

P3 - Continuous distributions

STAT 587 (Engineering) - Iowa State University

February 4, 2019

Cumulative distribution function

All properties of discrete random variables have direct counterparts for continuous random variables.

In particular,

Definition

The **cumulative distribution function** for a continuous random variable is

$$F_X(x) = P(X \leq x) = P(X < x)$$

since $P(X = x) = 0$ for any x .

we still have the properties

- $0 \leq F_X(x) \leq 1$ for all $x \in \mathbb{R}$
- F_X is monotone increasing, i.e. if $x_1 \leq x_2$ then $F_X(x_1) \leq F_X(x_2)$.
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$ and $\lim_{x \rightarrow \infty} F_X(x) = 1$.

Probability density function

Definition

The **probability density function (pdf)** for a continuous random variable is

$$f_X(x) = \frac{d}{dx}F_X(x)$$

and

$$F_X(x) = \int_{-\infty}^x f_X(t)dt.$$

Thus, the probability density function has the following properties

- $f_X(x) \geq 0$ for all x and
- $\int_{-\infty}^{\infty} f(x)dx = 1.$

Example

Let X be a random variable with probability density function

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

$f_X(x)$ defines a valid pdf because $f_X(x) \geq 0$ for all x and

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_0^1 3x^2 dx = x^3 \Big|_0^1 = 1.$$

The cdf is

$$F_X(x) = \begin{cases} 0 & x \leq 0 \\ x^3 & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}.$$

Expected value

Definition

Let X be a continuous random variable and h be some function. The **expected value** of a function of a continuous random variable is

$$E[h(X)] = \int_{-\infty}^{\infty} h(x) \cdot f_X(x) dx.$$

If $h(x) = x$, then

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx.$$

and we call this the **expectation** of X . We commonly use the symbol μ for this expectation.

Example (cont.)

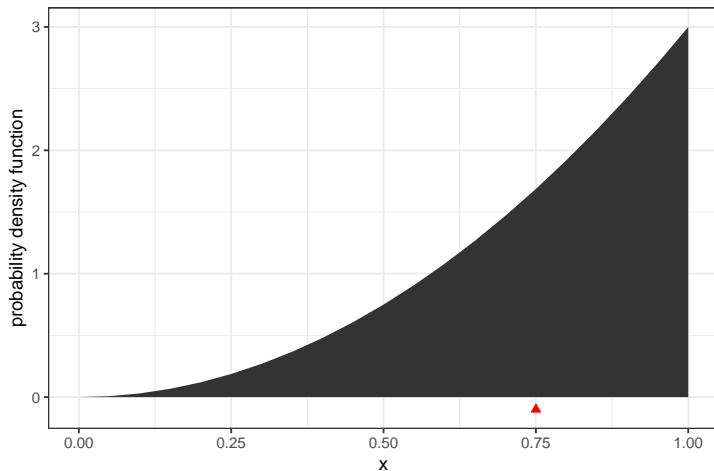
Let X be a random variable with probability density function

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The expected value is

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx \\ &= \int_0^1 3x^3 dx \\ &= 3 \frac{x^4}{4} \Big|_0^1 = \frac{3}{4}. \end{aligned}$$

Example - Center of mass



Variance

Definition

The **variance** of a random variable is defined as the expected squared deviation from the mean. For continuous random variables, variance is

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

where $\mu = E[X]$. The symbol σ^2 is commonly used for the variance.

Definition

The **standard deviation** is the positive square root of the variance

$$SD[X] = \sqrt{Var[X]}.$$

The symbol σ is commonly used for the standard deviation.

Example (cont.)

Let X be a random variable with probability density function

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

The variance is

$$\begin{aligned} \text{Var}[X] &= \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx \\ &= \int_0^1 \left(x - \frac{3}{4}\right)^2 3x^2 dx \\ &= \int_0^1 \left[x^2 - \frac{3}{2}x + \frac{9}{16}\right] 3x^2 dx \\ &= \int_0^1 3x^4 - \frac{9}{2}x^3 + \frac{27}{16}x^2 dx \\ &= \left[\frac{3}{5}x^5 - \frac{9}{8}x^4 + \frac{9}{16}x^3\right] \Big|_0^1 dx \\ &= \frac{3}{5} - \frac{9}{8} + \frac{9}{16} \\ &= \frac{3}{80}. \end{aligned}$$

Comparison of discrete and continuous random variables

For simplicity here and later, we drop the subscript X .

| | discrete | continuous |
|----------------|---|---|
| image | finite or countable | uncountable |
| pmf | $p(x) = P(X = x)$ | |
| pdf | | $p(x) = f(x) = F'(x)$ |
| cdf | $F(x) = P(X \leq x)$ $= \sum_{t \leq x} p(x)$ | $F(x) = P(X \leq x)$ $= \int_{-\infty}^x p(t)dt$ |
| expected value | $E[h(X)] = \sum_x h(x)p(x)$ | $E[h(X)] = \int_x h(x)p(x)dx$ |
| expectation | $\mu = E[X] = \sum_x x p(x)$ | $\mu = E[X] = \int_x x p(x)dx$ |
| variance | $Var[X] = E[(X - \mu)^2]$ $= \sum_x (x - \mu)^2 p(x)$ | $Var[X] = E[(X - \mu)^2]$ $= \int_x (x - \mu)^2 p(x)dx$ |

Note: we replace summations with integrals when using continuous as opposed to discrete random variables

Uniform

A **uniform** random variable on the interval (a, b) has equal probability for any value in that interval and we denote this $X \sim Unif(a, b)$. The pdf for a uniform random variable is

$$f(x) = \frac{1}{b-a} \mathbf{I}(a < x < b)$$

where $\mathbf{I}(A)$ is an indicator function that is 1 if A is true and 0 otherwise, i.e.

$$\mathbf{I}(A) = \begin{cases} 1 & A \text{ is true} \\ 0 & \text{otherwise.} \end{cases}$$

The expectation is

$$E[X] = \int_a^b \frac{1}{b-a} x \, dx = \frac{a+b}{2}$$

and the variance is

$$Var[X] = \int_a^b \frac{1}{b-a} \left(x - \frac{a+b}{2} \right)^2 dx = \frac{1}{12} (b-a)^2.$$

Example (cont.)

Pseudo-random number generators generate pseudo uniform values on $(0,1)$. These values can be used in conjunction with the inverse of the cumulative distribution function to generate pseudo-random numbers from any distribution.

The inverse of the cdf $F_X(x) = x^3$ is

$$F_X^{-1}(u) = u^{1/3}.$$

A uniform random number on the interval $(0,1)$ generated using the inverse cdf produces a random draw of X . So, in R

```
inverse_cdf = function(u) u^(1/3)
x = inverse_cdf(runif(1e6))
mean(x)
```

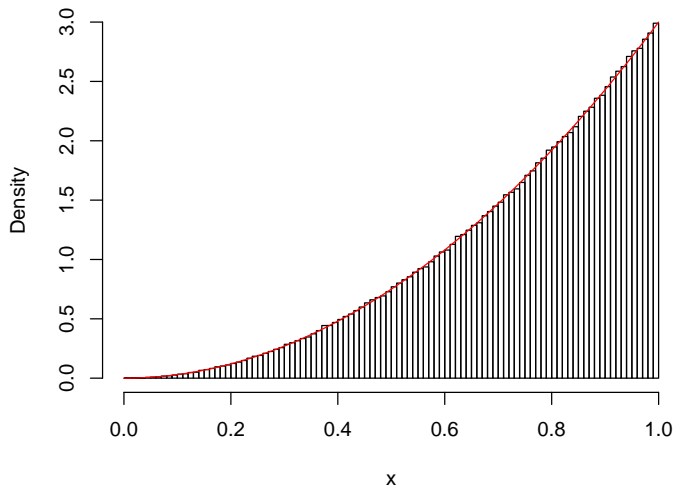
```
[1] 0.7502002
```

```
var(x); 3/80
```

```
[1] 0.03752111
```

```
[1] 0.0375
```

Histogram of x



Normal distribution

The **normal (or Gaussian) density** is a “bell-shaped” curve. The density has two parameters: mean μ and variance σ^2 and is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \quad \text{for } -\infty < x < \infty$$

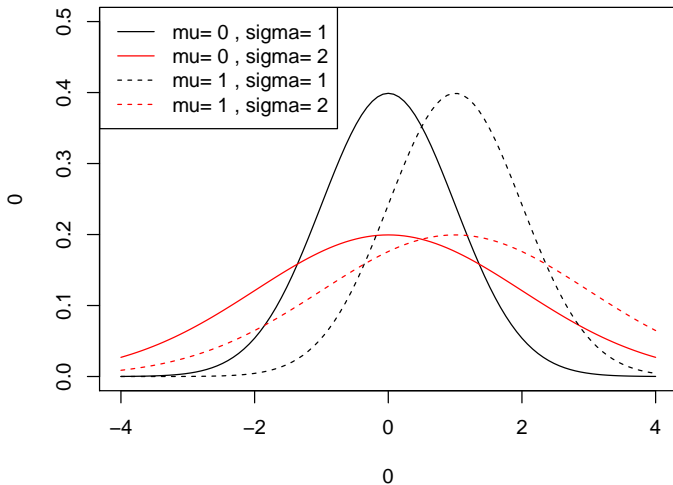
The expected value and variance of a normal distributed r.v. X are:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx = \dots = \mu \\ \text{Var}[X] &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \dots = \sigma^2. \end{aligned}$$

Thus, the parameters μ and σ^2 are actually the mean and the variance of the $N(\mu, \sigma^2)$ distribution.

There is no closed form cumulative distribution function for a normal random variable.

Example probability density functions



Properties of the normal distribution

Let $Z \sim N(0, 1)$, i.e. a **standard normal** random variable. (If you see Z without explanation, it is a standard normal random variable.) Then for constants μ and σ

$$X = \mu + \sigma Z \sim N(\mu, \sigma^2)$$

and

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

which is called **standardizing**.

Let $X_i \stackrel{\text{ind}}{\sim} N(\mu_i, \sigma_i^2)$. Then

$$Z_i = \frac{X_i - \mu_i}{\sigma_i} \stackrel{\text{iid}}{\sim} N(0, 1) \quad \text{for all } i$$

and

$$Y = \sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right).$$

Calculating the standard normal cdf

If $Z \sim N(0, 1)$, what is $P(Z \leq 1.5)$? Although the cdf does not have a closed form, very good approximations exist and are available as tables or in software, e.g.

```
pnorm(1.5) # default is mean=0, sd=1
```

```
[1] 0.9331928
```

If $Z \sim N(0, 1)$, then

- $P(Z \leq z) = \Phi(z)$
- $\Phi(z) = 1 - \Phi(-z)$ since a normal pdf is **symmetric** around its mean.

Calculating any normal cumulative distribution function

If $X \sim N(15, 4)$ what is $P(X > 18)$?

$$\begin{aligned}P(X > 18) &= 1 - P(X \leq 18) \\&= 1 - P\left(\frac{X-15}{2} \leq \frac{18-15}{2}\right) \\&= 1 - P(Z \leq 1.5) \\&\approx 1 - 0.933 = 0.067\end{aligned}$$

```
1-pnorm((18-15)/2)           # by standardizing  
  
[1] 0.0668072  
  
1-pnorm(18, mean=15, sd=2) # using the mean and sd arguments  
  
[1] 0.0668072
```

Manufacturing

Suppose you are producing nails that must be within 5 and 6 centimeters in length. If the average length of nails the process produces is 5.3 cm and the standard deviation is 0.1 cm. What is the probability the next nail produced is outside of the specification?

Let $X \sim N(\mu, \sigma^2)$ be the next nail produced with $\mu = 5.3$ cm and $\sigma = 0.1$ cm. We need to calculate

$$\begin{aligned} P(X < 5 \text{ or } X > 6) &= 1 - P(5 < X < 6) \\ &= 1 - [P(X < 6) - P(X < 5)] \quad \text{or} \\ &= P(X < 5) + [1 - P(X < 6)]. \end{aligned}$$

```
mu = 5.3
sigma = 0.1

pnorm(5, mean = mu, sd = sigma) + ( 1 - pnorm(6, mean = mu, sd = sigma) )

[1] 0.001349898
```