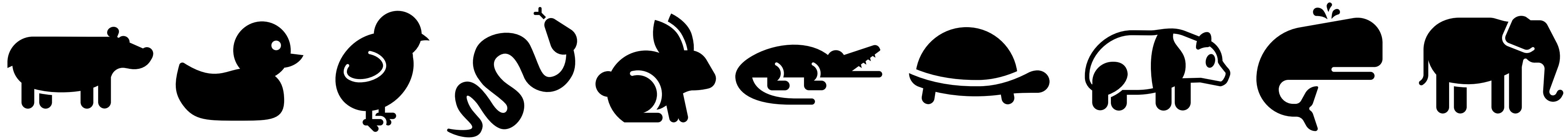
Continuous Probability

CSE 312 Winter 25
Lecture 14

Zoo!



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

$$p_X(k) = \frac{1}{b-a+1}$$
$$\mathbb{E}[X] = \frac{a+b}{2}$$
$$\text{Var}(X) = \frac{(b-a)(b-a+2)}{12}$$

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

$$p_X(0) = 1-p;$$
$$p_X(1) = p$$
$$\mathbb{E}[X] = p$$
$$\text{Var}(X) = p(1-p)$$

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

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$
$$\mathbb{E}[X] = np$$
$$\text{Var}(X) = np(1-p)$$

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

$$p_X(k) = (1-p)^{k-1} p$$
$$\mathbb{E}[X] = \frac{1}{p}$$
$$\text{Var}(X) = \frac{1-p}{p^2}$$

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

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$
$$\mathbb{E}[X] = \frac{r}{p}$$
$$\text{Var}(X) = \frac{r(1-p)}{p^2}$$

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

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$
$$\mathbb{E}[X] = n\frac{K}{N}$$
$$\text{Var}(X) = \frac{K(N-K)(N-n)}{N^2(N-1)}$$

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

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$
$$\mathbb{E}[X] = \lambda$$
$$\text{Var}(X) = \lambda$$

Zoo Takeaways

We skipped hypergeometric; slides are there, you can use the formulas!

- Drawing without replacement from an urn is the situation.

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?)

Where Are We?

What have we done over the past 5 weeks?

Counting

Combinations, permutations, indistinguishable elements, starts and bars, inclusion-exclusion…

Probability foundations

Events, sample space, axioms of probability, expectation, variance

Conditional probability

Conditioning, independence, Bayes' Rule

Refined our intuition

Especially around Bayes' Rule

What’s next?

Continuous random variables.

- So far our sample spaces have been countable. What happens if we want to choose a random real number?
- How do expectation, variance, conditioning, etc. change in this new context?
- Mostly analogous to discrete cases, but with integrals instead of sums.

Analysis when it’s inconvenient (or impossible) to exactly calculate probabilities.

- Central Limit Theorem (approximating discrete distributions with continuous ones)
- Tail Bounds/Concentration (arguing it’s unlikely that a random variable is far from its expectation)

A first taste of making predictions from data (i.e., a bit of ML)

Today

Continuous Probability

Probability Density Function

Cumulative Distribution Function

Goal for today is to get intuition on what's different in the continuous case. Your goal today is to start building up a gut-feeling of what's happening.

ASK QUESTIONS, (always, but today especially).

Continuous Random Variables

Continuous Random Variables

We'll need continuous probability spaces and continuous random variables to describe experiments that have uncountably-infinite sample spaces.

e.g. all real numbers

How long until the next bus shows up?

What location does a dart land?

Continuous Random Variables

Wait, we're computer scientists. Computers don't do real numbers, why should we?

Continuous random variables will be a useful model for enormous sample spaces. The math will be easier.

Example: polling a large population. The sample space is actually discrete. But we're going to round the result anyway. Make it continuous first for easier math, then round.

Why Need New Rules?

We want to choose a uniformly random real number between 0 and 1.

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

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

So we get $\frac{\infty}{\infty}$

The mathematical tools to get consistent answers from expressions like those is calculus.

Let's start with the pmf

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

We can't have a pmf quite like we did for discrete random variables. Let X be a random real number between 0 and 1.

$\mathbb{P}(X = .1) = \frac{1}{\infty}$??

Let's try to maintain as many rules as we can…

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

Use f_X instead of p_X to remember it's different .

The probability density function

For Continuous random variables, the analogous object is the "probability density function" we write $f_X(k)$ instead of $p_X(k)$

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

$\mathbb{P}(a \leq X \leq b) = c$

$$\int_a^b f_X(z)\, \mathrm{d}z = c$$

integrating is analogous to sum.

The probability density function

For Continuous random variables, the analogous object is the "probability density function" we write $f_X(k)$ instead of $p_X(k)$

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

$\mathbb{P}(a \leq X \leq b) = c$

$$\int_a^b f_X(z)\,\mathrm{d}z = c$$

integrating is analogous to sum.

Let's derive an example PDF together!
For a uniform random real number in [0,1]

The probability density function

For Continuous random variables, the analogous object is the "probability density function" we write $f_X(k)$ instead of $p_X(k)$

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

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

integrating is analogous to sum.

$\mathbb{P}(X \text{ is negative}) = 0$

$\mathbb{P}(.4 \leq X \leq .5) = .1$

$\int_{.4}^{.5} f_X(k)\,dk = .1$

The probability density function

For Continuous random variables, the analogous object is the "probability density function" we write $f_X(k)$ instead of $p_X(k)$

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

$\mathbb{P}(0 \le X \le 1) = 1$ | $\int_0^1 f_X(z)\, \mathrm{d}z = 1$ | integrating is analogous to sum.

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

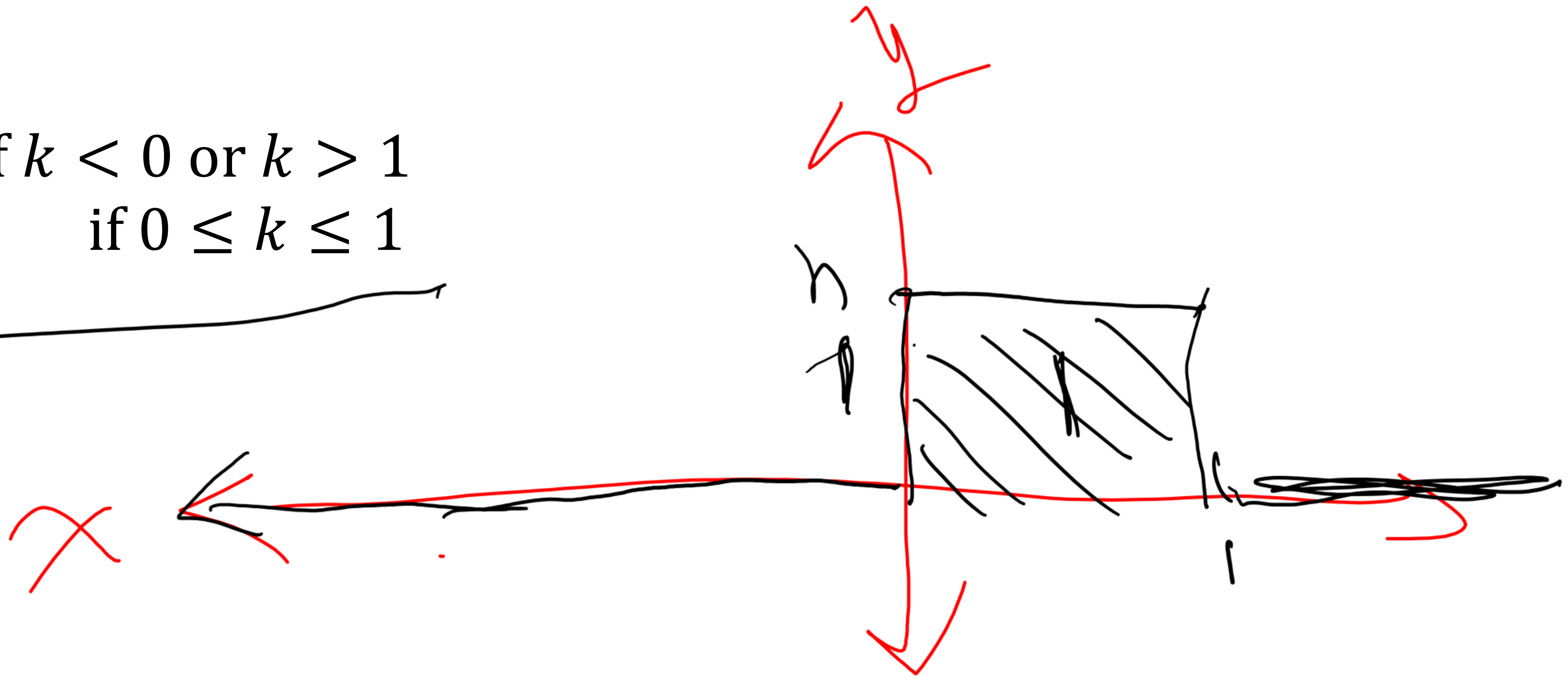
$\mathbb{P}(.4 \le X \le .5) = .1$ | $\int_{.4}^{.5} f_X(z)\, \mathrm{d}z = .1$

PDF for uniform

Let X be a uniform real number between 0 and 1.

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

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



Probability Density Function

So $\mathbb{P}(X = .1) = ??$

$f_X(.1) = 1$

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

This is different from $f_X(x)$

For continuous probability spaces:
Impossible events have probability 0,
but some probability 0 events might be possible.

So…what is $f_X(x)$???

Using the PDF

Let's look at a different pdf…

Compare the events: $X \approx .2$ and $X \approx .5$

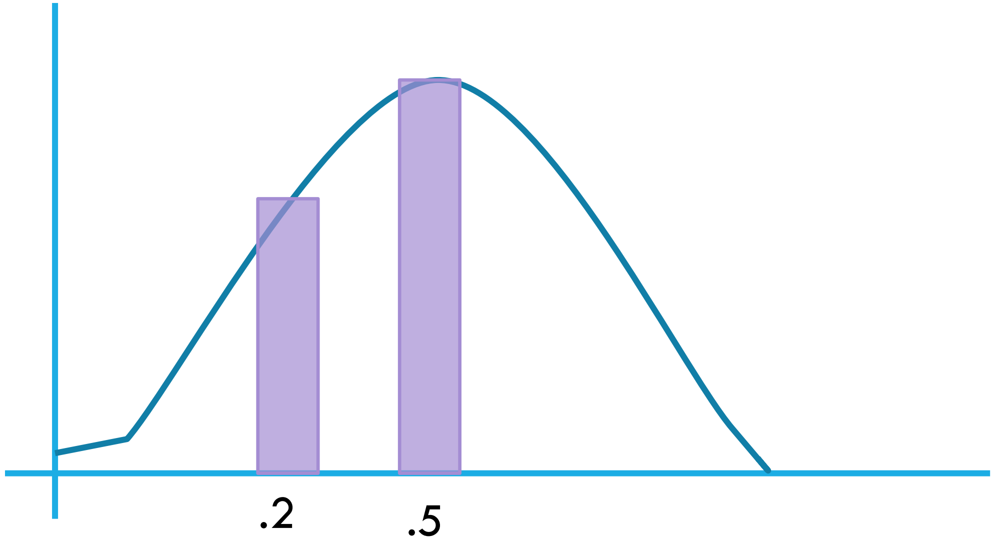
$\mathbb{P}(.2 - \epsilon/2 \leq X \leq .2 + \epsilon/2)$

