etherpad.wikimedia.org/p/312 for (anonymous) questions/comments!

Continuous Random Variables

CSE 312 24Su

Lecture 13

Fun fact!

You play a game where a fair coin is tossed until it comes up heads. The payoff is 2^n dollars, where n is the number of tosses. The expected value of this game is infinite, which is surprising because no one would realistically pay an extremely high entry fee to play. (convince yourself of this using the geometric distribution PMF and LOTUS!)

Logistics

- Please fill out **midterm feedback** form (closes tonight!)
- Some people have not taken the midterm, so do not discuss it yet
- HW3 grades will be released later today
- Midterm grades released early next week
- **HW4 released today evening**
  - Coding part

Zoo of Discrete Random Variables

Scenario: Negative Binomial

Run independent trials with probability p. How many trials do you need *until r successes*?

Example

You're playing a carnival game, and there are r little kids nearby who all want a stuffed animal. You can win a single game (and thus win one stuffed animal) with probability p (independently each time) How many times will you need to play the game before every kid gets their toy?

Try it

Run <u>independent</u> trials with probability $\boldsymbol{p}$. How many trials do you need *until* $\boldsymbol{r}$ *successes*?

X is the number of trials till (and including) the r'th success

What is the **support** of X?

What's the **PMF**?

i.e., what is the probability it takes exactly k trials till the r'th success?

Fill out the poll everywhere: pollev.com/cse312

Try it

Run <u>independent</u> trials with probability $\boldsymbol{p}$. How many trials do you need *until $\boldsymbol{r}$ successes*?

X is the number of trials till (and including) the r'th success

What is the **support** of X? $\Omega_X = \{r, r+1, r+2, \ldots\}$

What's the **PMF**?

i.e., what is the probability it takes exactly k trials till the r'th success?

Negative Binomial Analysis

Run independent trials with probability p
X is the number of trials till (and including) the r'th success

What's the **PMF**? Well how would we know $X = k$?

Negative Binomial Analysis

What's the **PMF**? Well how would we know $X = k$?

Of the first $k - 1$ trials, $r - 1$ must be successes.
And trial k must be a success.

1. We want **exactly $\boldsymbol{r - 1}$ of the first $\boldsymbol{k - 1}$ to be successes** – this sounds like a binomial! It's the $p_Y(r-1)$ where $Y \sim \mathrm{Bin}(k-1, r-1)$:
$\binom{k-1}{r-1}(1-p)^{k-1-(r-1)}p^{r-1} = \binom{k-1}{r-1}(1-p)^{k-r}p^{r-1}$

2. Multiply by p, probability **$\boldsymbol{k}$'th trial is success**

Total: $p_X(k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r$

Negative Binomial Analysis

X is the number of trials till we see r successes

To see r successes:

We do trials until we see success 1.

Then do trials until success 2.

…do trials until success r.

What's the **expectation** and **variance** (hint: linearity)?

How can we write X as a sum of random variables?

Fill out the poll everywhere: pollev.com/cse312

Negative Binomial Analysis

X is the number of trials till we see r successes

To see r successes:

We do trials until we see success 1.

Then do trials until success 2.

…do trials until success r.

The total number of flips is…the sum of geometric random variables!

Negative Binomial Analysis

Let $Z_1, Z_2, \dots, Z_r$ be independent copies of $\mathtt{Geo}(p)$

Z_i are called "independent and identically distributed" or "i.i.d.'

Because they are independent…and have identical pmfs.

$X \sim \text{NegBin}(r, p)$ $X = Z_1 + Z_2 + \cdots + Z_r.$

Negative Binomial Analysis

Let $Z_1, Z_2, \ldots, Z_r$ be independent copies of $\text{Geo}(p)$

Z_i are called "independent and identically distributed" or "i.i.d.'

Because they are independent…and have identical pmfs.

$X \sim \text{NegBin}(r, p)$ $X = Z_1 + Z_2 + \cdots + Z_r$.

$$\mathbb{E}[\boldsymbol{X}] = \mathbb{E}[\boldsymbol{Z_1} + \boldsymbol{Z_2} + \cdots \boldsymbol{Z_r}] = \mathbb{E}[\boldsymbol{Z_1}] + \mathbb{E}[\boldsymbol{Z_2}] + \cdots + \mathbb{E}[\boldsymbol{Z_r}] = \boldsymbol{r} \cdot \frac{\mathbf{1}}{\boldsymbol{p}}$$

Negative Binomial Analysis

Let $Z_1, Z_2, \dots, Z_r$ be independent copies of $\text{Geo}(p)$

Z_i are called "independent and identically distributed" or "i.i.d.'

Because they are independent…and have identical pmfs.

$X \sim \text{NegBin}(r, p)$ $X = Z_1 + Z_2 + \cdots + Z_r$.

$\mathbb{E}[X] = \mathbb{E}[Z_1 + Z_2 + \cdots Z_r] = \mathbb{E}[Z_1] + \mathbb{E}[Z_2] + \cdots + \mathbb{E}[Z_r] = r \cdot \frac{1}{p}$

$\mathbf{Var}(X) = \mathbf{Var}(Z_1 + Z_2 + \cdots + Z_r) = \mathbf{Var}(Z_1) + \mathbf{Var}(Z_2) + \cdots + Var(Z_r)$

$= r \cdot \frac{1-p}{p^2}$

Negative Binomial Analysis

Let $Z_1, Z_2, \ldots, Z_r$ be independent copies of $\text{Geo}(p)$

Z_i are called "independent and identically distributed" or "i.i.d.'

