DISCRETE RANDOM VARIABLES

MSc. Nguyễn Hoàng Huy Tú

November 18, 2023

Random Variable

Definition

A random variable is a variable that takes on numerical values realized by the outcomes in a random experiment.

We use capital letters, such as X, to denote the random variable and the corresponding lowercase letter, x, to denote a possible value.

Example:

- The face values of a single dice when a fair dice is rolled.
- The number of claims on a medical insurance policy in a particular year.
- The change in the price of a share of IBM common stock in a month.

Discrete Random Variable

Definition

A random variable is called a discrete random variable if the number of possible values is finite or infinite but countable.

Example:

- The number of heads when three fair coins are tossed independently.
- The number of defective items in a sample of 20 items from a large shipment.
- The number of customers arriving at a checkout counter in an hour.

Continuous Random Variable

Definition

A random variable is a continuous random variable if it can take any value in an interval.

Example:

- The yearly income for a family.
- The amount of oil imported into the United States in a particular month.
- Daily cell phone sales of a large company.

Exercise

Ex 1: For each of the following, indicate if a discrete or a continuous random variable provides the best definition:

a. The number of cars that arrive each day for repair in a two-person repair shop

b. The number of cars produced annually by General Motors

c. Total daily e-commerce sales in dollars

d. The number of passengers that are bumped from a specific airline flight 3 days before Christmas

Exercise

Ex 2: A presidential election poll contacts 2,000 randomly selected people. Should the number of people that support candidate A be analyzed using discrete or continuous probability models?

Ex 3: A salesperson contacts 20 people each day and requests that they purchase a specific product. Should the number of daily purchases be analyzed using discrete or continuous probability models?

Distribution Table

Definition

The possible values of a discrete random variable X are $x_1, x_2, \ldots, x_n$ with the corresponding probabilities $P(X = x_i) = p_i, i = 1, 2, \ldots, n$. distribution table is defined as follow:

X	x_1	x_2	$\ldots$	x_n
P	p_1	p_2	$\ldots$	p_n

For example, let X be the face values of a single die when a fair dice is rolled. Then the distribution table is

X	1	2	3	4	5	6
P	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Probability Distribution Function

Suppose that X is a discrete random variable and that x is one of its possible values. The probability that random variable X takes specific value x is denoted $P(X = x)$.

Definition

The probability distribution function, denoted by $P(x)$, of a discrete random variable X is defined by

$$P(x) = Pr(X = x), \forall x$$

Properties

Proposition

Let X be a discrete random variable with probability distribution $P(x)$. Then,

- $0 \leq P(x) \leq 1, \forall x$.
- $\sum_x P(x) = 1$.

Property 1 merely states that probabilities cannot be negative or exceed 1. Property 2 follows from the fact that the events "$X = x$", for all possible values of x, are mutually exclusive and collectively exhaustive.

Exercise

Ex 1: Show the probability distribution function (distribution table) of the number of heads when two fair coins are tossed independently.

Ex 2: A box has 10 products, including 4 defective products. Pick 2 products at random. Make a probability distribution table of the number of defective products taken out?

Ex 3: There are 2 packages. Package 1 has 3 good products and 2 bad products. Package 2 has 2 good products and 3 bad products. Randomly take 2 products from package 1 and 1 product from package 2. Establish a distribution table of the number of good products among the 3 selected products?

Exercise

Ex 4: The number of computers sold per day at Dan's Computer Works is defined by the following probability distribution:

X	0	1	2	3	4	5	6
P	0,05	0,10	0,20	0,20	0,20	0,15	0,10

Compute

a. $P(3 \leq X \leq 6)$

b. $P(X > 3)$

c. $P(X \leq 4)$

d. $P(2 < X \leq 5)$

Exercise

Ex 5: The probability that a university graduate will be offered no jobs within a month of graduation is estimated to be 5%. The probability of receiving one, two, and three job offers has similarly been estimated to be 43%, 31%, and 21%, respectively. Determine the following probabilities.

a. Make a distribution table.

b. Determine the probability that a graduate is offered fewer than two jobs.

c. Determine the probability that a graduate is offered more than one job.

