Chapter 3 Part 1: Random variables

Jess Kunke

MATH/STAT 394: Probability I (Summer 2022 A-term)

Outline

Outline

A philosophical point

My #1 goal: you know new things by the end of this class, not because I told or taught you but because you convinced yourself with some guidance from me

A philosophical point

My #1 goal: you know new things by the end of this class, not because I told or taught you but because you convinced yourself with some guidance from me

- Surprise 1: I hope that at the end of each lecture you have some (or sometimes many!) points of confusion

A philosophical point

My #1 goal: you know new things by the end of this class, not because I told or taught you but because you convinced yourself with some guidance from me

- Surprise 1: I hope that at the end of each lecture you have some (or sometimes many!) points of confusion
- Surprise 2: I don't want to completely clear up that confusion for you

A philosophical point

My #1 goal: you know new things by the end of this class, not because I told or taught you but because you convinced yourself with some guidance from me

- Surprise 1: I hope that at the end of each lecture you have some (or sometimes many!) points of confusion
- Surprise 2: I don't want to completely clear up that confusion for you
- Goal: help you to target areas of productive struggle
- Working through those points is the process of learning

A philosophical point

My #1 goal: you know new things by the end of this class, not because I told or taught you but because you convinced yourself with some guidance from me

- Surprise 1: I hope that at the end of each lecture you have some (or sometimes many!) points of confusion
- Surprise 2: I don't want to completely clear up that confusion for you
- Goal: help you to target areas of productive struggle
- Working through those points is the process of learning
- When you ask me questions, I often try to ask *you* the right questions so that you can answer the question yourself
- I try to design lectures and choose homework problems to give you a chance to work through your own learning process: figuring out the things you're confused about and identifying things you didn't realize you didn't understand

A philosophical point

Why?

A philosophical point

Why?

- You learn math and statistics more deeply
- You can tackle new problems better
- You can tackle problems that are only partially defined better (e.g. research or on-the-job problems)
- You learn other skills that I can't directly teach you
- You gain more confidence
- You learn how to learn more independently
- I even think you might have more fun this way because you're steering

Outline

Midterm logistics and rules

- Next week Friday July 8, 8:30-9:30am, in CMU 230 (our usual classroom)
- What I will give you: printed instructions and blank paper for you to write on
- What you should bring: a pencil or pen (perhaps multiple)
- What you can also bring: a half sheet of paper (cut an 8.5x11 sheet in half) with whatever handwritten notes you want to add to it on both sides
- Closed book: no calculators, computers, phones, textbook, notes, etc.

Midterm instructions

- Make sure to show your work, show your steps/logic
  - This is somewhat subjective as is life, but use your judgment
- Use notation from class, and define any variables/symbols that you use in your solution that were not in the problem (such as variables for particular events)
- For some of the problems the instructions will say to fully simplify your answer
- For the rest, you do not need to finish simplifying once you have plugged in the numbers, but you do need to have shown your work justifying why that calculation would lead to the correct answer
  - And sometimes simplifying can help you check your work

Tips for during the midterm exam

- Breathe- you did you preparation, and you know things!
- Before you start, quickly skim through the whole exam first to get a sense for what problems are easier/harder, how much you have to do, etc.
- Don't necessarily solve them in order
- Make sure to give some time to each problem
- Budget a couple minutes at the end so that if there are any problems you didn't have time to finish, you can write down what you know and some of the steps you think will be involved
- Make your work legible and your reasoning clear, but don't get too hung up making it perfect

Tips for preparing for the midterm exam

- The problems will be similar to the HW1 and HW2 problems
- Review the homework
  - Go through the solutions to understand what you missed
  - Try doing them again, fresh
  - Try similar problems from the textbook or that you make up
- Review the lecture slides/notes, both results and examples
- Answer the Chapter 1 and Chapter 2 recap questions I put in the slides (Chapter 2 recap is today); these will help refine your understanding
- About enumerating outcomes
  - On the midterm, you will not have time to solve problems by enumerating all possible outcomes in the sample space
  - The goal is to learn how to use the tools we covered in class to solve these more efficiently

