Continuous random variables

MATH/STAT 394: Probability I
Summer 2021 A Term

Introduction to Probability
D. Anderson, T.Seppäläinen, B. Valkó

§ 3.1

Aaron Osgood-Zimmerman

Department of Statistics

Logistics

- Two typos in the HW2 solutions (now fixed and re-uploaded)
  - First, problem 1.2, initially the solutions said B and C must be the empty set, but there are additional valid cases:

    $$P(A \cup (B \cup C)) = P(A \backslash (B \backslash C))$$

    when we have that the equality holds when B and C are both subsets of A and either $B = C$ or B is the empty set.
  - Secondly, problem 12.4 asked for $P(C^c \cap D^c)$ but the solution solved for $P(B^c \cap D^c)$
- Homework 3 is due Friday at 11:59pm
- Midterm will be a 24hr take home (+1.25 hrs) starting this Friday (July 9th) after lecture (10:45am) PST and will be due July 10th at noon PST
- Midterm covers up material from the first two weeks (up through lecture 13 on classical RV and conditional independence)

Recap

- $X \sim \mathrm{Ber}(p)$ for $p \in [0,1]$ *(flip of a coin)*, if
$$\mathbb{P}(X=1)=p \quad P(X=0)=1-p$$
- $X \sim \mathrm{Bin}(n,p)$ for $n \in \mathbb{N} \setminus \{0\}$, $p \in [0,1]$, *(how many tails in n flips of a coin)* if
$$\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \quad k=0,1,\dots,n$$
- $X \sim \mathrm{Geom}(p)$ for $p \in [0,1]$, *(how many flips before a tail)* if
$$\mathbb{P}(X=k)=(1-p)^{k-1}p$$
- $X \sim \mathrm{Hypergeom}(N, N_A, n)$, *(getting k red balls after sampling without replacement n balls from an urn that contains N_A red balls and $N-N_A$ blue balls)*
$$\mathbb{P}(X=k)=\frac{\binom{N_A}{k}\binom{N-N_A}{n-k}}{\binom{N}{n}}, \quad k=0,1,\dots,n,$$
- $X \sim \mathrm{Poisson}(\lambda)$, *(observing k events, given that the average number of events (expected over whatever units were observed) is λ*
$$\mathbb{P}(X=k)=\frac{\lambda^k e^{-\lambda}}{k!} \quad k=0,1,\dots$$

Recap

Conditional independence

- A_1, A_2 are independent given B if

$$\mathbb{P}(A_1 \cap A_2 \mid B) = \mathbb{P}(A_1 \mid B)\mathbb{P}(A_2 \mid B)$$

- This implies naturally $\mathbb{P}(A_2 \mid A_1, B) = \mathbb{P}(A_2 \mid B)$

Practice solution

Practice

At a lottery, there are 10 out of 100 tickets that have prices.

1. Consider picking 5 tickets with replacements, what is the prob. that you get exactly 2 prizes? (Namely you pick a ticket, look if you win or not and repeat that 5 times)
2. Consider picking 5 tickets without replacement, what is the prob. that you get exactly 2 prizes?

Solution

- With replacement, you have a binomial r.v.. The probability that you win each time is 10/100, so the probability that you get exactly 2 prizes is

$$\binom{5}{2}\left(\frac{1}{10}\right)^2\left(\frac{9}{10}\right)^3 \approx 0.073$$

- Without replacement, you have a hypergeometric r.v. The probability that you win 2 prizes is

$$\frac{\binom{10}{2}\binom{90}{3}}{\binom{100}{5}} \approx 0.070$$

Practice solution

Practice

Roll a fair die twice, define

$$A = \{\text{first die is a 2 or a 3}\}, B = \{\text{4 appears at least once}\}$$

