Intro to Continuous random variables

Lecture 3a (STAT 24400 F24)

Continuous random variables

Recall: a discrete random variable has finitely many or countably infinitely many possible values, characterized by its PMF

$$p(x) = \mathbb{P}(X = x)$$

(which is > 0 for at most countably infinitely many x) or its CDF: $F(x) = \mathbb{P}(X \le x)$.

Other random variables might take values in a continuous range
(which is uncountably infinitely many).

Definition: If a random variable X has no mass at any single value, i.e.,

$$\mathbb{P}(X = x) = 0 \qquad \text{for any } x \in \mathbb{R}$$

then it is a **continuous random variable**.

Density function for a continuous random variable

Just like for a discrete r.v., for a continuous r.v. we can describe its distribution via the CDF (cumulative distribution function),

$$F(x) = \mathbb{P}(X \le x)$$

However, for a continuous r.v., it is pointless to use a PMF, since $\mathbb{P}(X = x) = 0$ for every value x.

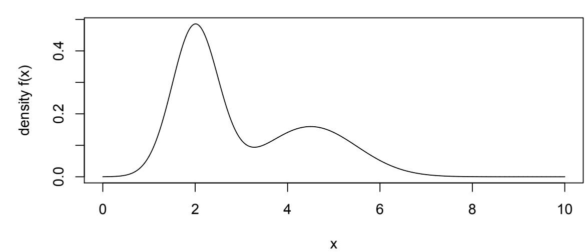
Instead we use a **probability density function** (a.k.a. the PDF), $f(x)$ (sometimes written as $f_X(x)$), which plays an analogous role.

The probability of r.v. X falling into an interval (a, b) is

$$\mathbb{P}(a < X < b) = \int_{x=a}^{b} f(x)\,\mathrm{d}x = \int_a^b f(x)dx$$

An example of density function

Example:



Intuitively, the PDF $f(x)$ is of larger magnitude or higher around x if X is more likely to take on values near x; and vice versa.

Properties of density functions

By the basic characteristics of probability, a density function should have the following properties:

- $f(x) \geq 0$ for all $x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$
- f is piecewise continuous

Density function & CDF

Relationship between density function & CDF:

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(-\infty < X \leq x)$$
$$= \mathbb{P}(-\infty < X < x) = \int_{t=-\infty}^{x} f_X(t)\,\mathrm{d}t$$

since X is continuous

Equivalently (by the fundamental theorem of calculus),

$$f_X(x) = F_X'(x)$$

(if f is continuous at x).

That is, density function is the derivative of the CDF.

(Notation: $\int_{t=-\infty}^{x} f_X(t)\,\mathrm{d}t = \lim_{s\to\infty} \int_{t=-s}^{x} f_X(t)\,\mathrm{d}t$)

Common distributions for continuous random variables

Some common continuous distribution:

- Uniform distribution
- Exponential distribution
- Gamma distribution
- Normal distribution

Uniform distribution

What do we mean when we "draw X at random from the interval $[a, b]$"?

- For a finite set (e.g. choose at random an integer between 1 and 10), the meaning is clear — every value should be equally likely.
- In the continuous case, we mean that the probability should be "evenly spread" across the interval.

Uniform$[a, b]$ distribution — parameters $a, b \in \mathbb{R}$ with $a < b$.

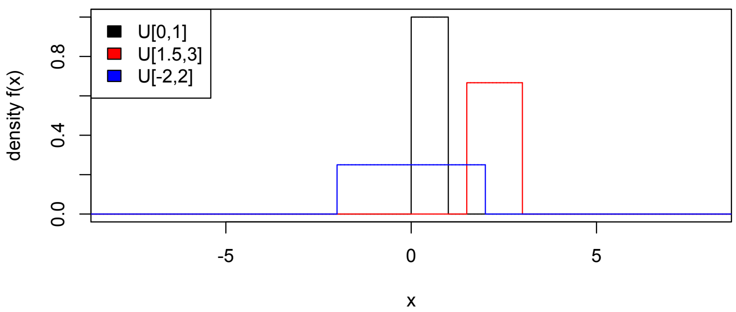
Density:

$$f(x) = \frac{1}{b-a} \quad \text{for } x \in [a, b]$$

CDF:

