Random Variables

Statistics for Data Science

CSE357 - Fall 2021

Random Variables

X: A mapping from Ω to $\mathbb{R}$ that describes the question we care about in practice.

Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

Example: **Ω** = 5 coin tosses = {<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>…}

Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

Example: **Ω** = 5 coin tosses = {<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>…}

We may just care about how many tails? Thus,

X(<HHHHH>) = 0

X(<HHHTH>) = 1

X(<TTTHT>) = 4

X(<HTTTT>) = 4

Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

Example: **Ω** = 5 coin tosses = {<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>…}

We may just care about how many tails? Thus,

X(<HHHHH>) = 0

X(<HHHTH>) = 1

X(<TTTHT>) = 4

X(<HTTTT>) = 4

X only has 6 possible values: 0, 1, 2, 3, 4, 5

Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

Example: **Ω** = 5 coin tosses = {<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>…}

We may just care about how many tails? Thus,

X(<HHHHH>) = 0

X(<HHHTH>) = 1

X(<TTTHT>) = 4

X(<HTTTT>) = 4

X only has 6 possible values: 0, 1, 2, 3, 4, 5

What is the probability that we end up with $k = 4$ tails?

$\mathbf{P}(X(\omega) = k)$ where $\omega \in \boldsymbol{\Omega}$

Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

Example: **Ω** = 5 coin tosses = {<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>…}

We may just care about how many tails? Thus,

X(<HHHHH>) = 0

X(<HHHTH>) = 1

X(<TTTHT>) = 4

X(<HTTTT>) = 4

X only has 6 possible values: 0, 1, 2, 3, 4, 5

What is the probability that we end up with k = 4 tails?

$\mathbf{P}(X = k) := \mathbf{P}(\ \{\omega : X(\omega) = k\}\ )$ where $\omega \in \mathbf{\Omega}$

Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

Example: **Ω** = 5 coin tosses = {<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>…}

We may just care about how many tails? Thus,

X(<HHHHH>) = 0

X(<HHHTH>) = 1

X(<TTTHT>) = 4

X(<HTTTT>) = 4

X only has 6 possible values: 0, 1, 2, 3, 4, 5

What is the probability that we end up with $k = 4$ tails?

$\mathbf{P}(X = k) := \mathbf{P}(\ \{\omega : X(\omega) = k\}\ )$ where $\omega \in \mathbf{\Omega}$

X(ω) = 4 for 5 out of 32 sets in **Ω**. Thus, assuming a fair coin, $\mathbf{P}(X = 4) = 5/32$

Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

Example: **Ω** = 5 coin tosses = $\{<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH> \ldots\}$

We may just care about how many tails? Thus,

$$X(<HHHHH>) = 0$$

$$X(<HHHTH>) = 1$$

$$X(<TTTHT>) = 4$$

$$X(<HTTTT>) = 4$$

X only has 6 possible values: 0, 1, 2, 3, 4, 5

What is the probability that we end up with $k = 4$ tails?

$$\mathbf{P}(X = k) := \mathbf{P}(\ \{\omega : X(\omega) = k\}\ ) \qquad \text{where } \omega \in \mathbf{\Omega}$$

$X(\omega)$ = 4 for 5 out of 32 sets in **Ω**. Thus, assuming a fair coin, $\mathbf{P}(X = 4) = 5/32$

(Not a variable, but a function that we end up notating a lot like a variable)

9-2-2021

Normal Distribution

Programming Statistics -- Numpy and Random Variables

Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

Example: **Ω** = 5 coin tosses = {**<HHHHH>, <HHHHT>, <HHHTH>, <HHHTH>**…}

We may just care about how many tails? Thus,

X(<HHHHH>) = 0

X(<HHHTH>) = 1

X(<TTTHT>) = 4

X(<HTTTT>) = 4

X is a *discrete random variable* if it takes only a countable number of values.

X only has 6 possible values: 0, 1, 2, 3, 4, 5

What is the probability that we end up with k = 4 tails?

$\mathbf{P}(X = k) := \mathbf{P}(\ \{\omega : X(\omega) = k\}\ )$ where $\omega \in \mathbf{\Omega}$

X(ω) = 4 for 5 out of 32 sets in **Ω**. Thus, assuming a fair coin, $\mathbf{P}(X = 4) = 5/32$

(Not a variable, but a function that we end up notating a lot like a variable)

Random Variables

$\mathbf{X}$: A mapping from Ω to $\mathbb{R}$ that describes the question we care about in practice.

$\mathbf{X}$ is a *continuous random variable* if it can take on an infinite number of values between any two given values.

$\mathbf{X}$ is a *discrete random variable* if it takes only a countable number of values.

Random Variables

$\mathbf{X}$: A mapping from $\mathbf{\Omega}$ to $\mathbb{R}$ that describes the question we care about in practice.

Example: $\mathbf{\Omega}$ = inches of snowfall = $[0, \infty) \subseteq \mathbb{R}$

$\mathbf{X}$ is a *continuous random variable* if it can take on an infinite number of values between any two given values.

Random Variables

$\mathbf{X}$: A mapping from $\mathbf{\Omega}$ to $\mathbb{R}$ that describes the question we care about in practice.

Example: $\mathbf{\Omega}$ = inches of snowfall = $[0, \infty) \subseteq \mathbb{R}$

$\mathbf{X}$ is a *continuous random variable* if it can take on an infinite number of values between any two given values.

X amount of inches in a snowstorm

$\mathbf{X}(\omega) = \omega$

Random Variables

$\mathbf{X}$: A mapping from $\mathbf{\Omega}$ to $\mathbb{R}$ that describes the question we care about in practice.

Example: $\mathbf{\Omega}$ = inches of snowfall = $[0, \infty) \subseteq \mathbb{R}$

$\mathbf{X}$ is a *continuous random variable* if it can take on an infinite number of values between any two given values.

X amount of inches in a snowstorm

$\mathbf{X}(\omega) = \omega$

What is the probability we receive (at least) a *inches?*

$\mathbf{P}(X \geq a) := \mathbf{P}(\ \{\omega : X(\omega) \geq a\}\ )$

What is the probability we receive between a *and* b *inches?*

$\mathbf{P}(a \leq X \leq b) := \mathbf{P}(\ \{\omega : a \leq X(\omega) \leq b\}\ )$

Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

Example: **Ω** = inches of snowfall = [0, ∞) ⊆ ℝ

X is a *continuous random variable* if it can take on an infinite number of values between any two given values.

X amount of inches in a snowstorm

X(ω) = ω

P(X = i) := 0, for all i ∈ **Ω**

(probability of receiving exactly i inches of snowfall is zero)

What is the probability we receive (at least) a inches?

P(X ≥ a) := **P**({ω : X(ω) ≥ a})

What is the probability we receive between a and b inches?

P(a ≤ X ≤ b) := **P**({ω : a ≤ X(ω) ≤ b})

X is a *continuous random variable* if it can take on an infinite number of values between any two given values.

X is a *continuous random variable* if there exists a function *fx* such that:

$$f_X(x) \geq 0, \text{for all } x \in X,$$

...

X is a *continuous random variable* if it can take on an infinite number of values between any two given values.

fx **: "probability density function" (pdf)**

X is a *continuous random variable* if there exists a function *fx* such that:

$$f_X(x) \geq 0, \text{for all } x \in X,$$

...

X is a *continuous random variable* if it can take on an infinite number of values between any two given values.

fx **: "probability density function" (pdf)**

X is a *continuous random variable* if there exists a function *fx* such that:

$$f_X(x) \geq 0, \text{for all } x \in X,$$

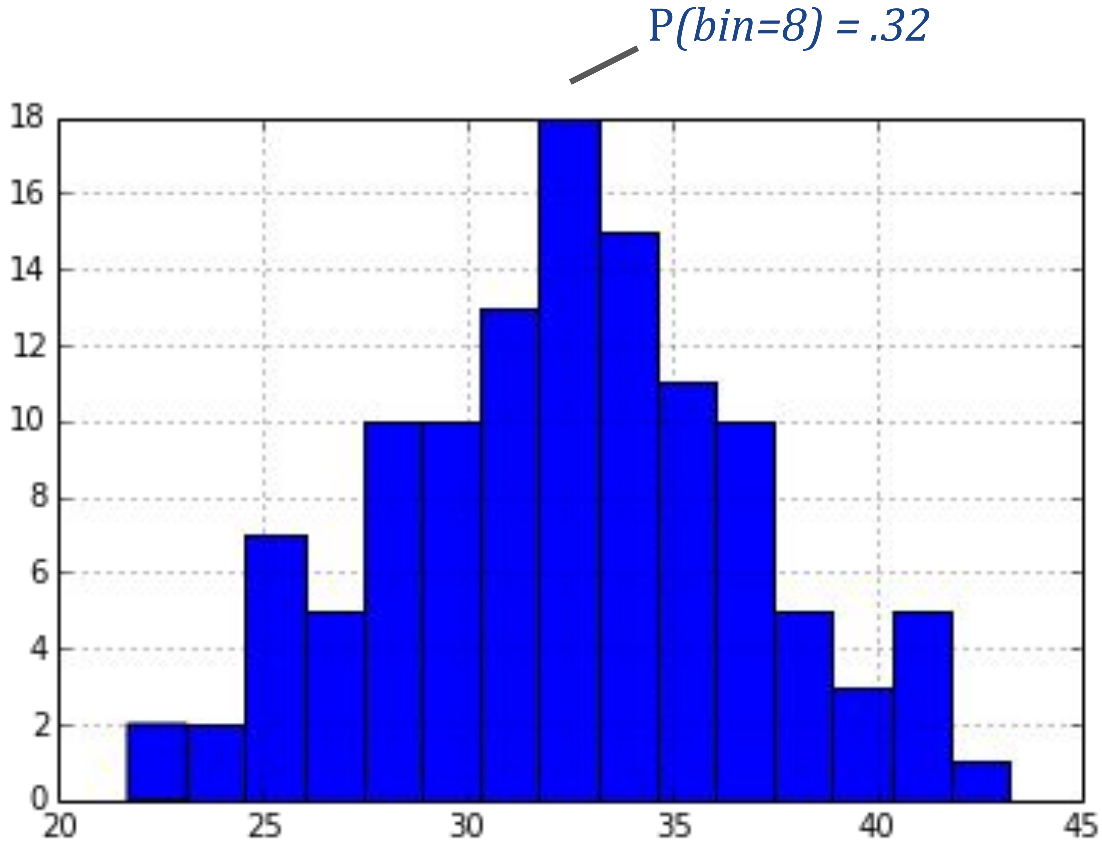
How to model?

