9-25: Continuous RVs; CSCI 3022 Fall 19

Opening Example:

A factory makes parts for a medical device company. 6% of those parts are defective. For each one of the problems below:

(i.) Define an appropriate random variable for the experiment.

(ii.) Give the values that the random variable can take on.

(ii.) Find the probability that the random variable equals 2.

(iv.) State any assumptions you need to make.

Problems:

1. Out of 10 parts, X are defective.
2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.
3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

Discrete Distributions Example

6% of those parts are defective. → independent & identically

1. Out of 10 parts, X are defective.

(i.) r.v.: $X \sim bin(10, .06)$

(ii.) Values of r.v.: $X \in \{0, 1, 2, \cdots 10\}$

(iii.) $P(X = 2)$: $\binom{10}{2} (.06)^2 \cdot (.94)^8$

# of ways the order of 10 parts could have 2x D

(iv.) Assumptions:

Discrete Distributions Example

6% of those parts are defective.

1. Out of 10 parts, X are defective.

(i.) r.v.:

$$X \sim bin(10, .06)$$

(ii.) Values of r.v.:

$$X \in \{0, 1, 2, \ldots, 10\}$$

(iii.) $P(X = 2)$:

$$\binom{10}{2} .06^2 .94^8$$

(iv.) Assumptions: Parts are *i.i.d.*

Discrete Distributions Example

6% of those parts are defective.

2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.

(i.) r.v.: $X \sim \text{Geom}(p)$

includes the first 'successful' trial

(ii.) Values of r.v.:

$X = \{0, 1, 2, \ldots, \infty\}$

(iii.) $P(X = 2)$: $.94^2 \cdot .06$

(iv.) Assumptions: i.i.d.

Discrete Distributions Example

6% of those parts are defective.

2. Upon observing an assembly line, X non-defective parts are observed before finding a defective part.

(i.) r.v.:

$$X + 1 \sim Geom(.06)$$

(ii.) Values of r.v.:

$$X \in \{0, 1, 2, \ldots, \infty\}$$

(iii.) $P(X = 2)$:

$$.94^2 .06^1$$

(iv.) Assumptions: Parts are *i.i.d.*

Discrete Distributions Example

6% of those parts are defective.

3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

(i.) r.v.: $X \sim Pois(10)$

(ii.) Values of r.v.: $X = \{0, 1, 2, \cdots, \infty\}$

(iii.) $P(X = 2)$: $f(x) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-10} 10^2}{2!}$ $\to 100$

(iv.) Assumptions: Poisson?

Discrete Distributions Example

6% of those parts are defective.

3. X is the number of defective parts made per day, where the rate of defective parts per day is 10.

(i.) r.v.:

$$X \sim Pois(10)$$

(ii.) Values of r.v.:

$$X \in \{0, 1, 2, \ldots, \infty\}$$

(iii.) $P(X = 2)$:

$$\frac{e^{-10} \cdot 10^2}{2!}$$

(iv.) Assumptions: Parts are... *Poisson*?

Announcements and Reminders

- Homework due Friday the 27th.
- Notebook day on Friday in class.

Last Time...: the blocks of discrete probability

1. Bernoulli: *one* binary outcome experiment. $f(x) = p^x (1-p)^{1-x}$
2. Binomial: binary outcome experiment success *count* in n tries.

   $$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

3. Geometric: Trials *until a success* of a binary outcome experiment.

   $$= (1-p)^{x-1} \cdot p$$

4. Negative Binomial: Trials until r binary outcome experiment *successes*.

   $$f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

5. Poisson: *counting* outcomes with a fixed rate λ.

Last Time...: the blocks of discrete probability

1. Bernoulli: *one* binary outcome experiment. $f(x) = p^x(1-p)^{1-x}$
2. Binomial: binary outcome experiment success *count* in n tries.
$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)}$$
3. Geometric: Trials *until a success* of a binary outcome experiment.
$$f(x) = (1-p)^{x-1}p$$
4. Negative Binomial: Trials until *r* binary outcome experiment *successes*.
$$f(x) = \binom{x-1}{r-1} p^r (1-p)^{(x-r)}$$
5. Poisson: *counting* outcomes with a fixed rate λ. $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Last Time...: the blocks of discrete probability