Upcoming deadlines and mid-course survey

- HW 2 due tonight (Fri July 1)
- Lectures: start Chapter 3 today, finish Chapter 3 next Wed July 6
- HW 3 (on Chapter 3) will be posted today, due Mon July 11 (after midterm)
- Midterm (Chapters 1-2) on Fri July 8

OH today 3-4 pm zoom

Anonymous mid-course feedback survey: click here or see Ed Discussion

Outline

Review of Chapter 2: for your practice

- Define or write out in math each of the following, and give an example of how you use each one in a problem (teach yourself fresh what these things are and how to use them)
  - Conditional probability
  - Multiplication rule
  - Law of total probability
  - Bayes' formula
  - (Mutual and pairwise) independence of events
  - Random variable (RV)
  - Discrete random variable
  - Probability mass function
  - Conditional independence
  - Independence of discrete random variables
- Give two examples of experiments that can be modeled as a series of independent trials (you will not be asked to prove whether random variables are identically distributed on the midterm)
- How do you prove that two random variables are independent? Not independent?
- Write the pmfs of the discrete distributions we covered and apply them to a simple example.

Outline

Recall: discrete random variables

- A random variable (RV for short) X is discrete if $\exists$ a finite or countably infinite set $\{k_1, k_2, ...\}$ of real numbers such that $\sum_i P(X = k_i) = 1$
    - This set is the set of possible values of X

Recall: discrete random variables

- A random variable (RV for short) X is discrete if $\exists$ a finite or countably infinite set $\{k_1, k_2, \ldots\}$ of real numbers such that $\sum_i P(X = k_i) = 1$
  - This set is the set of possible values of X
- The probability mass function (pmf) of a discrete random variable X is the function p (or p_X) defined by $p(k) = P(X = k)$ for all possible values k of X

Recall: discrete random variables

- A random variable (RV for short) X is discrete if $\exists$ a finite or countably infinite set $\{k_1, k_2, ...\}$ of real numbers such that $\sum_i P(X = k_i) = 1$
  - This set is the set of possible values of X
- The probability mass function (pmf) of a discrete random variable X is the function p (or p_X) defined by $p(k) = P(X = k)$ for all possible values k of X
  - The pmf completely determines the probability distribution of X: it tells us the probability of every possible value of X
  - Therefore, $\sum_k p(k) = 1$

Recall: discrete random variables

- A random variable (RV for short) X is discrete if $\exists$ a finite or countably infinite set $\{k_1, k_2, \ldots\}$ of real numbers such that $\sum_i P(X = k_i) = 1$
  - This set is the set of possible values of X
- The probability mass function (pmf) of a discrete random variable X is the function p (or p_X) defined by $p(k) = P(X = k)$ for all possible values k of X
  - The pmf completely determines the probability distribution of X: it tells us the probability of every possible value of X
  - Therefore, $\sum_k p(k) = 1$
  - How many probabilities do you have to define for a random variable that represents the number of heads in 5 tosses of the same coin?

list/table

$P(X=0) = \_\_\_$

$P(X=1) = \_\_\_$

$P(X=2) = \_\_\_$

$\vdots$

$Y \in \{\underline{0, 1, 2, 3, 4}, 5\}$

$P(\ ) = 0.75$ $\quad$ $P \doteq 0.25$

piecewise function

$$P(X=k) = \begin{cases} \_\_\_ & k=0 \\ \_\_\_ & k=1 \\ \vdots & \vdots \end{cases}$$

Degenerate random variables

$0,$

Definition

A random variable is **degenerate** if $\exists\ b \in \mathbb{R}$ such that $P(X = b) = 1$.

Example

Consider drawing a number N uniformly at random from the interval $[0, 10]$.