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0, & \text{for } x < a \\ \frac{x-a}{b-a}, & \text{for } x \in [a, b] \\ 1, & \text{for } x > b \end{cases}$$

Uniform distribution examples



(Note: vertical lines are for convenience only)

Exponential distribution

Exponential(λ) distribution—parameter $\lambda > 0$ is the "rate".
Density:

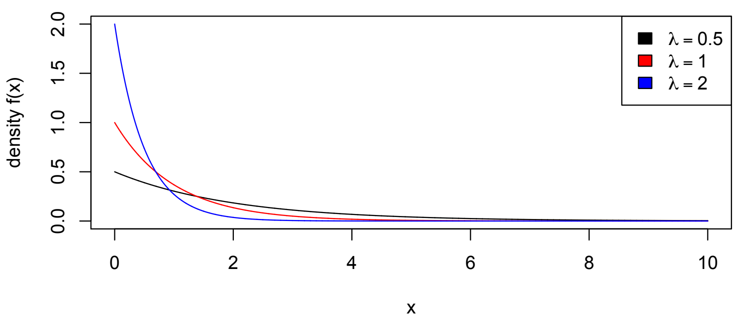
$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0$$

- This distribution is **memoryless**
  (recall the property of the geometric distribution in the discrete case)
- It's a natural model for many physical phenomena that are memoryless
  (e.g. half-life decay of a radioactive atom; time an ion passing through a channel)
- Larger λ means shorter time to decay

Calculate the CDF: for $x \geq 0$,

$$F(x) = \mathbb{P}(X \leq x) = \int_{t=0}^{x} f(t)\, dt = \int_{t=0}^{x} \lambda e^{-\lambda t}\, dt = \left[-e^{-\lambda t}\right]_{t=0}^{x} = 1 - e^{-\lambda x}$$

Exponential distribution examples



Gamma distribution

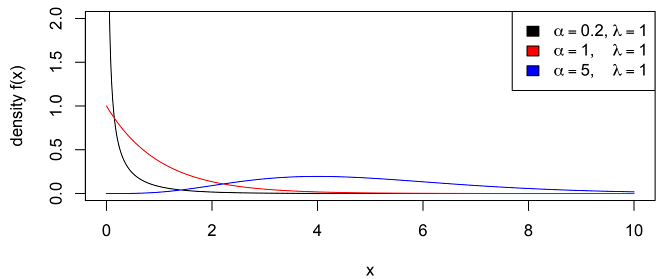
Gamma(α, λ) distribution — parameters $\alpha > 0$ ("shape"), $\lambda > 0$ ("rate")
Density:

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad \text{for } x \geq 0$$

- $\Gamma(\alpha)$ is a normalizing constant (so that density integrates to 1).
  $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$; $\Gamma(k) = (k-1)!$ for integers $k \geq 1$,
- Generalization of the exponential: Gamma$(1, \lambda)$ = Exponential(λ)
- This family allows for a change in the rate of decay
  - $\alpha < 1$ — an event that becomes increasingly less likely, with the density declining sharply after zero
  - $\alpha = 1$ — a constant rate of decay (= exponential distrib.)
  - $\alpha > 1$ — an event that becomes more likely as time goes on, and gives a shape that has a peak (mode) at some value above zero

Note: the text calls λ the "scale" but this does not agree with standard terminology; $\frac{1}{\lambda}$ is called a scale parameter sometimes.

