Lecture 3: Continuous Probability Distributions

Outline

- Normal probability distribution (aka Gaussian distribution): the empirical (or 68-95-99.7) rule
- Standard normal distribution: standard score or z-score
- Student's t distribution: t statistic or t score, degrees of freedom, α value, cumulative probability, T Distribution Calculator

Continuous Probability Distributions (1)

- If a random variable is a continuous variable, its probability distribution is called a **continuous probability distribution**.
- A continuous probability distribution differs from a discrete probability distribution in several ways.
  - The probability that a continuous random variable is exactly equal a specific value would be zero.
  - As a result, a continuous probability distribution cannot be expressed in tabular form.
  - Instead, an equation or formula is used to describe a continuous probability distribution.

Source: https://stattrek.com

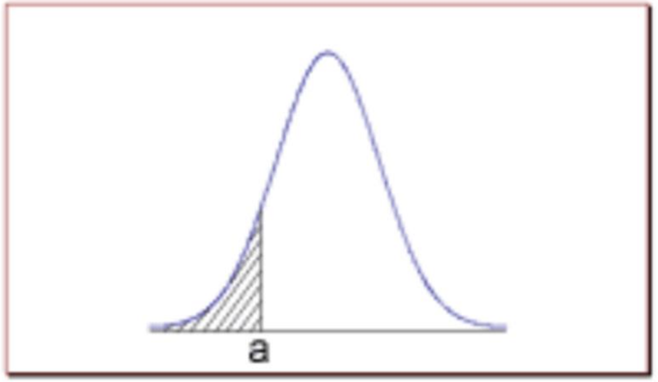
Continuous Probability Distributions (2)

- Most often, the equation used to describe a continuous probability distribution is called a **probability density function**. Sometimes, it is referred to as a **density function**, a **PDF**, or a **pdf**. For a continuous probability distribution, the density function has the following properties:
  - Since the continuous random variable is defined over a continuous range of values (called the **domain** of the variable), the graph of the density function will also be continuous over that range.
  - The area bounded by the curve of the density function and the x-axis is equal to 1, when computed over the domain of the variable.
  - The probability that a random variable assumes a value between *a* and *b* is equal to the area under the density function bounded by *a* and *b*.

Source: https://stattrek.com

Continuous Probability Distributions (3)

- For example, consider the probability density function shown in the graph below. The probability that *X* is less than or equal to *a* is equal to the area under the curve bounded by *a* and minus infinity - as indicated by the shaded area.



Source: https://stattrek.com

Continuous Probability Distributions (4)

Here are some common continuous probability distributions.

- Normal probability distribution (aka Gaussian distribution)
- Standard normal distribution
- Student's t distribution
- Chi-square distribution
- F distribution

Source: https://stattrek.com

Normal Distribution (1)

The normal distribution (as known as Gaussian distribution) is defined by the following equation:

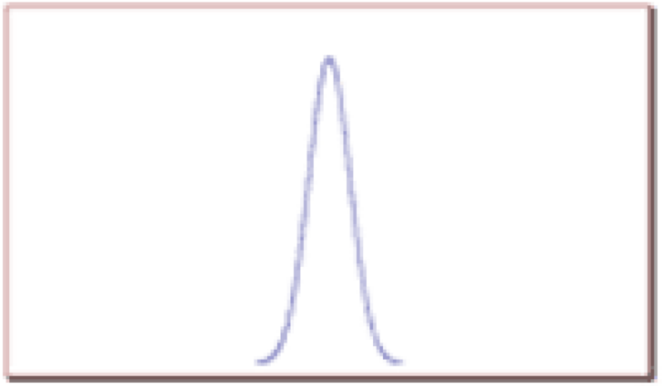
$$Y = \{ 1/[\sigma * \text{sqrt}(2\pi)] \} * e^{-(x - \mu)^2/2\sigma^2}$$