Because they are independent…and have identical pmfs.

$X \sim \text{NegBin}(r, p)$ $X = Z_1 + Z_2 + \cdots + Z_r$.

$$\mathbb{E}[X] = \mathbb{E}[Z_1 + Z_2 + \cdots Z_r] = \mathbb{E}[Z_1] + \mathbb{E}[Z_2] + \cdots + \mathbb{E}[Z_r] = r \cdot \frac{1}{p}$$

$$\mathbf{Var}(X) = \mathbf{Var}(Z_1 + Z_2 + \cdots + Z_r) = \mathbf{Var}(Z_1) + \mathbf{Var}(Z_2) + \cdots + Var(Z_r)$$

$$= r \cdot \frac{1-p}{p^2}$$

because they are independent

Negative Binomial

$X \sim \mathrm{NegBin}(\mathrm{r}, \mathrm{p})$

Parameters: r: the number of successes needed, p the probability of success in a single trial

X is the number of trials needed to get the r^{th} success.

PMF: $p_X(k) = \binom{k-1}{r-1}(1-p)^{k-r}p^r$

CDF: $F_X(k)$ is ugly, don't bother with it.

Expectation: $\mathbb{E}[X] = \frac{r}{p}$

Variance: $\mathrm{Var}(\mathrm{X}) = \frac{\mathrm{r}(1-\mathrm{p})}{p^2}$

Scenario: Hypergeometric

You have an urn with N balls, of which K are purple. You are going to draw n balls out of the urn **without** replacement uniformly at random. How many purple balls do we get in this sample?

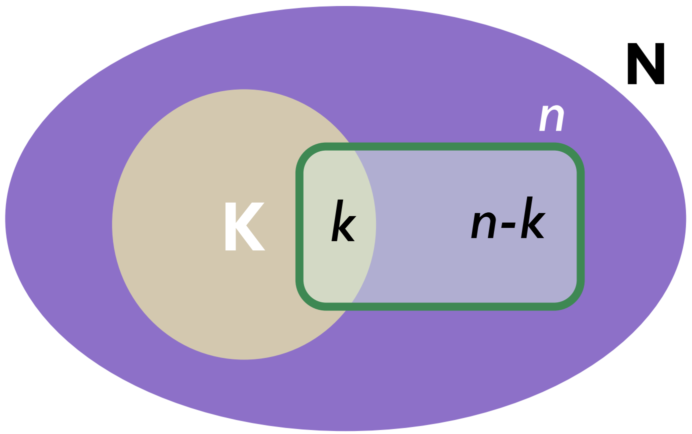
X is the number of purple balls in this sample of size n

Hypergeometric: Analysis (PMF)

You have an urn with N balls, of which K are purple. You are going to draw n balls out of the urn **without** replacement uniformly at random. How many purple balls do we get in this sample?

X is the number of purple balls in this sample of size n

If you draw out $\boldsymbol{n}$ balls, what is the probability you see $\boldsymbol{k}$ purple ones?



Hypergeometric: Analysis (PMF)

You have an urn with N balls, of which K are purple. You are going to draw n balls out of the urn **without** replacement uniformly at random. How many purple balls do we get in this sample?

X is the number of purple balls in this sample of size n

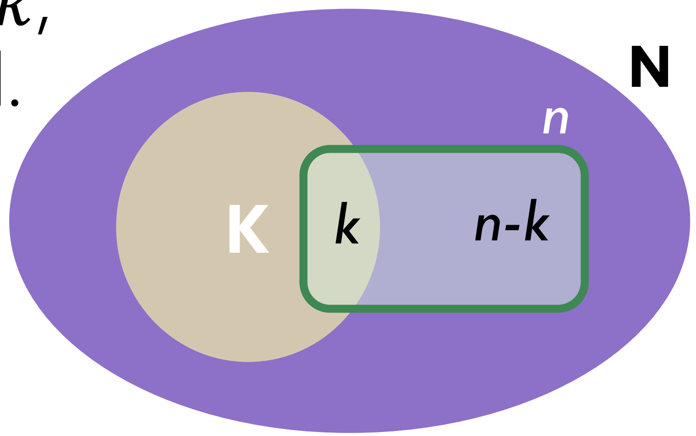
If you draw out n balls, what is the probability you see k purple ones?

> Of the K purple, we draw out k, choose which k will be drawn

> Of the $N-K$ other balls, we will draw out $n-k$, choose which $N-K-(n-k)$ will be removed.

Sample space all subsets of size n

$$\mathbb{P}(X=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$



Hypergeometric: Analysis (Expectation)

X is the number of purple balls in this sample of size n

Decompose: $X = D_1 + D_2 + \cdots + D_n$ *where* D_i *is the indicator that draw* i *is purple*

Apply LoE: $\mathbb{E}[D_1 + \cdots D_n] = \mathbb{E}[D_1] + \cdots + \mathbb{E}[D_n]$

Conquer: *What is* $\mathbb{E}[D_i] = \mathbb{P}(D_i = 1)$*?*

> $\mathbb{P}(D_1 = 1) = K/N$

> What about D_2? seems like it depends on whether the first was purple…

$$\mathbb{P}(D_2 = 1) = \frac{K-1}{N-1} \cdot \frac{K}{N} + \frac{K}{N-1} \cdot \frac{K-N}{N} = \frac{K(K-N+K-1)}{N(N-1)} = \frac{K}{N}$$

In general $\mathbb{P}(D_i = 1) = \frac{K}{N}$

It might feel counterintuitive, but it's true!

