No activity slide today! ☹

Continuous Random Variables

CSE 312 Summer 21
Lecture 14

Announcements

Homework 5 and Real World 2 have been released.

No Review Summary due tonight! Next one is due in a week.

Extra Office Hours sign up: https://forms.gle/BSsK7ivUNPJz6ckFA.

Please sign up by tonight!

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

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

e.g., all real numbers

What exact time does a bolt of lightning strike?

How long until the next bus shows up?

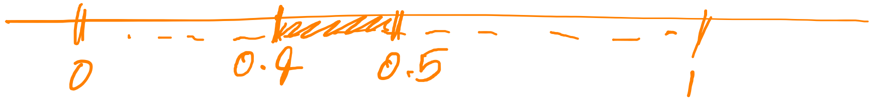
What location does a dart land?

Continuous Random Variables

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

$Y \rightarrow$ discrete

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.

$\mathbb{P}(X = .1) = \frac{1}{\infty}$?? $= 0$

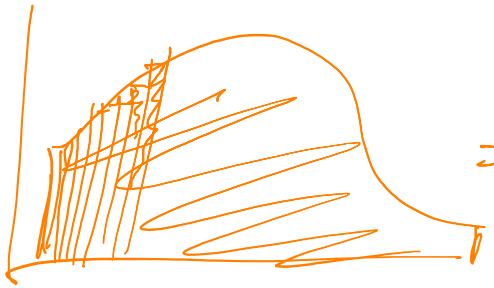
Let’s try to maintain as many rules as we can…

$p_Y(k) \geq 0$ was a requirement for discrete **(Non-negativity)**

We’ll keep that for continuous random variables.

$\sum_k p_Y(k) = 1$ for discrete **(Normalization)**

For continuous: $\int_{-\infty}^{\infty} f_X(k)\ \mathrm{d}k = 1$



$= 1$

The probability density function

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

Let's focus on making these events be correct:

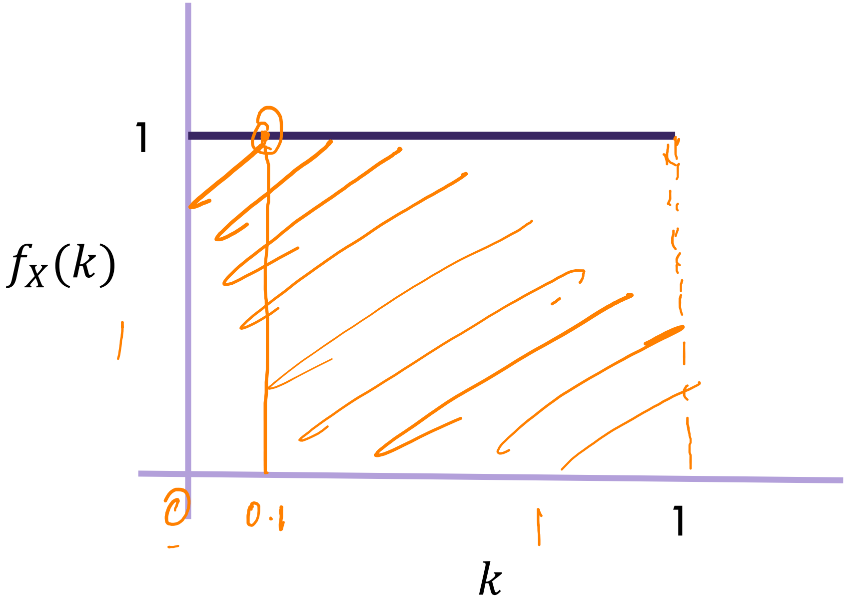
$\mathbb{P}(0 \leq X \leq 1) = 1$	$\boxed{\int_0^1 f_X(z)\,\mathrm{d}z = 1}$	integrating is analogous to sum.
$\mathbb{P}(X \text{ is negative}) = 0$	$\boxed{\int_{-\infty}^0 f_X(z)\,\mathrm{d}z = 0}$	Density $\neq$ Probability
$\mathbb{P}(.4 \leq X \leq .5) = .1$	$\boxed{\int_{.4}^{.5} f_X(z)\,\mathrm{d}z = .1}$	

PDF for uniform

Let X be a uniform real number between $\mathbf{0}$ and $\mathbf{1}$.

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

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



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, ← Same as discrete

but some probability 0 events might be possible.

