

Section 4: Random variables – continuous

STA 35C – Statistical Data Science III

Instructor: Akira Horiguchi

Fall Quarter 2025 (Sep 24 – Dec 12)
MWF, 12:10 PM – 1:00 PM, Olson 158
University of California, Davis

Based on Chapter 4 of textbook: <https://www.probabilitycourse.com/>

- Contains problems with solutions, and problems without solutions.

1 Introduction

2 Probability density function

3 Expected value and variance

4 Special Distributions

Introduction

- A *discrete* random variable has a countable range.
- A *continuous* random variable has a range in the form of an interval or a union of non-overlapping intervals on the real line (possibly the whole real line).
For any $a \in \mathbb{R}$, we have $P(X = a) = 0$.
- Need new tools. Overview: replace sums with integrals; replace PMFs with PDFs.

Example

Find the CDF of a real number chosen **uniformly at random** in the interval $[a, b]$.

- **Uniformly at random**: all intervals in $[a, b]$ that have the same length must have the same probability.
- For any interval subset $[c, d]$ of $[a, b]$ with $c < d$, uniformity assumption implies

$$P(X \in [c, d]) \propto (d - c). \quad (1)$$

- Since $P(X \in [a, b]) = 1$, we conclude *"proportional to"*

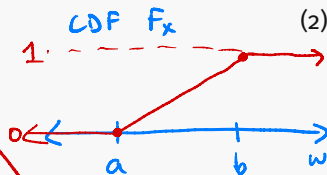
$$P(X \in [c, d]) = \frac{d - c}{b - a}. \quad (2)$$

To get the CDF, note that

- $F_X(w) = 0$ if $w < a$;
- $F_X(w) = 1$ if $w > b$;
- For $w \in [a, b]$, we have

$$F_X(w) = P(X \leq w) = P(X \in [a, w]) = \frac{w - a}{b - a}. \quad (3)$$

Thus the CDF looks like (no jumps)



In the previous example, the CDF is a continuous function (it has no jumps).

- Recall: jumps in the CDF correspond to values a for which $P(X = a) > 0$.
- “No jumps” is consistent with $P(X = a) = 0$ for all a .

Definition 1: continuous random variable

A random variable X is said to be *continuous* if its CDF $F_X: \mathbb{R} \rightarrow [0, 1]$ is continuous.

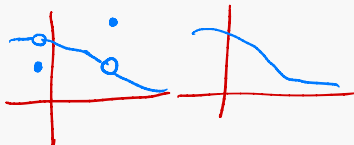
Probability density function

Motivation

To determine the distribution of a random variable of any type, we can provide its CDF.

- For a *continuous* r.v., we cannot use a PMF because $P(X = a) = 0$ for all $a \in \mathbb{R}$.
- We will instead define its *probability density function (PDF)*.
Similar to mass density in physics.
- For a continuous r.v. X and a value $a \in \mathbb{R}$, consider the quantity

$$\lim_{h \rightarrow 0^+} \frac{P(a < X \leq a+h)}{h} = \lim_{h \rightarrow 0^+} \frac{F_X(a+h) - F_X(a)}{h} = F_X'(a)$$



For simplicity, we will assume for this course that the CDF of a continuous random variable is differentiable except at possibly a few values in \mathbb{R} .

- “almost everywhere” in \mathbb{R}

Definition 2: probability density function (PDF)

Consider a continuous random variable X with an “absolutely continuous” CDF F_X . The function f_X defined by

$$f_X(a) = \frac{dF_X(a)}{da} = F'_X(a), \quad \text{if } F_X(a) \text{ is differentiable at } a \quad (4)$$

is called the *probability density function (PDF)* of X .

“absolutely continuous” allows us to obtain the CDF from a PDF (we’ll see this later)

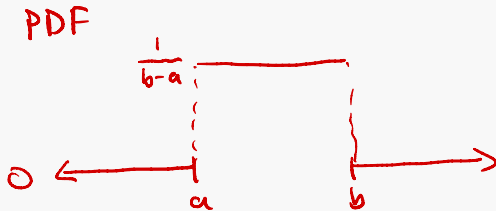
Example

Let's find the PDF of the uniform random variable discussed in slide 3.

- This random variable is said to have a *Uniform*(a, b) distribution.
- The CDF of X is given in slide 3. By taking the derivative, we obtain the PDF:


$$f_X(c) = \begin{cases} \frac{1}{b-a} & \text{if } a < c < b, \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

This function (the PDF) looks like



Obtain the CDF from a PDF

If we are given a PDF f_X , then the CDF is the function F defined by


$$F(c) = \int_{-\infty}^c f_X(u) du \quad (6)$$

- “absolutely continuous” is a stronger condition than “continuous”; allows us to use the fundamental theorem of calculus.
- For this course, assume that all CDFs of *continuous* r.v.s are absolutely continuous.