- Define a random variable X such that X is degenerate.

try defining a other X.

$$X = \begin{cases} 1 & N > 0 \\ 0 & N = 0 \end{cases}$$

$$P(N=k) = \begin{cases} 0 & k = 0, 1, \ldots \end{cases}$$

$$P(N=0) = 0.$$

$$X = \begin{cases} 1, & N \text{ is not an integer} \\ 0 & N \text{ is an integer} \end{cases}$$

$$P(N=0) + P(N=1) + \ldots + P(N=10)$$

$$\boxed{P(X=0)} = P(N \text{ is an int.}) = P(N=0,1,2,\ldots,10)$$

Continuous random variables: Motivation

- What about random variables (functions from Ω to $\mathbb{R}$) that can take uncountably infinitely many values?
  - e.g. temperature, height, time until an event occurs, ...
- What we would like is an easy way to compute $P(X \in [a, b])$ for any $a \leq b$
- Is there something like the pmf in this case?

Motivation

Example

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

- X is a *continuous* random variable, because the set of its possible values is uncountable
- How could I represent the probability dist?

Motivation

Example

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

- X is a *continuous* random variable, because the set of its possible values is uncountable
- How could I represent the probability dist?
- Suppose you have data on $N = 500$ 10-year-olds; split the data into n groups

  $$g_i = \{\text{childs with height between } h_i, h_i + \delta\ \} \quad \text{for e.g. } h_i = i/100,\ \delta = 1/100$$

  and count the number in each group
- This is called a *histogram*

Motivation

Example

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

- X is a *continuous* random variable, because the set of its possible values is uncountable
- How could I represent the probability dist?
- Suppose you have data on $N = 500$ 10-year-olds; split the data into n groups

  $$g_i = \{\text{childs with height between } h_i, h_i + \delta\ \} \quad \text{for e.g. } h_i = i/100,\ \delta = 1/100$$

  and count the number in each group
- This is called a *histogram*
- Increase n and $N \to$ histogram tends to a function, which could be integrated to give us probabilities

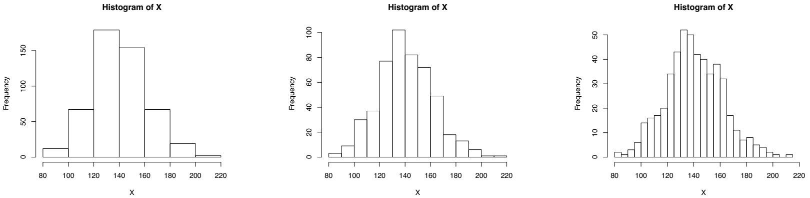


Figure: Histograms of decreasing bin width

Probability density function

Definition

A random variable X is called **continuous** is there exists a function f s.t.

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

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

Probability density function: properties

Definition for reference: $P(X \leq b) = \int_{-\infty}^{b} f(x)dx \quad$ for all $b \in \mathbb{R}$

- The probability of a single point is always zero!

$$P(X = c) = \int_{c}^{c} f(x)dx = 0$$

Probability density function: properties

Definition for reference: $P(X \leq b) = \int_{-\infty}^{b} f(x)dx$ for all $b \in \mathbb{R}$

- The probability of a single point is always zero!

$$P(X = c) = \int_{c}^{c} f(x)dx = 0$$

- Complement:

$$P(X > b) = \int_{b}^{+\infty} f(x)dx$$

Proof: HW 3

Probability density function: properties

Definition for reference: $P(X \leq b) = \int_{-\infty}^{b} f(x)dx$ for all $b \in \mathbb{R}$

- The probability of a single point is always zero!

$$P(X = c) = \int_c^c f(x)dx = 0$$

- Complement:

$$P(X > b) = \int_b^{+\infty} f(x)dx$$

  Proof: HW 3

- Notice that $P(X \geq b) = P(X > b) + \underbrace{P(X = b)}_{=0} = P(X > b)$

Probability density function: properties

Definition for reference: $P(X \leq b) = \int_{-\infty}^{b} f(x)dx$ for all $b \in \mathbb{R}$

- The probability of a single point is always zero!

