Statistical Methods 105

Department of Statistics and Operations Research

Edited by: Reem Alghamdi

August 2020

Chapter 1
Discrete Random Variable

Plan

Plan

1) Discrete Probability Distributions

Definition (Probability function)

The set of ordered pairs $(x, f(x))$ is a probability function, probability mass function, or probability distribution of the discrete random variable X if, for each possible outcome x,

1. $f(x) \geq 0$,
2. $\sum_{x \in X} f(x) = 1$,
3. $P(X = x) = f(x)$.

Definition (cumulative distribution function)

The cumulative distribution function $F(x)$ of a discrete random variable X with probability distribution $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \quad for \quad -\infty < x < +\infty.$$

Definition (Mean of a Random Variable)

Let X be a random variable with probability distribution $f(x)$. The mean, or expected value, of X is

$$\mu = E(x) = \sum_{x} x\, f(x).$$

Theorem

Let X be a random variable with probability distribution $f(x)$. The expected value of the random variable $g(X)$ is

$$\mu_{g(X)} = E[g(X)] = \sum_{x} g(x) f(x).$$

Example

Suppose that the number of cars X that pass through a car wash between 4:00 P.M. and 5:00 P.M. on any sunny Friday has the following probability distribution:

x	4	5	6	7	8	9
$f(x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

1. Find $E(X)$.
2. Let $g(X) = 2X - 1$ represent the amount of money, in dollars, paid to the attendant by the manager. Find the attendant's expected earnings for this particular time period.

Solution

Simple calculations yield:

x	4	5	6	7	8	9
$f(x)$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$.

$xf(x)$	$\frac{1}{3}$	$\frac{5}{12}$	$\frac{3}{2}$	$\frac{7}{4}$	$\frac{4}{3}$	$\frac{3}{2}$
$g(x)$	7	9	11	13	15	17
$g(x)f(x)$	$\frac{7}{12}$	$\frac{9}{12}$	$\frac{11}{4}$	$\frac{13}{4}$	$\frac{15}{6}$	$\frac{17}{6}$

1. $E(X) = \frac{1}{3} + \frac{5}{12} + \frac{3}{2} + \frac{7}{4} + \frac{4}{3} + \frac{3}{2} = 6.83$
2. The attendant's expected earnings for this particular time period is equal to:

$$E[g(X)] = \frac{7}{12} + \frac{9}{12} + \frac{11}{4} + \frac{13}{4} + \frac{15}{6} + \frac{17}{6} = 14.67.$$

Properties of the mean:

Theorem

Let X a random variable. If a and b are constants, then $E(aX + b) = aE(X) + b$.

Theorem

The expected value of the sum or difference of two or more functions of a random variable X is the sum or difference of the expected values of the functions. That is,

$$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)].$$

Example

Let X be a random variable with probability distribution as follows:

x	0	1	2	3
$f(x)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$

Find the expected value of $Y = (X-1)^2$.

Solution

Simple calculations yield:

x	0	1	2	3
$f(x)$	$\frac{1}{3}$	$\frac{1}{2}$	0	$\frac{1}{6}$
$g(x)$	1	0	1	4
$f(x)g(x)$	$\frac{1}{3}$	0	0	$\frac{2}{3}$

Therefore, the expected value of Y is equal to:

$$E(Y) = E[g(X)] = 1.$$

Another solution: by using the properties of the mean theorems, $E(Y) = E((X-1)^2) = E(X^2) - 2E(X) + 1.(Hint!)$

Theorems (Variance of Random Variable)

Let X be a random variable with probability distribution $f(x)$ and mean μ. The variance of X is

$$Var(X) = \sigma^2 = E[(X - \mu)^2] = \sum_x (x - \mu)^2 f(x).$$

or it can be written as:

$$\sigma^2 = E(X^2) - E(X)^2$$

$$Var(g(X)) = \sigma^2_{g(X)} = E[(g(x) - \mu_{g(X)})^2] = \sum_x (g(x) - \mu_{g(X)})^2 f(x).$$

The positive square root of the variance, σ, is called the standard deviation of X.

Properties of the Variance:

Theorem

Let X a random variable. If a and b are constants, then $Var(aX + b) = a^2 Var(X)$.