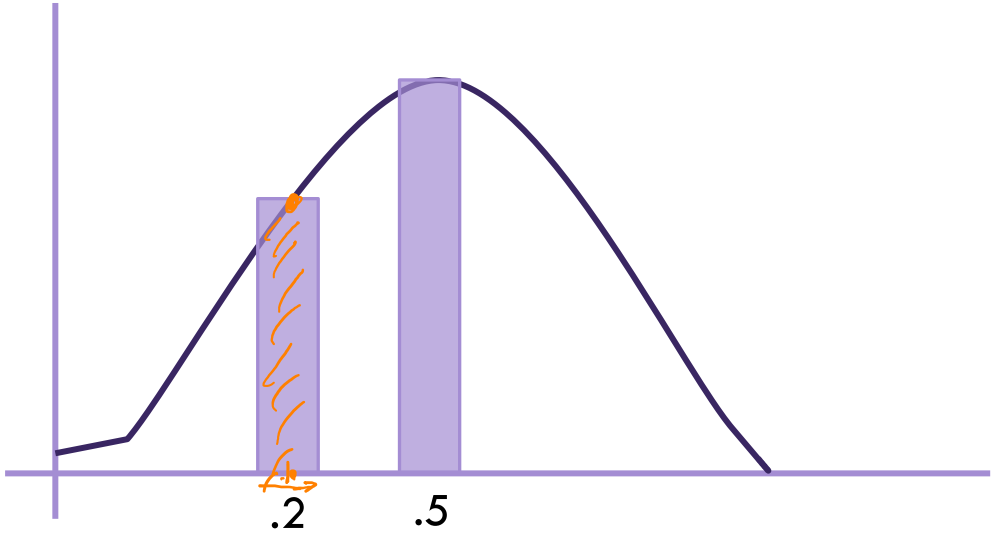
So…what is $f_X(x)$???

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



$f_X(.2) \cdot \epsilon$

$P(X \approx 0.5) = P\left(0.5 - \frac{\epsilon}{2} \leq X \leq 0.5 + \frac{\epsilon}{2}\right) = f_X(0.5) \cdot \epsilon$

What happens if we look at the ratio

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

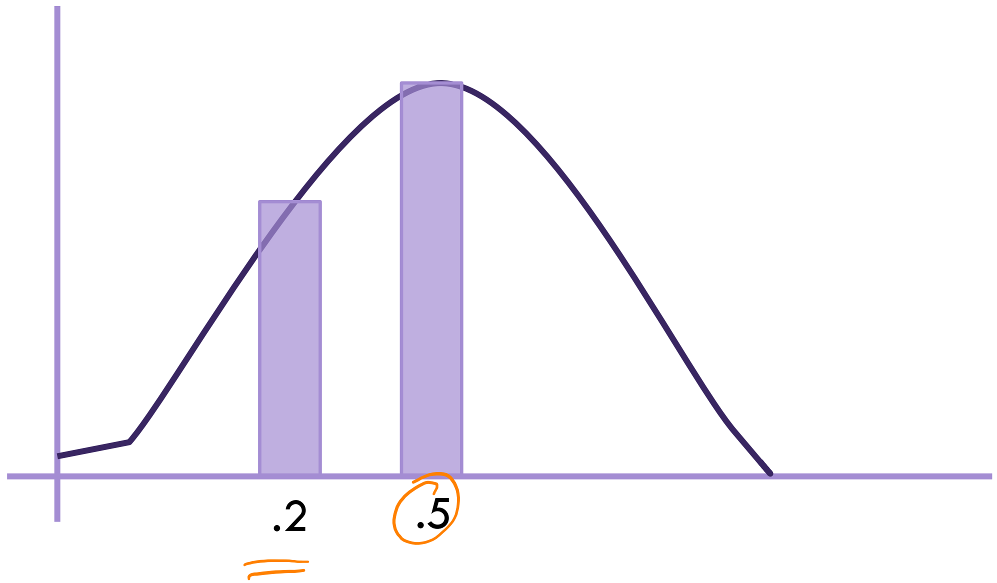
Using the PDF

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

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

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.

$$\int_{0.2}^{0.2} f_X(k)\,dk = 0$$

What about the CDF?

$F_X(k) = \sum_{x=-\infty}^{k} f_X(x)$

The Cumulative Distribution Function $F_X(k) = \mathbb{P}(X \leq k)$ is 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!

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

Continuous RVs

$$\int_{-\infty}^{b} f_X(z)dz - \int_{-\infty}^{a} f_X(z)dz$$

$F_X(k) = \mathbb{P}(X \le k) = \int_{-\infty}^{k} f_X(z)dz$

$\mathbb{P}(a \le X \le b) = \int_a^b f_X(z)dz = F_X(b) - F_X(a)$

$\mathbb{P}(X = k) = \int_k^k f_X(z)dz = 0$ (for any constant k)

Densities are not normalized to be between 0 and 1. Write out the pdf for a random real number between 0 and ½ to confirm this fact.

CDF is increasing, $\lim_{k \to -\infty} F_X(k) = 0$; $\lim_{k \to \infty} F_X(k) = 1$

Comparing Discrete and Continuous RVs

	Discrete Random Variables	**Continuous Random Variables**
Probability 0	Equivalent to impossible	All impossible events have probability 0, but not vice versa.
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?

Discrete

$E[Y] = \sum_k k \cdot p_X(k)$

For a random variable X, we define:

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

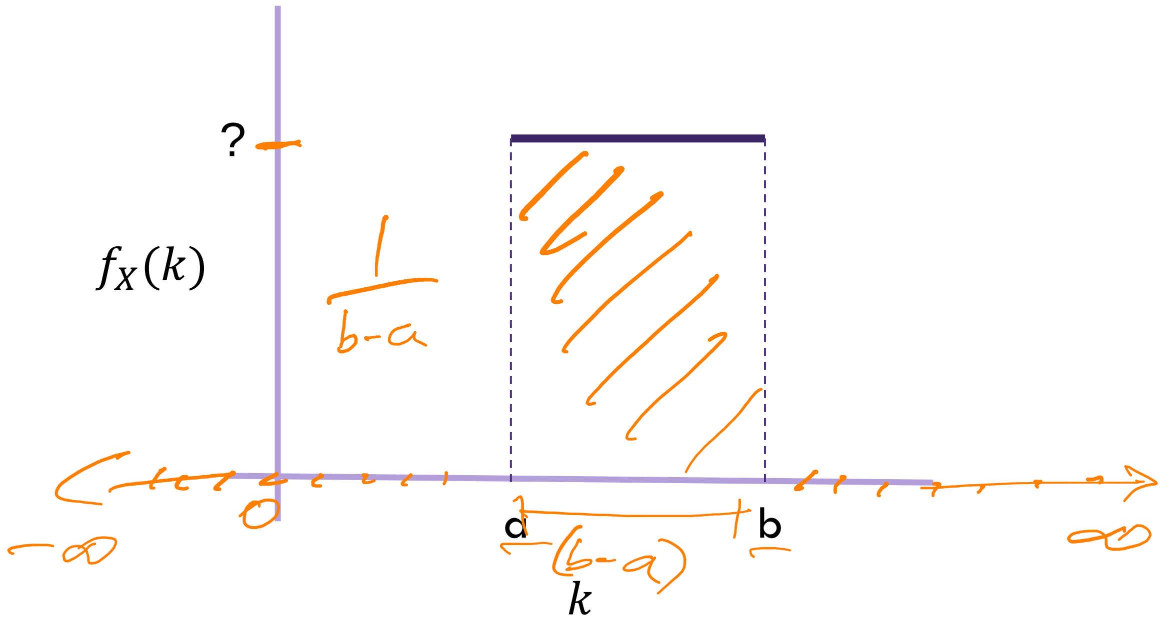
Just replace summing over the pmf with integrating the pdf.

It still represents the average value of X.

Let's calculate an expectation

Let X be a uniform random number between a and b

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



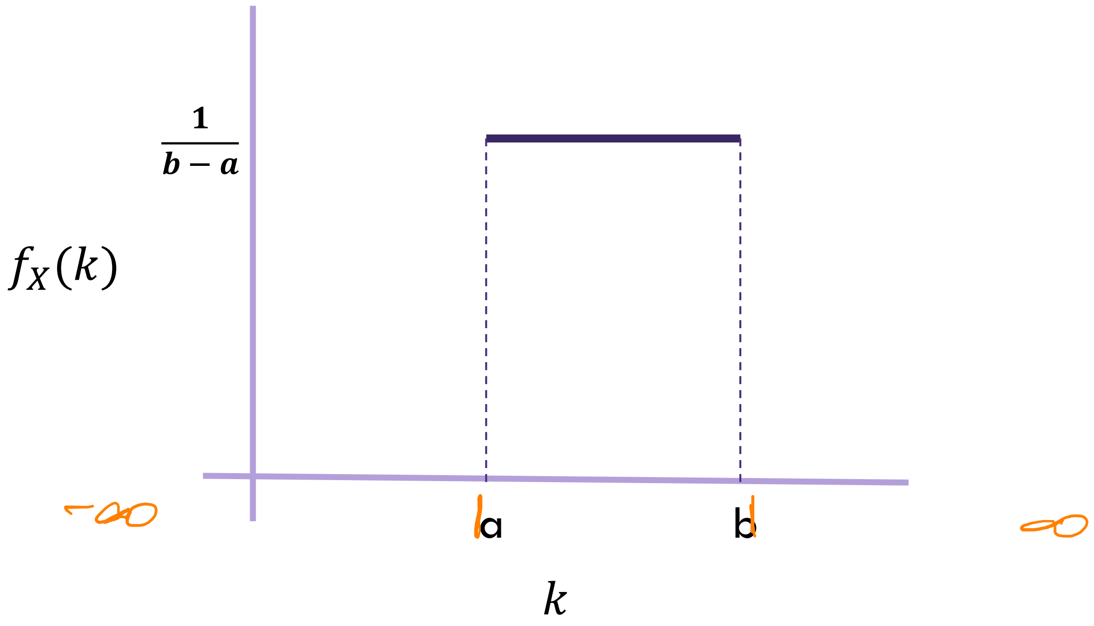
Let's calculate an expectation

Let X be a uniform random number between a and b

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

$b = \frac{1}{2}$ $\quad a = 0$

$\frac{1}{\frac{1}{2} - 0} = 2$



Let's calculate an expectation

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

$$\int_a^b x\,dx = \frac{x^2}{2}\Big|_a^b$$

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

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

$$= 0 + \int_a^b \frac{z}{b-a}\,\mathrm{dz} + 0$$

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

Linearity of Expectation

Still true!

$$\mathbb{E}[\boldsymbol{aX + bY + c}] = \boldsymbol{a}\mathbb{E}[\boldsymbol{X}] + \boldsymbol{b}\mathbb{E}[\boldsymbol{Y}] + \boldsymbol{c}$$

For all X, Y; even if they're continuous.

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

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

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

$$= a \underbrace{\int_{-\infty}^{\infty} k\, f_X(k) dk}_{E[X]} + b \underbrace{\int_{-\infty}^{\infty} f_X(k) dk}_{1}$$

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

Expectation of a function

$E[g(Y)] = \sum_k g(k) \cdot p_Y(k)$

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

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

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." LOTUS

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

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

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

$\int z^2 \, dz = \frac{z^3}{3}$

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

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

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

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

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

Let's assemble the variance

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

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

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

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

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

Variance

No surprises here

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

Continuous Uniform Distribution

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

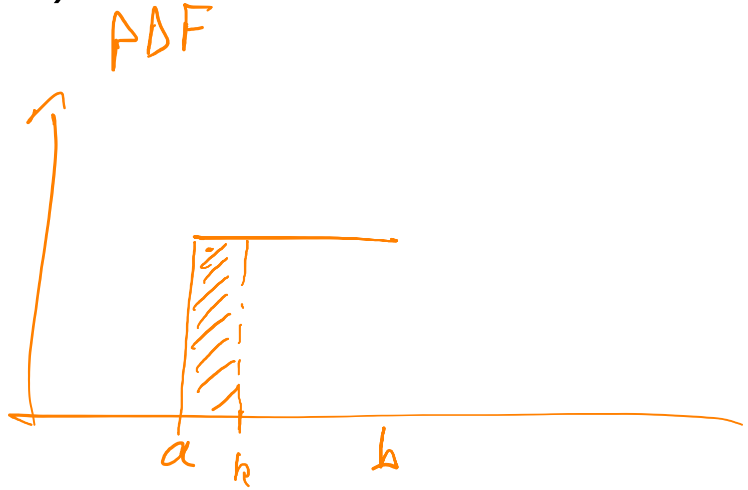
$$\text{PDF: } f_X(k) = \begin{cases} \frac{1}{b-a} & \text{if } \underline{a} \le k \le \underline{b} \\ 0 & \text{otherwise} \end{cases}$$