$$P(X = c) = \int_{c}^{c} f(x)dx = 0$$

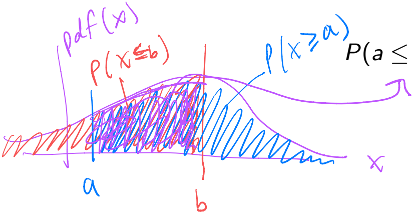
- Complement:

$$P(X > b) = \int_{b}^{+\infty} f(x)dx$$

  Proof: HW 3

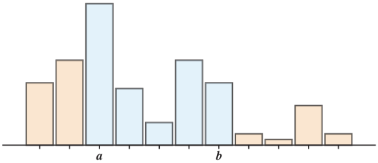
- Notice that $P(X \geq b) = P(X > b) + P(X = 0) = P(X > b)$
- The probability that X lies in an interval $[a, b]$ is

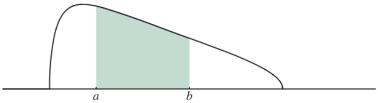
$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$



pmf vs pdf

Let's compare and contrast these pmf and pdf functions:





Left: pmf p_D of a discrete RV D

- Column at position a has height $p_D(a)$
- Sum of heights of blue bars $=$ $P(a \leq D \leq b)$ Σ
- $p_D(a) = P(D = a) \geq 0$

Right: pdf p_C of a continuous RV C

- Column at position a has height $p_C(a)$
- Sum of heights of green bars $=$ $P(a \leq C \leq b)$ $\int$
- $p_C(a)$ is not a probability! $p_C(a)$ generally > 0 while $P(C = a) = 0$

Example: Uniform distribution

We have already seen this classical continuous random variable, and now let's formalize it

$[2,5] = \{$all real numbers b/w 2 & 5$\}$

Example

Draw a number N uniformly at random from the interval $[2, 5]$.

- Define the pdf for this random variable N. What is $P(N > 3)$?
- Then generalize to the interval $[a, b]$ for any $a, b \in \mathbb{R}$, $a < b$.

$P(N > 3) = ?$

$f_N(n)$

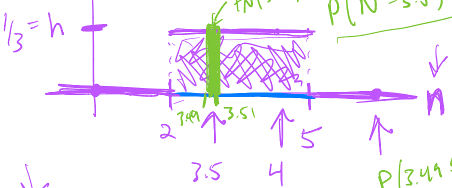
pdf(N)

$1 = A = hw = h \cdot 3 \Rightarrow h = 1/3$.

$1/3 = h$

$f_N(3.5) \neq 0$

$P(N = 3.5) = 0$



$P(3.49 \le N \le 3.51) \neq 0$

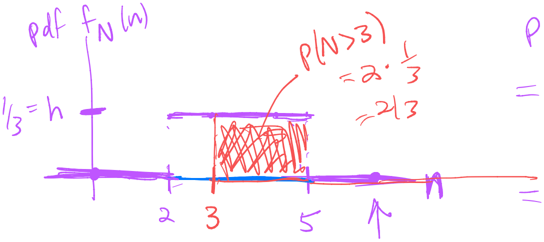
$P(N \in [2,5]) = 1$

$= P(2 \le N \le 5)$

$= \int_2^5 f_N(n)\,dn$

$=$ area under f_N from 2 to 5

$$f_N(n) = \begin{cases} 0 & n < 2 \\ 1/3 & 2 \le n \le 5 \\ 0 & n > 5 \end{cases}$$



$$f_N(n) = \begin{cases} 0 & n<2 \\ 1/3 & 2 \le n \le 5 \\ 0 & n>5 \end{cases}$$

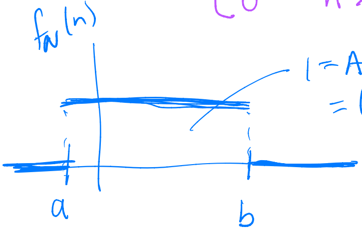
$P(N>3)$

$$= \int_3^\infty f_N(n)\,dn$$

$$= \int_3^5 \frac{1}{3}\,dn + \cancel{\int_5^\infty 0 \cdot dn}^{=0}$$

$$= \frac{n}{3}\Big|_{n=5} - \frac{n}{3}\Big|_{n=3} \quad \text{or} \quad \left[\frac{n}{3}\right]_{n=3}^{n=5}$$

$$= \frac{5}{3} - \frac{3}{3} = \frac{2}{3}.$$



$\Rightarrow h = \frac{1}{b-a}$

$a<b$

$$f_N(n) = \begin{cases} 0 & n<a \\ 1/(b-a) & a \le n \le b \\ 0 & n>b. \end{cases}$$

Notice that $\frac{1}{b-a}$ is not necessarily <1.

$[2, 2.1] \Rightarrow \frac{1}{0.1} = 10$

$$P(N \ge 3) = P(N>3) + \underbrace{P(N=3)}_{=0} = P(N>3)$$

b/c N is a continuous R.V.

Example: Uniform distribution

We have already seen this classical continuous random variable, and now let's formalize it

Example

Draw a number N uniformly at random from the interval $[2, 5]$.

- Define the pdf for this random variable N. What is $P(N > 3)$?
- Then generalize to the interval $[a, b]$ for any $a, b \in \mathbb{R}$, $a < b$.

Definition

A continuous RV X is **uniform** on $[a, b]$ for $a < b$ if X has the pdf

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

We write that $X \sim \text{Unif}([a, b])$. $X \sim Unif(a, b)$

Note: this is one type of uniform distribution. E.g. what about shooting an arrow at a target?

When is a function a probability density function?

How do we prove that some function f is a probability density function?

When is a function a probability density function?

How do we prove that some function f is a probability density function?

A function f is a pdf for some random variable out there (we may not have one in mind) if and only if f satisfies the following conditions:

1. $f(x) \geq 0 \ \forall x \in \mathbb{R}$,
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

When is a function a probability density function?

How do we prove that some function f is a probability density function?

A function f is a pdf for some random variable out there (we may not have one in mind) if and only if f satisfies the following conditions:

1. $f(x) \geq 0 \ \forall x \in \mathbb{R}$,
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

Example

Prove that the uniform distribution is a pdf.

$$f_N(n) = \begin{cases} 1/(b-a) & a \leq n \leq b \\ 0 & \text{else} \end{cases}$$

i) $f_N(n) = \frac{1}{b-a}$ or 0, & both are ≥ 0. ✓

ii) $\int_{-\infty}^{\infty} f_N(n)\,dn = \int_a^b \frac{1}{b-a}\,dn = \left[\frac{n}{b-a}\right]_{n=a}^{n=b} = \frac{b}{b-a} - \frac{a}{b-a}$

$$= \frac{b-a}{b-a} = 1. \checkmark$$

yes. $f_N(n)$ is a pdf.

Exponential random variable

Motivation

- Previously: the geometric distribution models the number of trials before a success (e.g. number of rolls until a 6 appears)
- Can we define a similar RV but for continuous time to model how long we wait for a bus to arrive?

Exponential random variable

Motivation

- Previously: the geometric distribution models the number of trials before a success (e.g. number of rolls until a 6 appears)
- Can we define a similar RV but for continuous time to model how long we wait for a bus to arrive?

Definition

A RV X is an **exponential** RV with parameter $\lambda > 0$ if it has a pdf of the form

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

We write that $X \sim \text{Expo}(\lambda)$.

Exponential random variable

Motivation

- Previously: the geometric distribution models the number of trials before a success (e.g. number of rolls until a 6 appears)
- Can we define a similar RV but for continuous time to model how long we wait for a bus to arrive?

Definition

A RV X is an **exponential** RV with parameter $\lambda > 0$ if it has a pdf of the form

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

We write that $X \sim \text{Expo}(\lambda)$.