Hypergeometric: Analysis (Expectation)

X is the number of purple balls in this sample of size n

Decompose: $X = D_1 + D_2 + \cdots + D_n$ *where* D_i *is the indicator that draw* i *is purple*

Apply LoE: $\mathbb{E}[D_1 + \cdots D_n] = \mathbb{E}[D_1] + \cdots + \mathbb{E}[D_n]$

Conquer: *after some thinking…* $\mathbb{E}[D_i] = \frac{K}{N}$

So, $\mathbb{E}[X] = n \cdot \frac{K}{N}$

Hypergeometric: Analysis (Expectation)

X is the number of purple balls in this sample of size n

Decompose: $X = D_1 + D_2 + \cdots + D_n$ *where* D_i *is the indicator that draw* i *is purple*

Apply LoE: $\mathbb{E}[D_1 + \cdots D_n] = \mathbb{E}[D_1] + \cdots + \mathbb{E}[D_n]$

Conquer: *after some thinking…* $\mathbb{E}[D_i] = \frac{K}{N}$

So, $\mathbb{E}[X] = n \cdot \frac{K}{N}$

Can we do the same for variance? Can we use linearity of variance?

Hypergeometric: Analysis (Expectation)

X is the number of purple balls in this sample of size n

Decompose: $X = D_1 + D_2 + \cdots + D_n$ *where* D_i *is the indicator that draw* i *is purple*

Apply LoE: $\mathbb{E}[D_1 + \cdots D_n] = \mathbb{E}[D_1] + \cdots + \mathbb{E}[D_n]$

Conquer: *after some thinking…* $\mathbb{E}[D_i] = \frac{K}{N}$

So, $\mathbb{E}[X] = n \cdot \frac{K}{N}$

Can we do the same for variance? Can we use linearity of variance?

No! The D_i are dependent. Even if they have the same probability.

Hypergeometric: Analysis

$\mathbb{E}[X]$

$= \mathbb{E}[D_1 + \cdots D_n] = \mathbb{E}[D_1] + \cdots + \mathbb{E}[D_n] = n \cdot \frac{K}{N}$

Can we do the same for variance?

No! The D_i are dependent. Even if they have the same probability.

Hypergeometric Random Variable

$X \sim \mathrm{HypGeo}(N, K, n)$

X is the number of success balls drawn in the sample.

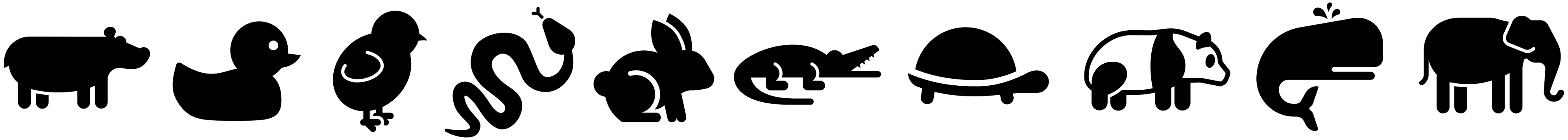
Parameters: A total of N balls in an urn, of which K are successes. Draw n balls without replacement.

PMF: $p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$

Expectation: $\mathbb{E}[X] = \frac{nK}{N}$

Variance: $\mathrm{Var}(X) = n \cdot \frac{K}{N} \cdot \frac{N-K}{N} \cdot \frac{N-n}{N-1}$

Zoo!