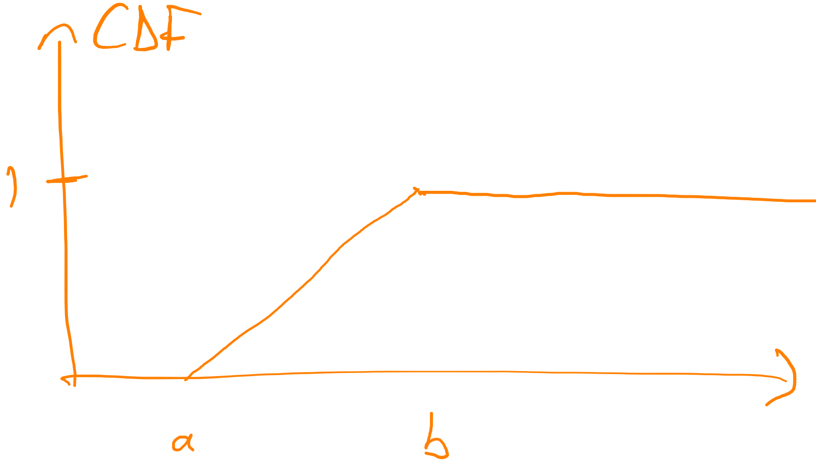
$$\text{CDF: } F_X(k) = \begin{cases} 0 & \text{if } k < a \\ \frac{k-a}{b-a} & \text{if } a \le k \le b \\ 1 & \text{if } k \ge b \end{cases}$$

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

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

$$F_X(b) = \frac{b-a}{b-a} = 1$$



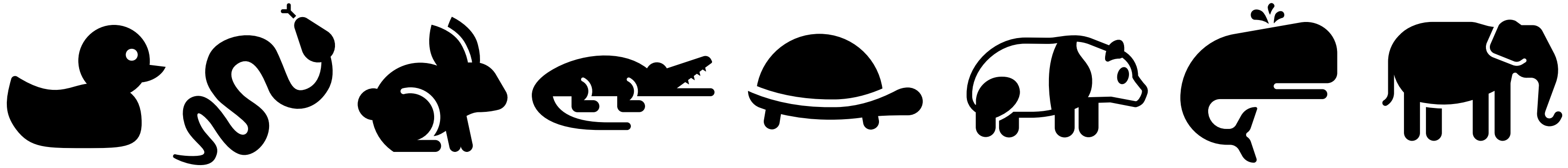


Continuous Zoo

$X \sim \text{Unif}(a, b)$	$X \sim \text{Exp}(\lambda)$	$X \sim \mathcal{N}(\mu, \sigma^2)$
$f_X(k) = \frac{1}{b-a}$ $\mathbb{E}[X] = \frac{a+b}{2}$ $\text{Var}(X) = \frac{(b-a)^2}{12}$	$f_X(k) = \lambda e^{-\lambda k}$ for $k \geq 0$ $\mathbb{E}[X] = \frac{1}{\lambda}$ $\text{Var}(X) = \frac{1}{\lambda^2}$	$f_X(k) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ $\mathbb{E}[X] = \mu$ $\text{Var}(X) = \sigma^2$

It's a smaller zoo, but it's just as much fun!

Discrete Zoo



$X \sim \mathrm{Unif}(a,b)$	$X \sim \mathrm{Ber}(p)$	$X \sim \mathrm{Bin}(n,p)$	$X \sim \mathrm{Geo}(p)$
$p_X(k) = \frac{1}{b-a+1}$	$p_X(0) = 1-p;$ $p_X(1) = p$	$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$	$p_X(k) = (1-p)^{k-1} p$
$\mathbb{E}[X] = \frac{a+b}{2}$	$\mathbb{E}[X] = p$	$\mathbb{E}[X] = np$	$\mathbb{E}[X] = \frac{1}{p}$
$\mathrm{Var}(X) = \frac{(b-a)(b-a+2)}{12}$	$\mathrm{Var}(X) = p(1-p)$	$\mathrm{Var}(X) = np(1-p)$	$\mathrm{Var}(X) = \frac{1-p}{p^2}$

$X \sim \mathrm{NegBin}(r,p)$	$X \sim \mathrm{HypGeo}(N,K,n)$	$X \sim \mathrm{Poi}(\lambda)$
$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$	$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$
$\mathbb{E}[X] = \frac{r}{p}$	$\mathbb{E}[X] = n\frac{K}{N}$	$\mathbb{E}[X] = \lambda$
$\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$	$\mathrm{Var}(X) = \frac{K(N-K)(N-n)}{N^2(N-1)}$	$\mathrm{Var}(X) = \lambda$