Cumulative Probability Distribution

Definition

The cumulative probability distribution, denoted by $F(x)$, of a random variable X is defined by

$$F(x) = P(X \leq x)$$

It represents the probability that X does not exceed the value x.

Example

Olaf Motors, Inc., is a car dealer in a small southern town. Based on an analysis of its sales history, the managers know that on any single day the number of Prius cars sold can vary from 0 to 5. The distribution table is as follow

X	0	1	2	3	4	5
$P(x)$	0,15	0,3	0,2	0,2	0,1	0,05
$F(x)$	0,15	0,45	0,65	0,85	0,95	1

The third row contains the cumulative distribution, $F(x)$, and it is easily determined by the second row.
This model could be used for planning the inventory of cars. For example, if there are only four cars in stock, Olaf Motors could satisfy customers' needs for a car 95% of the time.

Properties

Proposition

Let X be a discrete random variable with cumulative probability distribution $F(x)$. Then

i) $F(x_0) = \sum_{x \leq x_0} P(x)$.

ii) $0 \leq F(x) \leq 1, \forall x$.

iii) $F(x)$ is increasing. That is, if x_1 and x_2 are two numbers with $x_1 < x_2$, then $F(x_1) < F(x_2)$.

Expected Value

Definition

The expected value, denoted by $E(X)$, of a discrete random variable X is defined as

$$E(X) = \sum_x xP(x)$$

The expected value of a random variable is also called its mean and is denoted by μ.

Solution steps: Discrete Random Variable $\longrightarrow$ Distribution table $\longrightarrow$ Compute the expected value by the formula.

Example

Suppose that the probability distribution for the number of errors, X, on pages from business textbooks is as follows:

$$P(0) = 0,81, P(1) = 0,17, P(2) = 0,02$$

Find the mean number (expected value) of errors per page.

We have

$$E(X) = \sum_x xP(x) = 0 \times 0,81 + 1 \times 0,17 + 2 \times 0,02$$

Variance

Definition

Let X be a discrete random variable. The expectation of the squared deviations about the mean, denoted by σ^2, is called the variance and given by

$$\begin{aligned}\sigma^2 &= E[(X-\mu)^2] \\ &= \Sigma_x (x-\mu)^2 P(x) \\ &= (\Sigma_x x^2 P(x)) - \mu^2\end{aligned}$$

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

Example

A contractor is interested in the total cost of a project on which she intends to bid. She estimates that materials will cost $ 25,000 and that her labor will be $900 per day. The project will takes X days to complete. Given the probability distribution for completion times

X	10	11	12	13	14
P	0,1	0,3	0,3	0,2	0,1

a. Find the mean and variance for completion time X.
b. Let Y denote the total cost of project (in dollars). Find the mean, variance and standard deviation for total cost C.

Linear Functions of a Random Variable

Properties

Let X be a random variable with mean μ_X and the variance σ_X^2. If $Y = a + bX$, then

- $\mu_Y = E(a + bX) = a + b\mu_X$.
- $\sigma_Y^2 = Var(a + bX) = b^2\sigma_X^2$.

In particular,

- $E(a) = 0$, $Var(a) = 0$ where a is a constant number.
- $E(bX) = b\mu_X$ and $Var(bX) = b^2\sigma_X^2$.

Exercise

Ex 1: Given the probability distribution function:

X	0	1	2
P	0,25	0,5	0,25

a. Calculate the cumulative probability distribution.

b. Find the mean, variance and standard deviation.

Ex 2: A shipment of 20 parts contains 2 defectives. Two parts are chosen at random from the shipment and checked. Let the random variable Y denote the number of defectives found. Find the mean, variance and standard deviation of the random variable Y.

Ex 3:A very large shipment of parts contains 10% defectives. Two parts are chosen at random from the shipment and checked. Let the random variable Z denote the number of defectives found. Find the expected value, variance and standard deviation of this random variable.

Exercise

Ex 4:A factory manager is considering whether to replace a temperamental machine. A review of past records indicates the following probability distribution for the number of breakdowns of this machine in a week.

