FUNDAMENTALS OF STATISTICS (4ECON013C)

TUTORIAL 4

Continuous Probability Distributions

Learning outcomes:

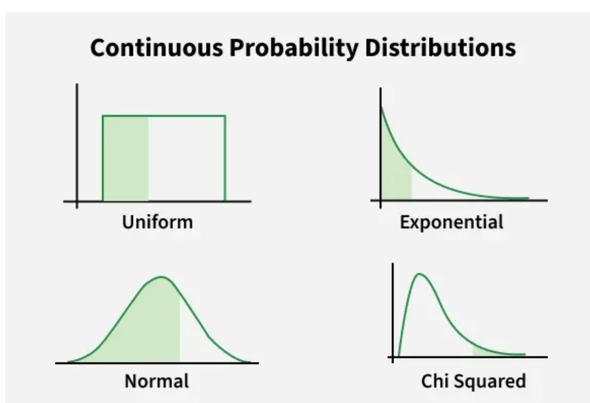
- understand the concept of a continuous random variable (r. v.).
- be able to find expected value and variance of continuous probability distributions.
- learn about probability density function of Uniform and Normal (Gaussian) distribution.
- understand Standard Normal (z) distribution.

Continuous random variable

- A continuous random variable can assume any value in an interval on the real line or in a collection of intervals.
- It is not possible to talk about the probability of the random variable assuming a particular value.
- Instead, we talk about the probability of the random variable assuming a value within a given interval.
- If f(x) is a continuous probability distribution where x ∈ $(-\infty,\ \infty)$, then $\int_{-\infty}^{\infty} f(x)dx = 1$.

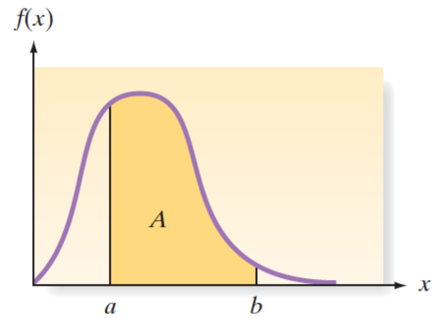
A **continuous probability distribution** describes the probability for a continuous random variable, which can take any value within a given range, such as a person's height or the temperature of a room. Because there are infinitely many possible values, the probability of the variable being exactly one specific value is zero.

Common Types of Continuous Probability Distributions:



Let X be a continuous r.v., then a probability distribution or **probability density function (pdf)** of X is a function f(x) such that for any two numbers a and b with $a \leq b$, we have

$$P(a \leq X \leq b) = \int_a^b f(x)dx = A$$



The probability that X is in the interval [a, b] can be calculated by integrating the pdf of the r.v. X.

Example 4.1. The following function is given:

$$f(x) = \begin{cases} 2x, & for\ \ 0 \leq x \leq 1 \\ 0, & otherwise \end{cases}$$

a. Prove that it is pdf.
b. Find the probability that X lies between 0.2 and 0.7.

Solution.

a. To be pdf, $\int_{-\infty}^{\infty} f(x)dx = 1$ should be true.

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^1 2x\,dx = [2\frac{x^2}{2}]_0^1 = [x^2]_0^1 = 1 - 0 = 1, \text{ so it is pdf.}$$

b. $P(0.2 \leq X \leq 0.7) = \int_{0.2}^{0.7} 2x\,dx = [x^2]_{0.2}^{0.7} = 0.7^2 - 0.2^2 = 0.45$

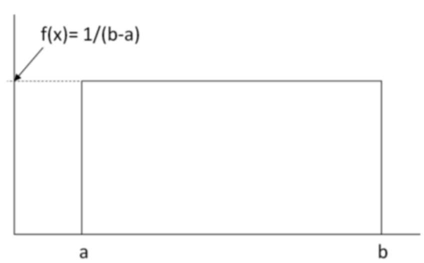
Mean and variance of continuous probability distributions.

$$E(x) = \mu = \int_{-\infty}^{\infty} x * f(x)dx$$

$$Var(x) = \sigma^2 = E(x^2) - [E(x)]^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

Uniform distribution.

A uniform distribution is a continuous random variable in which all values between a minimum value and a maximum value have the same probability.



The two parameters that define the Uniform Distribution are:

a = minimum, b = maximum

Example 4.2. A continuous **uniform** distribution is given as *X ~ U(a, b),* where X is the amount of time (in minutes) of waiting for a bus, a = 0, b = 15. Then the waiting times can be expressed as *X ~ U(0, 15)*.

The pdf is defined as:

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & otherwise \end{cases}$$

a. Find the E(x).
b. Find the σ.