Sanity check: Does the pdf of an exponential RV satisfy the conditions of being ~~a RV~~? the pdf of a RV?

i) $f(x) = \lambda e^{-\lambda x}$ or 0, & $\lambda e^{-\lambda x} > 0 \quad \forall x \in \mathbb{R}$. ✓

ii) $\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} \lambda e^{-\lambda x}\,dx =$

ii) $\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} \underline{\lambda e^{-\lambda x}}\,dx =$

$$\frac{d}{dx} e^{-\lambda x} = -\lambda e^{-\lambda x}$$

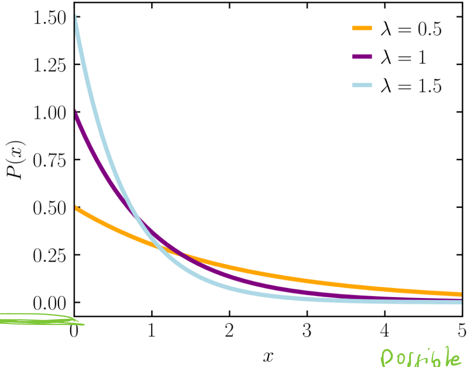
$$\left[-e^{-\lambda x}\right]_{x=0}^{x\to\infty}$$

$$= \lim_{x\to\infty}\left[-e^{-\lambda x}\right] - \left(-e^{-\lambda x}\right)\Big|_{x=0}$$

$$= 0 + 1 = 1. \checkmark$$

Exponential distribution

pdf



possible values

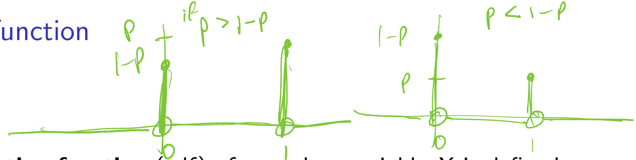
Figure: The exponential distribution Expo(λ) for different choices of λ. Image from wikipedia.

Outline

Motivation and definition

- Discrete RVs have a pmf while continuous RVs have a pdf
- Can we define a type of function that represents the probability distribution of any type of random variable?

Cumulative distribution function



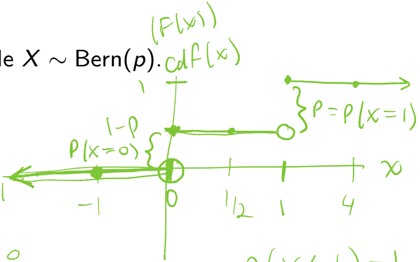
Definition

The **cumulative distribution function** (cdf) of a random variable X is defined by

$$F(s) = P(X \leq s)\ \forall\, s \in \mathbb{R}.$$

Example

Graph and define the cdf of a random variable $X \sim \text{Bern}(p)$.



$$\Rightarrow \quad X = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p \end{cases}$$

$$\text{pmf}(x) = P(X = x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \end{cases}$$

$$P(X \leq -1) = 0 \qquad P(X \leq 0) = \underbrace{P(X < 0)}_{0} + \underbrace{P(X = 0)}_{1-p} = 1-p$$

$$P(X \leq 1) = 1$$

we've now included all possible outcomes

$$P(X \leq 4) = 1$$

$$P(X \leq 1/2) = \underbrace{P(X \leq 0)}_{1-p} + \underbrace{P(0 < X \leq 1/2)}_{0} = 1-p$$

better way to visualize pmf:

pmf, pdf, cdf: more examples

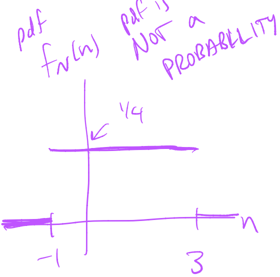
What other examples can we think of?