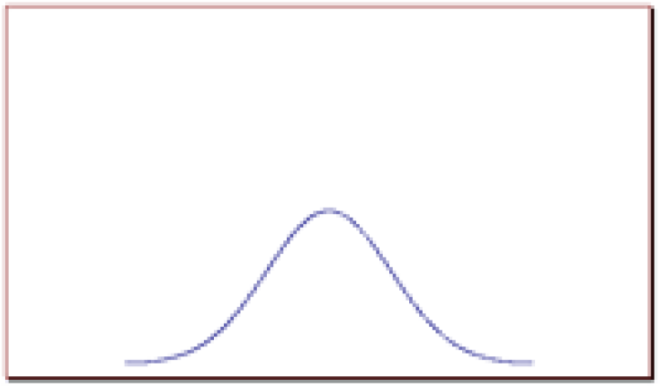
where X is a normal random variable, μ is the mean, σ is the standard deviation, π is approximately 3.14159, and e is approximately 2.71828. The normal equation is the probability density function for the normal distribution.

Source: https://stattrek.com

Normal Distribution: The Normal Curve (2)

The graph of the normal distribution depends on two factors - the mean and the standard deviation. The mean of the distribution determines the location of the center of the graph, and the standard deviation determines the height and width of the graph. All normal distributions look like a symmetric, bell-shaped curve, as shown below.





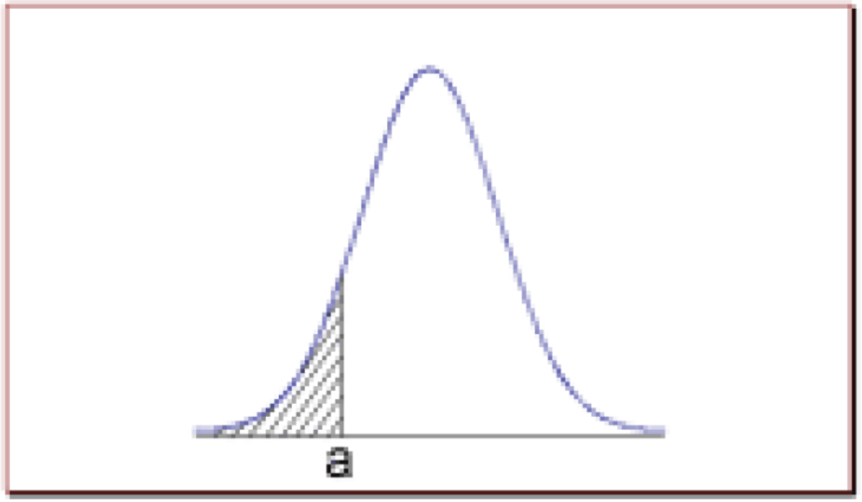
When the standard deviation is small, the curve is tall and narrow (see the left graph above); and when the standard deviation is big, the curve is short and wide (see the right graph above).

Source: https://stattrek.com

Normal Distribution: Probability and the Normal Curve (3)

The normal distribution is a continuous probability distribution. This has several implications for probability.

- The total area under the normal curve is equal to 1.
- The probability that a normal random variable *X* equals any particular value is 0.
- The probability that *X* is greater than *a* equals the area under the normal curve bounded by a and plus infinity (as indicated by the non-shaded area in the figure below).
- On the other hand, the probability that *X* is less than *a* equals the area under the normal curve bounded by *a* and minus infinity (as indicated by the shaded area in the figure below).



Source: https://stattrek.com

Normal Distribution: Probability and the Normal Curve (4)

Additionally, every normal curve (regardless of its mean or standard deviation) conforms to the following "rule".

- About 68% of the area under the curve falls within 1 standard deviation of the mean.
- About 95% of the area under the curve falls within 2 standard deviations of the mean.
- About 99.7% of the area under the curve falls within 3 standard deviations of the mean.

Collectively, these points are known as the **empirical rule** or the **68-95-99.7 rule**. Clearly, given a normal distribution, most outcomes will be within 3 standard deviations of the mean.