$X \sim \mathrm{Unif}(a,b)$

$$p_X(k) = \frac{1}{b-a+1}$$

$$\mathbb{E}[X] = \frac{a+b}{2}$$

$$\mathrm{Var}(X) = \frac{(b-a)(b-a+2)}{12}$$

$X \sim \mathrm{Ber}(p)$

$$p_X(0) = 1-p; \quad p_X(1) = p$$

$$\mathbb{E}[X] = p$$

$$\mathrm{Var}(X) = p(1-p)$$

$X \sim \mathrm{Bin}(n,p)$

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\mathbb{E}[X] = np$$

$$\mathrm{Var}(X) = np(1-p)$$

$X \sim \mathrm{Geo}(p)$

$$p_X(k) = (1-p)^{k-1} p$$

$$\mathbb{E}[X] = \frac{1}{p}$$

$$\mathrm{Var}(X) = \frac{1-p}{p^2}$$

$X \sim \mathrm{Poi}(\lambda)$

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$\mathbb{E}[X] = \lambda$$

$$\mathrm{Var}(X) = \lambda$$

$X \sim \mathrm{NegBin}(r,p)$

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

$$\mathbb{E}[X] = \frac{r}{p}$$

$$\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$$

$X \sim \mathrm{HypGeo}(N,K,n)$

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

$$\mathbb{E}[X] = n\frac{K}{N}$$

$$\mathrm{Var}(X) = \frac{K(N-K)(N-n)}{N^2(N-1)}$$

Discrete Zoo of Random Variables

- **Uniform:** Every integer between a and b are equally likely
  $\mathrm{Un}if(a, b)$
- **Bernoulli:** Whether there is success in one trial
  $\mathrm{Ber}(p)$ is 1 with probability p and 0 otherwise
- **Binomial:** Number of successes in n independent trials
  $\mathrm{Bin}(n, p)$ - n independent trials, probability p of success on each trial
- **Geometric:** Number of trials till first success
  $\mathrm{Geo}(p)$ - probability p of success on each trial
- **Poisson:** Number of successes in a time interval
  $\mathrm{Poi}(\lambda)$ - average number of successes in the time interval
- **Negative Binomial**: Number of trials till r'th success
  $\mathrm{NegBin}(r, p)$ - probability p of success on each trial, want trials till the r'th success
- **Hypergeometric:** Number of successes when drawing a sample
  $\mathrm{HypGeo}(N, K, n)$ - drawing a sample of n items from a set of N with K successes

Zoo Takeaways

You can do relatively complicated counting/probability calculations much more quickly than you could week 1!

You can now explain why your problem is a zoo variable and save explanation on homework (and save yourself calculations in the future).

Don't spend extra effort memorizing…but be careful when looking up Wikipedia articles.

- The exact definitions of the parameters can differ (is a geometric random variable the number of failures before the first success, or the total number of trials including the success?)

Continuous Random Variables

Goal for today is to get intuition on what's different in the continuous case

ASK QUESTIONS (always! but today especially ☺)

Discrete Random Variables

The kind that we've been working with up till now!

The support has *finite* or *countably infinite values*

e.g., number of successes, number of trials till success, attendance at a class are all discrete because they take on a set of finite or countably infinite values

Some random experiments have uncountably-infinite sample spaces

> *How long until the next bus shows up?*

> *Throwing a dart on a board (what location does the dart land?)*

Continuous Random Variables

Random variables with a support of *uncountably-infinite values*

> *e.g., RVs that take on any **real** number in some interval(s) like distance, height, time, etc.*

Why Need New Rules?

Random Experiment: choose a random <u>*real*</u> number between 0 and 1

What's the probability the number is between 0.4 *and* 0.5*?*

> For discrete spaces, we'd ask for $\frac{|E|}{|\Omega|}$

> So we get $\frac{\infty}{\infty}$ ☹

When working with continuous random experiments, we'll almost always use continuous random variables to interact with these spaces

X is the number we choose. We want $\mathbb{P}(0.4 \leq X \leq 0.5)$

Probability Density Function (PDF)

Analogous to PMF in a discrete random variable

How to describe probability of a *single* value?

For **discrete random variables**, we defined the PMF: $p_Y(k) = \mathbb{P}(Y = k)$.

How to describe probability of a *single* value?

For **discrete random variables**, we defined the PMF: $p_Y(k) = \mathbb{P}(Y = k)$.

Can we use the same for continuous random variables?

What is the probability that a person's height is exactly 5.678123589 feet

What is the probability that a bus arrives exactly 6 hours, 23 minutes and 2976427909 seconds from now?

$$p_X(0.135) = \mathbb{P}(X = 0.135) =$$

How to describe probability of a *single* value?

For **discrete random variables**, we defined the PMF: $p_Y(k) = \mathbb{P}(Y = k)$.

Can we use the same for continuous random variables?

What is the probability that a person's height is exactly 5.678123589 feet

What is the probability that a bus arrives exactly 6 hours, 23 minutes and 2976427909 seconds from now?

$p_X(0.135) = \mathbb{P}(X = 0.135) = \frac{1}{\infty} = 0$ ***No!***

How to describe probability of a *single* value?

For *discrete* random variables, we defined the PMF: $p_Y(k) = \mathbb{P}(Y = k)$.

Can we use the same for continuous random variables?

What is the probability that a person's height is exactly 5.678123589 feet

What is the probability that a bus arrives exactly 6 hours, 23 minutes and 2976427909 seconds from now?

$$p_X(0.135) = \mathbb{P}(X = 0.135) = \frac{1}{\infty} = 0$$ ***No!***

Instead, we use the probability density function (PDF)

$\boldsymbol{f_X(k)}$ is the density (not probability!) of the continuous random variable X at the value k

Probability *Density* Function

For *discrete* **random variables**, we defined the PMF: $p_Y(k) = \mathbb{P}(Y = k)$.

For *continuous* random variables, we use the probability density function

$\boldsymbol{f_X(k)}$ is the density (not probability!) of the continuous random variable X at the value k

Discrete (PMF)	Continuous (PDF)
$p_Y(k) \geq 0$	$f_X(k) \geq 0$
$\sum_{k \in \Omega_Y} p_Y(k) = 1$	$\int_{-\infty}^{\infty} f_X(k)\ \mathrm{d}k = 1$

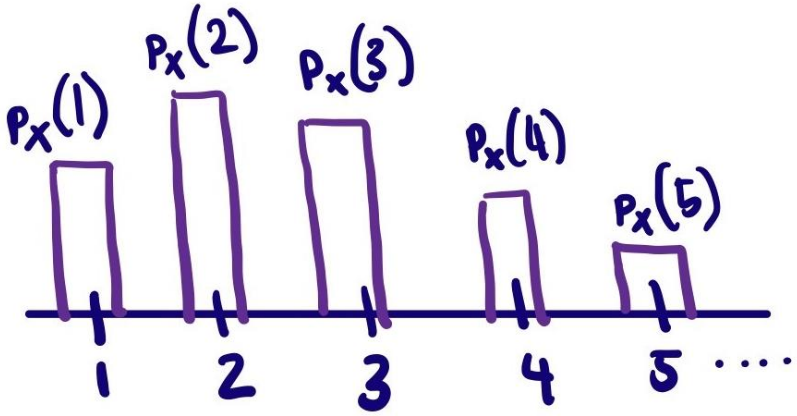
Probability *Density* Function

For *discrete* random variables, we defined the PMF: $p_Y(k) = \mathbb{P}(Y = k)$.

