Statistical Methods MATH2715 info

Teaching material is all online!

- On Minerva `http://minerva.leeds.ac.uk`
- On GitHub `https://github.com/luisacutillo78/Statistical-Methods-Lecture-Notes`

Resources

- Mathematical Statistics and Data Analysis - 3rd ed. (by J. A. Rice);
- Introduction to Statistics - Online Edition -D.M.Lane et al.
- http://www1.maths.leeds.ac.uk/statistics/R/Rintro.pdf;
- https://www.datacamp.com/courses/free-introduction-to-r.

Random Variables: Continuous Distributions

Definition. A *continuous random variable* is a r.v. that can take on a *continuum* of values.

Density function $f(x)$

The role of the pmf is taken by a **density function**, $f(x)$, which has the following properties:

- $f(x) \geq 0$
- $f(x)$ is piecewise continuous
- $\int_{-\infty}^{+\infty} f(x) dx = 1$
- $Pr(a < X < b) = \int_a^b f(x), \forall a < b.$

Random Variables: Continuous Distributions

Consequence

The probability that a continuous rv takes on a particular value is 0:

$$Pr(c < X < c) = \int_{c}^{c} f(x) = 0.$$

NOTE THAT. This is not true for discrete rv!

Cumulative Distribution Function $F(x)$

The cdf $F(x) = P(X \leq x)$ of a continuous rv $f(x)$, can be expressed as:

$$\int_{-\infty}^{x} f(x)$$

From the fund. Theor. of calc., if $f(x)$ is continuous at x, $f(x) = F'(x)$.

Properties of F

Consequence

The cdf of a continuous rv X can be use to evaluate the probability that X falls into an interval:

$$Pr(a \leq X \leq b) = \int_a^b f(x) = F(b) - F(a).$$

Quantiles

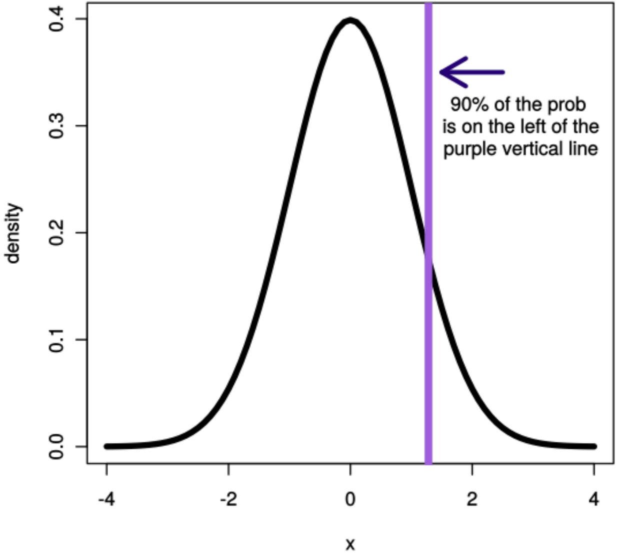
Definition. The pth quantile of F is defined to be the value:

$$x_p \ s.t. F(x_p) = p$$

Special values

N.T. when $p = 1/2$,$p = 1/4$ or $p = 3/4$, then x_p corresponds to the **median**, the *lower* or the *upper* **quartile** of F.

Properties of F



Uniform Distribution

Definition

A Uniform rv on $[0,1]$ describes the experiment "Pick a number *at random* between 0 and 1". The **Uniform Density** in $[0,1]$ is:

$$f(x) = \begin{cases} 1, & \text{if } 0 \leq x \leq 1 \\ 0, & \text{if } x < 0 \text{ or } x > 1 \end{cases}$$

Similarly, on a generic interval $[a,b]$, the Uniform Density is:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{if } x < a \text{ or } x > b \end{cases}$$

Uniform Distribution

Given that the cdf $F(x) = P(X \leq x)$ of a continuous rv $f(x)$, can be expressed as:

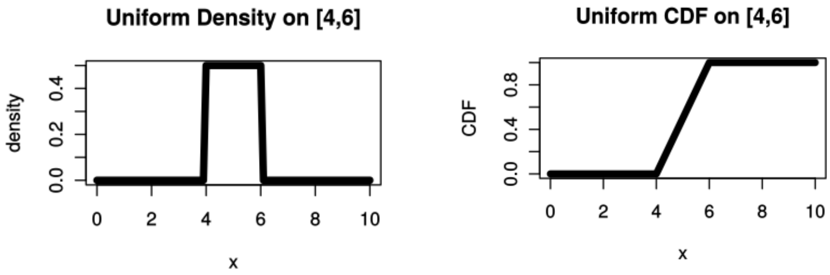
$$\int_{-\infty}^{x} f(x)$$

CDF

The CDF of a Uniform rv on $[0, 1]$ is:

$$F(x) = \begin{cases} 0, & if \ x \leq 0 \\ x, & if \ 0 \leq x \leq 1 \\ 1, & if \ x \geq 1 \end{cases}$$

Uniform Distribution



Exponential Distribution

Like the Poisson Distribution, the exponential density depends on a **single parameter** *lambda*.

Exponential Density Function

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & if \;\; x \geq 0 \\ 0, & if \;\; x < 0 \end{cases}$$

Exponential CDF

$$F(x) = \int_{-\infty}^{x} f(x) = \begin{cases} 1 - e^{-\lambda x}, & if \;\; x \geq 0 \\ 0, & if \;\; x < 0 \end{cases}$$

Exponential Distribution

The Exponential Distribution is used to model lifetimes or waiting times ($x \sim t$). E.g. : Suppose we model the lifetime of a fridge as an exponential rv. Assume that

- the fridge lasted a time s
- we want to compute the probability that it will last t more units: $P(T > t + s | T > s)$

Property: Memoryless

$$\begin{aligned} P(T > t + s | T > s) &= \frac{P(T > t + s \text{ and } T > s)}{P(T > s)} \\ &= \frac{P(T > t + s)}{P(T > s)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} \\ &= e^{-\lambda t} \end{aligned}$$

Exponential Distribution

Question

Is the exponential distribution a good model for human lifetime?

Exponential Distribution

Question

Is the exponential distribution a good model for human lifetime?

No!

The **memoryless** property implies that it is not a good model for human lifetime! Indeed, the probability that a 14 year old will live at least 5 more years must be very different from the probability that a 93 year old will live at least 5 more years!

Normal Distribution

A Normal random variable is often used as a generic symmetric random variable; i.e., the *bell-shaped* curve. Its density function depends on two parameters,μ and σ. The parameter μ corresponds to the mean and σ to the standard deviation.

Normal Density Function

If $X \sim N(\mu, \sigma^2)$. is a normal distribution with parameters μ and σ^2, its density function is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.$$

Standard Normal Density Function

A Normal rv is called a standard normal $Z \sim N(0, 1)$ if $\mu = 0$ and $\sigma = 1$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

Functions of a Normal Random Variable

Question

Suppose that $X \sim N(\mu, \sigma^2)$. What is the distribution of $Y = aX + b$?

SOLUTION: Y is still a Normal rv! Let's work it out.

Functions of a Normal Random Variable

Question

Suppose that $X \sim N(\mu, \sigma^2)$. What is the distribution of $Y = aX + b$?

SOLUTION: Y is still a Normal rv! Let's work it out.

Proposition 1

If $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, then $Y \sim N(a\mu + b, a^2\sigma^2)$.

Functions of a Generic Random Variable

Proposition 2

Let $Z = F(X)$; then Z has a Uniform distribution in $[0, 1]$.

Proposition 2

Let $Z = F(X)$; then Z has a Uniform distribution in $[0, 1]$.

Proposition 3

Let U be a Uniform in $[0, 1]$, and let $X = F^{-1}(U)$. Then the cdf of X is F.

R Examples

- I will introduce some R coding examples.
- I developed a library on Microsoft Azure for this module.
- `https://notebooks.azure.com/luisacutillo/libraries/MATH2715`
- Please create a free account to access my library (no need to do it now!).
- You will find a README file with instructions.
- You can Clone my notebooks if you want to run them on line (No need to install anything!)
- You can download my notebooks as .R files, if you want if you want a local run.
- In any case You need to create your own copy! If you clone it, please use a specific ID as a name.