Lecture 2:
Random Variable

When the value of a variable is determined by a chance event, that variable is called a **random variable**.

Source: https://stattrek.com

Outline

- Definition of random variable
- Discrete vs. continuous random variables
- Discrete vs. continuous probability distributions
- Discrete random variable: discrete probability distribution, cumulative probability, mean, variance
- Continuous random variable: continuous probability distribution, probability density function (pdf)
- Examples for discrete probability distributions: e.g., Binomial probability distribution, Poisson probability distribution

Discrete vs. Continuous Random Variables

- **Discrete**. Within a range of numbers, discrete variables can take on only certain values. Suppose, for example, that we flip a coin and count the number of heads, which is a discrete random variable.
- **Continuous**. Continuous variables, in contrast, can take on any value within a range of values. For example, suppose we randomly select an individual from a population. We measure the age of that person. Then, age is a continuous random variable.

Source: https://stattrek.com

Probability Distributions: Discrete vs. Continuous

- All probability distributions can be classified as discrete probability distributions or as continuous probability distributions, depending on whether they define probabilities associated with discrete variables or continuous variables.
- If a variable can take on any value between two specified values, it is called a **continuous variable**; otherwise, it is called a **discrete variable**.
- Some examples will clarify the difference between discrete and continuous variables.
- Suppose the fire department mandates that all fire fighters must weigh between 150 and 250 pounds. The weight of a fire fighter would be an example of a continuous variable; since a fire fighter's weight could take on any value between 150 and 250 pounds.
- Suppose we flip a coin and count the number of heads. The number of heads could be any integer value between 0 and plus infinity. However, it could not be any number between 0 and plus infinity. We could not, for example, get 2.5 heads. Therefore, the number of heads must be a discrete variable.
- Just like variables, probability distributions can be classified as discrete or continuous.

Source: https://stattrek.com

Discrete Variables: Finite vs. Infinite

Some references state that continuous variables can take on an infinite number of values, but discrete variables cannot. This is incorrect.

- In some cases, discrete variables can take on only a finite number of values. For example, the number of aces dealt in a poker hand can take on only five values: 0, 1, 2, 3, or 4.
- In other cases, however, discrete variables can take on an infinite number of values. For example, the number of coin flips that result in heads could be infinitely large.

When comparing discrete and continuous variables, it is more correct to say that continuous variables can always take on an infinite number of values; whereas some discrete variables can take on an infinite number of values, but others cannot.

Source: https://stattrek.com

Test Your Understanding (Polling 4)

Problem 1

Which of the following is a discrete random variable?

I. The average height of a randomly selected group of boys.
II. The number of the first-prize winners in a Hong Kong Mark Six lottery.
III. The number of US presidential elections in the 20th century.

(A) I only
(B) II only
(C) III only
(D) I and II
(E) II and III

Source: https://stattrek.com

Test Your Understanding (Polling 4)

The correct answer is B.

The number of the first-prize winners in a Hong Kong Mark Six lottery results from a random process, but it can only be a whole number - not a fraction; so it is a discrete random variable. The average height of a randomly-selected group of boys could take on any value between the height of the smallest and tallest boys, so it is not a discrete variable. And the number of US presidential elections in the 20th century can take on only certain values but does not result from a random process; so it is not a random variable.

Source: https://stattrek.com

Probability Distribution

- In probability theory and statistics, a probability distribution is the mathematical function that gives the probabilities of occurrence of different possible outcomes for an experiment.
- The probability distribution of a discrete random variable can always be represented by a table. For example, suppose you flip a coin two times. This simple exercise can have four possible outcomes: HH, HT, TH, and TT. Now, let the variable X represent the number of heads that result from the coin flips. The variable X can take on the values 0, 1, or 2; and X is a discrete random variable.
- The table in the next slide shows the probabilities associated with each possible value of X. The probability of getting 0 heads is 0.25; 1 head, 0.50; and 2 heads, 0.25. Thus, the table is an example of a probability distribution for a discrete random variable.

Source: https://stattrek.com

Probability Distribution (continued)

Number of heads, x	Probability, P(x)
0	0.25
1	0.50
2	0.25

- **Note:** Given a probability distribution, you can find cumulative probabilities. For example, the probability of getting 1 or fewer heads [$P(X \leq 1)$] is $P(X = 0) + P(X = 1)$, which is equal to 0.25 + 0.50 or 0.75.