$N \sim \text{Unif}[-1, 3]$

1) define the pdf & graph it

$$f_N(n) = \begin{cases} 1/4 & -1 \le n \le 3 \\ 0 & \text{else} \end{cases}$$

2) how can we construct (define & graph) the cdf

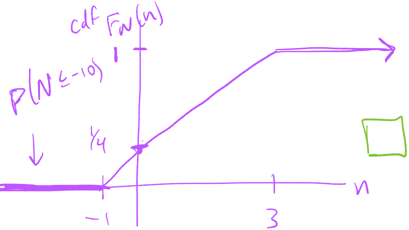


$$F_N(n) = P(N \le n) \, ? \qquad = \int_{-\infty}^{n} f_N(x)\,dx = \begin{cases} 0 & n < -1 \\ \boxed{\frac{n+1}{4}} & -1 \le n \le 3 \\ 1 & n > 3 \end{cases}$$

$$P(N \le 0) = \int_{-\infty}^{0} f_N(n)\,dn = \int_{-1}^{0} \frac{1}{4}\,dn$$

$$= \left[\frac{n}{4}\right]_{n=-1}^{n=0} = \frac{0}{4} - \left(-\frac{1}{4}\right) = \frac{1}{4}$$

$$\boxed{} : \int_{-1}^{n} \frac{1}{4}\,dx = \left[\frac{x}{4}\right]_{x=-1}^{x=n} = \frac{n}{4} - \left(-\frac{1}{4}\right) = \frac{n+1}{4}$$



pmf, pdf, cdf: more examples

What similarities, differences, patterns did you notice?

Properties of the cdf

- The cdf is always strictly non-decreasing: $F(a) \leq F(b)$ for all $a, b \in \mathbb{R}$ with $a < b$.
  - Note: a function is monotonic if it is either strictly non-decreasing or strictly non-increasing

Properties of the cdf

- The cdf is always strictly non-decreasing: $F(a) \leq F(b)$ for all $a, b \in \mathbb{R}$ with $a < b$.
  - Note: a function is monotonic if it is either strictly non-decreasing or strictly non-increasing
- $\lim_{t \to -\infty} F(t) = 0, \quad \lim_{t \to +\infty} F(t) = 1$

Properties of the cdf

- The cdf is always strictly non-decreasing: $F(a) \leq F(b)$ for all $a, b \in \mathbb{R}$ with $a < b$.
  - Note: a function is monotonic if it is either strictly non-decreasing or strictly non-increasing

- $\lim_{t \to -\infty} F(t) = 0, \quad \lim_{t \to +\infty} F(t) = 1$

- The cdf is right-continuous
  - For a function F, we denote, if it exists,

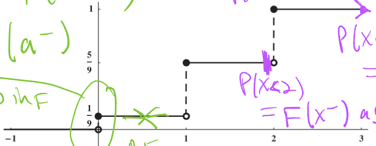
$$F(t+) = \lim_{s \to t^+} F(s) := \lim_{\substack{s \to t \\ s > t}} F(s) \qquad \text{Right limit (from above)}$$

$$F(t-) = \lim_{s \to t^-} F(s) := \lim_{\substack{s \to t \\ s < t}} F(s) \qquad \text{Left limit (from below)}$$

  - For any given $t \in \mathbb{R}$, $F(t) = F(t+)$

e.g. $X = \#$ heads from 2 coin flips

$P(X=a) = P(X \leq a) - P(X<a)$
$= F(a) - F(a^-)$
$=$ size of jump in F at a



$P(X=1) = P(X \leq 1) - P(X<1) = 5/9 - 1/9 = 4/9$

Figure: cdf of $X \sim \text{Bin}(2, 2/3)$. Black: right limit. White: left limit.

Going from pdf or pmf to cdf

We can summarize the relationships we've seen so far:

Discrete: pmf to cdf
Sum all the probabilities of possible values of X that are less than or equal to s:

$$F(s) = P(X \leq s) = \sum_{k:\, k \leq s} P(X = k)$$

Continuous: pdf to cdf
Integrate the density function over the interval $(-\infty, s)$:

$$F(s) = P(X \leq s) = \int_{-\infty}^{s} f(x)dx$$

Going from cdf to pdf or pmf

- We saw how to go from pmf or pdf $\rightarrow$ cdf
- What about the other way around?

– take derivative (continuous case)

– look at graph, look for jumps (discrete case)

Going from cdf to pdf or pmf

- We saw how to go from pmf or pdf $\rightarrow$ cdf
- What about the other way around?