continuous random variable

Discretize them!
(group into discrete bins)

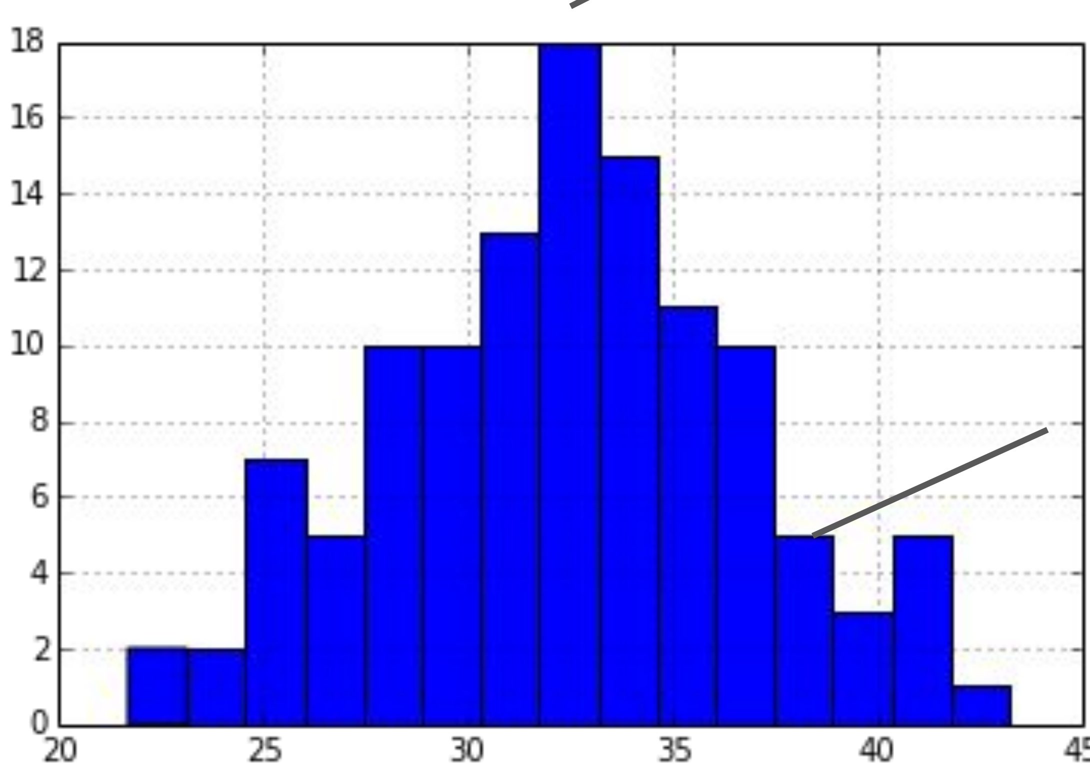
How to model?