Source: https://stattrek.com

Test Your Understanding (Polling 9)

Problem

An average light bulb manufactured by Philips lasts 300 days with a standard deviation of 50 days. Assuming that bulb life is normally distributed, what is the probability that a Philips light bulb will last at most 365 days (i.e., normal random variable = 365)?

(A) 0.80
(B) 0.85
(C) 0.90
(D) 0.95

Hint: Please use the on-line Normal Distribution Calculator (https://homepage.divms.uiowa.edu/~mbognar/applets/normal.html or https://stattrek.com/online-calculator/normal.aspx)

Source: https://stattrek.com

Test Your Understanding (Polling 9)

Solution*:* The correct answer is **C**. Given a mean score of 300 days and a standard deviation of 50 days, we want to find the cumulative probability that bulb life is less than or equal to 365 days. Thus, we know the following:

- The value of the normal random variable is 365 days.
- The mean is equal to 300 days.
- The standard deviation is equal to 50 days.

We enter these values into the Normal Distribution Calculator and compute the cumulative probability. The answer is: $P(X \leq 365) = 0.903$. Hence, there is a 90% chance that a light bulb will burn out within 365 days.

Source: https://stattrek.com

Test Your Understanding (Polling 10)

Problem

Suppose scores on an IQ test are normally distributed. If the test has a mean of 100 and a standard deviation of 10, what is the probability that a person who takes the test will score between 90 and 110?

(A) 0.58
(B) 0.68
(C) 0.78
(D) 0.88

Hint: Please use the on-line Normal Distribution Calculator (https://homepage.divms.uiowa.edu/~mbognar/applets/normal.html or https://stattrek.com/online-calculator/normal.aspx)

Source: https://stattrek.com

Test Your Understanding (Polling 10)

***Solution**:* The correct answer is **B**. Here, we want to know the probability that the test score falls between 90 and 110. The "trick" to solving this problem is to realize the following:

$$P(\, 90 < X < 110 \,) = P(\, X < 110 \,) - P(\, X < 90 \,)$$

We use the Normal Distribution Calculator to compute both probabilities on the right side of the above equation.

- To compute P(X < 110), we enter the following inputs into the calculator: The value of the normal random variable is 110, the mean is 100, and the standard deviation is 10. We find that P(X < 110) is 0.84.
- To compute P(X < 90), we enter the following inputs into the calculator: The value of the normal random variable is 90, the mean is 100, and the standard deviation is 10. We find that P(X < 90) is 0.16.

We use these findings to compute our final answer as follows:

$$\begin{aligned} P(\, 90 < X < 110 \,) &= P(\, X < 110 \,) - P(\, X < 90 \,) \\ &= 0.84 - 0.16 \\ &= 0.68 \end{aligned}$$

Thus, about 68% of the test scores will fall between 90 and 110.

Source: https://stattrek.com

Standard Normal Distribution

- The **standard normal distribution** is a special case of the normal distribution. It is the distribution that occurs when a normal random variable has a mean of zero and a standard deviation of one.
- The normal random variable of a standard normal distribution is called a **standard score** or a **z-score**. Every normal random variable *X* can be transformed into a *z* score via the following equation:

$$z = (X - \mu) / \sigma$$

where *X* is a normal random variable, μ is the mean of *X*, and σ is the standard deviation of *X*.

Source: https://stattrek.com

The Normal Distribution as a Model for Real-World Events

Often, phenomena in the real world follow a normal (or near-normal) distribution. This allows researchers to use the normal distribution as a model for assessing probabilities associated with real-world phenomena. Typically, the analysis involves two steps.

- Transform raw data. Usually, the raw data are not in the form of z-scores. They need to be transformed into z-scores, using the transformation equation presented earlier: $z = (X - \mu) / \sigma$.
- Find probability. Once the data have been transformed into z-scores, you can use **standard normal distribution tables**, **online calculators** (e.g., Stat Trek's free normal distribution calculator), or **handheld graphing calculators** to find probabilities associated with the z-scores.

The problem in the next section demonstrates the use of the normal distribution as a tool to model real-world events.