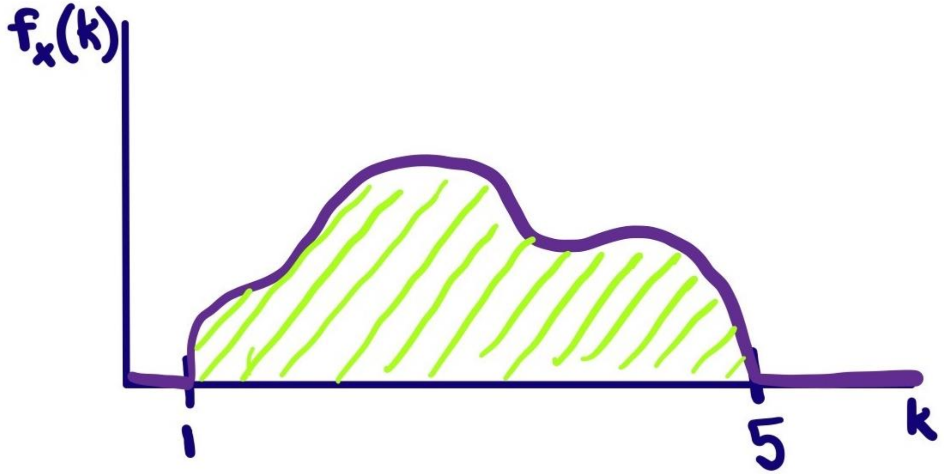
For *continuous* random variables, we use the probability density function

$\boldsymbol{f_X(k)}$ is the density (not probability!) of the continuous random variable X at the value k

e.g., PMF for a discrete RV, $\boldsymbol{X}$



e.g., PDF for a continuous RV, $\boldsymbol{X}$



Probability *Density* Function

For *discrete* random variables, we defined the PMF: $p_Y(k) = \mathbb{P}(Y = k)$.

For *continuous* random variables, we use the probability density function

$\boldsymbol{f_X(k)}$ is the density of the continuous random variable X at the value k

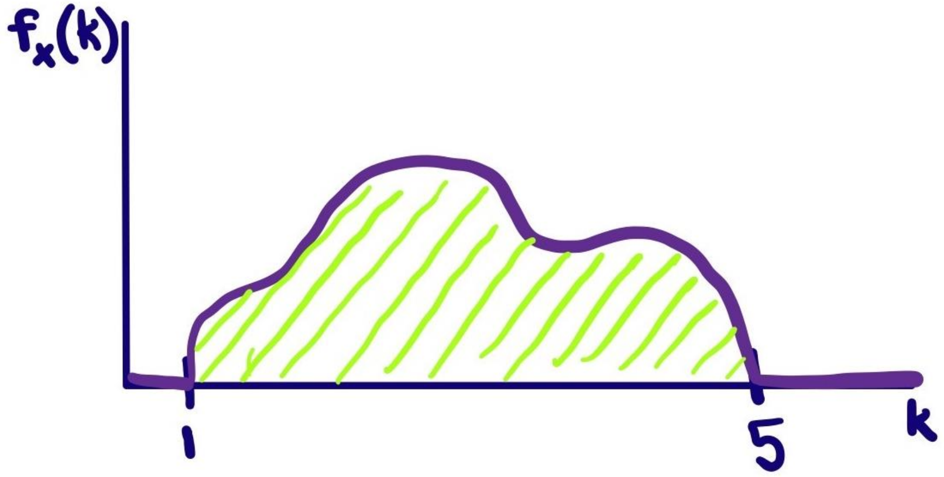
How do we use the PDF?

To compute probabilities of events!

$\mathbb{P}(a \le X \le b) = \int_a^b f_X(z)\,\mathrm{d}z$

integrating is analogous to summing

e.g., PDF for a continuous RV, $\boldsymbol{X}$



Probability *Density* Function (*example)*

X is a **uniform real number** in $[0,1]$. Let's derive the PDF for X!

Idea: Make it "work right" for **events** since single outcomes don't make sense.

$\mathbb{P}(0 \leq X \leq 1) = \qquad = \int$

$\mathbb{P}(X \text{ is negative}) = \qquad = \int$

$\mathbb{P}(.4 \leq X \leq .5) = \qquad = \int$

Probability *Density* Function (*example*)

X is a **uniform real number** in $[0,1]$. Let's derive the PDF for X!

