 1$ is the probability of success in a single Bernouli's trial.

The geometric distribution is also constructed from independent Bernoulli trials, but from an infinite sequence. Let X denote the trial number on which the first success occurs.

4. Geometric Distribution:

The random variable X is called the geometric random variable with parameters p if its probability mass function is of the form

$$f(x) = (1-p)^{x-1}p, \ x = 0, 1, \ldots, \infty,$$

where $0 < p < 1$ is the probability of success in a single Bernouli's trial.

If X has a geometric distribution we denote it as $X \sim \text{GEO}(p)$.

EXAMPLE: : If X is the number of tosses needed until the first head when tossing a coin.

2. The probability of winning in a certain lottery is said to be about $1/9$. If it is exactly $1/9$, the distribution of the number of tickets a person must purchase up to and including the first winning ticket is a geometric random variable with $p = 1/9$.

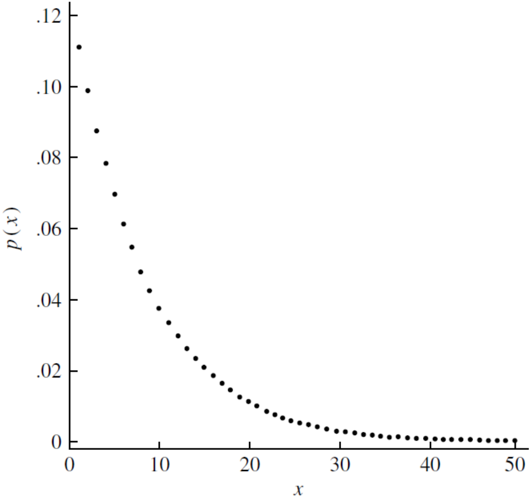

FIGURE : The probability mass function of a geometric random variable with $p = \frac{1}{9}$.

Theorem: If X is a geometric random variable with parameter p, then the mean, variance and moment generating functions are respectively given by

$$E(X) = \mu_X = \frac{1}{p}$$

$$Var(X) = \sigma_X^2 = \frac{1-p}{p^2}$$

$$M_X(t) = \frac{pe^t}{1-(1-p)e^t}, \text{ if } t < log(1-p).$$

Let X denote the trial number on which the rth success occurs. Here r is a positive integer greater than or equal to one. This is equivalent to saying that the random variable X denotes the number of trials needed to observe the rth successes.

Let X denote the trial number on which the rth success occurs. Here r is a positive integer greater than or equal to one. This is equivalent to saying that the random variable X denotes the number of trials needed to observe the rth successes.

5. Negative Binomial (or Pascal) Distribution:

The random variable X is called the negative distribution random variable with parameters p if its probability mass function is of the form

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r, \; x = r, r+1 \ldots, \infty,$$

where $0 < p < 1$ is the probability of success in a single Bernouli's trial.

Let X denote the trial number on which the rth success occurs. Here r is a positive integer greater than or equal to one. This is equivalent to saying that the random variable X denotes the number of trials needed to observe the rth successes.

5. Negative Binomial (or Pascal) Distribution:

The random variable X is called the negative distribution random variable with parameters p if its probability mass function is of the form

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r, \; x = r, r+1 \ldots, \infty,$$

where $0 < p < 1$ is the probability of success in a single Bernouli's trial.

If X has a negative binomial distribution we denote it as $X \sim \text{NBIN}(p)$.

EXAMPLE: What is the probability that the second head is observed on the 3rd independent flip of a coin?

EXAMPLE: What is the probability that the second head is observed on the 3rd independent flip of a coin?

In this case $p = \frac{1}{2}$

$$P(X = 3) = f(3) = \binom{2}{1} p^2 (1 - p)$$

EXAMPLE: The distribution of the number of tickets purchased up to and including the second winning ticket is negative binomial: $P(X = k) = (k-1)p^2(1-p)^{k2}$.

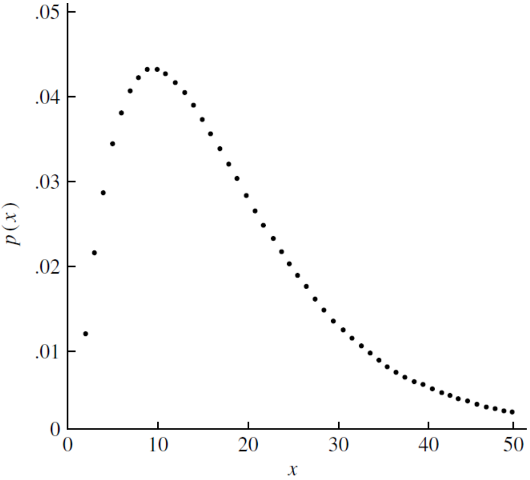

FIGURE The probability mass function of a negative binomial random variable with $p = \frac{1}{9}$ and $r = 2$.

We shall now investigate the limiting form of the binomial distribution when $n \to \infty$, $p \to 0$, while np remains constant. Letting this constant be λ, that is, $np = \lambda$ and, hence, $p = \lambda/n$, we can write

$$f(x; n, p) = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{n(n-1)(n-2)\ldots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= 1\frac{(1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{x-1}{n})}{x!}(\lambda)^x \left(\left(1 - \frac{\lambda}{n}\right)^{-n/\lambda}\right)^{-\lambda} \left(1 - \frac{\lambda}{n}\right)^{-x}$$

Finally, if we let $n \to \infty$ while x and λ remain fixed, we find that

$$1(1-\frac{1}{n})(1-\frac{2}{n})\dots(1-\frac{x-1}{n}) \to 1$$

$$\left(1-\frac{\lambda}{n}\right)^{-x} \to 1$$

$$\left(1-\frac{\lambda}{n}\right)^{-n/\lambda} \to e.$$

and, hence, that the limiting distribution becomes

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!},\ x = 1, 2, \dots, \infty,$$

5. Poisson Distribution:

A random variable X is said to have a Poisson distribution if its probability mass function is given by

0

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \; x = 1, 2, \ldots, \infty,$$

where $0 < \lambda$ is a parameter.

We denote such a random variable by $X \sim \text{POI}(\lambda)$.

5. Poisson Distribution:

A random variable X is said to have a Poisson distribution if its probability mass function is given by

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \; x = 1, 2, \ldots, \infty,$$

where $0 < \lambda$ is a parameter.

We denote such a random variable by $X \sim \mathrm{POI}(\lambda)$.

Example given below is taken from the book "*Mathematical Statistics and Data Analysis by John A. Rice*".

EXAMPLE A Two dice are rolled 100 times, and the number of double sixes, X, is counted. The distribution of X is binomial with $n = 100$ and $p = \frac{1}{36} = .0278$. Since n is large and p is small, we can approximate the binomial probabilities by Poisson probabilities with $\lambda = np = 2.78$. The exact binomial probabilities and the Poisson approximations are shown in the following table:

k	Binomial Probability	Poisson Approximation
0	.0596	.0620
1	.1705	.1725
2	.2414	.2397
3	.2255	.2221
4	.1564	.1544
5	.0858	.0858
6	.0389	.0398
7	.0149	.0158
8	.0050	.0055
9	.0015	.0017
10	.0004	.0005
11	.0001	.0001

The approximation is quite good. ∎

THEOREM: If X is a Poisson's random variable with parameter λ, then the mean, variance and moment generating functions are respectively given by

$$E(X) = \mu_X = \lambda$$
$$\mathrm{Var}(X) = \sigma_X^2 = \lambda$$
$$M_X(t) = e^{\lambda(e^t - 1)}.$$

Table of Contents

SOME SPECIAL CONTINUOUS DISTRIBUTIONS

1. Uniform Distribution

A random variable X is said to be uniform on the interval $[a, b]$ if its probability density function is of the form

$$f(x) = \frac{1}{b-a}, \; a \leq x \leq b$$

where a and b are constants.

We denote a random variable X with the uniform distribution on the interval $[a, b]$ as $X \sim \text{UNIF}(a, b)$.

APPLICATION: Random number generation.