- Are A, B independent?
- Are A, B conditionally independent given that

$$C = \{\text{the sum of the dice is a 6}\}?$$

Solution

- $\mathbb{P}(A) = \frac{12}{36}$, $\mathbb{P}(B) = \frac{11}{36}$, $\mathbb{P}(A \cap B) = \frac{2}{36} \neq \frac{12 \cdot 11}{36^2} = \mathbb{P}(A)\mathbb{P}(B)$, so no they are not independent
- $C = \{(3,3), (2,4), (4,2), (5,1), (1,5)\}$
- Now

$$\mathbb{P}(A \mid C) = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} = \frac{2/36}{5/36} = \frac{2}{5} \quad \mathbb{P}(B \mid C) = \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(C)} = \frac{2/36}{5/36} = \frac{2}{5}$$

$$\mathbb{P}(A \cap B \mid C) = \frac{A \cap B \cap C}{\mathbb{P}(C)} = \frac{1/36}{5/36} = \frac{1}{5}$$

- So, $\mathbb{P}(A \mid C)\mathbb{P}(B \mid C) \neq \mathbb{P}(A \cap B \mid C)$: they are not conditionally indep. given C.

Outline

Continuous random variables

Discrete Random variables

Reminders

- A r.v. X is discrete if there exists a countable set $\mathcal{X} = \{k_1, k_2, k_3, \dots\}$ s.t.

$$\sum_{k \in \mathcal{X}} \mathbb{P}(X = k) = 1$$

- **Interpretation:** we assign non-zero weights to some elements in a finite/countable infinite set.
- The probability mass function (p.m.f.) of X, defined as

$$p_X : k \to \mathbb{P}(X = k)$$

  entirely characterizes the probability distribution of X.
- Namely it gives access to $\mathbb{P}(X \in B)$ for any $B \subseteq \mathbb{R}$ as

$$\mathbb{P}(X \in B) = \sum_{k \in \mathcal{X} \cap B} p_X(k)$$

Discrete random variables

Example

Roll a die 5 times, what is the prob. that the numbers of 4 you get is btw 3.2 and 5.6?

Solution

- Let S be the r.v. associated to the number of 4 you get in the 5 rolls
- $S \sim \text{Bin}(5, 1/6)$ and the possible values for S are $\mathcal{S} = \{0, 1, 2, 3, 4, 5\}$
- The prob. of interest is

$$\mathbb{P}(S \in [3.2, 5.6]) = \mathbb{P}(S = 4) + \mathbb{P}(S = 5) = p_S(4) + p_S(5) = \sum_{k \in \mathcal{S} \cap [3.2, 5.6]} p_S(k)$$

Continuous random variables

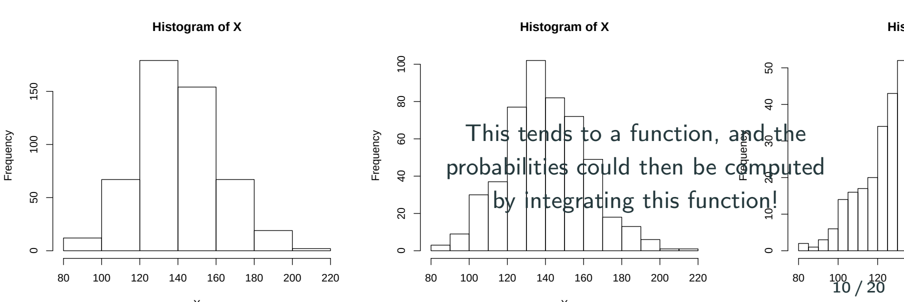
Motivation

- What about functions from Ω to $\mathbb{R}$ that can take any value e.g. in $[0, 1]$?
- For example, the amount of water you take every day to shower
- What we would like is an easy way to compute $\mathbb{P}(X \in [a, b])$ for any $a \leq b$
- What is analagous the p.m.f. in this case?

Histograms

Example
Let X be the height of a random 10-year-old child

- X is a *continuous* random variable, because it can take any real value.
- How could I represent the prob. dist?
- Say you have data of 500 10-year-olds. You can split the data into N groups
  $g_i = \{$childs with height between $h_i, h_i + \delta$ $\}$
  e.g. with $N = 100$, $h_i = i/100$, $\delta = 1/100$ and count the number in each group.
- This is called a *histogram*. Below histograms for smaller and smaller intervals.



This tends to a function, and the probabilities could then be computed by integrating this function!

Continuous random variables

Definition

A r.v. X is **continuous** *is there exists a function f s.t.*

$$\mathbb{P}(X \leq b) = \int_{-\infty}^{b} f(x)dx \quad \textit{for any } b \in \mathbb{R}$$

The function f is then called the ***probability density function*** *(p.d.f.) of X.*

Consequences:

- A function f is a p.d.f. of some r.v. if and only if

$$f(x) \geq 0 \quad \text{for all}^1\ x \in \mathbb{R} \qquad \text{(non-negative function)}$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1, \qquad \text{(normalized function)}$$

because $\int_{-\infty}^{+\infty} f(x)dx = \mathbb{P}(X \leq +\infty) = 1$ and $\mathbb{P}([a,b]) \geq 0$ for any $a \leq b$

[1] In fact, it can take a countable number of negative values. To be rigorous, from the definition, we have that there exists no $a < b$ such that $f(x) < 0$ for all $x \in [a,b]$. So there may be $c \in R$ s.t. $f(c) < 0$ but it is an isolated point that won't matter in the computations see later.

Continuous random variables

Consequences

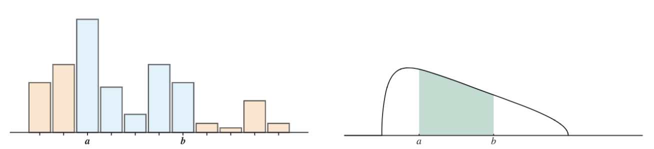
- Complement
$$\mathbb{P}(X > b) = \int_b^{+\infty} f(x)dx$$
because $\mathbb{P}(X > b) = 1 - \mathbb{P}(X \le b) = \int_{-\infty}^{+\infty} f(x)dx - \int_{-\infty}^{b} f(x)dx = \int_b^{+\infty} f(x)dx$
- Probability of a point is always zero!
$$\mathbb{P}(X = c) = \int_c^{c} f(x)dx = 0$$
because
$\mathbb{P}(X = c) = \mathbb{P}(X \le c) - \mathbb{P}(X > c) = \int_{-\infty}^{c} f(x)dx - \int_c^{+\infty} f(x)dx = \int_c^{c} f(x)dx$
- Probability of an interval for $a \le b$,
$$\mathbb{P}(a \le X \le b) = \int_a^{b} f(x)dx$$
because $\mathbb{P}(a \le X \le b) = \mathbb{P}(X \ge a) - \mathbb{P}(X \le b) = \mathbb{P}(X > a) - \mathbb{P}(X \le b)$ since $\mathbb{P}(X = a) = 0$

Continuous random variables

More generally, if X is a r.v. with p.d.f. f, then for any Borel set[2],

$$\mathbb{P}(X \in B) = \int_{x \in B} f(x)dx$$

Illustration



Left: p.m.f. of a discrete r.v., in blue prob. that $X \in [a, b]$

Right: p.d.f. of a continuous r.v., in blue prob. that $X \in [a, b]$

Figure from Introduction to probability, D. Anderson, T. Seppäläinen, B. Valkò

[2] Definition: a Borel set is a combination of intersections/unions of intervals

Probability density function

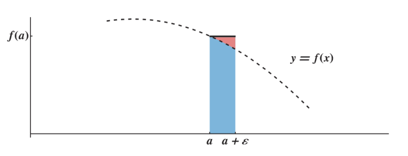
Beware:

- The value $f(x)$ of a p.d.f. **is not** a probability!
- $f(x)$ only gives probabilities by integration.
- Fundamentally it can measure infinitesimal probabilities as follows

Lemma
Suppose that a r.v. X has a continuous p.d.f. f, then for small $\varepsilon > 0$,

$$\mathbb{P}(a \le X \le a+\varepsilon) \approx f(a)\varepsilon$$

Proof Comes from the fact that $\lim_{\varepsilon\to 0} \frac{\mathbb{P}(a\le X\le a+\varepsilon)}{\varepsilon} = \lim_{\varepsilon\to 0} \int_a^{a+\varepsilon} f(x)dx = f(a)$.
The limit is true by continuity of f.



In blue, the true value of $\mathbb{P}(a \le X \le a+\varepsilon)$ in blue+red the approx. done.

Uniform random variable

We have already seen a classical continuous random variable

Definition

*A r.v. X has a **uniform distribution** on $[a, b]$ for $a < b$ if X has a p.d.f.*

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{if } x \notin [a, b] \end{cases}$$

We denote it $X \sim \mathrm{Unif}([a, b])$.

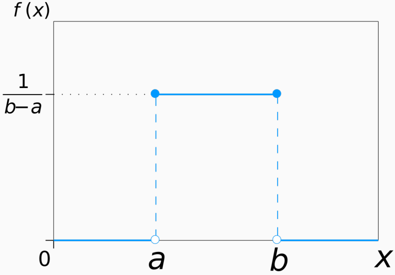
Sanity check

Does the p.d.f. of a uniform r.v. satisfy the conditions of being a r.v.?

- $f(x) \geq 0$ for all x
- $\int_{-\infty}^{+\infty} f(x)dx = \int_a^b \frac{1}{b-a}dx = 1$

Uniform random variable

Illustration



p.d.f. of $X \sim \mathrm{Unif}([a,b])$

image taken from Wikipedia

Note:

Do we retrieve our intuition about uniform experiments on an interval?

Yes, for $a \leq c \leq d \leq b$,

$$\mathbb{P}(c \leq X \leq d) = \int_c^d f(x)dx = \int_c^d \frac{1}{b-a}dx = \frac{d-c}{b-a} \left(= \frac{\mathrm{size}([c,d])}{\mathrm{size}([a,b])} \right)$$

Exponential Random variable

Motivation

- We have seen that the geometric distribution models the number of trials before getting a positive outcome.
- We might want a similar r.v. but for continuous time to model for example the time to wait for a bus to come?

Definition

A r.v. X has an ***Exponential*** *dist. with parameter $\lambda > 0$ if it has a p.d.f. of the form*

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \textit{if } x \geq 0 \\ 0 & \textit{if } x < 0 \end{cases}$$

We denote it $X \sim \mathrm{Exp}(\lambda)$.

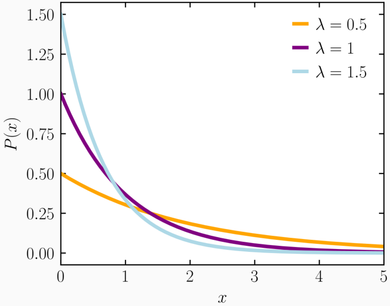
Sanity check

Does the p.d.f. of an exponential r.v. satisfy the conditions of being a r.v.?

- $f(x) \geq 0$ for all x
- $\int_{-\infty}^{+\infty} f(x)dx = \int_0^{+\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^{+\infty} = 0 - (-1) = 1$

Exponential distribution

Illustration



p.d.f. of $X \sim \text{Exp}(\lambda)$

image taken from Wikipedia

Exponential distribution

Exercise

Suppose the length of a phone call in minutes is modeled by an exponential r.v. with param. 1/10.

What is the prob. that the phone call lasts more than 8 min?

Solution Denote $X \sim \mathrm{Exp}(0.1)$ the length of the phone call, then

$$\mathbb{P}(X \geq 8) = \int_8^{+\infty} 0.1 \cdot e^{-0.1x} dx = \left[-e^{-0.1x}\right]_8^{+\infty} = e^{-8/10} \approx 0.45$$

Practice next lecture

Practice

Determine if there exist some values a, b, c such that the following functions satisfy the conditions to be a p.d.f.

$$f_1(x) = \frac{a}{x^2+1}$$

$$f_2(x) = \begin{cases} b\cos(x) & \text{if } x \in [0, 2\pi] \\ 0 & \text{if } x \notin [0, 2\pi] \end{cases}$$

$$f_3(x) = \begin{cases} cx^{-4} & \text{if } x \geq 1 \\ 0 & \text{if } x \leq 1 \end{cases}$$