STATISTICS FOR DATA SCIENCE

Random Variables

Prof. Uma D

Department of Computer Science and Engineering

STATISTICS FOR DATA SCIENCE

Continuous Random Variables

Prof. Uma D

STATISTICS FOR DATA SCIENCE

Topics to be covered...

- Continuous Random Variable
- Probability Density Function
- Cumulative Distribution Function
- Mean and Variance

Continuous Random variables

- A continuous random variable is one which takes an **infinite number of possible** values.

- Continuous random variables are usually **measurements**.

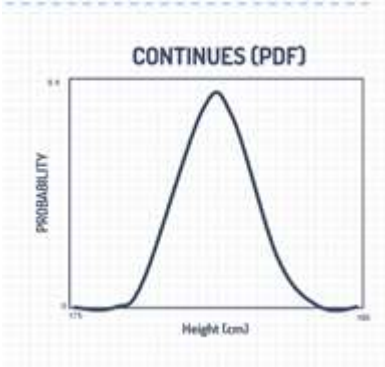
Examples

- height
- weight
- the amount of sugar in an orange
- the time required to run a mile.

STATISTICS FOR DATA SCIENCE

Probability Density Function

- A random variable is **continuous** if its probabilities are given by areas under a curve.



- The curve is called a **probability density function** (**pdf**) for the random variable. Sometimes the **pdf** is called the **probability distribution**.

- The function **f(x)** is the probability density function of *X*.

- Let *X* be a continuous random variable with probability density function $f(x)$. Then $\int_{-\infty}^{\infty} f(x)\,dx = 1$

STATISTICS FOR DATA SCIENCE

Continuous Random Variables

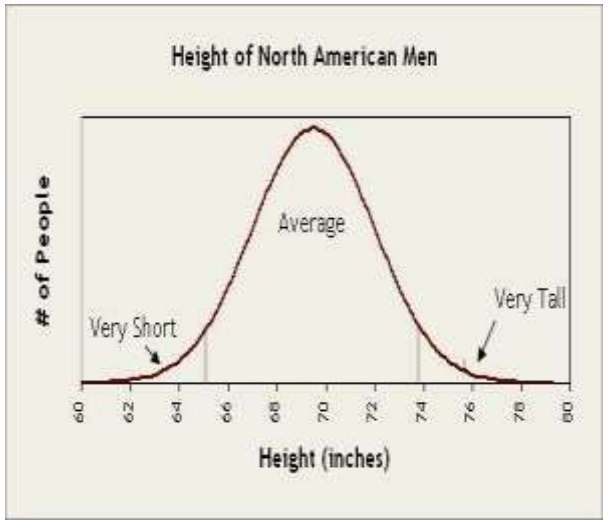
We model Continuous Random Variables with a **curve f(x)** called a probability density function(**pdf**).

f(x) is function that represents the height of the curve at point x.

Values where the curve is **high** are **more likely to occur**.

For Continuous Random Variables, probabilities are areas under the curve – hence found using integration.

Looks like a **smooth histogram**.



Source: www.slideshare.net

Probability Density Function of a C.R.V.

Probability Density Function of a C.R.V.

Cumulative Distribution Function of a C.R.V.

Percentile and Median of a C.R.V.

Mean and Variance of a C.R.V.

Problem 1

Suppose for a random variable X:

$f(x) = cx^3$ for $2 <= x <= 4$ and

0 otherwise.

a) What value of c makes this a legitimate probability distribution?
b) What is P(X > 3) .
c) Find P(X <= 2.7).
d) What is the median of this distribution?
e) Find mean and variance of this distribution.
f) What is the cumulative distribution function?

Solution:

a) What value of c makes this a legitimate probability distribution?

Solution:

b) What is P(X > 3) .

Solution:

c) Find P(X <= 2.7).

Solution:

d) What is the median of this distribution?

Solution:

e) Find mean and variance of this distribution.

Solution:

e) Find mean and variance of this distribution.

Solution:

f) What is the cumulative distribution function?

Problem 1 - Solution

a) c = 1/60

b) P(X> 3) = 0.729

c) P(X <= 2.7) = 0.155

d) Median = 3.415

e) Mean = 248/75 = 3.3
Variance = 11.2 – sq(3.3) = 0.31

f) CDF = $(x^4 – 2^4)/ 240$

Problem

Do It Yourself !!!

Let X be a random variable with PDF given by

$$f(x)=\begin{cases} x/250 & 20 \le x \le 30 \\ 0 & \text{otherwise} \end{cases}$$

1) Find $P(X \ge 25)$.
2) Find E(X) and Var(X).
3) Find CDF.
4) Find median.
5) Find 60th percentile.

THANK YOU

Prof. Uma D

Department of Computer Science and Engineering