From Eq. (6) we also get F is the CDF of X (b)

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f_X(u) du \quad (7)$$

In particular, since $P(-\infty < X \leq \infty) = 1$, we must also have

$$\int_{-\infty}^{\infty} f_X(u) du = 1, \quad (8)$$

i.e., *the area under the PDF curve must be equal to one.*

Summary of properties of a PDF

Consider a continuous random variable X with PDF f_X . We have

1. $f_X(c) \geq 0$ for all $c \in \mathbb{R}$.
2. Eq. (8) — *area under PDF curve = 1*
3. Eq. (7)
4. More generally, for a set A , we have $P(X \in A) = \int_A f_X(u) du$.

For item 4, the set A must satisfy some mild conditions which are almost always satisfied in practice. E.g., the set A could be a union of some disjoint intervals:

$$P(X \in \overbrace{[0, 2] \cup [3, 4]}) = \underbrace{\int_0^2 f_X(u) du} + \underbrace{\int_3^4 f_X(u) du}.$$

Example

Let X be a continuous random variable with the following PDF

$$f_X(u) = \begin{cases} \underline{c} e^{-u} & u \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

where c is a positive constant.

1. Find c .
2. Find the CDF of X .
3. Find $P(1 < X < 3)$.

$$\begin{aligned} 1. \quad 1 &= \int_{-\infty}^{\infty} f_X(u) du = \int_0^{\infty} c e^{-u} du \\ \Rightarrow c &= \frac{1}{\int_0^{\infty} e^{-u} du} = \frac{1}{1} = 1 \end{aligned}$$

$$2. \quad F_X(b) = \int_{-\infty}^b f_X(u) du = \begin{cases} 0 & b < 0 \\ \int_0^b e^{-u} du = 1 - e^{-b} & , \quad b \geq 0 \end{cases}$$

$$3. \quad P(1 < X < 3) = F_X(3) - F_X(1) = e^{-1} - e^{-3}.$$

$$P(1 \leq X \leq 3) = \underbrace{P(X=1)}_0 + P(1 < X < 3) + \underbrace{P(X=3)}_0$$

If X is a continuous random variable, we can define the *range of X* as

$$R_X = \{c \mid \underline{f_X(c)} > 0\},$$

i.e., the set of real numbers c for which the PDF f_X is larger than zero.

Expected value and variance

Definition

The *expected value* of a random variable is the weighted average of all of its *values*, where the *weights* are the probabilities that these values occur.

Definition 3: Expected value $E(\cdot)$

Let X be a continuous random variable. Then the *expected value* of X is defined as

$$E(X) = \int_{-\infty}^{\infty} \underbrace{u}_{\text{value}} \underbrace{f_X(u)}_{\text{weight}} du \quad (9)$$

Expected Value of a Function of a Continuous Random Variable

Law of the unconscious statistician (LOTUS) for continuous random variables

$$E(g(X)) = \int_{-\infty}^{\infty} g(u) f_X(u) du \quad (10)$$

The “Expectation is linear” properties from slide 21 of Sec 3 also hold for continuous random variables.

Same expressions as with discrete random variables, but this time using Eq. (9) for the expected value:

$$\text{Var}(X) = E[(X - EX)^2] = E[X^2] - (EX)^2 \quad (11)$$

- The variance properties in slides 28 and 29 of Sec 3 also hold for continuous random variables.

Special Distributions

Uniform distribution

A random variable X is **uniformly** distributed on an interval $M = (a, b)$, with $b > a$, if the PDF has the form

$$f_X(x) = \frac{1}{b-a} \quad \text{for all } x \in (a, b).$$

- Such distributions occur when all (uncountably many) possible outcomes are equally likely.
- The interval M can also instead be $[a, b)$, or $(a, b]$, or $[a, b]$.
- Here we also write $X \sim U(M)$ or $X \sim \text{Unif}(M)$.
- Nine random draws in (3, 5) in R: `runif(n=9, min=3, max=5)`
- Expected value and variance

$$EX = \frac{a+b}{2} \quad \text{and} \quad \text{Var}(X) = \frac{(b-a)^2}{12}.$$

End of 10/13
lecture

Normal distribution

A random variable X is **normally** distributed with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, if the PDF has the form

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for all } x \in \mathbb{R}.$$

- This distribution appears often in this class, in future classes, and in life!
- We write $X \sim N(\mu, \sigma^2)$. We also call it **Gaussian** distributed.
- Expected value (location parameter) and variance (squared scale parameter)

$$E(X) = \mu \quad \text{and} \quad \text{Var}(X) = \sigma^2.$$

- If $X \sim N(0, 1)$, the distribution of X is said to be **standard normal**.
- Nine random draws in R: `rnorm(n=9, mean=2, sd=1)`

PDF of $X \sim N(0, 1)$, $Y \sim N(2, 1)$, $Z \sim N(0, 3)$