Source: https://stattrek.com

Discrete Probability Distributions (1)

- If a random variable is a discrete variable, its probability distribution is called a **discrete probability distribution**.
- An example will make this clear. Suppose you flip a coin two times. This simple statistical experiment can have four possible outcomes: HH, HT, TH, and TT. Now, let the random variable *X* represent the number of Heads that result from this experiment. The random variable *X* can only take on the values 0, 1, or 2, so it is a discrete random variable.
- The probability distribution for this statistical experiment appears below.

Number of heads, x	Probability, P(x)
0	0.25
1	0.50
2	0.25

Source: https://stattrek.com

Mean and Variance of Random Variables

Just like variables from a data set, random variables are described by measures of central tendency (like the mean) and measures of variability (like variance). This lesson shows how to compute these measures for discrete random variables.

Source: https://stattrek.com

Mean of a Discrete Random Variable

The mean of the discrete random variable X is also called the **expected value** of X. Notationally, the expected value of X is denoted by $E(X)$. Use the following formula to compute the mean of a discrete random variable.

- $E(X) = \mu_x = \Sigma [x_i * P(x_i)]$

where x_i is the value of the random variable for outcome i, μ_x is the mean of random variable X, and $P(x_i)$ is the probability that the random variable will be outcome i.

Source: https://stattrek.com

Variance of a Discrete Random Variable

The equation for computing the variance of a discrete random variable is shown below.

$$\sigma^2 = \Sigma \{ [x_i - E(x)]^2 * P(x_i) \}$$

where x_i is the value of the random variable for outcome i, $P(x_i)$ is the probability that the random variable will be outcome i, $E(x)$ is the expected value of the discrete random variable x.

Source: https://stattrek.com

Standard Deviation of a Discrete Random Variable

The equation for computing the standard deviation of a discrete random variable is shown below.

$$\text{sqrt } [\sigma^2] = \sigma = \text{sqrt } [\Sigma \{ [x_i - E(x)]^2 * P(x_i) \}]$$

where x_i is the value of the random variable for outcome i, $P(x_i)$ is the probability that the random variable will be outcome i, $E(x)$ is the expected value of the discrete random variable x.

Source: https://stattrek.com

Test Your Understanding (Pollings 5 & 6)

The number of adults, x (1,2,3 or 4), living in homes on a randomly selected city block is described by the following probability distribution.

Number of adults, x	1	2	3	4
Probability, P(x)	0.25	0.50	0.15	0.10

Source: https://stattrek.com

Test Your Understanding (Polling 5)

What is the mean of the probability distribution?

(A) 0.50
(B) 0.62
(C) 0.79
(D) 0.89
(E) 2.10

Source: https://stattrek.com

Test Your Understanding (Polling 5)

Solution

The correct answer is E. Computations are shown below, beginning with the mean (expected value).

$$E(X) = \Sigma [x_i * P(x_i)]$$
$$= 1*0.25 + 2*0.50 + 3*0.15 + 4*0.10 = 2.10$$

Source: https://stattrek.com

Test Your Understanding (Polling 6)

What is the standard deviation of the probability distribution?

(A) 0.50
(B) 0.62
(C) 0.79
(D) 0.89
(E) 2.10

Source: https://stattrek.com

Test Your Understanding (Polling 6)

Solution

The correct answer is D. Now that we know the expected value (i.e., E(x)=2.10), we find the variance.

$\sigma^2 = \Sigma \{ [x_i - E(x)]^2 * P(x_i) \}$
$= (1 - 2.1)^2 * 0.25 + (2 - 2.1)^2 * 0.50 + (3 - 2.1)^2 * 0.15 + (4 - 2.1)^2 * 0.10$
$= (1.21 * 0.25) + (0.01 * 0.50) + (0.81) * 0.15) + (3.61 * 0.10)$
$= 0.3025 + 0.0050 + 0.1215 + 0.3610 = 0.79$

And finally, the standard deviation is equal to the square root of the variance; so the standard deviation is sqrt(0.79) or 0.889.

Source: https://stattrek.com

Test Your Understanding (Polling 7)

Problem 1

The number of adults, x (1,2,3,4,5…), living in homes on a randomly selected city block is described by the following probability distributi

Number of adults, x	Probability, P(x)
1	0.25
2	0.50
3	0.15
4 or more	???

Source: https://stattrek.com

Test Your Understanding (Polling 7)

What is the probability that 4 or more adults reside at a randomly selected home?

(A) 0.10
(B) 0.15
(C) 0.25
(D) 0.50
(E) 0.90

Source: https://stattrek.com

Test Your Understanding (Polling 7)

Solution

The correct answer is A. The sum of all the probabilities is equal to 1. Therefore, the probability that four or more adults reside in a home is equal to 1 - (0.25 + 0.50 + 0.15) or 0.10.