Solution.

a. $E(x) = \int_0^{15} x * \frac{1}{15-0} dx = \frac{1}{15} * \frac{x^2}{2}\Big|_0^{15} = \frac{1}{30}(15^2 - 0^2) = \mathbf{7.5}$

Indeed, the expected value of a uniform distribution is $\mathbf{E(x) = \frac{a+b}{2}}$.

b. $E(x^2) = \int_0^{15} x^2 * \frac{1}{15} dx = \frac{1}{15} * \frac{x^3}{3}\Big|_0^{15} = \frac{1}{45}(15^3 - 0^3) = 75$

$\sigma^2 = E(x^2) - \mu^2 = 75 - 7.5^2 = 18.75$

The general formula of variance for Uniform distribution is $\boldsymbol{\sigma^2} = \frac{(\boldsymbol{b-a})^2}{\mathbf{12}}$

Then, $\sigma = \sqrt{18.75} = \mathbf{4.33}$

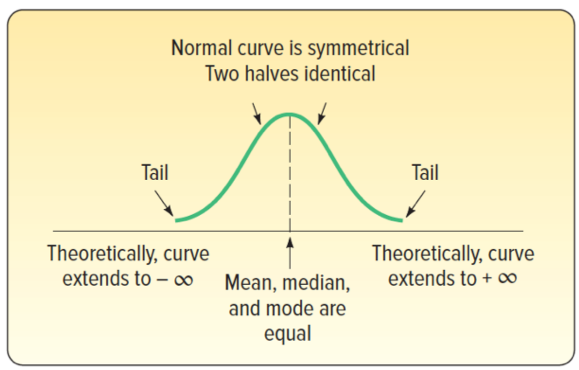
The normal (Gaussian) distribution

The normal distribution is the most important distribution in statistics and it is essential in many statistical methods you use in the future whether in academia or in industry. We can call it the "king" of the probability distributions.

The probability distribution curve and its main properties are given below:

Properties of a Normal Distribution

1. The mean, median, and mode are equal.
2. The normal curve is bell-shaped and symmetric about the mean.
3. The total area under the curve is equal to one.
4. The normal curve approaches, but never touches the *x* axis as it extends farther and farther away from the mean.



If X is a random variable that has normal distribution, then we can use the following notation: $X \sim N(\mu, \sigma^2)$. Here, the mean and variance are the parameters of the Normal distribution. The probability density function (pdf) of Normal distribution takes the following form:

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} * e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}}$$

To calculate the probabilities related to normal distribution, one needs to take integral from the above pdf, but luckily, we do not have to do this because we have a **standard normal distribution** (**z distribution**). Using the standardizing formula as following:

$$\boxed{\mathbf{z} = \frac{x-\mu}{\sigma}}$$

we can use the z table to find the cumulative probabilities.

Example 4.3.

The scores on a verbal reasoning test are normally distributed with a population mean of $\mu = 75$ and a population standard deviation of $\sigma = 14$.

a. What is the probability that a randomly chosen person scores less than 85?
b. What is the probability that a randomly chosen person scores more than 70?
c. What is the probability that a randomly chosen person scores between 70 and 85?
d. What is the 95th percentile score?

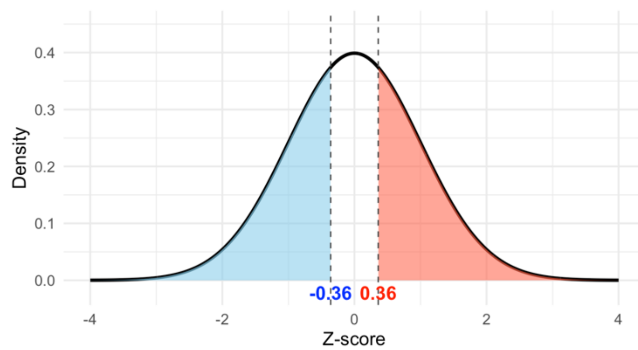
Solution. Let suppose X is the score of a person, then $X \sim N(75, 14^2)$

a. $P(X < 85) = P(z < \frac{85-75}{14}) = P(z < 0.71) = \mathbf{0.7611}$

z	.00	.01	.02
0.6	.7257	.7291	.7324
0.7	.7580	.7611	.7642
0.8	.7881	.7910	.7939

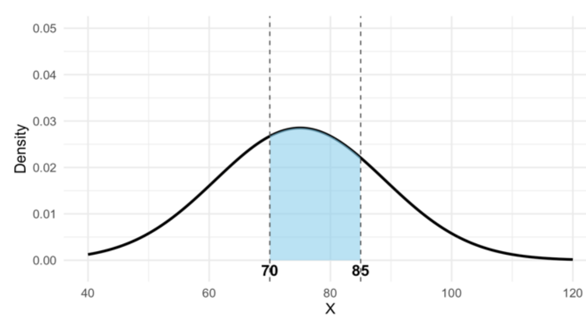
b. $P(X > 70) = P(z > \frac{70-75}{14}) = P(z > -0.36) = P(z < 0.36) =$
$= \mathbf{0.6406}$