The underlying pieces of discrete RVs:

1. The random variable X takes inputs/events in the (discrete) sample space Ω and maps them to a (discrete) finite or infinite set of probability values $a_1, a_2, a_3, \ldots$.
2. We find probabilities in the probability mass function or probability density function
$$f(x) = P(X = x).$$
3. We can find cumulative probabilities or probability on ranges of outcomes in the cumulative density function
$$F(x) = P(X \leq x) = \boxed{\sum_{X \leq x} f(x).}$$

won't be a sum if cont's.

Continuous RVs

Many real-life random processes must be modeled by random variables that can take on continuous (non-discrete) values. Some example:

1. Peoples' heights: $X \in \{[1\text{ in}, 120\text{ in}]\}$
2. Final grades in a class: $X \in \{[0, 100]\}$
3. Time between people checking out at a store : $t \in \{(0, \infty)\}$

Continuous RVs

Many real-life random processes must be modeled by random variables that can take on continuous (non-discrete) values. Some example:

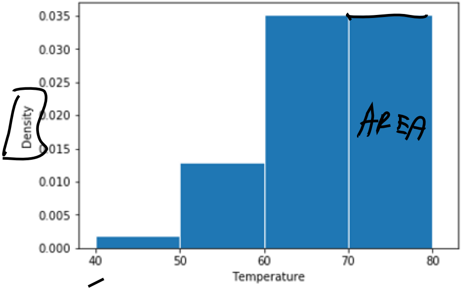
1. Peoples' heights: $X \in \{[0, 7.5ft]\}$
2. Final grades in a class: $X \in \{[0, 100]\}$
3. Time between people checking out at a store : $t \in \{[0, \infty]\}$

Store just closes;)

More precision!

Suppose your friend asks you what the high temperature will be today. They want to know the probability it will be between 70F and 80F, so they can wear shorts.

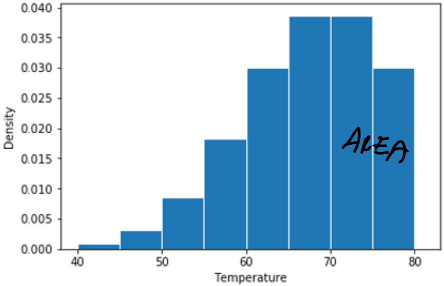
Probability:
Add up the share of outcomes between 70F and 80F!



More precision!

Suppose your friend asks you what the high temperature will be today. They want to know the probability it will be between 70F and 80F, so they can wear shorts.

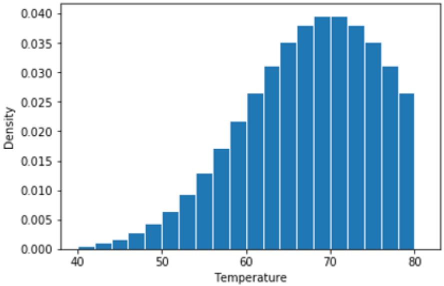
Probability:
Add up the share of outcomes between 70F and 80F!



More precision!

Suppose your friend asks you what the high temperature will be today. They want to know the probability it will be between 70F and 80F, so they can wear shorts.

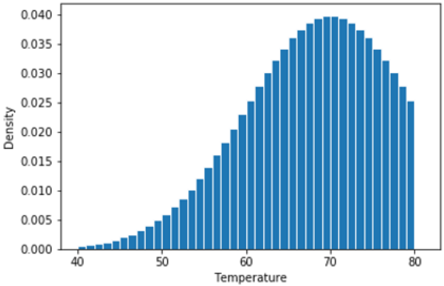
Probability:
Add up the share of outcomes between 70F and 80F!



More precision!

Suppose your friend asks you what the high temperature will be today. They want to know the probability it will be between 70F and 80F, so they can wear shorts.