Source: https://stattrek.com

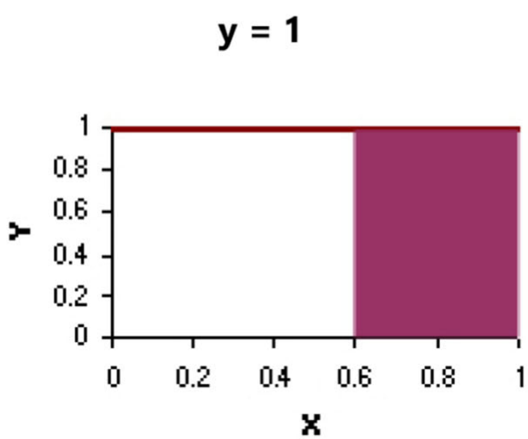
Continuous Probability Distributions (1)

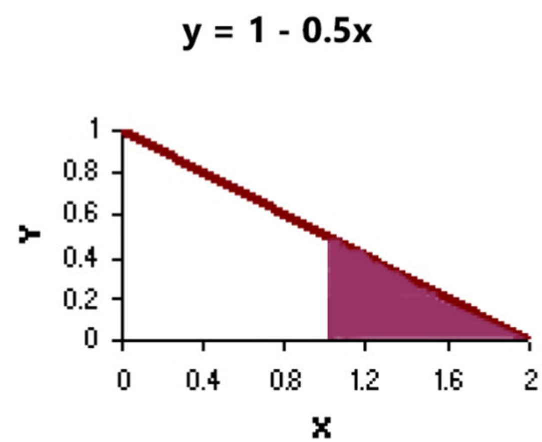
The probability distribution of a continuous random variable is represented by an equation, called the **probability density function** (pdf). All probability density functions satisfy the following conditions:

- The random variable Y is a function of X; that is, $y = f(x)$.
- The value of y is greater than or equal to zero for all values of x.
- The total area under the curve of the function is equal to one.

Source: https://stattrek.com

Continuous Probability Distributions (2)





Note that the area under the curve is equal to 1 for both charts.

Source: https://stattrek.com

Continuous Probability Distributions (3)

- The probability that a continuous random variable falls in the interval between *a* and *b* is equal to the area under the pdf curve between *a* and *b*. For example, in the left chart above, the shaded area shows the probability that the random variable *X* will fall between 0.6 and 1.0. That probability is 0.40. And in the second chart, the shaded area shows the probability of falling between 1.0 and 2.0. That probability is 0.25.
- **Note:** With a continuous distribution, there are an infinite number of values between any two data points. As a result, the probability that a continuous random variable will assume a particular value is always zero. For example, in both of the above charts, the probability that variable *X* will equal *exactly* 0.4 is zero.

Source: https://stattrek.com

Examples for Discrete Probability Distributions

Here are some common discrete probability distributions.

- Binomial probability distribution
- Poisson probability distribution
- Hypergeometric probability distribution
- Multinomial probability distribution
- Negative binomial distribution

Note: With a discrete probability distribution, each possible value of the discrete random variable can be associated with a non-zero probability. Thus, a discrete probability distribution can always be presented in tabular form.

Source: https://stattrek.com

Binomial Probability Distribution: Binomial Experiment

A **binomial experiment** is a statistical experiment that has the following properties:

- The experiment consists of *n* repeated trials.
- Each trial can result in just two possible outcomes. We call one of these outcomes a success and the other, a failure.
- The probability of success, denoted by *P*, is the same on every trial.
- The trials are independent; that is, the outcome on one trial does not affect the outcome on other trials.

Source: https://stattrek.com

Binomial Probability Distribution: Example for Binomial Experiment

Consider the following statistical experiment. You flip a coin 2 times and count the number of times the coin lands on heads. This is a binomial experiment because:

- The experiment consists of repeated trials. We flip a coin 2 times.
- Each trial can result in just two possible outcomes - heads or tails.
- The probability of success is 0.5 - same on every trial.
- The trials are independent; that is, getting heads on one trial does not affect whether we get heads on other trials.

