STATISTICS FOR DATA SCIENCE

Continuous Random Variables

Prof. Uma D
Prof. Suganthi S
Prof. Silviya Nancy J

Department of Computer Science and Engineering

STATISTICS FOR DATA SCIENCE

Continuous Random Variables

Prof. Uma D
Prof. Suganthi S
Prof. Silviya Nancy J

Topics to be covered...

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

Continuous Random variables

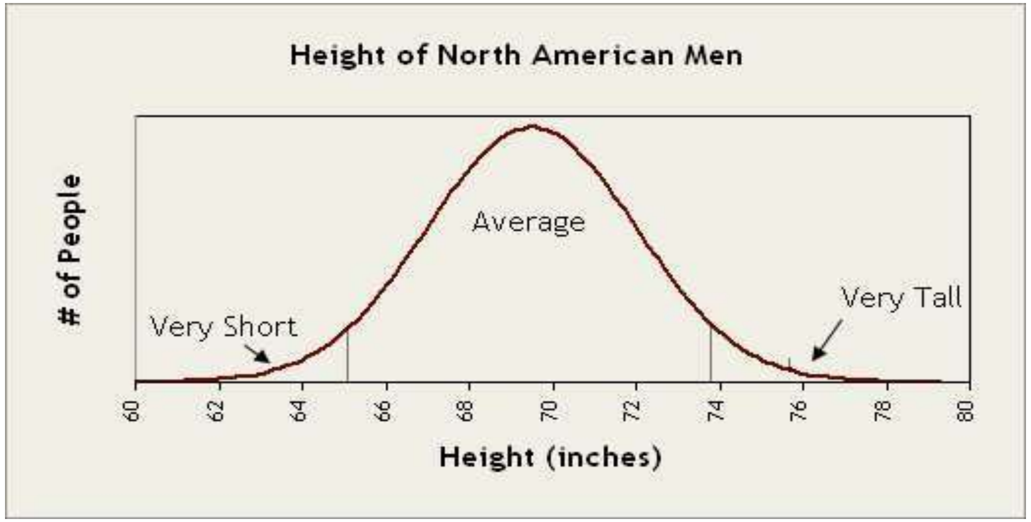
- A continuous random variable is one which takes an infinite number of possible values.
- Continuous random variables are usually measurements.

Examples

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

Continuous Random variables

- A **continuous random variable** *X* takes on all values in an interval of numbers.
- The probability distribution of *X* is described by a **density curve**.



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

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

Computing Probabilities with PDF

- Let X be a continuous random variable with probability density function $f(x)$. Let a and b be any two numbers, with $a < b$. Then

$$P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = P(a < X < b) = \int_a^b f(x)\,dx$$

In addition,

$$P(X \le b) = P(X < b) = \int_{-\infty}^{b} f(x)\,dx$$

$$P(X \ge a) = P(X > a) = \int_a^{\infty} f(x)\,dx$$

Example

A hole is drilled in a sheet-metal component, and then a shaft is inserted through the hole. The shaft clearance is equal to the difference between the radius of the hole and the radius of the shaft. Let the random variable X denote the clearance, in millimeters. The probability density function of X is

$$f(x) = \begin{cases} 1.25(1 - x^4) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Components with clearances larger than 0.8 mm must be scrapped. What proportion of components are scrapped?

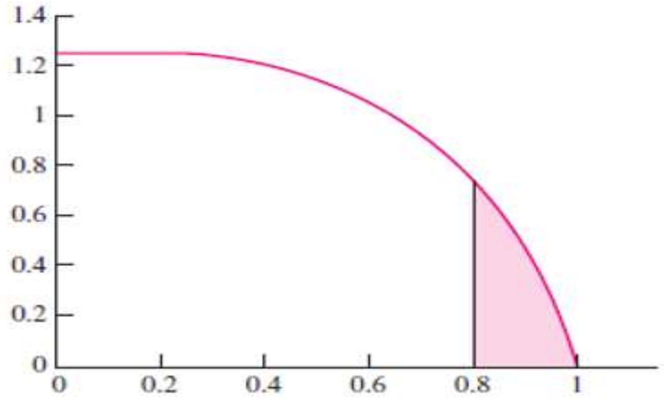


FIGURE 2.13 Graph of the probability density function of X, the clearance of a shaft. The area shaded is equal to $P(X > 0.8)$.

Example

Solution:

This area is given by

$$
\begin{aligned}
P(X > 0.8) &= \int_{0.8}^{\infty} f(x)\,dx \\
&= \int_{0.8}^{1} 1.25(1 - x^4)\,dx \\
&= 1.25\left(x - \frac{x^5}{5}\right)\Bigg|_{0.8}^{1} \\
&= 0.0819
\end{aligned}
$$

Cumulative Distribution function of a CRV

- Let X be a continuous random variable with probability density function $f(x)$.

- The **cumulative distribution function** of X is the function

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

Example

Find the Cumulative distribution function F(x) and plot it for the earlier example.

Solution:

The probability density function of X is given by $f(t) = 0$ if $t \leq 0$, $f(t) = 1.25(1 - t^4)$ if $0 < t < 1$, and $f(t) = 0$ if $t \geq 1$. The cumulative distribution function is given by $F(x) = \int_{-\infty}^{x} f(t)\, dt$. Since $f(t)$ is defined separately on three different intervals, the computation of the cumulative distribution function involves three separate cases.

If $x \leq 0$:

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)\, dt \\ &= \int_{-\infty}^{x} 0\, dt \\ &= 0 \end{aligned}$$

Example

If $0 < x < 1$:

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(t)\,dt \\
&= \int_{-\infty}^{0} f(t)\,dt + \int_{0}^{x} f(t)\,dt \\
&= \int_{-\infty}^{0} 0\,dt + \int_{0}^{x} 1.25(1 - t^4)\,dt \\
&= 0 + 1.25\left(t - \frac{t^5}{5}\right)\Bigg|_{0}^{x} \\
&= 1.25\left(x - \frac{x^5}{5}\right)
\end{aligned}
$$

If $x > 1$:

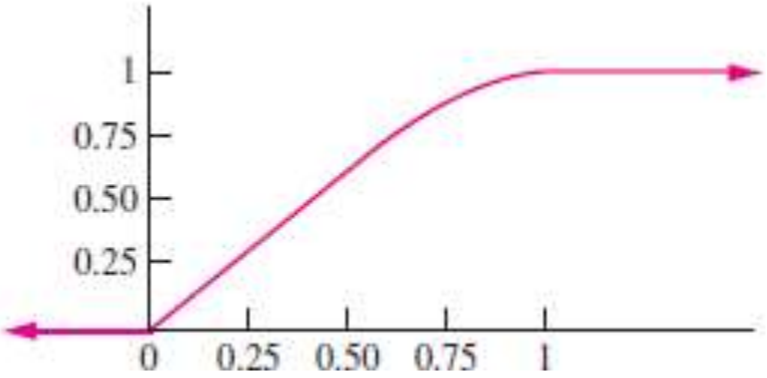
$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(t)\,dt \\
&= \int_{-\infty}^{0} f(t)\,dt + \int_{0}^{1} f(t)\,dt + \int_{1}^{x} f(t)\,dt \\
&= \int_{-\infty}^{0} 0\,dt + \int_{0}^{1} 1.25(1 - t^4)\,dt + \int_{1}^{x} 0\,dt \\
&= 0 + 1.25\left(t - \frac{t^5}{5}\right)\Bigg|_{0}^{1} + 0 \\
&= 0 + 1 + 0 \\
&= 1
\end{aligned}
$$

Example

Therefore

$$F(x) = \begin{cases} 0 & x \le 0 \\ 1.25\left(x - \dfrac{x^5}{5}\right) & 0 < x < 1 \\ 1 & x \ge 1 \end{cases}$$

A plot of $F(x)$ is presented here.



Mean for Continuous Random Variable

- Let X be a continuous random variable with probability density function .
- The **mean** of X is given by,

$$\mu_X = \int_{-\infty}^{\infty} x f(x)\, dx$$

- The mean of X is sometimes called the expectation, or expected value, of X and may also be denoted by $E(X)$ or by μ.

Variance for Continuous Random Variable

- Let X be a continuous random variable with probability density function, then

The variance of X is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\, dx$$

An alternate formula for the variance is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu_X^2$$

Example

For the earlier example, find mean and variance for the random variable.

Solution:

We compute the mean by using

$$\begin{aligned}
\mu_X &= \int_{-\infty}^{\infty} x f(x)\,dx \\
&= \int_0^1 x[1.25(1 - x^4)]\,dx \\
&= 1.25\left(\frac{x^2}{2} - \frac{x^6}{6}\right)\Bigg|_0^1 \\
&= 0.4167
\end{aligned}$$

Example

Solution:

We compute the variance by using

$$\begin{aligned}
\sigma_X^2 &= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu_X^2 \\
&= \int_0^1 x^2[1.25(1 - x^4)]\,dx - (0.4167)^2 \\
&= 1.25\left(\frac{x^3}{3} - \frac{x^7}{7}\right)\Bigg|_0^1 - (0.4167)^2 \\
&= 0.0645
\end{aligned}$$

Problem 1

Suppose for a random variable X:

$$f(x) = \begin{cases} cx^3 & \text{for } 2 <= x <= 4 \quad \text{and} \\ 0 & \text{otherwise.} \end{cases}$$

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 = $(x^4 - 2^4)/ 240$

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

THANK YOU

Prof. Uma D
Prof. Suganthi S
Prof. Silviya Nancy J

Department of Computer Science and Engineering