Idea: Make it "work right" for **events** since single outcomes don't make sense.

$$\mathbb{P}(0 \leq X \leq 1) = \ \ 1 \ \ = \int_0^1 f_X(z)\,\mathrm{d}z$$

$$\mathbb{P}(X \text{ is negative}) = 0 = \int_{-\infty}^0 f_X(z)\,\mathrm{d}z$$

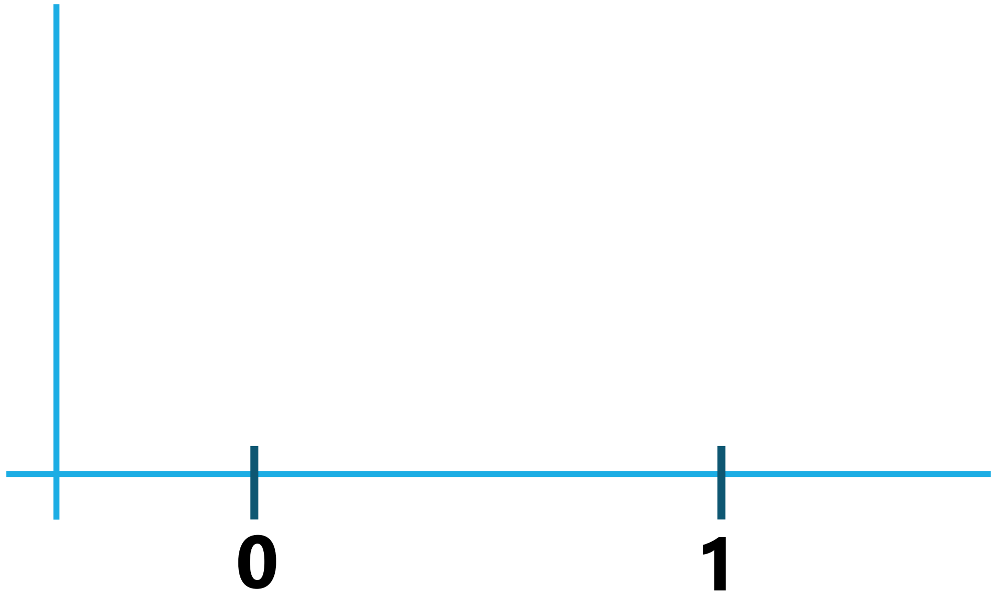
$$\mathbb{P}(.4 \leq X \leq .5) = 0.1 = \int_{.4}^{.5} f_X(z)\,\mathrm{d}z$$

Probability *Density* Function (*example*)

X is a **uniform real number** in $[0,1]$. Let's derive the PDF for X!

What should $f_X(k)$ be to make all those events integrate to the right values?

$$f_X(k) = \begin{cases} 0 & \text{if } k < 0 \text{ or } k > 1 \\ 1 & \text{if } 0 \leq k \leq 1 \end{cases}$$



Probability vs. Density

Key idea: integrating the PDF gives us probabilities

But the PDF itself does not give us probabilities

The number that best represents $\mathbb{P}(X = .1)$ is 0

But, this is different from $f_X(.1) = 1$

For continuous probability spaces:

> *Impossible events have probability* 0

> *But even though probability of a specific value is* 0*, it's still possible*

What exactly do the values in the PDF mean?

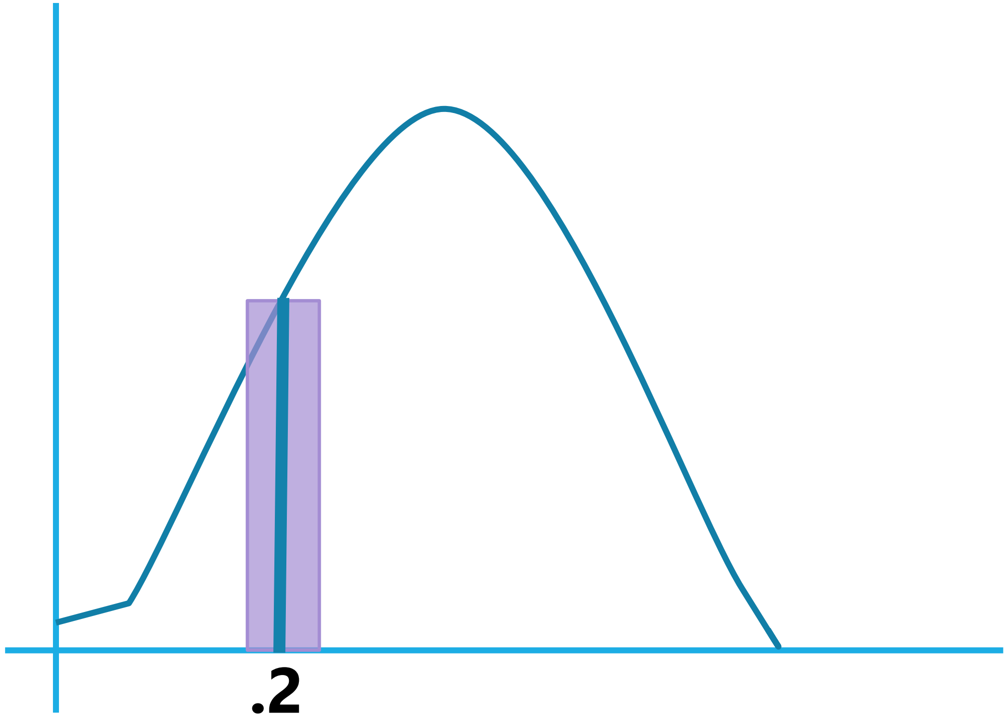
Let's look at the event $X \approx .2$

For a *very* small value of ϵ,

$$\mathbb{P}(X \approx .2) = \mathbb{P}(.2 - \epsilon/2 \leq X \leq .2 + \epsilon/2)$$

$$= \int_{.2-\epsilon/2}^{-2+\epsilon/2} f_X(z)\,\mathrm{dz} \approx \mathbf{height} \cdot \mathbf{width}$$

$$= f_X\,(.2) \cdot \epsilon$$



What happens if we look at the ratio

$$\frac{\mathbb{P}(X \approx .2)}{\mathbb{P}(X \approx .5)}$$

What exactly do the values in the PDF mean?