What will the pdf give? $\int_{.2-\epsilon/2}^{-2+\epsilon/2} f_X(z)\,\mathrm{d}z$

$f_X(.2) \cdot \epsilon$

What happens if we look at the ratio

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



Using the PDF

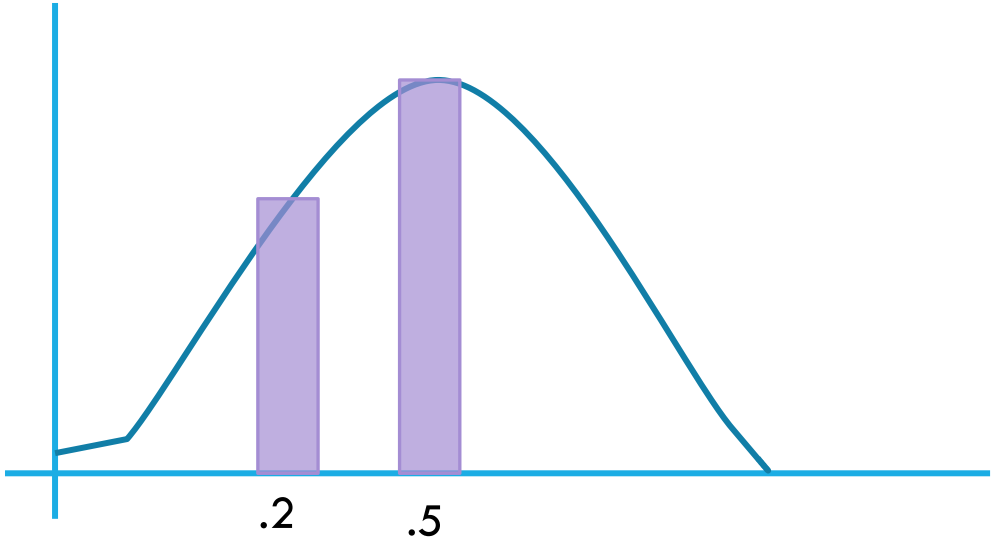
Let's look at a different pdf…

Compare the events: $X \approx .2$ and $X \approx .5$

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

What will the pdf give? $\int_{.2-\epsilon/2}^{.2+\epsilon/2} f_X(z)\, \mathrm{d}z$

$f_X(.2) \cdot \epsilon$



What happens if we look at the ratio

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

So what's the pdf?

It's the number that when integrated over gives the probability of an event.

Equivalently, it's number such that:

-integrating over all real numbers gives 1.

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

CDFs

What's a CDF?

The Cumulative Distribution Function $F_X(k) = \mathbb{P}(X \leq k)$ 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.
$\mathbb{E}[X]$	$\sum_{\omega} X(\omega) \cdot p_X(\omega)$	$\int_{-\infty}^{\infty} z \cdot f_X(z)\, \mathrm{d}z$
$\mathbb{E}[g(X)]$	$\sum_{\omega} g(X(\omega)) \cdot p_X(\omega)$	$\int_{-\infty}^{\infty} g(z) \cdot f_X(z)\, \mathrm{d}z$
$\mathrm{Var}(X)$	$\mathbb{E}[X^2] - (\mathbb{E}[X])^2$	$\mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \int_{-\infty}^{\infty} (z - \mathbb{E}[X])^2 f_X(z)\, \mathrm{d}z$

What about expectation?

For a random variable X, we define:

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

Just replace summing over the pmf with integrating the pdf.

It still represents the average value of X.

Expectation of a function

For any function g and any continuous random variable, X:

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

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

We're going to treat this as a definition.

Technically, this is really a theorem; since $f()$ is the pdf of X and it only gives relative likelihoods for X, we need a proof to guarantee it "works" for $g(X)$.

Sometimes called "Law of the Unconscious Statistician."

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

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

Let's calculate an expectation

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

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

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

Let's calculate an expectation

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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