Corollary

Setting $a = 1$, then $Var(X + b) = Var(X)$.

Corollary

Setting $b = 0$, then $Var(aX) = a^2 Var(X)$.

Example

Calculate the variance of $g(X) = 2X + 3$, where X is a random variable with probability distribution:

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

Solution

Simple calculations yield

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$
$g(x)$	3	5	7	9
$g(x)f(x)$	$\frac{3}{4}$	$\frac{5}{8}$	$\frac{7}{2}$	$\frac{9}{8}$

Therefore, The expected value of $g(X)$ is equal to

$$E[g(X)] = \frac{3}{4} + \frac{5}{8} + \frac{7}{2} + \frac{9}{8} = 6$$

So, the variance of $g(X) = 2X + 3$ is equal to

$$\sigma^2 = (3-6)^2 * \frac{1}{4} + (5-6)^2 * \frac{1}{8} + (7-6)^2 * \frac{1}{2} + (9-6)^2 * \frac{1}{8} = 4,$$

and the standard deviation of of $g(X)$ is equal to: $\sigma = \sqrt{4} = 2.$

Another solution: By using the properties of the variance

$$Var(2X+3) = 2^2 Var(X) = 4Var(X) = 4[E(X^2) - E(X)^2]$$

.

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$
x^2	0	1	4	9

Therefore, $E[X^2] = \frac{1}{8} + 2 + \frac{9}{8} = \frac{26}{8}$, and $E[X] = \frac{1}{8} + 1 + \frac{3}{8} = \frac{12}{8}$.
Then, the variance of $g(X) = 2X + 3$ is equal to

$$\sigma^2 = 4\Big[\frac{26}{8} - (\frac{12}{8})^2\Big] = 4.$$

Plan

Definition (Discrete Uniform Random Variable)

A random variable X is called discrete uniform if has a finite number of specified outcomes, say $x_1, x_2, \ldots, x_k$ and each outcome is equally likely. Then, the discrete uniform mass function is given by:

$$f(x) = P(X = x) = \begin{cases} \frac{1}{k}, & x = x_1, x_2, \ldots, x_k \\ 0, & \text{otherwise.} \end{cases}$$

Note: k is called the parameter of the distribution.

Theorem

The expected value (mean) and variance of the discrete uniform distribution are:

$$\mu = E(X) = \sum_{i=1}^{k} \frac{x_i}{k}, \text{ and } \sigma^2 = \frac{1}{k} \sum_{i=1}^{k} [x_i - E(X)]^2.$$

Example

Suppose that you select a ball from a box contains 6 balls labeled $1, 2, \cdots, 6$. Let $X =$ the number that is observed when selecting a ball. Find $E(X)$ and $Var(X)$.

Solution

The probability distribution of X is:

$$P(X = x) = \begin{cases} \frac{1}{6}, & x = x_1, x_2, \ldots, x_6 \\ 0, & \text{otherwise.} \end{cases}$$

The expected value:

$$\mu = E(X) = \sum_{i=1}^{k} \frac{x_i}{k} = \frac{1+2+3+4+5+6}{6} = 3.5.$$

The variance:

$$\sigma^2 = \frac{1}{k} \sum_{i=1}^{k} [x_i - E(X)]^2 = \frac{(1-3.5)^2 + \cdots + (6-3.5)^2}{6} = 2.92.$$

Definition (Bernouilli Process)

The process is said to be a Bernoulli process if:

- The outcomes of process is either success or failure.
- The probability of success is $P(X = 1) = p$ and the probability of failure is $P(X = 0) = 1 - p = q$.

Strictly speaking, trials of random experiment are called binomial trials if satisfy the following conditions:

1. The experiment consists of finite number of repeated trials.
2. Each trial has exactly two outcomes: success or failure.
3. The repeated trials are independent.
4. The probability of success remains the same in each trial.

Definition (Binomial Distribution)

The binomial distribution is defined based on the Bernoulli process. It is made up of n independent Bernoulli processes. Suppose that $X_1, X_2, \cdots, X_n$ are independent Bernoulli random variables, then $Y = \sum Xi$ will conform binomial distribution. The probability mass function of the binomial random variable X is given by:

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

We denote the binomial distribution with the parameters n and p, by $Bin(n, p)$ and $\binom{n}{x} = \frac{n!}{x!(n-x)!}$.