Note here that you are given z table of only positive z scores. But using symmetric property of z distribution, you can find the cumulative probabilities for negative z scores easily. Here, the areas shaded in light-blue and red are equal.

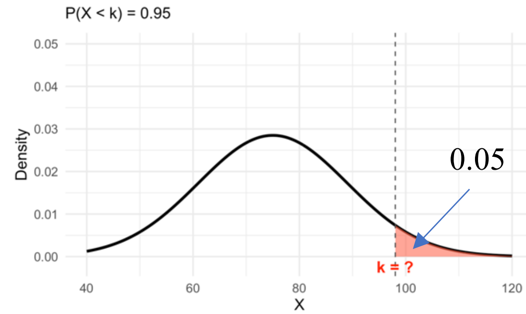


c. $P(70 < X < 85) = P(X < 85) - P(X < 70) = P(z < 0.71) - P(z < -0.36) = 0.7611 - (1 - 0.6406) = \mathbf{0.4017}$

Note that $P(z < -0.36) = P(z > 0.36) = 1 - P(z \leq 0.36)$.



d. P(X < k) = 0.95, we need to find k.



z score for the lower tail cumulative probability of 0.95 is 1.645 (between 1.64 and 1.65).

$1.645 = \frac{k-75}{14}$ => k = **98.03**

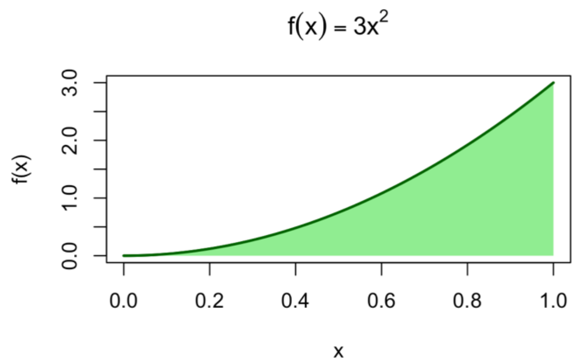
TASKS

1. Decide whether the statement makes sense (or is clearly true) or does not make sense (or is clearly false). Explain clearly.

- **a.** The probability that a continuous random variable X equals exactly 5 is greater than 0.
- **b.** The mean and median of a uniform distribution *U(a,b)* are always the same.
- **c.** Scores on a statistics test are normally distributed with a mean of 75 and a standard deviation of 75.
- **d.** Birth weights (in grams) in Uzbekistan are normally distributed with a mean of 3.32 kg and a standard deviation of 400 g.
- **e.** Scores on a standard test of depth perception are normally distributed with two different modes.
- **f.** SAT scores are normally distributed with a mean of 1518 and a standard deviation of 325.

2. You are given the following function:

$$f(x) = \begin{cases} 3x^2, & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$$



- **a.** Prove that this is pdf.
- **b.** Find the expected value of x.
- **c.** Find the variance of x.
- **d.** Find the probability that x is between 0.2 and 0.8.

3. An unprincipled used-car dealer sells a car to an unsuspecting buyer, even though the dealer knows that the car will have a major breakdown within the next 6 months. The dealer provides a warranty of 45 days on all cars sold. Let x represent the length of time until the breakdown occurs. Assume that x is a uniform random variable with values between 0 and 6 months. Calculate the probability that the breakdown occurs while the car is still under warranty (assume there are 30 days in a month).

4. According to the Insurance Institute of America, a family of four spends between \$400 and \$3,800 per year on all types of insurance. Suppose the money spent is uniformly distributed between these amounts.