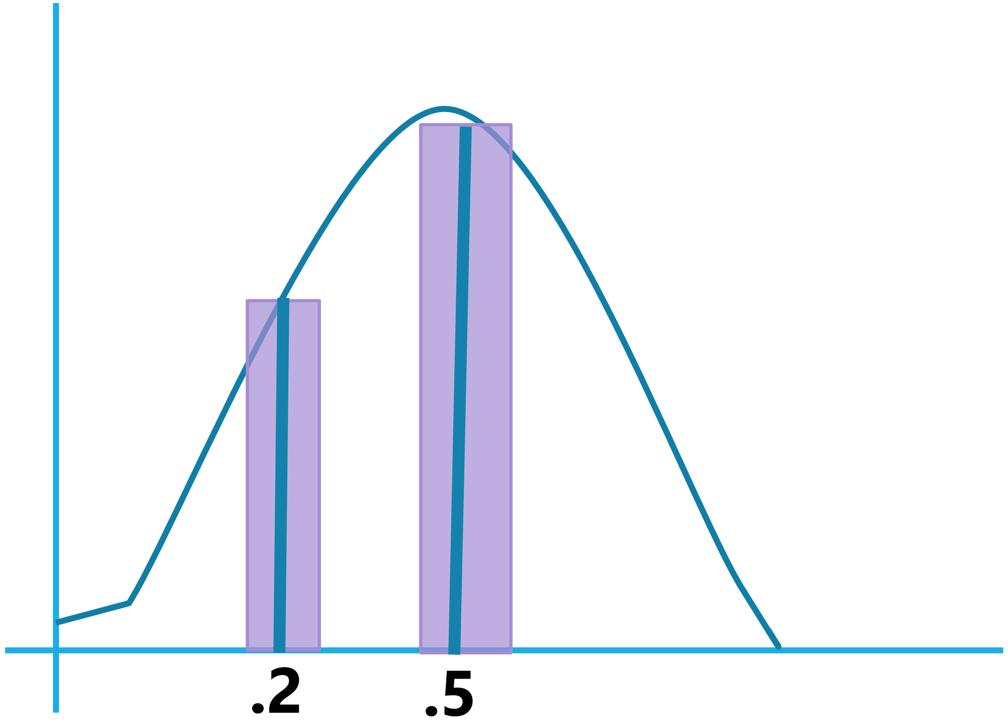
Let's look at the event $X \approx .2$

For a *very* small value of ϵ,

$\mathbb{P}(X \approx .2) = \mathbb{P}(.2 - \epsilon/2 \leq X \leq .2 + \epsilon/2)$

$= \int_{.2-\epsilon/2}^{-2+\epsilon/2} f_X(z)\, \mathrm{dz} \approx \text{height} \cdot \text{width}$

$= f_X\,(.2) \cdot \epsilon$



What happens if we look at the ratio

$$\frac{\mathbb{P}(X\approx.2)}{\mathbb{P}(X\approx.5)} = \frac{\mathbb{P}\left(.2-\frac{\epsilon}{2}\leq X\leq.2+\frac{\epsilon}{2}\right)}{\mathbb{P}\left(.5-\frac{\epsilon}{2}\leq X\leq.5+\frac{\epsilon}{2}\right)} = \frac{\epsilon f_X(.2)}{\epsilon f_X(.5)} = \frac{f_X(.2)}{f_X(.5)}$$

What exactly do the values in the PDF mean?

The number that when integrated over gives the probability of an event.

Equivalently, it's number such that:

-it is always non-negative

-integrating over all real numbers gives 1.

-comparing $f_X(k)$ and $f_X(\ell)$ gives the relative chances of $\boldsymbol{X}$ being near $\boldsymbol{k}$ or ℓ.

Cumulative Distribution Function (CDF)

What's a CDF?

The Cumulative Distribution Function $F_X(k) = \mathbb{P}(\boldsymbol{X} \leq \boldsymbol{k})$ analogous to the CDF for discrete variables.

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

So how do I get from CDF to PDF? Taking the derivative!

$$\frac{\mathrm{d}}{\mathrm{d}k} F_X(k) = \frac{\mathrm{d}}{dk}\left(\int_{-\infty}^{k} f_X(z)\,\mathrm{d}z\right) = f_X(k)$$

Comparing Discrete and Continuous

	Discrete Random Variables	Continuous Random Variables
Probability 0	Equivalent to impossible	All impossible events have probability 0, but not conversely.
Relative Chances	PMF: $p_X(k) = \mathbb{P}(X = k)$	PDF $f_X(k)$ gives chances relative to $f_X(k')$
Events	Sum over PMF to get probability	Integrate PDF to get probability
Convert from CDF to PMF	Sum up PMF to get CDF. Look for "breakpoints" in CDF to get PMF.	Integrate PDF to get CDF. Differentiate CDF to get PDF.
$\mathbb{E}[X]$	$\sum_{\omega} X(\omega) \cdot f_X(\omega)$	$\int_{-\infty}^{\infty} z \cdot f_X(z)\, \mathrm{d}z$
$\mathbb{E}[g(X)]$	$\sum_{\omega} g\big(X(\omega)\big) \cdot f_X(\omega)$	$\int_{-\infty}^{\infty} g(z) \cdot f_X(z)\, \mathrm{d}z$
$\mathrm{Var}(X)$	$\mathbb{E}[X^2] - (\mathbb{E}[X])^2$	$\mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{\infty} (z - \mathbb{E}[X])^2 f_X(z)\, \mathrm{d}z$

Expectation and Variance

What about expectation?

