Stat 230 Introductory Statistics
Continuous Distribution

University of British Columbia Okanagan

Background

- We have already seen **discrete** probability distributions.
  - Bernoulli
  - Binomial
  - Hypergeometric
  - Poisson
- We can now extend the concepts from last lectures to **continuous** random variables.

Background

- A **discrete** random variable maps the sample space to finite or countably infinite set (e.g. the number of times we switch a light- bulb on and off before it dies, X = 0,1,2,3,...)

- A **continuous** random variable maps the sample space to an un- countable set (e.g. the lifetime of a lightbulb $X = \{x : x \geq 0\}$)

Definitions

Definition 4.1 (probability density function (pdf))

A function with values $f(x)$, defined over the set of all real numbers, is called a **probability density function (pdf)** of continuous random variable X if and only if

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

for any real constants a and b with $a \leq b$.

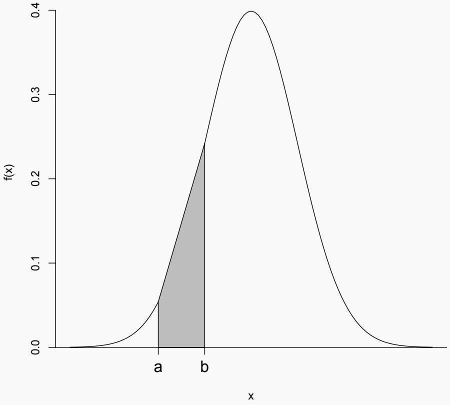
Note that $f(x) \neq P(X = x)$.

It is only once we integrate, that we get probabilities

Graphical presentation of probability densities

A pdf $f(x)$ integrated from a to b (with $a \leq b$), gives the probability that the continuous random variable will take on a value on the interval $[a, b]$.

- Graphically the $P(a \leq X \leq b)$ is represented by shaded region below.
- $P(X = x) = 0 \;\forall\; x$



Theorem 4.2

A function can serve as a probability density of a continuous random variable X if its values, $f(x)$, satisfy the conditions

(1) $f(x) \geq 0$, *for* $-\infty < x < \infty$.

(2) $\int_{-\infty}^{\infty} f(x)dx = 1$.

Definition 4.3 (Cumulative distribution function (cdf))

For a continuous random variable X that has the probability density (or pdf) at t as $f(t)$, the **cumulative distribution**, of X is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \qquad \forall -\infty < x < \infty.$$

Probabilities

- For X a continuous random variable, the following hold:

$$
\begin{aligned}
&P(X = x) = 0 \quad \text{for all } x,\\
&P(X \leq x) = P(X < x) \quad \text{for all } x,\\
&P(a \leq X \leq b) = F(b) - F(a) = \int_a^b f(x)dx,
\end{aligned}
$$

  where a and b are constants.

- This will all make sense after some examples and exercises.

Example 4.1

Find the constant c for the following pdf:

$$f(x) = \begin{cases} cx & \text{if } 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

Example 4.1

Find c and $P(0.2 < X < 0.5)$ using the pdf from Example 4.1.

Example 4.2

The pdf of a continuous variable X is given by

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find $F(x)$, the cdf of X, and $P(0.2 < X < 0.5)$ using the cdf.

Definition 4.4

The expected value of a continuous RV is defined as follows;

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)dx$$
$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

where $f(x)$ is the pdf of X.

We often denote $\mathbb{E}[X]$ by the Greek letter μ.

Definition 4.5

The variance of continuous RV X is

$$\mathbb{V}\text{ar}[X] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx.$$

Example 4.3

Find $\mathbb{E}(X)$ and $\mathbb{V}\text{ar}[X]$ for a random variable X with pdf:

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

Exercises I

Exercise 4.1 The pdf of a continuous variable X is given by

$$f(x) = \begin{cases} cx^2 & \text{if } 0 \leq x \leq 2 \\ 0 & \text{elsewhere} \end{cases}$$

where c is constant.

4.1.1 Find c.

4.1.2 Find the cdf $F(x)$.

4.1.3 What is $P(1 \leq X \leq 1.5)$?

4.1.4 Find $\mathbb{E}[X]$.

4.1.5 Find $\mathbb{V}\mathrm{ar}[X]$.

Exercises II

Exercise 4.2 The cdf of a continuous variable X is given by

$$F(x) = \begin{cases} 1 - e^{-x} & \text{if } x > 0 \\ 0 & \text{elsewhere} \end{cases}$$

4.2.1 What is $P(X \leq 2.6)$?
4.2.2 What is $P(1 < X < 4)$?

Background

- The normal distribution is the most frequently used continuous probability distribution.
- Many measurements can be well approximated by a normal distribution.
- The normal distribution is characterized by a *bell-shaped curve*.
- The normal distribution is also called the Gaussian distribution (after Johann Carl Friedrich Gauss).

Density Function

Definition 4.6 (Gaussian/Normal Distribution)

If a random variable X follows a normal distribution with parameters μ and σ, then we write $X \sim N(\mu, \sigma^2)$, and the pdf of X is

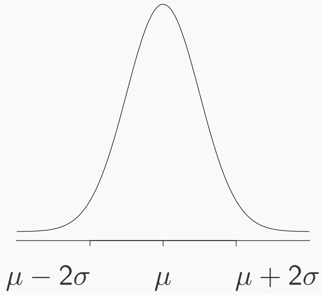
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty, \tag{1}$$

where

- μ is the **mean** of X.
- σ is the **standard deviation** of X.
- σ^2 is the **variance** of X.