Gamma distribution examples



Examples:

- $\alpha < 1$ — how long a person works at a company
  (quitting is more likely early during the adjustment period)
- $\alpha = 1$ — time-to-decay of a radioactive atom (memoryless)
- $\alpha > 1$ — lifespan of an organism (death more likely when older)

Normal distribution

Normal(μ, σ^2) distribution (also written as N(μ, σ^2))

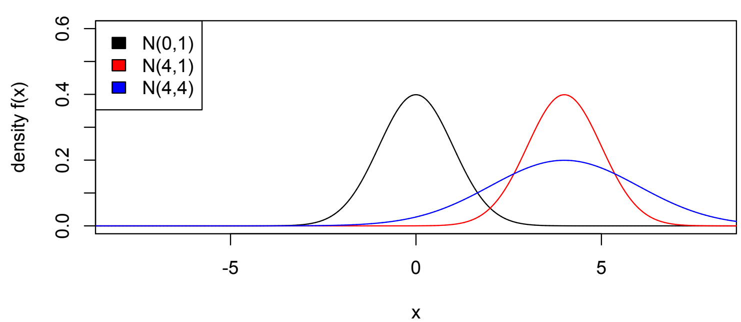
Parameters:

- $\mu \in \mathbb{R}$ — "mean"
- $\sigma > 0$ — "standard deviation" ($\sigma^2 > 0$ is called the "variance")

Density:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \qquad \text{for } x \in \mathbb{R}$$

Normal distribution examples



Standard Normal

- N$(0, 1)$ is called the **standard normal distribution**, with density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad \text{for } x \in \mathbb{R}$$

- Caution of notations:
  Different texts or software may use N(μ, σ^2) or N(μ, σ) — be careful!
  (i.e., does N$(1, 3)$ indicate $\sigma = 3$ or $\sigma^2 = 3$?)

Properties of normal distribution

- The normal distribution has a symmetric "bell curve" shape.
  The tails are very "thin" :
  the density becomes very low very quickly as x moves away from the mean.
- Often used to approximate distribution of some naturally occurring phenomenon or measured quantities such as height
- Most variables we can measure are not normally distributed
- However, there is an important connection between the normal distribution and the process of sampling from a population:

  > An average of a randomly chosen sample is approximately normally distributed, even if the original distribution is not.
  > (Central Limit Theorem, later in the course)

  — The main reason that normal distribution is very important in statistics.

Remarks on densities — 'Support" of PDF

X is supported on the interval $I \Leftrightarrow f(x) = 0$ for any $x \notin I$.

It may be convenient to think of f as a function with domain either I or $\mathbb{R}$.

For example for the Uniform$[a,b]$ distribution, we might write

$$f(x) = \frac{1}{b-a} \quad \text{on the support } x \in [a,b]$$

or equivalently we might write

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

Similarly for the CDF:

$$F(x) = \frac{x-a}{b-a} \qquad \text{on the support } x \in [a,b]$$

which is equivalent to

$$F(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \le x \le b, \\ 1, & x > b \end{cases}$$

Remarks on densities – end points of intervals

For any continuous distribution, point mass is always zero:

$$\mathbb{P}(X = x) = \int_{t=x}^{x} f(t)dt = 0 \quad \text{for any } x$$

Consequently, an event has the same probability on closed or open intervals,

e.g.

$\mathbb{P}(X \ge a) = \mathbb{P}(X > a)$

$\mathbb{P}(a \le X \le b) = \mathbb{P}(a \le X < b) = \mathbb{P}(a < X < b)$

Thus we can ignore closed vs open endpoints of intervals for continuous r.v.'s, when calculating event probabilities.

Remarks on densities – piecewise continuity

The PDF $f(x)$ of a continuous random variable may not be continuous.

Example (exercise)

A r.v. X has PDF $f(x) = \begin{cases} c, & x \in [0,1] \\ c(x-1), & x \in (1,2] \\ 0, & \textit{elsewhere}. \end{cases}$

- Note that $f(x)$ is not continuous at $x = 0, 1, 2$.
- What is the value of c?

$$\int_{-\infty}^{\infty} f(x)dx = \int_{x=0}^{1} c\,dx + \int_{x=1}^{2} c(x-1)dx = \cdots = 1 \quad \Rightarrow \quad c = \frac{2}{3}$$

- What is the CDF of X?