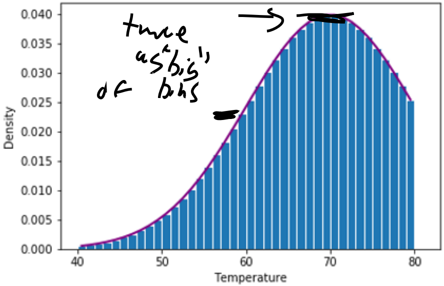
Probability:
Add up the share of outcomes between 70F and 80F!



More precision!

Suppose your friend asks you what the high temperature will be today. They want to know the probability it will be between 70F and 80F, so they can wear shorts.

Probability:
Add up the share of outcomes between 70F and 80F!

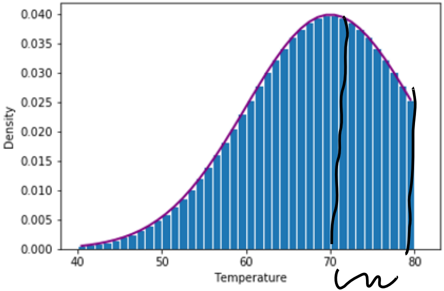


More precision!

Suppose your friend asks you what the high temperature will be today. They want to know the probability it will be between 70F and 80F, so they can wear shorts.

Probability:

Integrate up the share of outcomes between 70F and 80F!

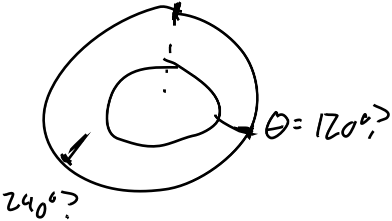


limit of the sum of the rectangles as we draw more.

Continuous Distributions

Example:

Consider the reference line connecting the valve stem on a tire to the center point. $0°$



Let X be the angle measured clockwise to the location of an imperfection. The pdf for X is:

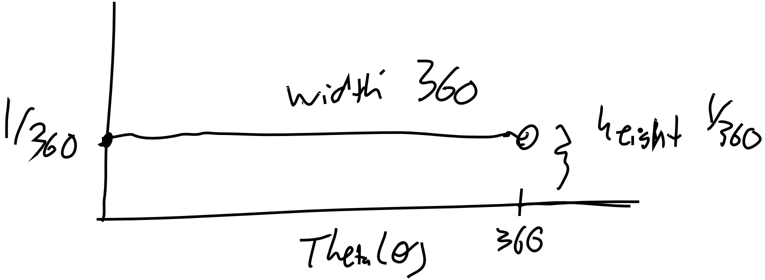
$$f(x) = \begin{cases} \frac{1}{360} & 0 \leq X < 360 \\ 0 & else \end{cases}$$

Continuous Distributions

Example, cont'd:

$$f(x) = \begin{cases} \frac{1}{360} & 0 \leq X < 360 \\ 0 & else \end{cases}$$

Graphically, the pdf of X is:



Continuous Distributions

$f(x) \geq 0$

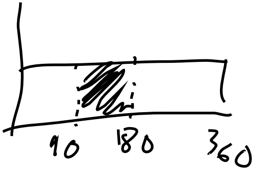
$\Rightarrow$ "sum" of $f(x) = 1$

Example, cont'd:

How can we show that:

1. the total area of the pdf of x is 1?

Area of rectangle: 1.



2. How do we calculate $P(90 \leq X \leq 180)$?

$\int_{90}^{180} f(x)\,dx = \frac{180-90}{360}$ width & 1/height $= 1/4$

3. What is the probability that the angle of occurrence is within 90 of the reference line? (The reference line is at 0 degrees.)

$P(0 \leq X \leq 90 \text{ OR } 270 \leq X \leq 360) = P(0 \leq X \leq 90) + P(270 \leq X \leq 360)$
$= 1/4 + 1/4$

Continuous Distributions

Example, cont'd:

How can we show that:

1. the total area of the pdf of x is 1?

$$\int_0^{360} f(x)\, dx = 1?$$

2. How do we calculate $P(90 \leq X \leq 180)$?

$$\int_{90}^{180} f(x)\, dx = \ldots?$$

3. What is the probability that the angle of occurrence is within 90 of the reference line? (The reference line is at 0 degrees.)

$$P(X < 90 \textbf{ OR } X > 270) = \int_0^{90} f(x)\, dx + \int_{270}^{360} f(x)\, dx = \ldots?$$

Uniform Distribution

$f(x) \neq P(X = x)$

The previous problem was an example of the uniform distribution.