The Cumulative Distribution Function (CDF)

The cumulative distribution function (CDF) of binomial distribution $Bin(n, p)$ is:

$$F_X(x) = P(X \leq x) = \sum_{i=0}^{x} \binom{n}{i} p^i q^{n-i}.$$

Theorem

The mean and variance of the binomial distribution $Bin(n, p)$ are

$$\mu = n\,p \; and \; \sigma^2 = n\,p\,q.$$

Example

If the mean and the variance of a binomial distribution are 10 and 5 respectively, then:

1. Determine the probability mass function.
2. Calculate the probability $P(X = 0)$.
3. Calculate the probability $P(X \geq 2)$.

Solution

1. By solving $E(X) = np = 10$ and $Var(X) = np(1-p) = 5$, we get $p = 0.5$ and $n = 20$. The probability mass function is:

$$P(X = x) = \binom{20}{x}\left(0.5\right)^{x}\left(0.5\right)^{20-x}, x = 0, 1, \cdots, 20$$

2. $P(X = 0) = \binom{20}{0}\left(0.5\right)^{0}\left(0.5\right)^{20} = 0.5^{20}$.
3. $P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - 0.00002 = 0.99998$.

Example

Suppose that the probability that a person dies when he or she contracts a certain disease is 0.4. A sample of 10 persons who contracted this disease is randomly chosen. Find the following:

1. The probability that exactly 4 persons will die.
2. The probability that less than 3 persons will die.
3. The probability that more than 8 persons will die.
4. The expected number of persons who will die.
5. The variance of the number of persons who will die.

<u>Solution</u>

The probability mass function is:

$$P(X = x) = \binom{10}{x}\left(0.4\right)^{x}\left(0.6\right)^{10-x}, \; x = 0, 1, \cdots, 10$$

1. $P(X = 4) = \binom{10}{4}(0.4)^4(0.6)^{10-4} = 0.251$

2.

$$\begin{aligned} P(X < 3) =& P(X = 0) + P(X = 1) + P(X = 2) \\ =& \sum_{x=0}^{2} \binom{10}{x} (0.4)^x (0.6)^{10-x} = 0.167. \end{aligned}$$

3.

$$\begin{aligned} P(X > 8) =& P(X = 9) + P(X = 10) \\ =& \sum_{x=9}^{10} \binom{10}{x} (0.4)^x (0.6)^{10-x} = 0.0017. \end{aligned}$$

4. $E(X) = np = (10)(0.4) = 4$
5. $Var(X) = np(1-p) = (10)(0.4)(0.6) = 2.4$

Definition

The probability distribution of the hypergeometric random variable X describes the probability of K successes (random draws for which the object drawn has a specified feature) in n draws, without replacement, from a finite population of size N that contains exactly K objects with that feature, where each draw is either a success or a failure.

There are two methods of selection:

1. Selection with replacement: If we select the elements of the sample at random and with replacement, then $X \sim Bin(n, p)$; where $p = \frac{K}{N}$
2. Selection without replacement: When the selection is made without replacement, the random variable X has a hypergeometric distribution with parameters N, n, and K. and we write $X \sim h(x; N, n, K)$.

The probability mass function for hypergeometric random variable X is:

$$f(x) = P(X = x) = \begin{cases} \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} & x = 0, 1, 2, \cdots, min(K, n) \\ 0, & \text{otherwise.} \end{cases}$$

Theorem

The mean and variance of the hypergeometric distribution $h(x; N, n, K)$ are

$$\mu = n\frac{K}{N} \text{ and } \sigma^2 = n\frac{K}{N}\left(1 - \frac{K}{N}\right)\frac{N-n}{N-1}.$$

Example

Suppose there are 50 officers, 10 female officers and 40 male officers. Suppose 20 of them will be selected for promotion. Let X represent the number of female promotions. Find:

1. The probability that exactly one female is found in the sample.
2. The expected value (mean) and the variance of the number of females in the sample.