continuous random variable



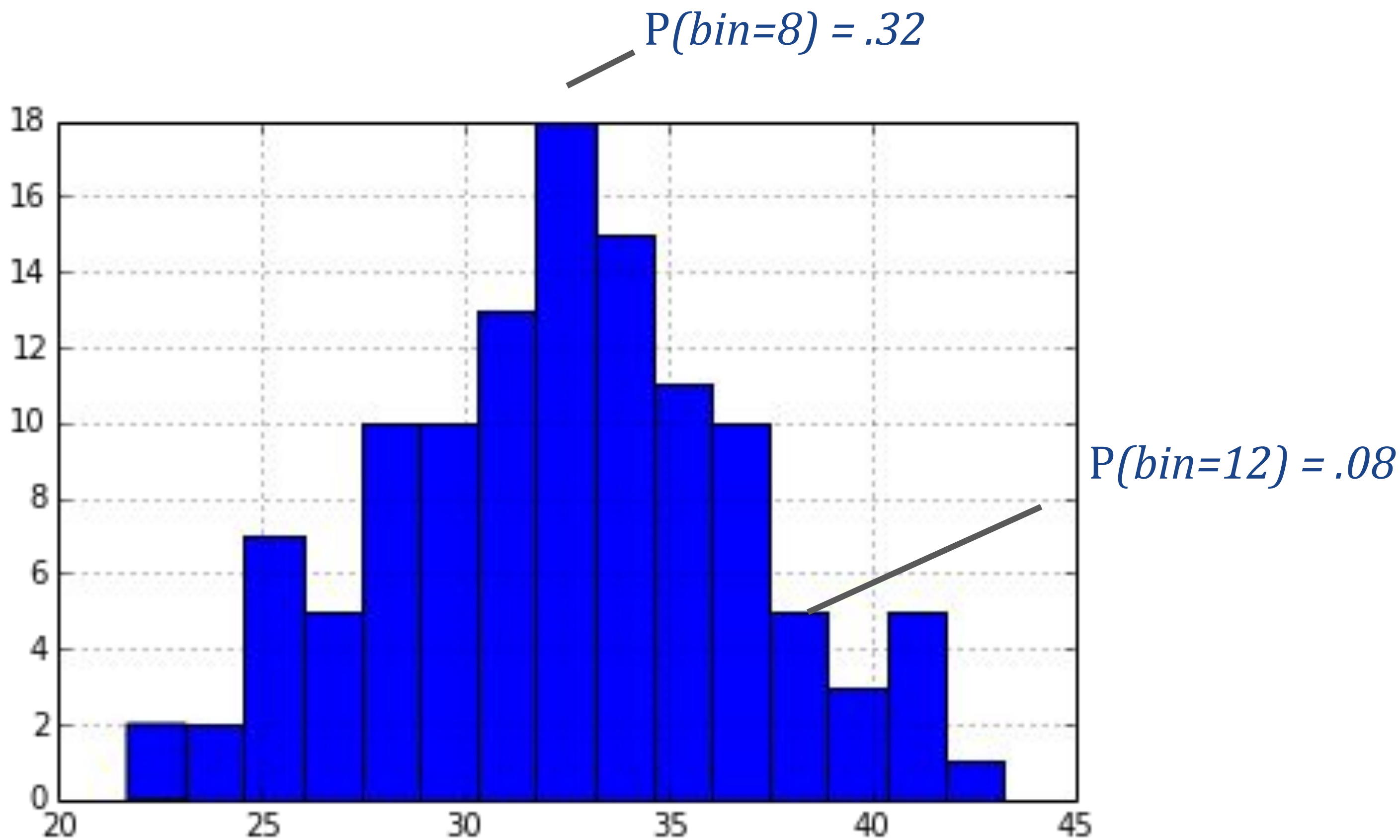
continuous random variable

P(bin=8) = .32



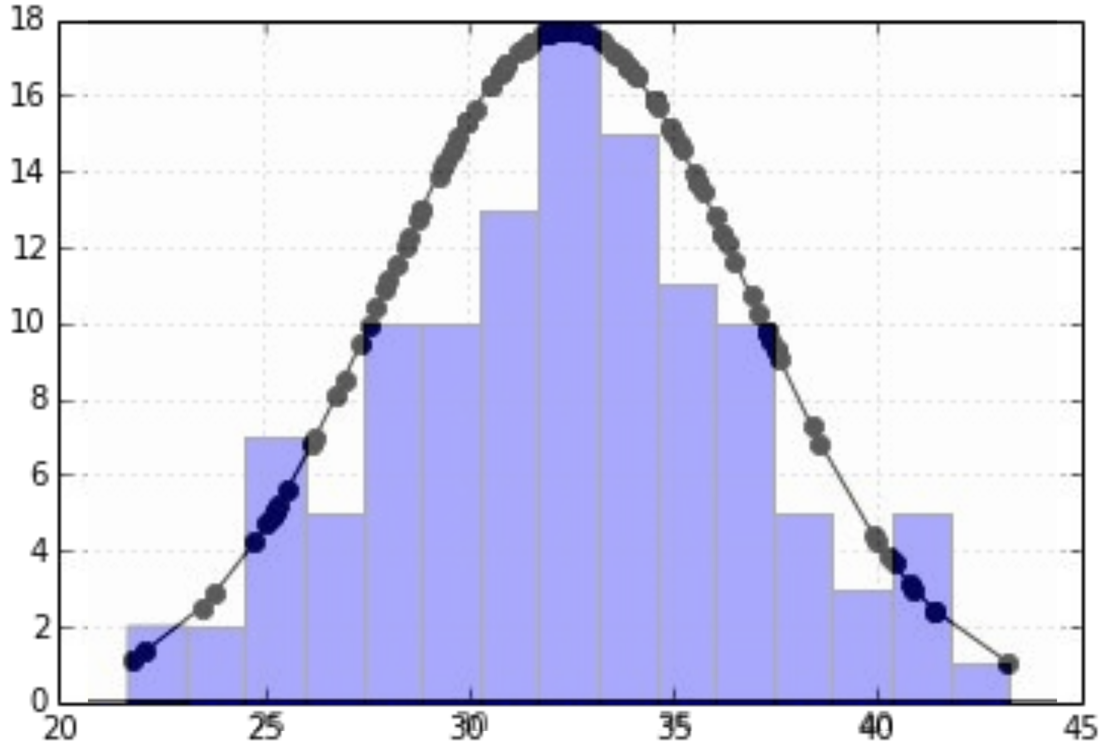
P(bin=12) = .08

continuous random variable

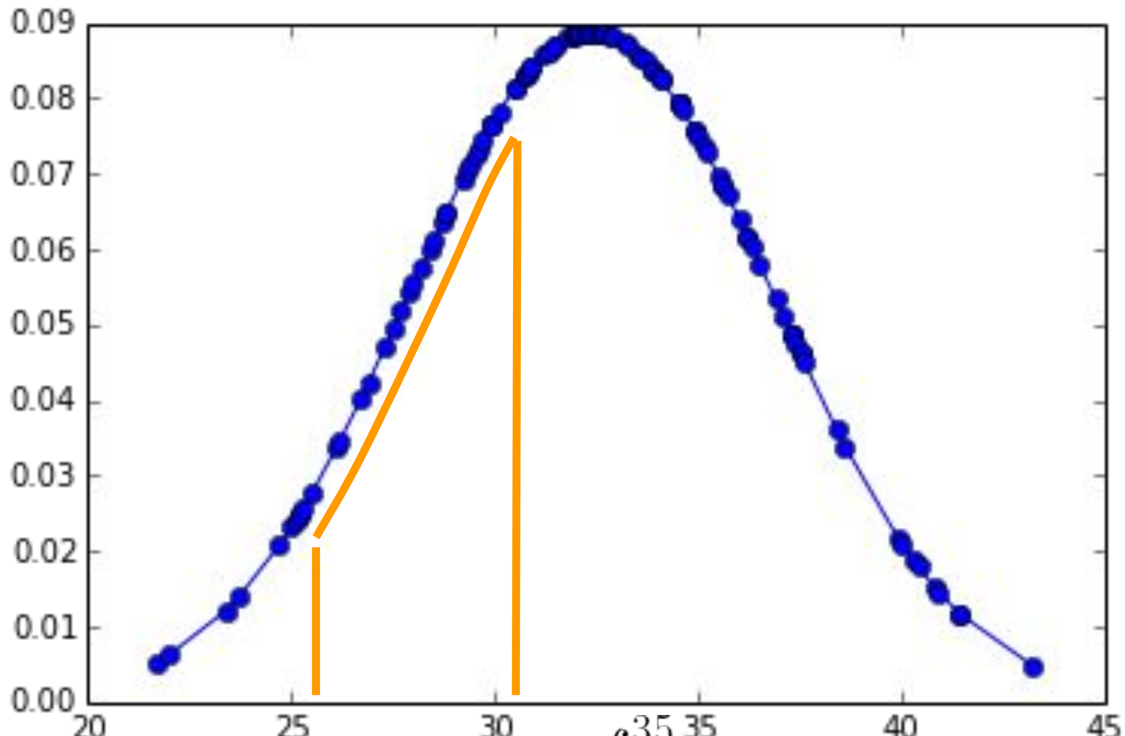


But aren't we throwing away information?

continuous random variable



Continuous Distribution



$$P(25 < X < 35) = \int_{25}^{35} fx(x)dx$$

Continuous Distribution

fx **: “probability density function” (pdf)**

X is a *continuous random variable* if there exists a function *fx* such that:

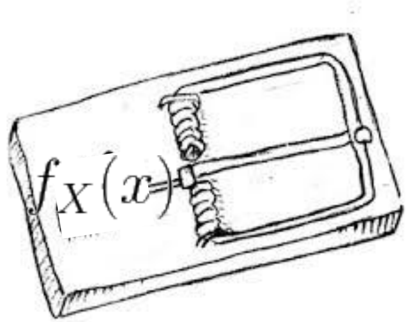
$$f_X(x) \geq 0, \text{for all } x \in X,$$

$$\int_{-\infty}^{\infty} f_X(x)dx = 1, \quad \text{and}$$

$$\mathrm{P}(a < X < b) = \int_a^b f_X(x)dx$$

Continuous Distribution

Common Trap

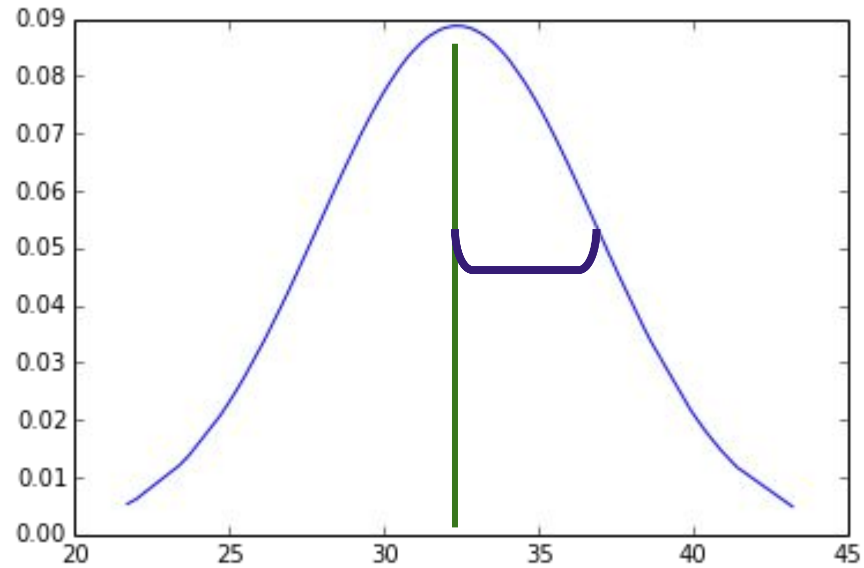


- $f_X(x)$ does not yield a probability
  - $\int_a^b f_X(x)dx$ does
  - x may be anything ($\mathbb{R}$)
    - thus, $f_X(x)$ may be > 1

Continuous Distribution

Common *pdfs*: Normal(μ, σ^2)

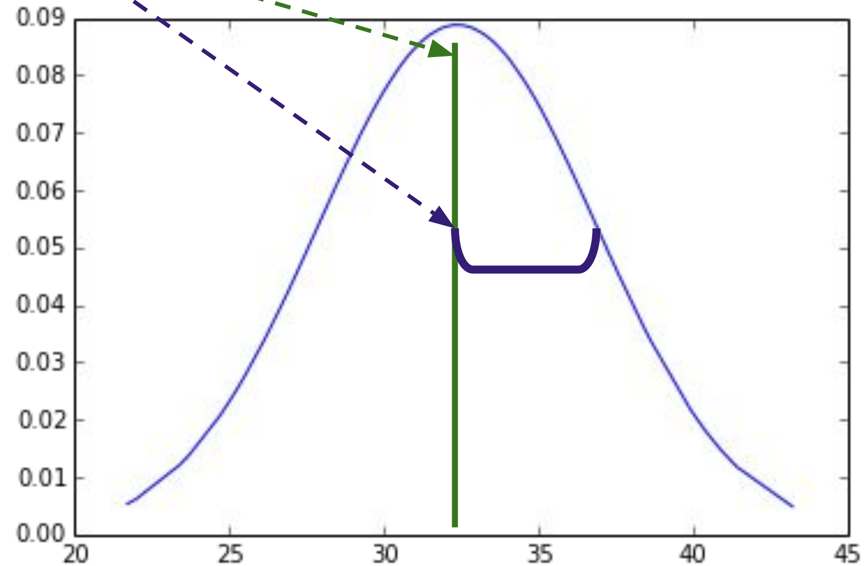
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Continuous Distribution

Common *pdfs*: Normal(μ, σ^2)

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



μ: mean (or “center”)

= expectation

σ^2: variance,

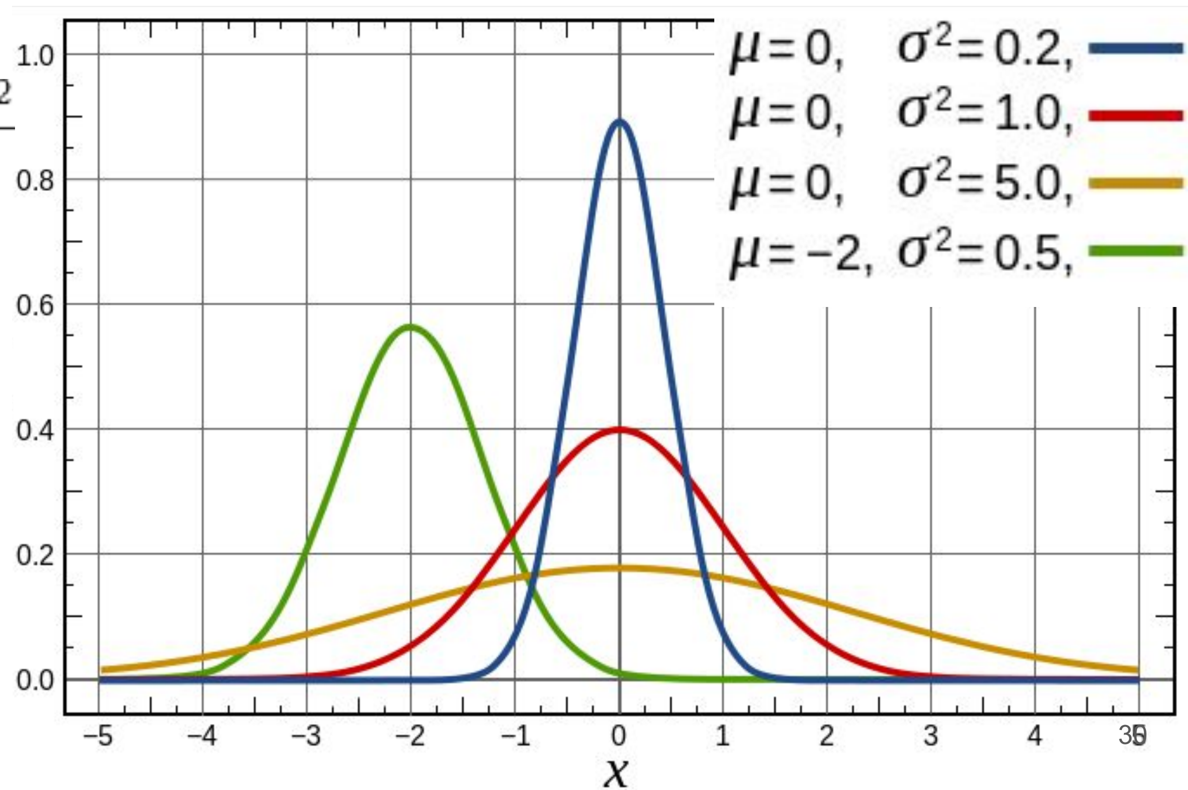
σ: standard deviation

Continuous Distribution

Common *pdfs*: Normal(μ, σ^2)

Credit: Wikipedia

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



μ: mean (or "center")

= expectation

σ^2: variance,

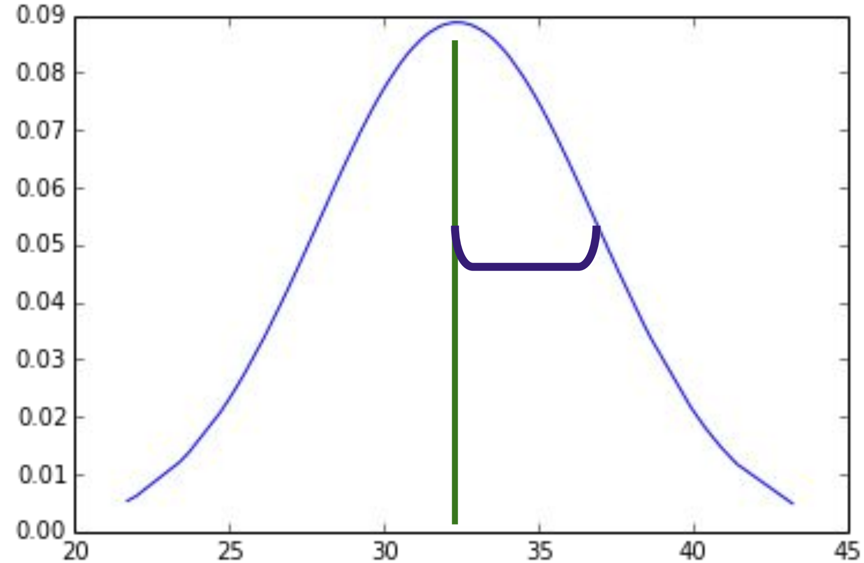
σ: standard deviation

Continuous Distribution

Common *pdfs*: Normal(μ, σ^2)

$X \sim \text{Normal}(\mu, \sigma^2)$, examples:

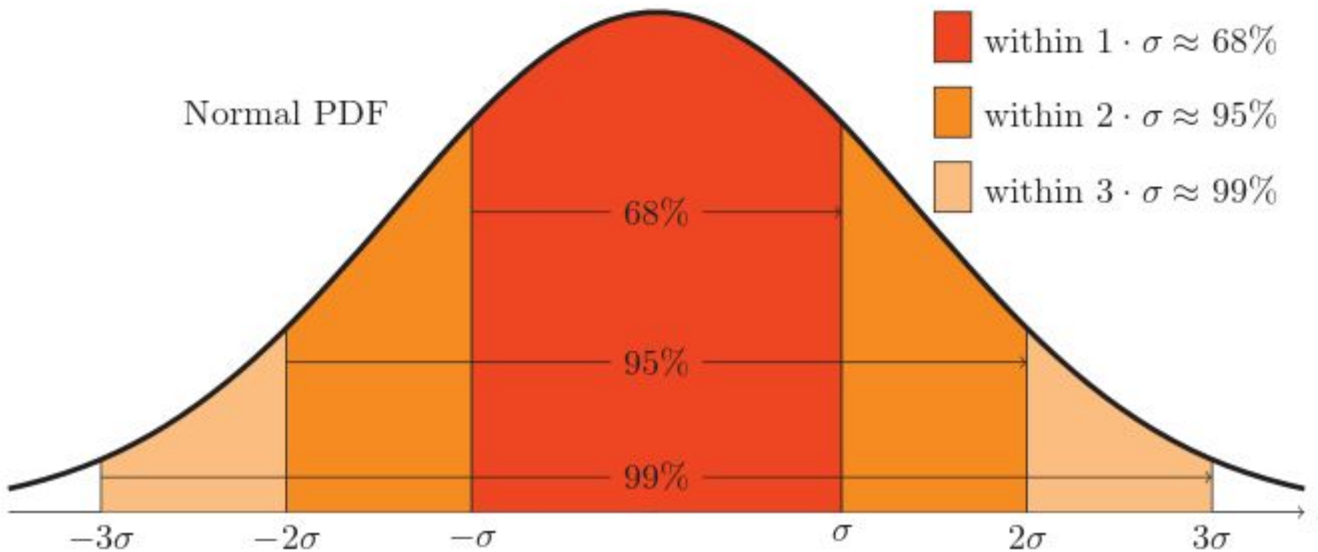
- height
- intelligence/ability
- **measurement error**
- averages (or sum) of
  lots of random variables



Continuous Distribution

Common *pdf*s: Normal(0, 1)

$$P(-1 \leq Z \leq 1) \approx .68, \quad P(-2 \leq Z \leq 2) \approx .95, \quad P(-3 \leq Z \leq 3) \approx .99$$



Credit: MIT Open Courseware: Probability and Statistics

Continuous Distribution

Common *pdfs*: Normal(0, 1) ("standard normal")

How to "standardize" any normal distribution:

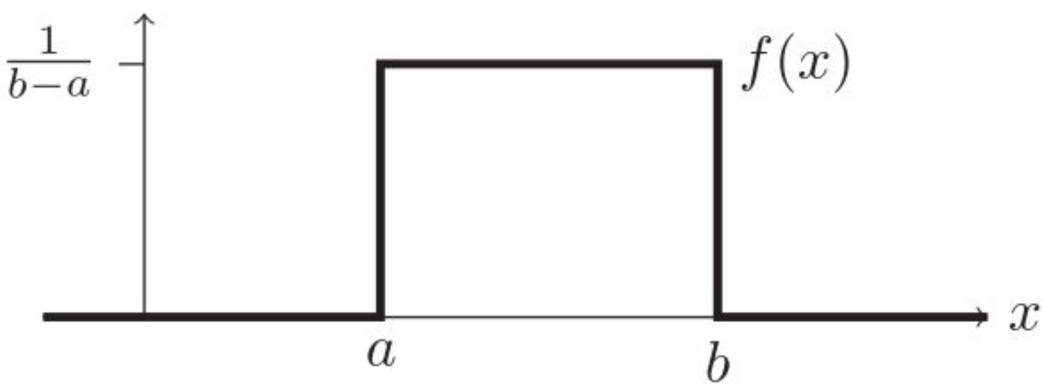
1. subtract the mean, μ (aka "mean centering")
2. divide by the standard deviation, σ

$z = (x - \mu) / \sigma$, (aka "z score")

Continuous Distribution

Common *pdfs*: Uniform(a, b)

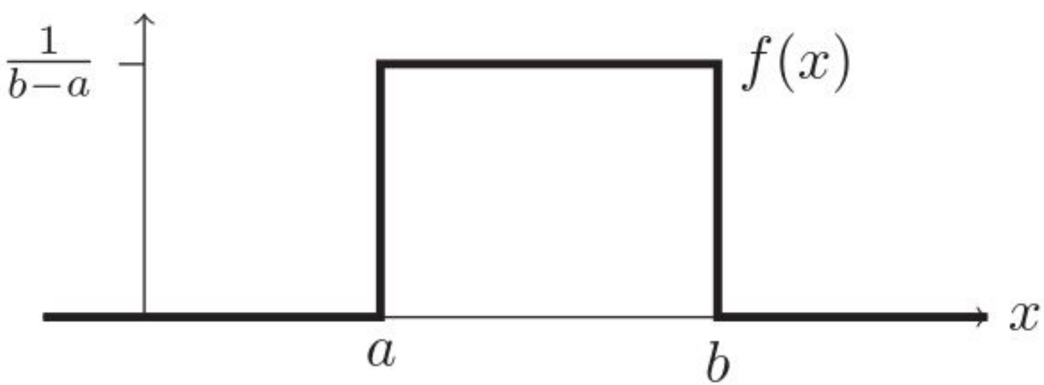
$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$



Continuous Distribution

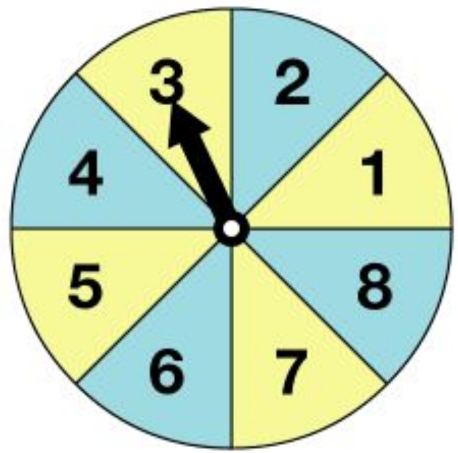
Common *pdfs*: Uniform(a, b)

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$



X ~ Uniform(a, b), examples:

- spinner in a game
- random number generator
- analog to digital rounding error



Continuous Distribution

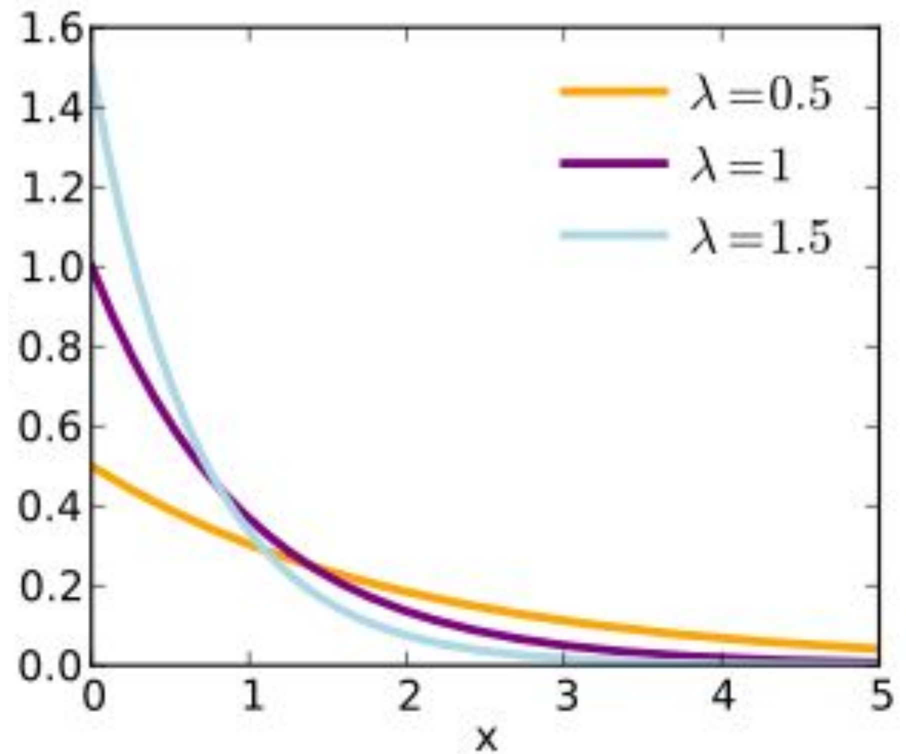
Common *pdf*s: Exponential(λ)