For a **discrete** random variable X, we have $\mathbb{E}[X] = \sum_{k \in \Omega_X} k \cdot p_X(k)$

For a **continuous** random variable X, we define:

$\mathbb{E}[X] = \int_{-\infty}^{\infty} z \cdot f_X(z)\, \mathrm{d}z$

Just replace summing over the PMF with integrating the PDF.

It still represents the average value of X.

Expectation of a function

For a **discrete** random variable X, we have $\mathbb{E}[g(X)] = \sum_{k \in \Omega_X} g(k) \cdot p_X(k)$

For any function g and any **continuous** random variable, X:
$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(z) \cdot f_X(z)\, \mathrm{d}z$

Again, analogous to the discrete case; just replace summation with integration and pmf with the pdf.

Linearity of Expectation

Still true!

$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y\} + c$
For all X, Y; even if they're continuous.

Won't show you the proof – for just $\mathbb{E}[aX + b]$, it's

$$\mathbb{E}[aX + b] = \int_{-\infty}^{\infty} [aX(k) + b] f_X(k)\,\mathrm{d}k$$

$$= \int_{-\infty}^{\infty} aX(k) f_X(k) dk + \int_{-\infty}^{\infty} b f_X(k) dk$$

$$= a\int_{-\infty}^{\infty} X(k) f_X(k) dk + b\int_{-\infty}^{\infty} f_X(k) dk$$

$$= a\mathbb{E}[X] + b$$

Variance

No surprises here

$$\mathbf{Var}(\boldsymbol{X}) = \mathbb{E}\left[\boldsymbol{X}^2\right] - (\mathbb{E}[\boldsymbol{X}])^2 = \int_{-\infty}^{\infty} \boldsymbol{f}_{\boldsymbol{X}}(\boldsymbol{k})(\boldsymbol{k} - \mathbb{E}[\boldsymbol{X}])^2 \, \mathbf{d}\boldsymbol{k}$$

Let's calculate an expectation

Let X be a uniform random number between a and b.

$\mathbb{E}[X] = \int_{-\infty}^{\infty} z \cdot f_X(z)\, \mathrm{d}z$

$= \int_{-\infty}^{a} z \cdot 0\, \mathrm{d}z + \int_{a}^{b} z \cdot \frac{1}{b-a} \mathrm{d}z + \int_{b}^{\infty} z \cdot 0\, \mathrm{d}z$

$= 0 + \int_{a}^{b} \frac{z}{b-a}\, \mathrm{d}z + 0$

$= \frac{z^2}{2(b-a)}\Big|_{z=a}^{b} = \frac{b^2}{2(b-a)} - \frac{a^2}{2(b-a)} = \frac{b^2-a^2}{2(b-a)} = \frac{(b+a)(b-a)}{2(b-a)} = \frac{a+b}{2}$

What about $\mathbb{E}[g(X)]$

Let $X \sim \text{Unif}(a, b)$, what about $\mathbb{E}[X^2]$?

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} z^2 f_X(z) \mathrm{dz}$$

$$= \int_{-\infty}^{a} z^2 \cdot 0 \ \mathrm{dz} + \int_{a}^{b} z^2 \cdot \frac{1}{b-a} \mathrm{dz} + \int_{b}^{\infty} z^2 \cdot 0 \ \mathrm{dz}$$

$$= 0 + \int_{a}^{b} z^2 \cdot \frac{1}{b-a} \mathrm{dz} + 0$$

$$= \frac{1}{b-a} \cdot \frac{z^3}{3} \bigg|_{z=a}^{b} = \frac{1}{b-a} \left(\frac{b^3}{3} - \frac{a^3}{3} \right) = \frac{1}{3(b-a)} \cdot (b-a)(a^2 + ab + b^2)$$

$$= \frac{a^2+ab+b^2}{3}$$

Let’s assemble the variance

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

$$= \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{4(a^2+ab+b^2)}{12} - \frac{3(a^2+2ab+b^2)}{12}$$

$$= \frac{a^2-2ab+b^2}{12}$$

$$= \frac{(a-b)^2}{12}$$

Continuous Uniform Distribution

$X \sim \text{Unif}(a, b)$ (uniform real number between a and b)

PDF: $f_X(k) = \begin{cases} \frac{1}{b-a} & \text{if } a \le k \le b \\ 0 & \text{otherwise} \end{cases}$

CDF: $F_X(k) = \begin{cases} 0 & \text{if } k < a \\ \frac{k-a}{b-a} & \text{if } a \le k \le b \\ 1 & \text{if } k \ge b \end{cases}$

$\mathbb{E}[X] = \frac{a+b}{2}$

$\text{Var}(X) = \frac{(b-a)^2}{12}$