Example

Consider a random variable Z with cdf

$$F(z) = \begin{cases} 0 & z < 1, \\ \frac{1}{7} & 1 \leq z < 2, \\ \frac{4}{7} & 2 \leq z < 5, \\ \frac{5}{7} & 5 \leq z < 8, \\ 1 & z \geq 8. \end{cases}$$

Graph this cdf and define the corresponding pmf $p_Z(z)$ or pdf $f_Z(z)$.

Do discrete and continuous RVs partition the space of possible RVs?

- If F is piece-wise constant
  $\Longrightarrow$ it is the cdf of a **discrete RV**
- If F is continuous
  $\Longrightarrow$ it is the cdf of a **continuous RV**
- If F is discontinuous and not piece-wise constant
  $\Longrightarrow$ neither discrete nor continuous RV
  - but we can still compute probabilities using the cdf
  - e.g. mixtures of distributions

The cdf exists for **any** RV

Why did we introduce the cdf?

Theoretical reason

- We only need $P(X \leq t)$ for any t to compute any prob. measure
- Therefore the cdf is sufficient for our purposes

Why did we introduce the cdf?

Theoretical reason

- We only need $P(X \leq t)$ for any t to compute any prob. measure
- Therefore the cdf is sufficient for our purposes

Practical reason

- The cdf itself is a prob. so we can use classical rules of prob. to manipulate it
- On the other hand the pdf is just a function and sometimes it is not practical or does not exist

Why did we introduce the cdf?

Example

Let $X \sim \text{Expo}(\lambda)$, $Y \sim \text{Expo}(\mu)$ be independent.
What is the pdf of $M = \min(X, Y)$?
Recall that $F_X(t) = 1 - e^{-\lambda t}$.

- We have that the event $\{\min\{X, Y\} > t\}$ is equivalent to the event $\{X > t\} \cap \{Y > t\}$.
- Therefore

$$
\begin{aligned}
1 - F_M(t) = P(\min(X, Y) > t) &= P(X > t, Y > t) \\
&= P(X > t)P(Y > t) \qquad \text{(by independence)} \\
&= e^{-\lambda t} e^{-\mu t} = e^{-(\lambda+\mu)t}
\end{aligned}
$$

- So the pdf of M is given as

$$f_M(x) = F_M'(x) = (1 - e^{-(\lambda+\mu)x})' = (\lambda + \mu)e^{-(\lambda+\mu)x}$$

- In fact we recognize that $M \sim \text{Expo}(\lambda + \mu)$.

Tips on mastering probability distributions

Wikipedia pages on probability distributions are a great resource for practice!

- Check out the distributions from class (binomial, uniform, exponential, etc.)
- Shows pmf/pdf, cdf, and lots of other properties
- Presents definitions and applications, connections to other distributions, and sometimes some history
- You can explore some new distributions you haven't seen before too

Outline

Expectation

Motivation

- Given a RV, we have numerous tools to compute probabilities
- We said that we also sometimes want to know what kind of result we expect on average, a “typical value” for a given RV
- e.g. if you flip a coin n times, what is the average number of tails you should get?
- In probability, this “average” number is called an **expectation** and it is a central object

Intuition

Example

At a casino, suppose

- you lose 1\$ 90% of the time,
- you gain 10\$ 9% of the time, and
- you gain 100\$ 1% of the time.

What is your expected net gain?

Intuition

Example

At a casino, suppose

- you lose 1\$ 90% of the time,
- you gain 10\$ 9% of the time, and
- you gain 100\$ 1% of the time.

What is your expected net gain?

- First understand that the average is a **number** not a probability
- Then

$$\text{expected net gain} = \underbrace{(-1)}_{\text{net gain}} \cdot \underbrace{\frac{90}{100}}_{\text{frequency}} + 10 \cdot \frac{9}{100} + 100 \cdot \frac{1}{100} = 1$$

Expectation

Definition

The **expectation** or **mean** of a discrete random variable Y is defined by

$$E(Y) = \sum_k kP(X = k).$$

Expectation is often written with square brackets, $E[Y]$.