$$f_X(x) = \lambda e^{-\lambda x}, x > 0$$

λ: rate or inverse scale

β: scale $\left(\lambda = \frac{1}{\beta} \right)$

Credit: Wikipedia



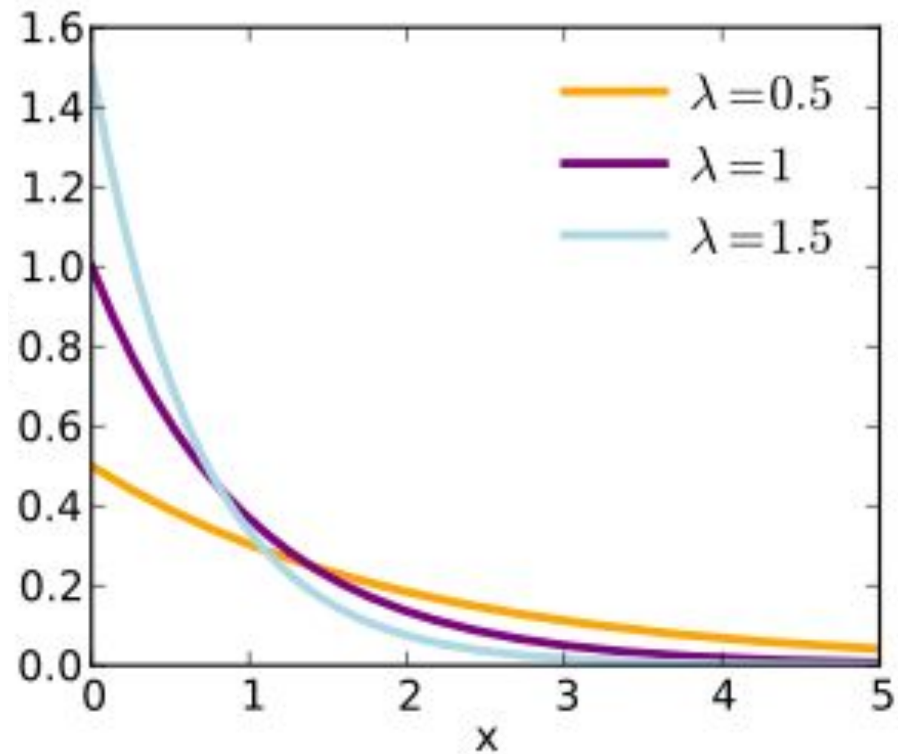
Continuous Distribution

Common *pdf*s: Exponential(λ)

$X \sim \text{Exp}(\lambda)$, examples:

- lifetime of electronics
- waiting times between rare events (e.g. waiting for a taxi)
- recurrence of words across documents

Credit: Wikipedia



Continuous Distribution: CDF

For a given random variable X, the *cumulative distribution function* (CDF), *Fx:* $\mathbb{R} \rightarrow [0, 1]$, is defined by:

$$F_X(x) = \mathrm{P}(X \leq x)$$

fx **:**
probability density function (pdf)

$$f_X(x) \geq 0, \text{ for all } x \in X,$$

$$\int_{-\infty}^{\infty} f_X(x)dx = 1, \quad \text{and}$$

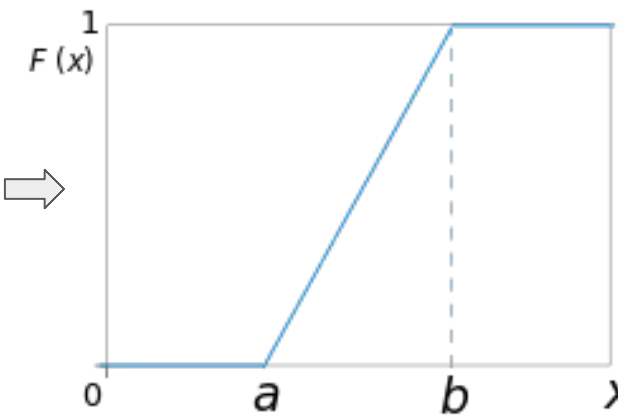
$$\mathrm{P}(a < X < b) = \int_a^b f_X(x)dx$$

Continuous Distribution: CDF

For a given random variable X, the *cumulative distribution function* (CDF), *Fx:* $\mathbb{R} \rightarrow [0, 1]$, is defined by:

$$F_X(x) = \mathrm{P}(X \leq x)$$

Uniform ⇨

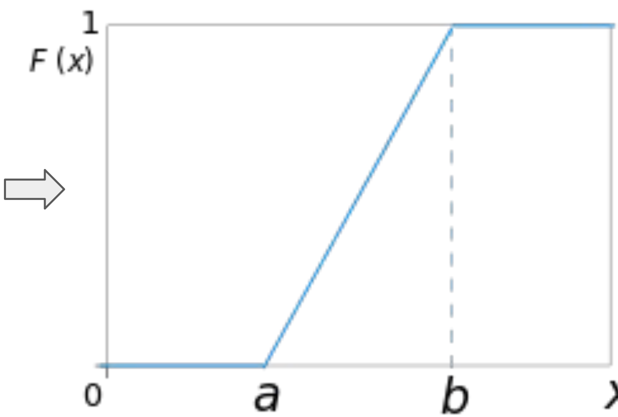


Continuous Distribution: CDF

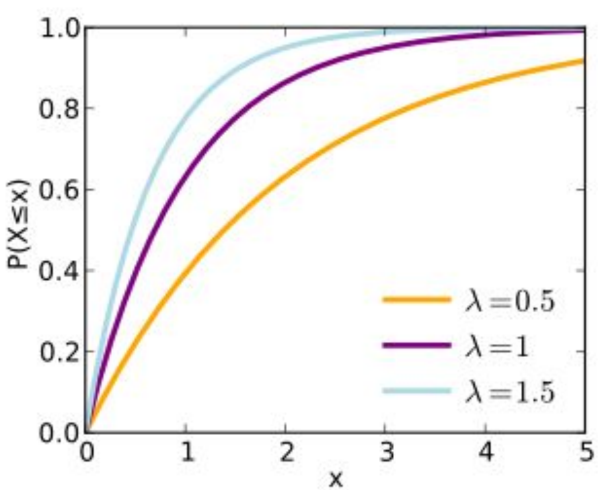
For a given random variable X, the *cumulative distribution function* (CDF), *Fx:* $\mathbb{R} \rightarrow [0, 1]$, is defined by:

$$F_X(x) = \mathrm{P}(X \leq x)$$

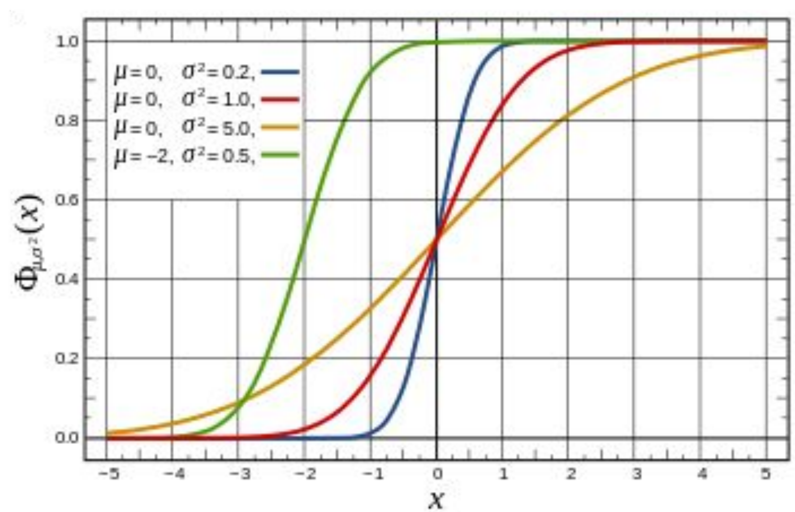
Uniform



Exponential



Normal

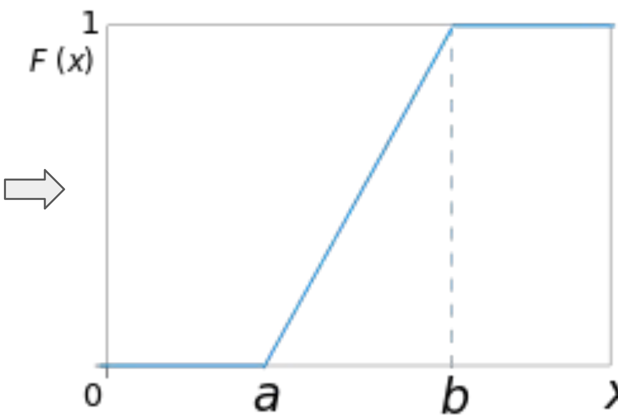


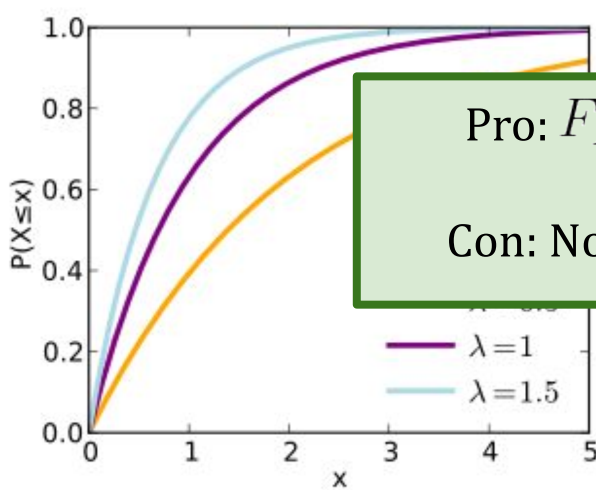
Continuous Distribution: CDF

For a given random variable X, the *cumulative distribution function* (CDF), *Fx:* $\mathbb{R} \rightarrow [0, 1]$, is defined by:

$$F_X(x) = \mathrm{P}(X \leq x)$$

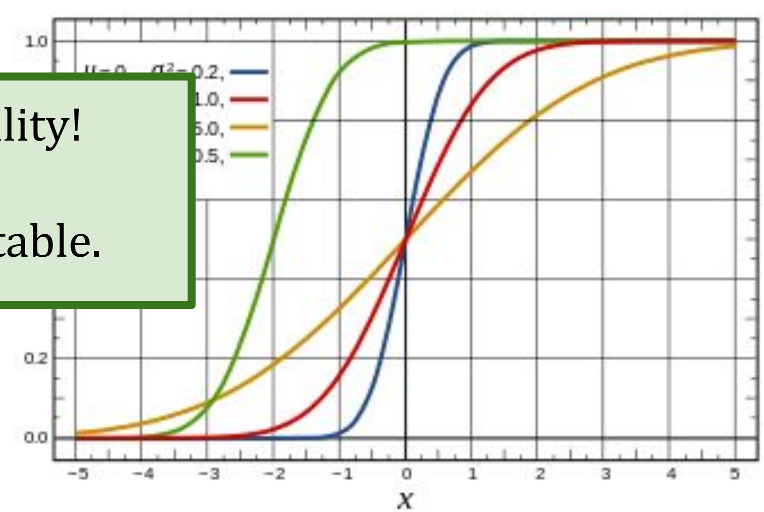
Uniform ⇨





Pro: $F_X(x)$ yields a probability!