Solution

- Note that the binomial distribution doesn't apply here, as the officers are without replacement once they are drawn. In other words, the trials are not independent events.
- X has a hypergeometric distribution with $N = 50, n = 20$, and $K = 10$; i.e. $X \sim h(x; N, n, K) = h(x; 50, 20, 10)$.

$$P(X = x) = \begin{cases} \frac{\binom{10}{x}\binom{50-10}{20-x}}{\binom{50}{20}} & x = 0, 1, 2, \cdots, 10 \\ 0, & \text{otherwise.} \end{cases}$$

1. The probability that exactly one female is found in the sample is:

$$f(1) = P(X = 1) = \frac{\binom{10}{1}\binom{40}{19}}{\binom{50}{20}} = 0.0279$$

2. The expected value (mean) is $E(X) = n\frac{K}{N} = 20 \times \frac{10}{50} = 4$.
3. The variance is
$\sigma^2 = n\,\frac{K}{N}\left(1 - \frac{K}{N}\right)\frac{N-n}{N-1} = 20 \times \frac{10}{50} \times (1 - \frac{10}{50}) \times \frac{50-20}{50-1} = 1.9592$

(Binomial Approximation Theorem)

If n is small compared to K, then a binomial distribution $Bin(n, p = \frac{K}{N})$ can be used to approximate the hypergeometric distribution $h(x; N, n, K)$.

Example

A manufacturer of automobile tires reports that among a shipment of 5000 sent to a local distributor, 1000 are slightly blemished. If one purchases 10 of these tires at random from the distributor, what is the probability that exactly 3 are blemished?

Solution

Since $K = 1000$ is large relative to the sample size $n = 10$, we shall approximate the desired probability by using the binomial distribution. The probability of obtaining a blemished tire is 0.2. Therefore, the probability of obtaining exactly 3 blemished tires is

$$h(3; 5000, 10, 1000) \approx Bin(10, p = \frac{1000}{5000}) = \binom{10}{3} (0.2)^3 (0.8)^7 = 0.2013.$$

2.4)Poisson Distribution

Definition

Let X the number of outcomes occurring during a given time interval. X is called a Poisson random variable, with parameter λ, when its probability mass function is given by

$$P(x,\lambda) = P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \;\; x = 0, 1, 2, \ldots$$

where e is an irrational number approximately equal to 2.71828 and λ is the average number of occurrences per interval unit.

Theorem

If a random variable X has a Poisson distribution. Then both the mean and the variance of X are λ.

$$\mu = \lambda \;\; and \;\; \sigma^2 = \lambda$$

Example

The mean number of accidents per month at a certain intersection is 3.

1. What is the probability that in any given month 4 accidents will occur at this intersection?
2. What is the probability that more than 4 accidents will occur in any given month at the intersection?
3. What is the probability that 4 accidents will occur in 5 months?

Solution

1. $f(4) = P(X = 4) = e^{-3}\frac{3^4}{4!} = 0.168.$
2. $P(X > 4) = 1 - P(X \leq 4) = 1 - [P(X = 0) + \cdots + P(X = 4)] = 1 - [\sum_{x=0}^{4} e^{-3}\frac{3^x}{x!}] = 0.1847.$
3. Since the average number of accidents at a certain intersection per month is 3, thus the average number of accidents in 5 months is 15. Let X represent the number of accidents in 5 months, $f(x) = P(X = x) = e^{-15}\frac{15^x}{x!}, \quad x = 0, 1, 2, \ldots.$ Then, $f(4) = P(X = 4) = e^{-15}\frac{15^4}{4!} = 0.00065.$

Theorem (Approximation)

Let X be a binomial random variable with probability distribution $B(n,p)$. When n is large ($n \to +\infty$), and p small ($p \to 0$), then the poisson distribution can be used to approximate the binomial distribution $B(n,p)$ by taking $\lambda = np$.

Example

In a certain industrial facility, accidents occur infrequently. It is known that the probability of an accident on any given day is 0.005 and accidents are independent of each other.

1. What is the probability that in any given period of 400 days there will be an accident on one day?
2. What is the probability that there are at most three days with an accident?

Solution

Let X be a binomial random variable with $n = 400$ and $p = 0.005$. Thus, $np = 2$. Use the Poisson approximation,

1. $$P(X = 1) = e^{-2}\, 2^1 = 0.271.$$
2. $$P(X \leq 3) = \sum_{x=0}^{3} e^{-2} \frac{2^x}{x!} = 0.857.$$