Definition: *Uniform Distribution*

A continuous rv X is said to have a *uniform distribution* on the interval $[a, b]$ if the pdf of X is:

$$f(x) = \begin{cases} c = \frac{1}{b-a} & \text{from } a \text{ to } b \\ 0 & \text{else} \end{cases}$$

NOTATION: We write $X \sim U[a,b]$ to indicate that X is a uniform rv with lower bound a and upper bound b.

Uniform Distribution

The previous problem was an example of the uniform distribution.
Definition: *Uniform Distribution*
A continuous rv X is said to have a *uniform distribution* on the interval $[a, b]$ if the pdf of X is:

$$f(x) = \frac{1}{b-a}; \qquad x \in [a, b]$$

NOTATION: We write $\underline{X \sim U(a, b)}$ to indicate that X is a uniform rv with lower bound a and upper bound b.

Exponential Distribution

The family of exponential distributions provides probability models that are very widely used in engineering and science disciplines to describe time-to-event data.

It can be thought of as a continuous analogue to the Poisson distribution, but instead of events-per-time, it measure time-per-events.

Examples: 1) bus times
2) Checkout times.

Exponential Distribution

λ is still a rate: $\frac{\text{count}}{\text{unit time}}$

Definition: *Exponential Distribution*
A continuous rv X is said to have an *exponential distribution* with rate parameter λ if the pdf of X is:

$$f(x) = \lambda e^{-\lambda x} \qquad x \in (0, \infty)$$

NOTATION: We write $X \sim exp(\lambda)$ to indicate that X is an exponential rv with rate λ.

Exponential Distribution

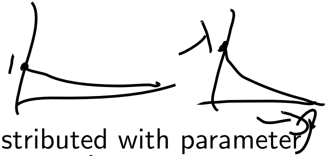
Definition: *Exponential Distribution*
A continuous rv X is said to have an *exponential distribution* with rate parameter λ if the pdf of X is:

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

NOTATION: We write $\underline{X \sim exp(\lambda)}$ to indicate that X is an exponential rv with rate λ.

Exponential Distribution

e^{-x}:



Example:

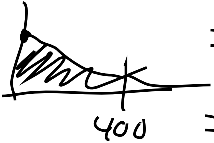
Suppose a light bulb's lifetime is exponentially distributed with parameter $\lambda = 1/1000$.

1. What are the units for λ?

$\frac{1}{1000} = \frac{\text{count}}{\text{time}} = \frac{\text{burnouts}}{\cancel{\text{days}}\ \text{hours}}$

2. What is the probability that the lifetime of the light bulb lasts less than 400 hours?

$P(X \le 400)$:



$= \int_0^{400} f(x)\,dx$

$= \int_0^{400} \lambda e^{\boxed{-\lambda x}}\,dx$

$u = -\lambda x$

$du = -\lambda\,dx$

$= \int -e^u\,du$

$= -e^u = \boxed{-e^{-\lambda x}}\Big|_{x=0}^{x=400}$

$= -e^{-\lambda \cdot 400} - -e^{0}$

$= \boxed{1 - e^{-2/5}} \approx 1/3$

3. What is the probability that the lifetime of the light bulb lasts more than 5 hours?

$P(X > 5) = \int_5^\infty f(x)\,dx = 1 - \int_0^5 f(x)\,dx$

$1 - P(X \le 5)$

$= -e^{-\lambda x}\Big|_5^\infty = 0 - -e^{-\lambda \cdot 5} = e^{-5/1000} \approx 1$

Exponential Distribution

Example:

Suppose a light bulb's lifetime is exponentially distributed with parameter $\lambda = 1/1000$.