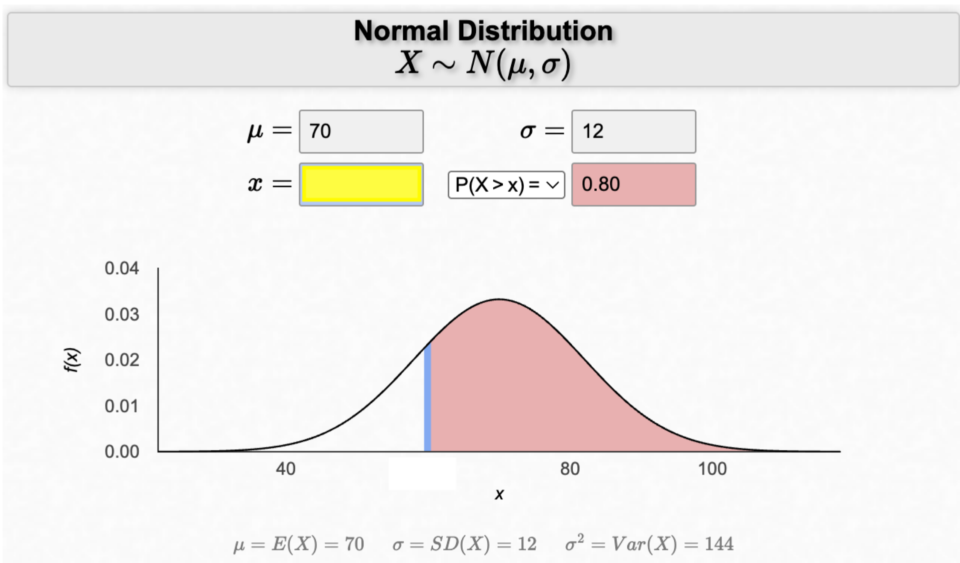
- **a.** If we select a family at random, what is the probability they spend less than \$2,000 per year on insurance?

b. What is the 90th percentile of the family spending on insurance?

c. If you are told that the family spent more than \$1,500, what is the probability that this family spent \$2500 at most?

5. The annual income of full-time employees in a certain town is normally distributed with a mean \$11,000 and standard deviation \$1,500. Given that there are 150,000 people in full-time employment in this town, estimate the number of people whose annual income is between \$8,000 and \$16,000.

6. Find the value of x.



7. The following website provides the average IQ (IIT) scores by country for 2024:

https://worldpopulationreview.com/country-rankings/average-iq-by-country

The standard deviation is approximately equal to 15.

a. If a person in Uzbekistan scores 125 or higher on an IQ test, they are considered *highly intelligent*. Find the percentage of people in Uzbekistan who belong to this category.

b. Suppose a national scholarship program in Uzbekistan accepts only the top 5% of students based on IQ test performance. What is the minimum IQ cutoff for eligibility?

c. If Uzbekistan's average IQ rose by 2 points, how would the percentage of people above 120 change?

d. Suppose X is an IQ score of an individual in a specific country where $X \sim N(\mu, 15^2)$.

If $P(X > 99) = 0.6554$, then which country is this?

8. X is distributed normally, $P(X \geqslant 59.1) = 0.0281$ and $P(X \geqslant 29.2) = 0.9345$. Find the mean and standard deviation of the distribution.

HOMEWORK

9. Suppose 8-week old babies' smiling times, in seconds, follow a uniform distribution between zero and 23 seconds.

a. What is the probability a randomly chosen 8-week-old baby smiles between 2 and 18 seconds?

b. Find the 80th percentile for an eight-week-old baby's smiling time.

c. Find the probability that a random eight-week-old baby smiles more than 12 seconds knowing that the baby smiles more than eight seconds.

10. Ace Heating and Air Conditioning Service finds that the amount of time a repairman needs to fix a furnace is uniformly distributed between 1.5 and four hours. Let x = the time needed to fix a furnace. Find the 30th percentile of furnace repair times.

11. The amount of time, in minutes, that a person must wait for a bus is uniformly distributed between zero and 15 minutes.

a. What is the probability that a person waits less than 12.5 minutes?

b. On average, how long must a person wait? Find the mean, μ, and the standard deviation, σ.

c. Ninety percent of the time, the time a person must wait falls below what value?

12. The number of newspapers sold daily at a kiosk is normally distributed with a mean of 350 and a standard deviation of 30.

a. Find the probability that fewer than 300 newspapers are sold on a particular day.

b. Find the probability that the number of newspapers sold is between 300 and 360 on a particular day.

c. How many newspapers should the newsagent stock each day such that the probability of running out on any particular day is 5%?

13. Assume that the hourly cost to operate a commercial airplane follows the normal distribution with a mean of $2,500 per hour and a standard deviation of $300. What is the operating cost for the lowest 15% of the airplanes?

14. The annual commissions earned by sales representatives of Machine Products Inc., a manufacturer of light machinery, follow the normal probability distribution. The mean yearly amount earned is $40,000 and the standard deviation is $5,000.

a. What percent of the sales representatives earn more than $42,000 per year?

b. What percent of the sales representatives earn less than $34,000 per year?