Source: https://stattrek.com

Test Your Understanding (Polling 11)

Problem

Molly earned a score of 940 on a national achievement test. The mean test score was 850 with a standard deviation of 100. What proportion of students had a higher score than Molly? (Assume that test scores are normally distributed.) Please use the z-score transformation equation.

(A) 0.10
(B) 0.18
(C) 0.50
(D) 0.82
(E) 0.90

Source: https://stattrek.com

Test Your Understanding (Polling 11)

Solution

The correct answer is **B**. As part of the solution to this problem, we assume that test scores are normally distributed. In this way, we use the normal distribution to model the distribution of test scores in the real world. Given an assumption of normality, the solution involves three steps.

- First, we transform Molly's test score into a z-score, using the z-score transformation equation. $z = (X - \mu) / \sigma = (940 - 850) / 100 = 0.90$
- Then, using **an online calculator** (e.g., Stat Trek's free normal distribution calculator), a **handheld graphing calculator**, or the **standard normal distribution table**, we find the cumulative probability associated with the z-score. In this case, we find $P(Z < 0.90) = 0.8159$.
- Therefore, the $P(Z > 0.90) = 1 - P(Z < 0.90) = 1 - 0.8159 = 0.1841$.

Thus, we estimate that 18.41 percent of the students tested had a higher score than Molly.

Source: https://stattrek.com

Student's t Distribution (1)

The **Student's t distribution** (aka, **t-distribution**) is a probability distribution that is used to estimate population parameters when the sample size is small (say less than or equal to 30) and/or when the population variance is unknown.

Why Use the t Distribution?

According to the central limit theorem, the sampling distribution of a statistic (like a sample mean) will follow a normal distribution, as long as the sample size is sufficiently large. Therefore, when we know the standard deviation of the population, we can compute a z-score, and use the normal distribution to evaluate probabilities with the sample mean.

Source: https://stattrek.com

Student's t Distribution (2)

- But sample sizes are sometimes small, and often we do not know the standard deviation of the population. When either of these problems occur, statisticians rely on the distribution of the **t statistic** (also known as the **t score**), whose values are given by:

  $$t = [\ \bar{x} - \mu\ ]\ /\ [\ s\ /\ \mathrm{sqrt}(\ n\ )\ ]$$

  where $\bar{x}$ is the sample mean, μ is the population mean, *s* is the standard deviation of the sample, and *n* is the sample size. The distribution of the *t* statistic is called the **t distribution** or the **Student's t distribution**.

- The t distribution allows us to conduct statistical analyses on certain data sets that are not appropriate for analysis, using the normal distribution.

Source: https://stattrek.com

Student's t Distribution: Degrees of Freedom (4)

- There are actually many different t distributions. The particular form of the t distribution is determined by its **degrees of freedom**. The degrees of freedom refers to the number of independent observations in a set of data.
- When estimating a mean score from a single sample, the number of independent observations is equal to the sample size minus one. Hence, the distribution of the *t* statistic from samples of size 8 would be described by a t distribution having 8 - 1 or 7 degrees of freedom. Similarly, a t distribution having 15 degrees of freedom would be used with a sample of size 16.

Source: https://stattrek.com

Probability and the Student t Distribution (5)

- When a sample of size *n* is drawn from a population having a normal (or nearly normal) distribution, the sample mean can be transformed into a t statistic, using the equation presented previously. We repeat that equation below:

  t = [$\bar{x}$ - μ] / [*s* / sqrt(*n*)]

  where $\bar{x}$ is the sample mean, μ is the population mean, *s* is the standard deviation of the sample, *n* is the sample size, and degrees of freedom are equal to *n* - 1.

- The t statistic produced by this transformation can be associated with a unique cumulative probability. This cumulative probability represents the likelihood of finding a sample mean less than or equal to $\bar{x}$, given a random sample of size *n*.

- The easiest way to find the probability associated with a particular t statistic is to use the T Distribution Calculator (https://stattrek.com/online-calculator/t-distribution.aspx), a free tool provided by Stat Trek.