Number of breakdowns	0	1	2	3	4
Probability	0,1	0,26	0,42	0,16	0,06

a. Find the mean and standard deviation of the number of weekly breakdowns.

b. It is estimated that each breakdown costs the company \$1,500 in lost output. Find the mean and standard deviation of the weekly cost to the company from breakdowns of this machine.

Exercise

Ex 5: An investor is considering three strategies for a \$1,000 investment. The probable returns are estimated as follows:

- Strategy 1: A profit of \$10,000 with probability 0.15 and a loss of \$1,000 with probability 0.85.
- Strategy 2: A profit of \$1,000 with probability 0.50, a profit of \$500 with probability 0.30, and a loss of \$500 with probability 0.20.
- Strategy 3: A certain profit of \$400.

Which strategy has the highest expected profit? Explain why you would or would not advise the investor to adopt this strategy.

Binomial Distribution

Definition

Suppose that

- a random experiment can result in two possible outcomes, “success” and “failure,”
- and that p is the probability of a success in a single trial.

Let X be the the number of resulting successes in n independent trials. The probability distribution of X is called binomial distribution.

Formulas of Binomial Distribution

Probability Distribution Function

$$P(k) = P(X = k) = C_k^n p^k (1 - p)^{n-k}$$

Expected Value and Variance

$$\mu_X = E(X) = np$$

and

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

Example: Sales of Airline Seats

Suppose that you are in charge of marketing airline seats for a major carrier. Four days before the flight date you have 16 seats remaining on the plane. You know from past experience data that 80% of the people that purchase tickets in this time period will actually show up for the flight.

a. If you sell 20 extra tickets, what is the probability that you will overbook the flight?

b. If you sell 18 extra tickets, what is the probability that you have at least 1 empty seat?

Exercise

Ex 1: A production manager knows that 5% of components produced by a particular manufacturing process have some defect. Six of these components, whose characteristics can be assumed to be independent of each other, are examined.

a. What is the probability that none of these components has a defect?

b. What is the probability that one of these components has a defect?

c. What is the probability that at least two of these components have a defect?

Exercise

Ex 2: A notebook computer dealer mounts a new promotional campaign. Purchasers of new computers may, if dissatisfied for any reason, return them within 2 days of purchase and receive a full refund. The cost to the dealer of such a refund is \$100. The dealer estimates that 15% of all purchasers will, indeed, return computers and obtain refunds. Suppose that 50 computers are purchased during the campaign period.

a. Find the mean and standard deviation of the number of these computers that will be returned for refunds.

b. Find the mean and standard deviation of the total refund costs that will accrue as a result of these 50 purchases.

Example

The following discrete random variables have the Poisson distribution.

- The number of failures in a large computer system during a day.
- The number of replacement orders for a part received by a firm in a month.
- The number of ships arriving at a loading facility during a 6-hour loading period.
- The number of delivery trucks to arrive at a central warehouse in an hour.

Poisson Distribution

Definition

Let X be the number of occurrences in a given continuous interval (such as time, surface area, or length). Then the probability distribution of X is called the Poisson distribution.

$$P(k) = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \text{for} \quad k = 0, 1, 2, \ldots$$

where

- $\lambda =$ the expected number of occurrences per time or space unit.
- $e = 2,71828$ (the base for natural logarithms)

Moreover,

$$\mu_X = \lambda \quad \text{and} \quad \sigma_X^2 = \lambda$$

Example

Andrew Whittaker, computer center manager, reports that his computer system experienced three component failures during the past 100 days.

a. What is the probability of no failures in a given day?

b. What is the probability of one or more component failures in a given day?

c. What is the probability of at least two failures in a 3-day period?

Exercise

Ex 1: Customers arrive at a busy checkout counter at an average rate of 3 per minute. If the distribution of arrivals is Poisson, find the probability that in any given minute there will be 2 or fewer arrivals.

Ex 2: The number of accidents in a production facility has a Poisson distribution with a mean of 31.2 per month.
a. For a given month what is the probability there will be fewer than 2 accidents?
b. For a given month what is the probability there will be more than 3 accidents?

Hypergeometric Distribution

SELF-STUDY