Source: https://stattrek.com

Binomial Probability Distribution: Notation

- x: The number of successes that result from the binomial experiment.
- n: The number of trials in the binomial experiment.
- P: The probability of success on an individual trial.
- Q: The probability of failure on an individual trial. (This is equal to 1 - P.)
- $n!$: The factorial of n (also known as n factorial) where $n! = 1 \times 2 \times 3 \times \ldots \times n$.
- b(x; n, P): Binomial probability - the probability that an n-trial binomial experiment results in exactly x successes, when the probability of success on an individual trial is P.

Source: https://stattrek.com

Binomial Probability Distribution

Given x, n, and P, we can compute the binomial probability:

$$b(x; n, P) = \{ n! / [x! (n - x)!] \} * P^x * (1 - P)^{n-x}$$

The binomial distribution has the following properties:

- The mean of the distribution (μ_x) is equal to $n * P$.
- The variance (σ^2_x) is $n * P * (1 - P)$.
- The standard deviation (σ_x) is sqrt[$n * P * (1 - P)$].

Source: https://stattrek.com

Example for Binomial Probability Distribution

- Recall that a **binomial random variable** is the number of successes *x* in *n* repeated trials of a binomial experiment. The probability distribution of a binomial random variable is called a **binomial probability distribution**.
- Suppose we flip a coin two times and count the number of heads (successes). The binomial random variable is the number of heads, which can take on values of 0, 1, or 2. The binomial distribution with P(x) = 0.5 is presented below.

Number of heads, x	Probability, P(x)
0	0.25
1	0.50
2	0.25

Source: https://stattrek.com

Example for Binomial Probability Distribution

Given x, n, and P, we can compute the binomial probability:

$$b(x; n, P) = \{ n! / [x! (n - x)!] \} * P^x * (1 - P)^{n-x}$$

Now $n = 2$, $P = 0.5$

$$b(0; 2, 0.5) = \{ 2! / [0! (2 - 0)!] \} * 0.5^0 * (1 - 0.5)^{2-0} = 0.25$$

$$b(1; 2, 0.5) = \{ 2! / [1! (2 - 1)!] \} * 0.5^1 * (1 - 0.5)^{2-1} = 0.5$$

$$b(2; 2, 0.5) = \{ 2! / [2! (2 - 2)!] \} * 0.5^2 * (1 - 0.5)^{2-2} = 0.25$$

Test Your Understanding (Polling 8)

Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours?

(A) 0.124
(B) 0.161
(C) 0.312
(D) 0.537
(E) 0.902

Source: https://stattrek.com

Test Your Understanding (Polling 8)

Solution

The correct answer is B. This is a binomial experiment in which the number of trials is equal to 5, the number of successes is equal to 2, and the probability of success on a single trial is 1/6 or about 0.167. Therefore, the binomial probability is:

$$b(2; 5, 0.167) = \{ 5! / [2! (5 - 2)!] \} * (0.167)^2 * (0.833)^3 = 0.161$$

Source: https://stattrek.com

More examples on Binomial Probability Distribution: Cumulative Binomial Probability

- A **cumulative binomial probability** refers to the probability that the binomial random variable falls within a specified range (e.g., is greater than or equal to a stated lower limit and less than or equal to a stated upper limit).
- For example, we might be interested in the cumulative binomial probability of obtaining 45 or fewer heads in 100 tosses of a coin (see Example 1 below).

Source: https://stattrek.com

More examples on Binomial Probability Distribution: Cumulative Binomial Probability

Example 1

What is the probability of obtaining 45 or fewer heads in 100 tosses of a coin?

Solution: To solve this problem, we compute 46 individual probabilities, using the binomial formula. The sum of all these probabilities is the answer we seek. Thus,

$$b(x \leq 45; 100, 0.5) = b(x = 0; 100, 0.5) + b(x = 1; 100, 0.5) + \ldots + b(x = 45; 100, 0.5) = 0.184$$

Source: https://stattrek.com

More examples on Binomial Probability Distribution: Cumulative Binomial Probability

Example 2

The probability that a student is accepted to a prestigious college is 0.3. If 5 students from the same school apply, what is the probability that at most 2 are accepted?

Solution: To solve this problem, we compute 3 individual probabilities, using the binomial formula. The sum of all these probabilities is the answer we seek. Thus,

$$\begin{aligned} b(x \leq 2; 5, 0.3) &= b(x = 0; 5, 0.3) + b(x = 1; 5, 0.3) + b(x = 2; 5, 0.3) \\ &= 0.1681 + 0.3601 + 0.3087 = 0.8369 \end{aligned}$$

Source: https://stattrek.com

Poisson Distribution: Poisson experiment (1)

A **Poisson experiment** is a statistical experiment that has the following properties:

- The experiment results in outcomes that can be classified as successes or failures.
- The average number of successes (μ) that occurs in a specified region is known.
- The probability that a success will occur is proportional to the size of the region.
- The probability that a success will occur in an extremely small region is virtually zero.

Source: https://stattrek.com

Poisson Distribution: Common examples of Poisson processes (2)

- customers calling a help center
- visitors to a website
- radioactive decay in atoms
- photons arriving at a space telescope
- movements in a stock price

Source: https://stattrek.com

Poisson Distribution: Notation (3)

- The following notation is helpful, when we talk about the Poisson distribution.
- *e*: A constant equal to approximately 2.71828. (Actually, *e* is the base of the natural logarithm system.)
- μ: The mean number of successes that occur in a specified region.
- *x*: The actual number of successes that occur in a specified region.
- P(*x*; μ): The **Poisson probability** that <u>exactly</u> *x* successes occur in a Poisson experiment, when the mean number of successes is μ.

Source: https://stattrek.com

Poisson Distribution (4)

- A **Poisson random variable** is the number of successes that result from a Poisson experiment. The probability distribution of a Poisson random variable is called a **Poisson distribution**.
- Given the mean number of successes (μ) that occur in a specified region, we can compute the Poisson probability:

$$P(x; \mu) = (e^{-\mu}) (\mu^x) / x!$$

where x is the actual number of successes that result from the experiment.

- The Poisson distribution has the following properties:
  - The mean of the distribution is equal to μ .
  - The variance is also equal to μ .

Source: https://stattrek.com

Poisson Distribution: Poisson Distribution Example (5)

The average number of homes sold by the Centaline Property in Hong Kong is 2 homes per day. What is the probability that exactly 3 homes will be sold tomorrow?

Solution: This is a Poisson experiment in which we know the following:

- $\mu = 2$; since 2 homes are sold per day, on average.
- $x = 3$; since we want to find the likelihood that 3 homes will be sold tomorrow.
- e = 2.71828; since *e* is a constant equal to approximately 2.71828.

Source: https://stattrek.com

Poisson Distribution: Poisson Distribution Example (6)

- We plug these values into the Poisson formula as follows:

$$P(x; \mu) = (e^{-\mu}) (\mu^{x}) / x!$$

$$P(3; 2) = (2.71828^{-2}) (2^{3}) / 3!$$

$$= (0.13534) (8) / 6$$

$$= 0.180$$

- Thus, the probability of selling 3 homes tomorrow is 0.180 .

Source: https://stattrek.com

Poisson Distribution: Cumulative Poisson Probability (7)

- A **cumulative Poisson probability** refers to the probability that the Poisson random variable is greater than some specified lower limit and less than some specified upper limit.

Source: https://stattrek.com

Poisson Distribution: Cumulative Poisson Example (8)

Suppose the average number of lions seen daily on YouTube is 5. What is the probability that tourists will see fewer than four lions one day on YouTube?

Solution: This is a Poisson experiment in which we know the following:

- μ = 5; since 5 lions are seen per day, on average.
- x = 0, 1, 2, or 3; since we want to find the likelihood that tourists will see fewer than 4 lions; that is, we want the probability that they will see 0, 1, 2, or 3 lions.
- e = 2.71828; since *e* is a constant equal to approximately 2.71828.

Source: https://stattrek.com

Poisson Distribution: Cumulative Poisson Example (9)

To solve this problem, we need to find the probability that tourists will see 0, 1, 2, or 3 lions. Thus, we need to calculate the sum of four probabilities: P(0; 5) + P(1; 5) + P(2; 5) + P(3; 5). To compute this sum, we use the Poisson formula:

$$P(x \leq 3, 5) = P(0; 5) + P(1; 5) + P(2; 5) + P(3; 5)$$

$$P(x \leq 3, 5) = [\ (e^{-5})(5^0) / 0!\ ] + [\ (e^{-5})(5^1) / 1!\ ] + [\ (e^{-5})(5^2) / 2!\ ] + [\ (e^{-5})(5^3) / 3!\ ]$$

$$= [\ (0.006738)(1) / 1\ ] + [\ (0.006738)(5) / 1\ ] + [\ (0.006738)(25) / 2\ ] + [\ (0.006738)(125) / 6\ ]$$

$$= [\ 0.0067\ ] + [\ 0.03369\ ] + [\ 0.084224\ ] + [\ 0.140375\ ]$$

$$= 0.2650$$

Thus, the probability of seeing at no more than 3 lions is 0.2650.

Source: https://stattrek.com