Normal Distribution I

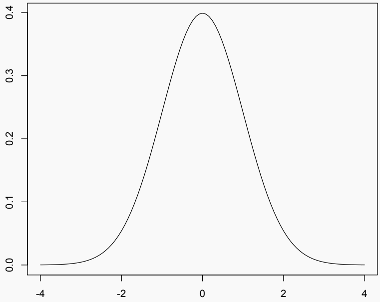
- The normal distribution is usually characterized by its mean μ and standard deviation σ.



- The normal distribution is symmetric about the mean.
- 95% of the density of the distribution lies within two (1.96 to be precise) standard deviations of the mean.

Standard Normal Distribution

- The **standard normal distribution** is a special case of the normal distribution when the mean (μ) is 0 and standard deviation (σ) is 1.



- 95% of the density of the distribution (the area under the curve) lies been ± 1.96.

- The normal or Gaussian distribution in Eq (1) cannot be integrated directly to obtain $P(X \leq x)$.

- The integration has to been done numerically using the standard normal which has the pdf

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \text{ for } -\infty < z < \infty$$

- The cumulative distribution function of a standard normal random variable is denoted as $\Phi(x)$, that is

$$\Phi(x) = P(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}z^2} dz$$

- Note: we can always transform a normal distribution to a standard normal distribution.

Normal Distribution

Theorem 4.7

Let $X \sim N(\mu, \sigma^2)$ and

$$Z = \frac{X - \mu}{\sigma}.$$

Then $Z \sim N(0, 1)$. This transformation is referred to as ***standardization****.*

- Therefore, to compute $P(X \leq x)$ we can use the fact that

$$\begin{aligned} P(X \leq x) &= P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{x-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

- Consult the standard normal tables (Tables in Canvas) that store the value of $P(Z \leq z)$ for many values of z.

Use Standard Normal Table to Calculate Probability

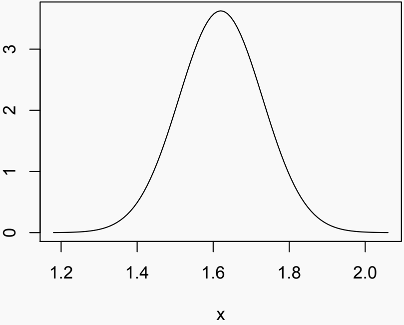
	Second decimal place of Z									
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015

Normal Distribution

Example 4.4

Consider the example of the height of women in Ireland, which we assume is normally distributed with mean 1.62m and standard deviation 0.11m.

- The median and mode of the height of women is 1.62m.
- Theoretically, the proportion of women shorter than 1.52m is the same as the proportion of women taller than 1.72m.
- 95% of women are between 1.40m and 1.84m tall.



Normal Distribution

Let us see some examples of how we use Z-Tables to find probabilities

Example 4.5

Let $X \sim N(2, 0.15^2)$. Compute $P(X \leq 2.1)$.

Example 4.6

Let $X \sim N(2, 0.15^2)$. Compute $P(X > 2.1)$

Exercises

Exercise 4.3 The height of women in a particular region follows a normal distribution with mean 1.55m and standard deviation 0.17m.

4.3.1 What proportion of women are smaller than 1.61m?

4.3.2 What proportion of women are taller than 1.8m?

4.3.3 What proportion of women are between 1.45m and 1.61m tall?

Exercise 4.4 The wingspans of the males of a certain species of bird of prey form a normal distribution with mean 162.50cm and standard deviation 6.0cm. What is the probability that the wingspan of a randomly selected male will exceed 170cm?

Exponential Distributions

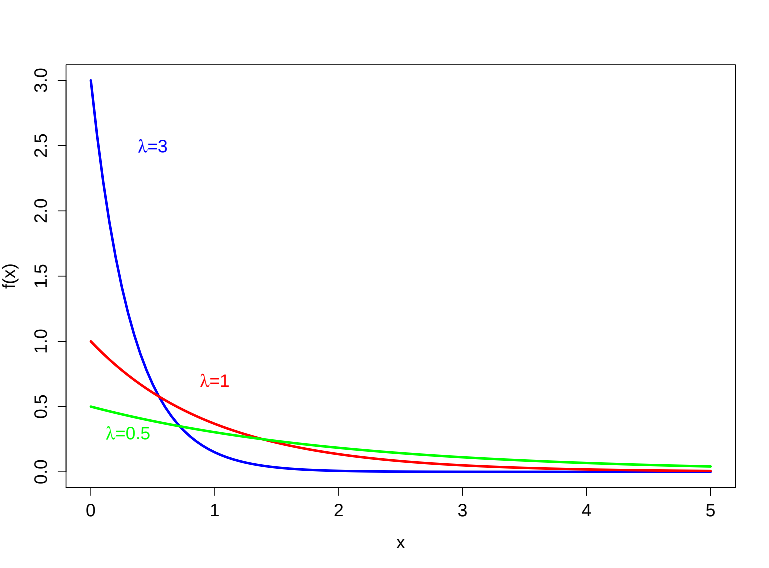
- Let X be a continuous random variable with density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

  then X arises from an exponential distribution.

- λ is called a rate parameter and $\lambda > 0$.

Visually: Exponential



Cumulative Distribution Function of Exponential Random Variable

The CDF for the exponential distribution does exist in closed form...

Example 4.7

Find the CDF for an exponential distribution.

Expected value and Variance of Exponential Random Variable

Example 4.8

Find the expected value and variance of exponential distribution

Memoryless Exponential random Variable

Example 4.9

We say that a nonnegative random variable X is memoryless if

$$P(X > s + t \mid X > t) = P(X > s)$$