Source: https://stattrek.com

Notation and t Statistics (6)

- Statisticians use t_α to represent the t statistic that has a cumulative probability of (1 - α). For example, suppose we were interested in the t statistic having a cumulative probability of 0.95. In this example, α would be equal to (1 - 0.95) or 0.05. We would refer to the t statistic as $t_{0.05}$.
- Of course, the value of $t_{0.05}$ depends on the number of degrees of freedom. For example, with 2 degrees of freedom, $t_{0.05}$ is equal to 2.92; but with 20 degrees of freedom, $t_{0.05}$ is equal to 1.725.
- Note: Because the t distribution is symmetric about a mean of zero, the following is true.

$$t_\alpha = -t_{1-\alpha} \quad \text{and} \quad t_{1-\alpha} = -t_\alpha$$

Thus, if $t_{0.05} = 2.92$, then $t_{0.95} = -2.92$.

- Also note: We will talk more about t statistic for the topic of confidence interval.

Source: https://stattrek.com

Test Your Understanding (Polling 12)

Problem

Philips manufactures light bulbs. The CEO claims that an average Acme light bulb lasts 300 days. A researcher randomly selects 15 bulbs for testing. The sampled bulbs last an average of 290 days, with a standard deviation of 50 days. If the CEO's claim were true, what is the probability that 15 randomly selected bulbs would have an average life of no more than 290 days?

(A) 0.13
(B) 0.23
(C) 0.33
(D) 0.43

Hint: Please use the on-line T Distribution Calculator (https://stattrek.com/online-calculator/t-distribution.aspx)

Source: https://stattrek.com

Test Your Understanding (Polling 12)

Solution: The correct answer is **B**. The first thing we need to do is compute the t statistic, based on the following equation:

$$t = [\ \bar{x} - \mu\ ] / [\ s / \text{sqrt}(\ n\ )\ ]$$
$$= (\ 290 - 300\ ) / [\ 50 / \text{sqrt}(\ 15)\ ]$$
$$= -10 / 12.909945 = -0.7745966$$

where $\bar{x}$ is the sample mean, μ is the population mean, s is the standard deviation of the sample, and n is the sample size.

Now, we are ready to use the T Distribution Calculator. Since we know the t statistic, we select "T score" from the Random Variable dropdown box. Then, we enter the following data:

- The degrees of freedom are equal to 15 - 1 = 14.
- The t statistic is equal to - 0.7745966.

The calculator displays the cumulative probability: 0.226. Hence, if the true bulb life were 300 days, there is a 22.6% chance that the average bulb life for 15 randomly selected bulbs would be less than or equal to 290 days.

Source: https://stattrek.com

Test Your Understanding (Polling 13)

Problem

Suppose scores on an IQ test are normally distributed, with a population mean of 100. Suppose 20 people are randomly selected and tested. The standard deviation in the sample group is 15. What is the probability that the average test score in the sample group will be at most 110?

(A) 0.696

(B) 0.796

(C) 0.896

(D) 0.996

Hint: Please use the on-line T Distribution Calculator (https://stattrek.com/online-calculator/t-distribution.aspx)

Source: https://stattrek.com

Test Your Understanding (Polling 13)

Solution:

The correct answer is **D**. To solve this problem, we will work directly with the raw data from the problem. We will not compute the t statistic; the T Distribution Calculator will do that work for us. Since we will work with the raw data, we select "Sample mean" from the Random Variable dropdown box. Then, we enter the following data:

- The degrees of freedom are equal to 20 - 1 = 19.
- The population mean equals 100.
- The sample mean equals 110.
- The standard deviation of the sample is 15.

We enter these values into the T Distribution Calculator. The calculator displays the cumulative probability: 0.996. Hence, there is a 99.6% chance that the sample average will be no greater than 110.

Source: https://stattrek.com

EE1004 Teaching and Learning Survey (23 Mar 2021)

https://cityuhk.questionpro.com/t/AR91DZlmAr