Con: Not intuitively interpretable.



Continuous Distribution

How to decide which pdf is best for my data?

Look at a *non-parametric* curve estimate:
(If you have lots of data)

- Histogram
- Kernel Density Estimator

Continuous Distribution

How to decide which pdf is best for my data?

Look at a *non-parametric* curve estimate:
(If you have lots of data)

- Histogram
- **Kernel Density Estimator**

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)$$

K: kernel function, *h:* bandwidth

(for every data point, draw *K* and add to density)

Continuous Distribution

How to decide which pdf is best for my data?

Look at a *non-parametric* curve estimate:

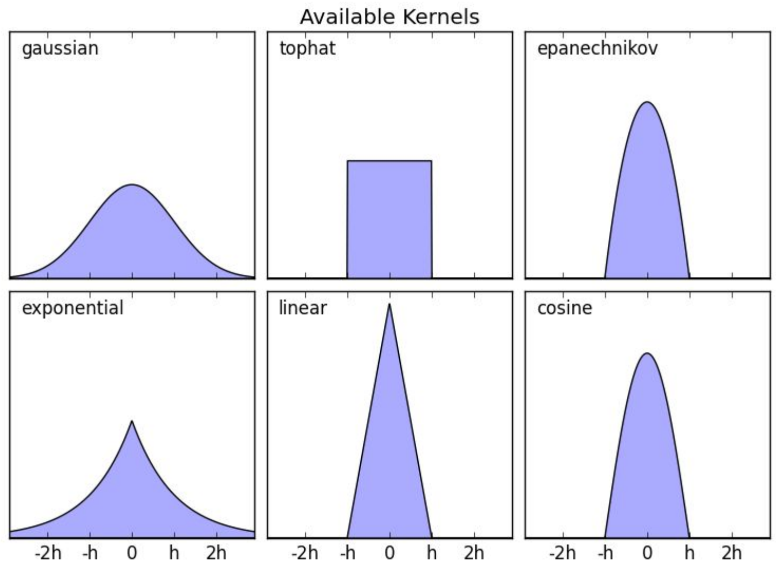
(If you have lots of data)

- Histogram
- **Kernel Density Estimator**

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - X_i}{h}\right)$$

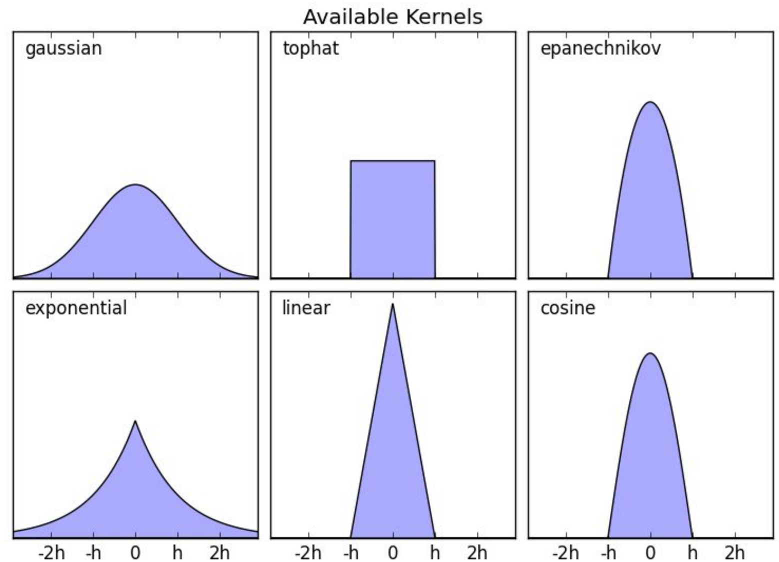
K: kernel function, *h:* bandwidth

(for every data point, draw *K* and add to density)



Continuous Distribution

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)$$

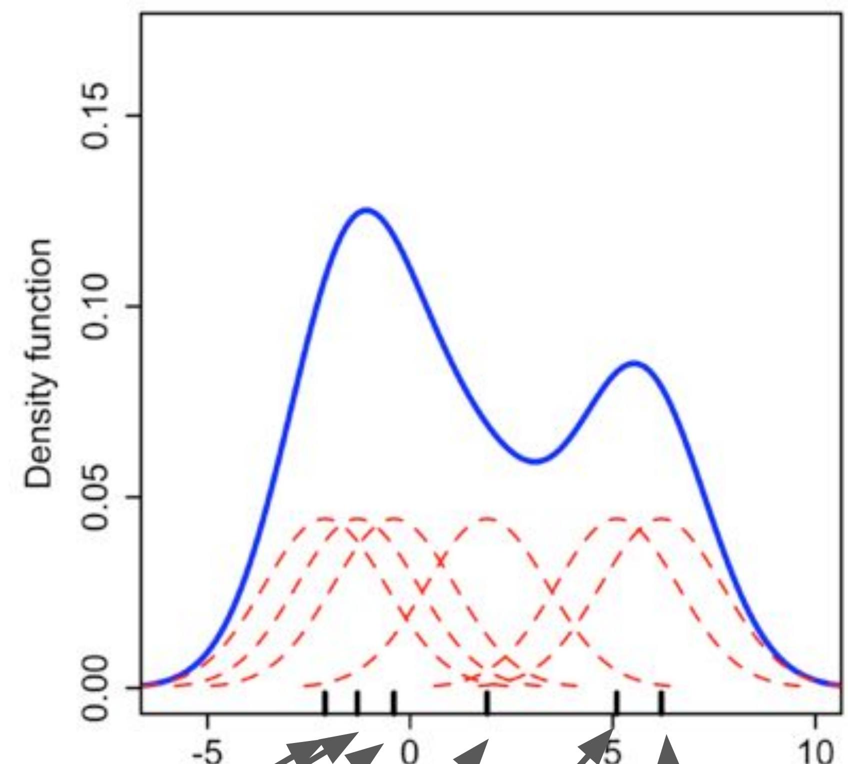


Continuous Distribution

just like a pdf, this function takes in an x and returns the appropriate y on an estimated distribution curve

to figure out y for a given x, take the sum of where each where each kernel (a density plot for each data point in the original X) puts that x.

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)$$



Random Variables

X: A mapping from **Ω** to ℝ that describes the question we care about in practice.

X is a *continuous random variable* if it can take on an infinite number of values between any two given values.

X is a *discrete random variable* if it takes only a countable number of values.

Amount of snowfall

Amount of sales of a blue case

Discrete Distribution

For a given *discrete* random variable X, *probability mass function* (*pmf*), *fx:* $\mathbb{R} \to [0, 1]$, is defined by:

$$f_X(x) = \mathrm{P}(X = x)$$

X is a *discrete random variable* if it takes only a countable number of values.

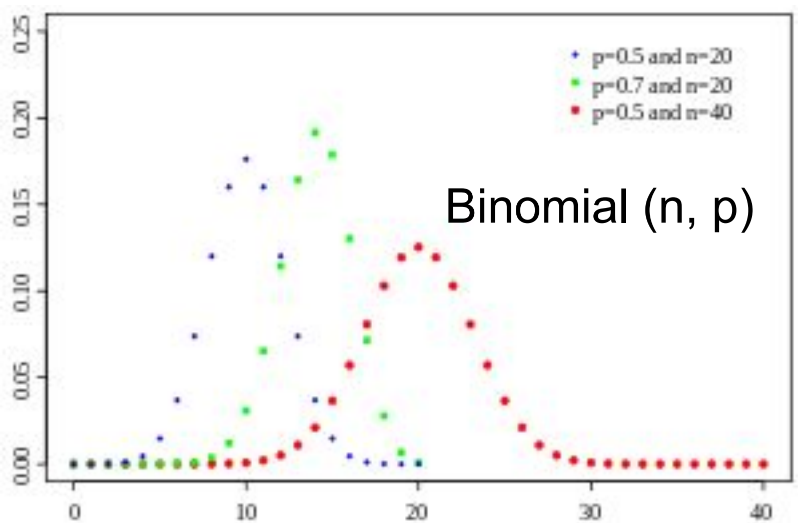
Amount of sales of a blue case

Was a single sale a blue case: {0, 1}

Discrete Distribution

For a given *discrete* random variable X, *probability mass function* (*pmf*), *fx*: $\mathbb{R} \rightarrow [0, 1]$, is defined by:

$$f_X(x) = \mathrm{P}(X = x)$$



X is a *discrete random variable* if it takes only a countable number of values.

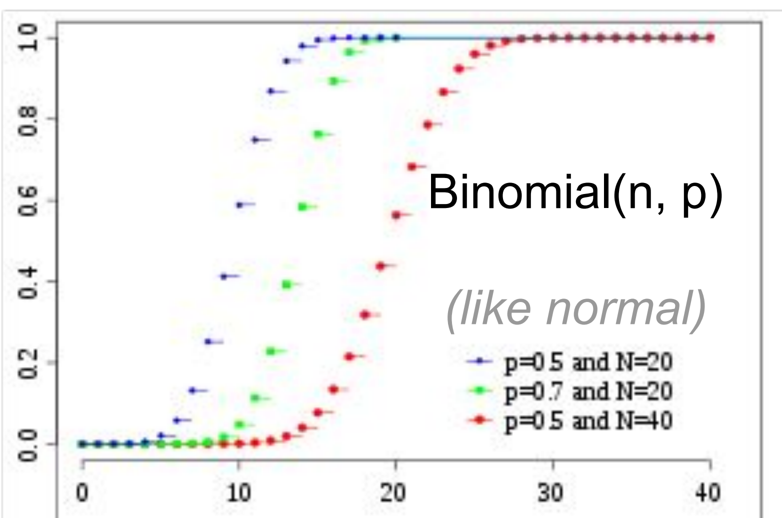
Amount of sales of a blue case

Was a single sale a blue case: {0, 1}

Discrete Distribution

For a given random variable X, the *cumulative distribution function* (CDF), *Fx:* $\mathbb{R} \to [0, 1]$, is defined by:

$$F_X(x) = \mathrm{P}(X \leq x)$$



X is a *discrete random variable* if it takes only a countable number of values.

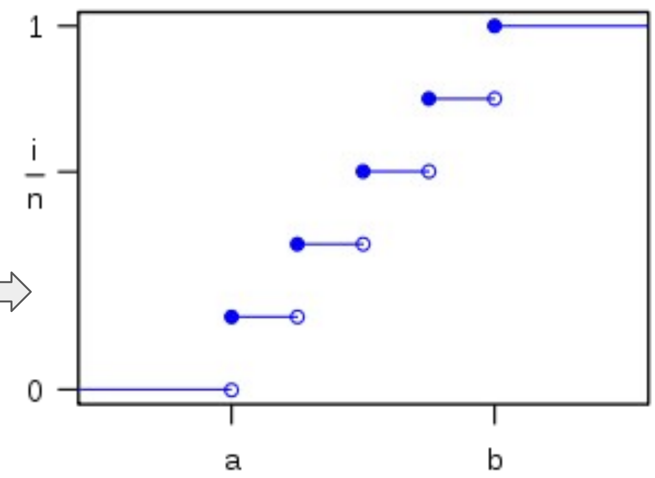
Amount of sales of a blue case

Discrete Distribution

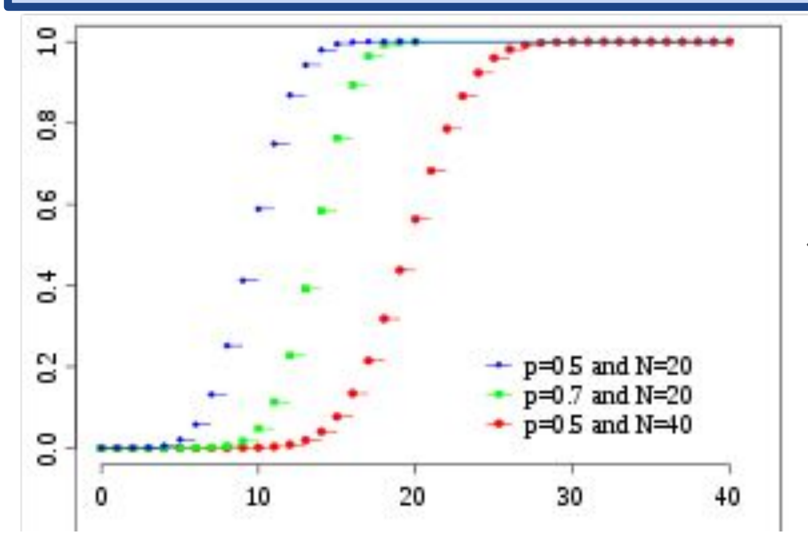
For a given random variable X, the *cumulative distribution function* (CDF), *Fx:* $\mathbb{R} \rightarrow [0, 1]$, is defined by:

$$F_X(x) = \mathrm{P}(X \leq x)$$

Discrete Uniform ⇨



X is a *discrete random variable* if it takes only a countable number of values.



⇦ Binomial (n, p)

(like normal)

Discrete Distribution

For a given random variable X, the *cumulative distribution function* (CDF), *Fx:* $\mathbb{R} \rightarrow [0, 1]$, is defined by:

$$F_X(x) = \mathrm{P}(X \leq x)$$

For a given discrete random variable X, *probability mass function* (*pmf*), *fx:* $\mathbb{R} \rightarrow [0, 1]$, is defined by:

$$f_X(x) = \mathrm{P}(X = x)$$

X is a *discrete random variable* if it takes only a countable number of values.

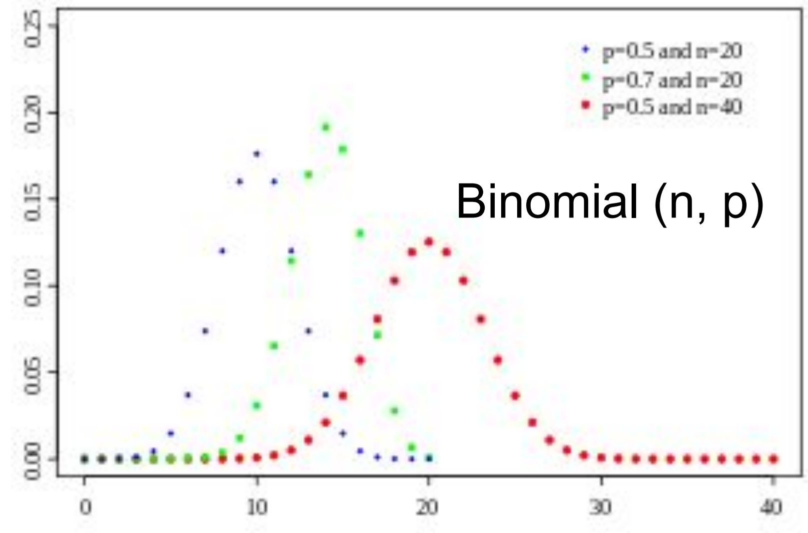
Discrete Distribution

For a given random variable X, the *cumulative distribution function* (CDF), *Fx:* $\mathbb{R} \rightarrow [0, 1]$, is defined by:

$$F_X(x) = \mathrm{P}(X \leq x)$$

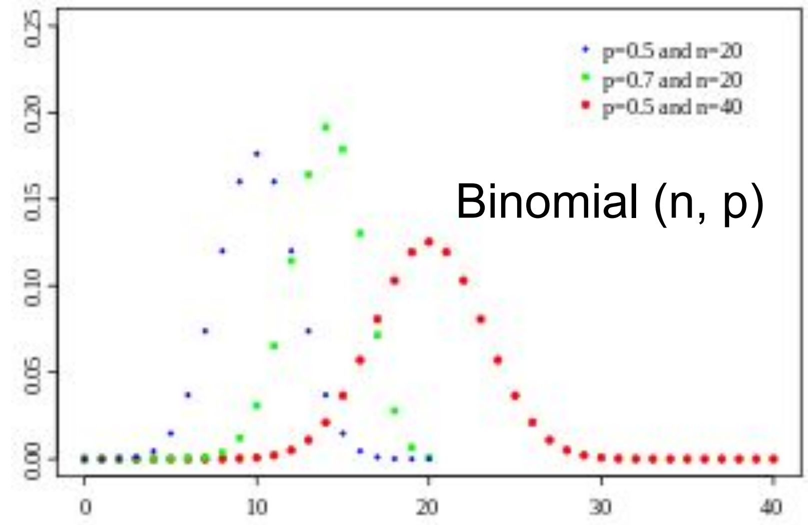
For a given discrete random variable X, *probability mass function* (*pmf*), *fx:* $\mathbb{R} \rightarrow [0, 1]$, is defined by:

$$f_X(x) = \mathrm{P}(X = x)$$



X is a *discrete random variable* if it takes only a countable number of values.

Discrete Distribution



For a given random variable X, the *cumulative distribution function* (CDF), *Fx:* $\mathbb{R} \rightarrow [0, 1]$, is defined by:

$$F_X(x) = \mathrm{P}(X \leq x)$$

For a given discrete random variable X, *probability mass function* (*pmf*), *fx:* $\mathbb{R} \rightarrow [0, 1]$, is defined by:

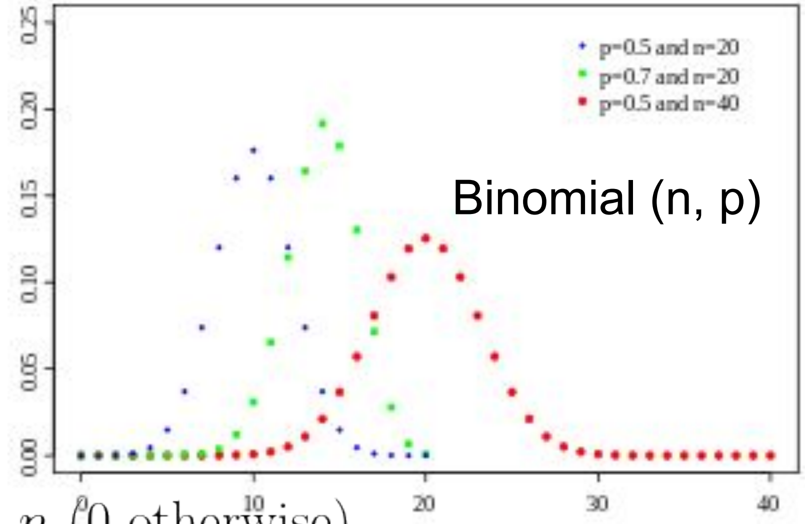
$$f_X(x) = \mathrm{P}(X = x)$$

X is a *discrete random variable* if it takes only a countable number of values.

$$\sum_i f_X(x) = 1$$

$$F_X(f) = \mathrm{P}(X \leq x) = \sum_{x_i \leq x} f_X(x)$$

Discrete Distribution



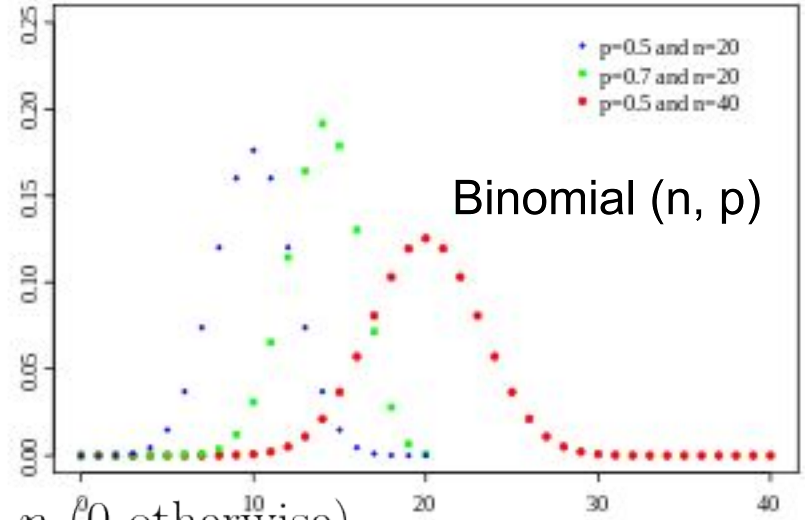
Common Discrete Random Variables

- Binomial(n, p)

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ if } 0 \leq x \leq n \text{ (0 otherwise)}$$

example: number of heads after n coin flips (p, probability of heads)

Discrete Distribution



Common Discrete Random Variables

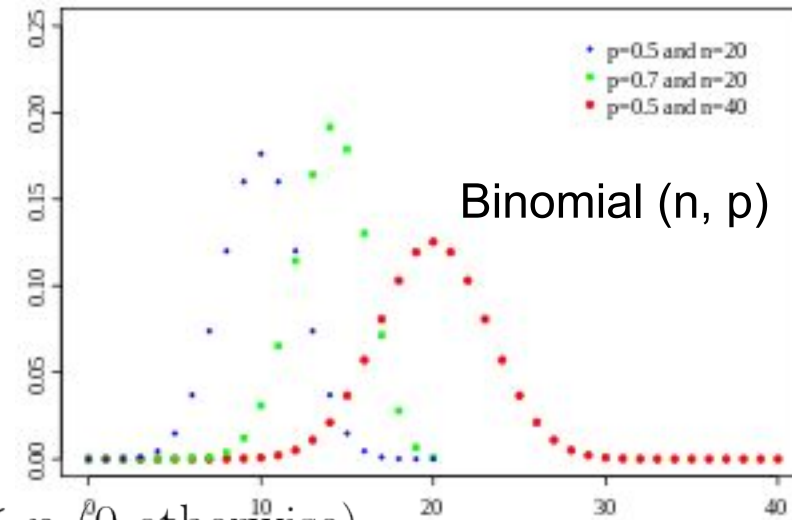
- Binomial(n, p)

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ if } 0 \le x \le n \text{ (0 otherwise)}$$

example: number of heads after n coin flips (p, probability of heads)

<u>Parameters:</u>
n: number of "trials"
p: probability of the event

Discrete Distribution



Common Discrete Random Variables

- Binomial(n, p)

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ if } 0 \le x \le n \text{ (0 otherwise)}$$

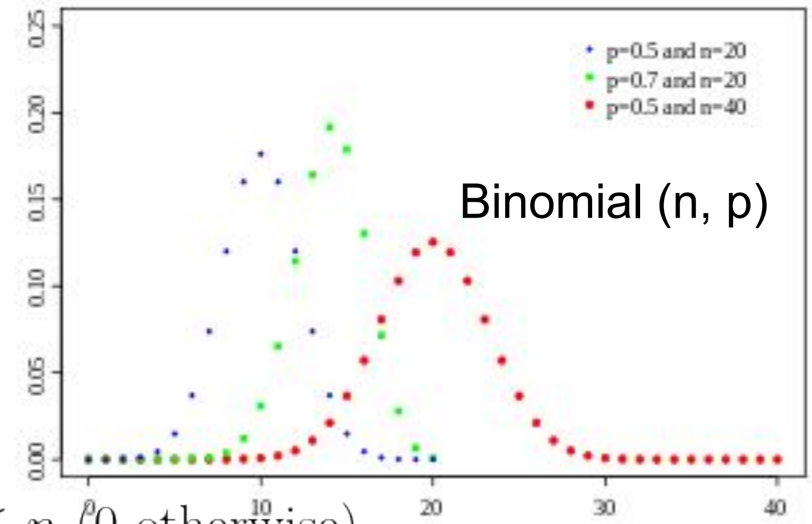
example: number of heads after n coin flips (p, probability of heads)

<u>Parameters:</u>
n: number of "trials"
p: probability of the event

<u>binomial coefficient:</u>
"n choose x": total number of ways to have x successes of the event.

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Discrete Distribution



Common Discrete Random Variables

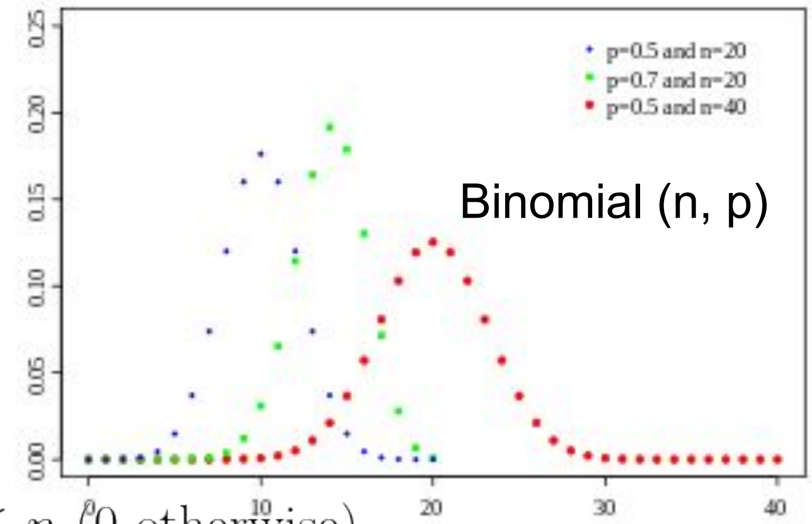
- Binomial(n, p)

  $$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ if } 0 \le x \le n \text{ (0 otherwise)}$$

  example: number of heads after n coin flips (p, probability of heads)
- Bernoulli(p) = Binomial(1, p)

  example: one trial of success or failure

Discrete Distribution



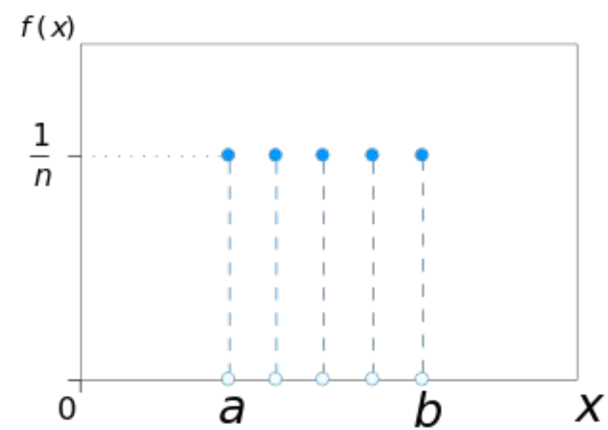
Common Discrete Random Variables

- Binomial(n, p)

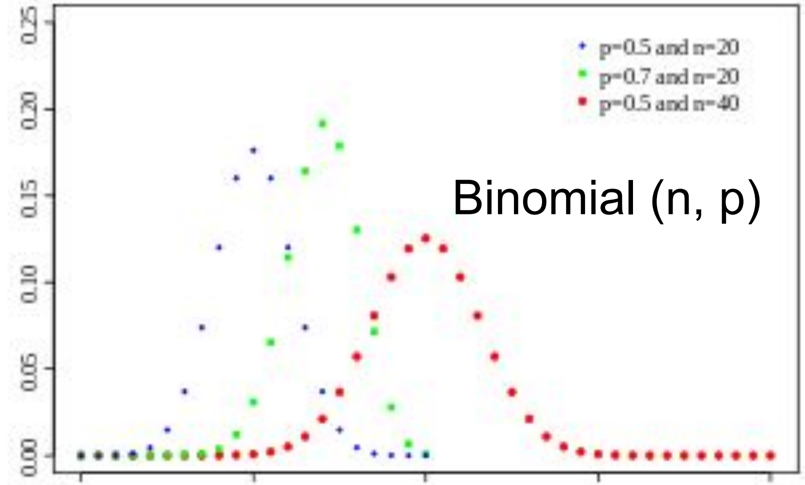
$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ if } 0 \le x \le n \text{ (0 otherwise)}$$

  example: number of heads after n coin flips (p, probability of heads)
- Bernoulli(p) = Binomial(1, p)

  example: one trial of success or failure
- Discrete Uniform(a, b)



Discrete Distribution



Common Discrete Random Variables

- Binomial(n, p)

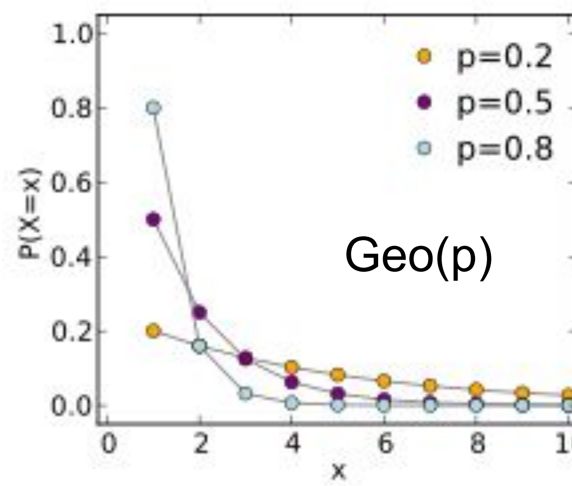
  $$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ if } 0 \leq x \leq n \text{ (0 otherwise)}$$

  example: number of heads after n coin flips (p, probability of heads)
- Bernoulli(p) = Binomial(1, p)

  example: one trial of success or failure
- Discrete Uniform(a, b)
- Geometric(p)

  $P(X = k) = p(1 - p)^{k-1}, \ k \geq 1$

  example: coin flips until first head



RV Review

- Continuous random variable
  - PDFs, the notion of density
  - normal, uniform, exponential
  - CDFs
  - kernel density estimation
- Discrete random variables
  - PMFs
  - binomial, Bernoulli, uniform, geometric