1. What are the units for λ?
   Same as Poisson: outcomes per time; so maybe burnouts per hour?

2. What is the probability that the lifetime of the light bulb lasts less than 400 hours?

$$P(X < 400) = \int_0^{400} \lambda e^{-\lambda x} = -e^{-\lambda x}|_0^{400} = 1 - e^{2/5}$$

3. What is the probability that the lifetime of the light bulb lasts more than 5 hours?

$$P(X > 5) = \int_5^{\infty} \lambda e^{-\lambda x} = -e^{-\lambda x}|_5^{\infty} = 0 - -e^{1/2000} \approx 1$$

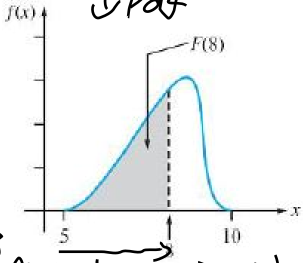
Cumulative Density Function

Definition: *Cumulative Density Function*
The *cumulative distribution function* (cdf) is denoted with $F(x)$. For a continuous r.v. X with pdf $f(x)$, $F(x)$ is defined for every real number x by:

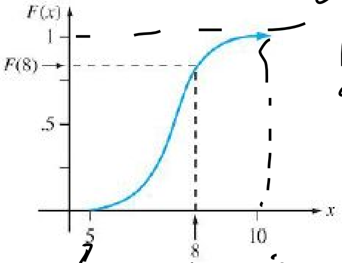
Same

$$\boxed{F(x) = P(X \le x)} = \int_{-\infty}^{x} f(t)\,dt$$

↓pdf



$\int_5^8 f(x)\,dx = F(8)$

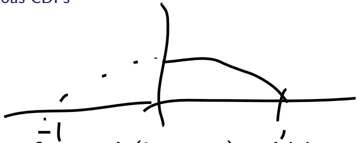


end at $F(x) \to 1$ as x grows.

0

x

Continuous CDFs



Example:
The distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous rv X with pdf

$$f(x) = \begin{cases} \frac{3}{2}(1-x^2) & 0 \leq X < 1 \\ 0 & else \end{cases}$$

1. What is the cdf of sales for any x?

$$F(x) = \int_0^x f(t)\,dt = \int_0^x \frac{3}{2} - \frac{3t^2}{2}\,dt = \frac{3t}{2} - \frac{t^3}{2}\Big|_0^x$$

$$= \frac{3x}{2} - \frac{x^3}{2}$$

2. Find the probability that X is less than .25?
3. X is greater than .75?
4. $P(.25 < X < .75)$?

Continuous CDFs

Example:

The distribution of the amount of gravel (in tons) sold by a particular construction supply company in a given week is a continuous rv X with pdf

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \leq X < 1 \\ 0 & else \end{cases}$$

1. What is the cdf of sales for any x?
   $F(x) = P(X \leq x) = \int_0^x \frac{3}{2}(1 - t^2)\,dt$
   $F(x) = \frac{3x}{2} - \frac{x^3}{2}$
2. Find the probability that X is less than .25? $F(.25)$
3. X is greater than .75? $1 - F(.75)$
4. $P(.25 < X < .75)$? $F(.75) - F(.25)$

Continuous CDFs

Wait, we've seen this before...
Recall: *The Fundamental Theorem of Calculus*.
Suppose F is an anti-derivative of f. Then:

1. $$\frac{d}{dx}\int_a^x f(t)\,dt = f(x);$$

   a.k.a.
   $$\frac{d}{dx}F(x) = f(x);$$

2. $$\int_a^b f(x)\,dx = F(B) - F(A).$$

Percentiles of a Distribution

Definition: The median $\tilde{x}$ of a continuous distribution is the 50th percentile or $.5$ quantile of the distribution.

How can we express this in terms of $f(x), F(x)$?

Notation:

Visually:

Percentiles of a Distribution

Definition: The median $\tilde{x}$ of a continuous distribution is the 50th percentile or $.5$ quantile of the distribution.

How can we express this in terms of $f(x), F(x)$?

Notation:

$\tilde{x}$ satisfies $F(\tilde{x}) = .5$, or

$$.5 = \int_{\infty}^{\tilde{x}} f(x)\,dx$$

Visually: