Continuous Random Variables

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 Random Variables

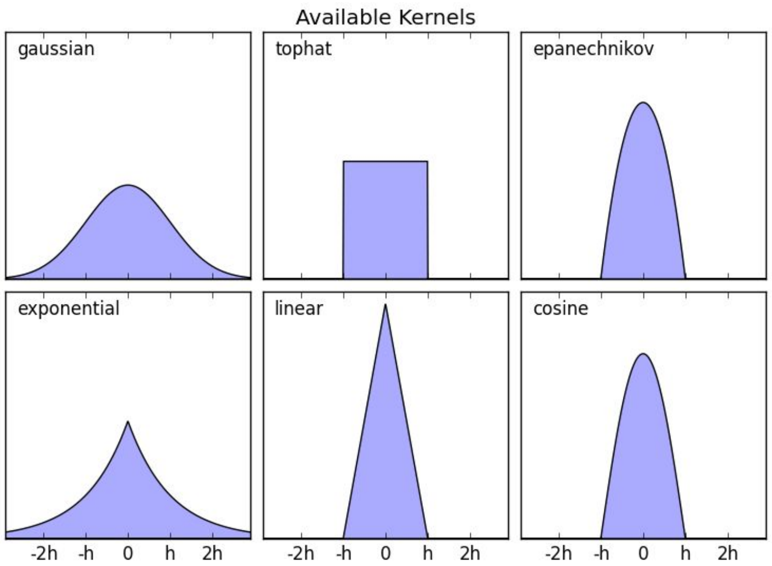
How to decide which pdf is best for my data?

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

- Histogram
- **Kernel Density Estimator**

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

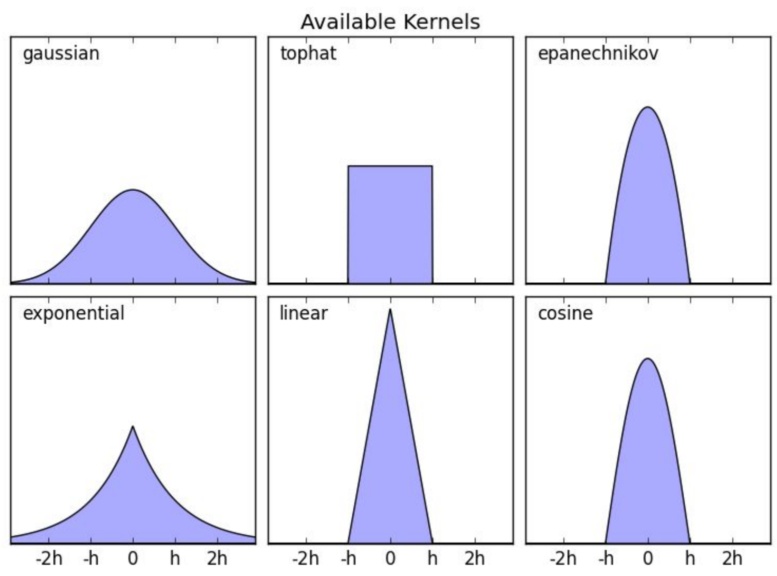
K: kernel function, *h:* bandwidth



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

Continuous Random Variables

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

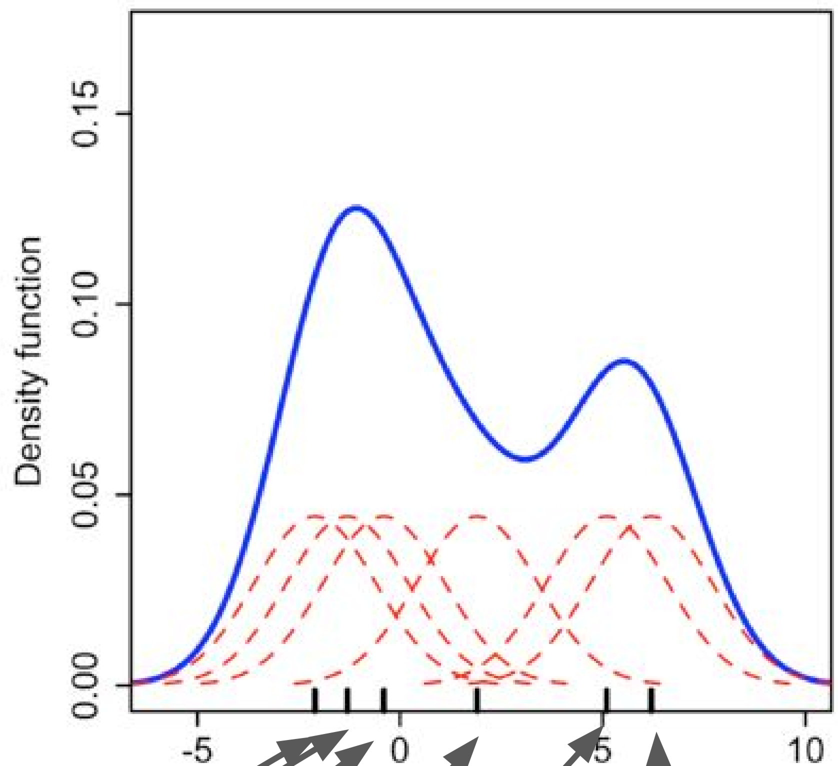


Continuous Random Variables

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.

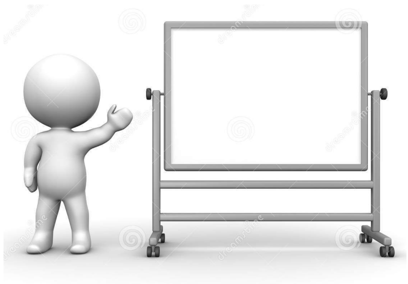
$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{x - X_i}{h}\right)$$



Continuous Random Variables

Analogies

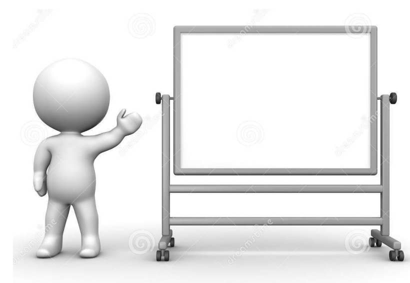
- Funky dartboard Credit: MIT Open Courseware: Probability and Statistics



Continuous Random Variables

Analogies

- Funky dartboard



- Random number generator

Cumulative Distribution Function

- Random number generator

Cumulative Distribution Function

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

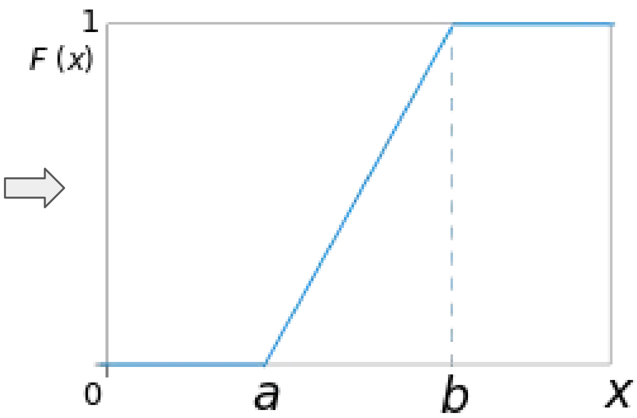
$$F_X(x) = \mathrm{P}(X \leq x)$$

Cumulative Distribution Function

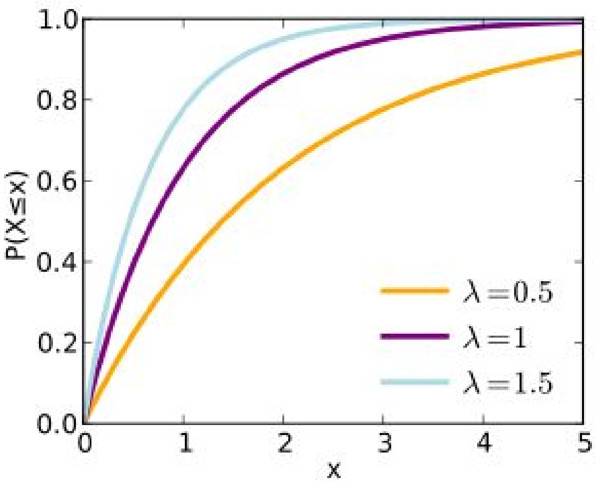
For a given random variable X, the *cumulative distribution function* (CDF), *Fx:* $\mathbb{R} \to [0, 1]$, is defined by:

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

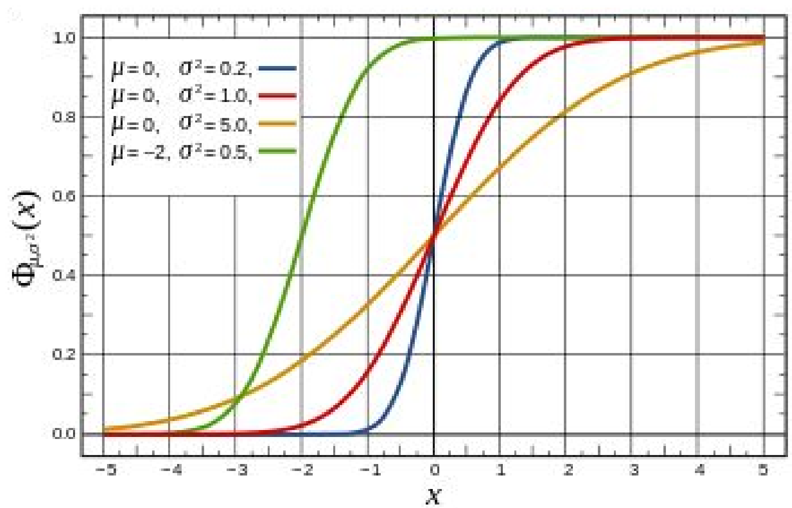
Uniform



Exponential



Normal

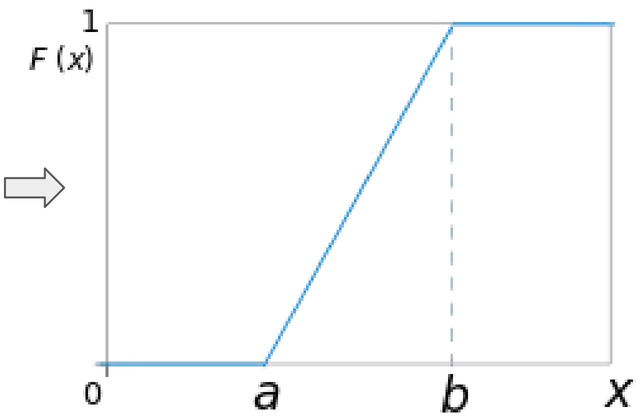


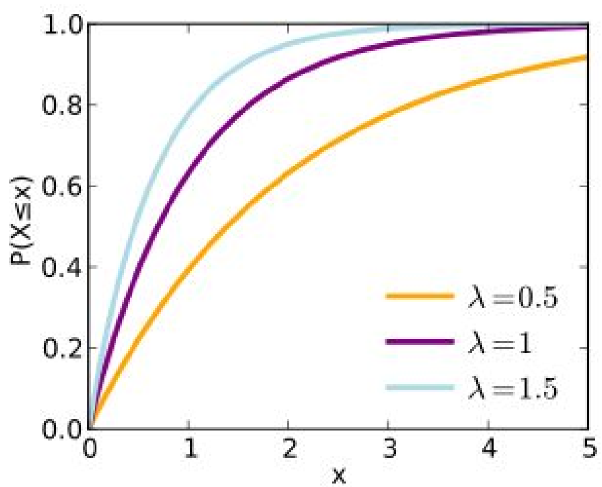
Cumulative Distribution Function

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

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

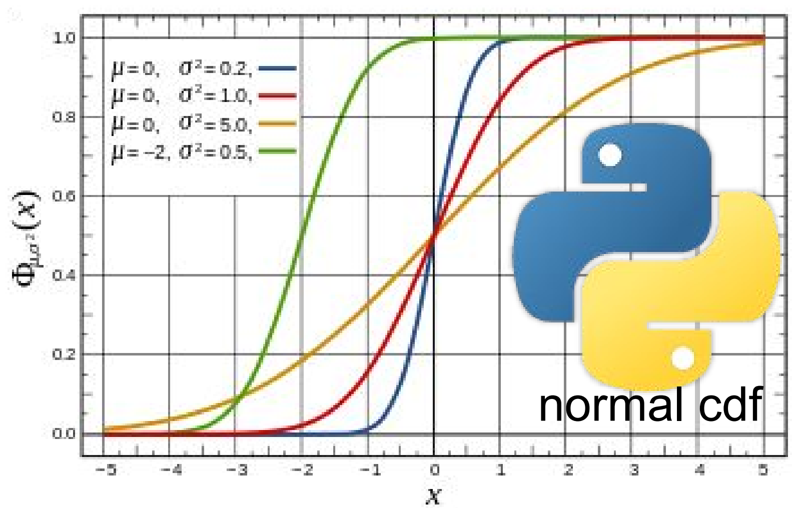
Uniform ⇨





⇦ Exponential

Normal ⇨

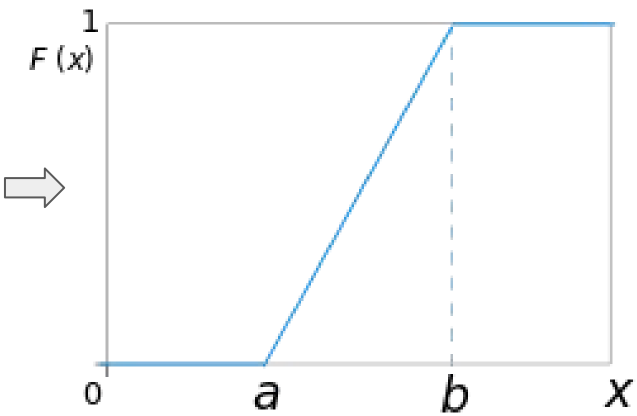


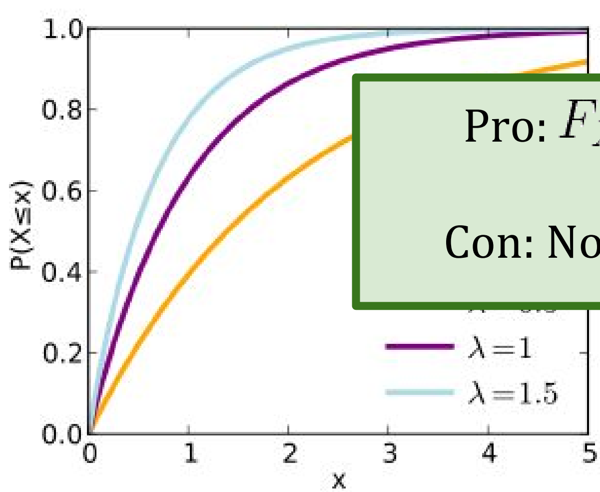
Cumulative Distribution Function

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

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

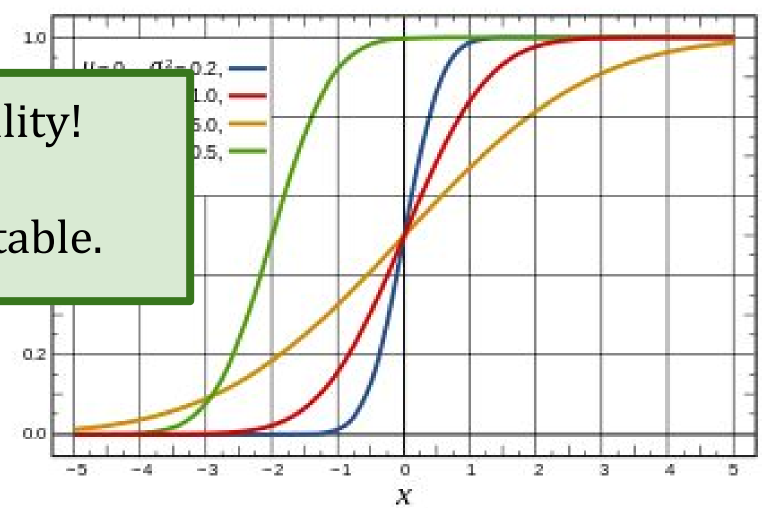
Uniform ⇨





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

Con: Not intuitively interpretable.



Random Variables, Revisited

$\mathbf{X}$: A mapping from $\boldsymbol{\Omega}$ 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.

Discrete Random Variables

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

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

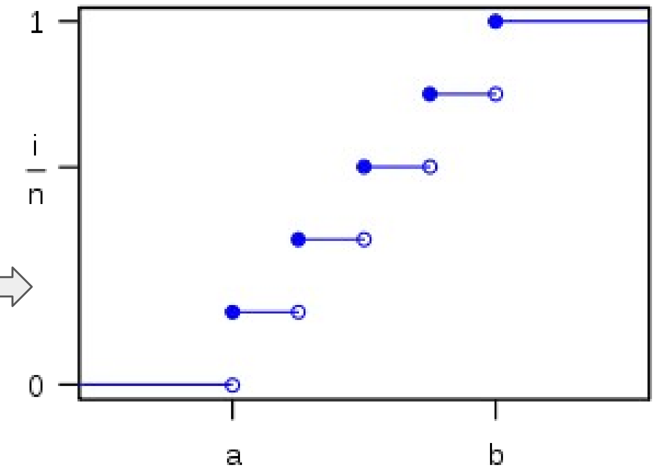
X is a *discrete random variable* if it takes only a countable number of values.

Discrete Random Variables

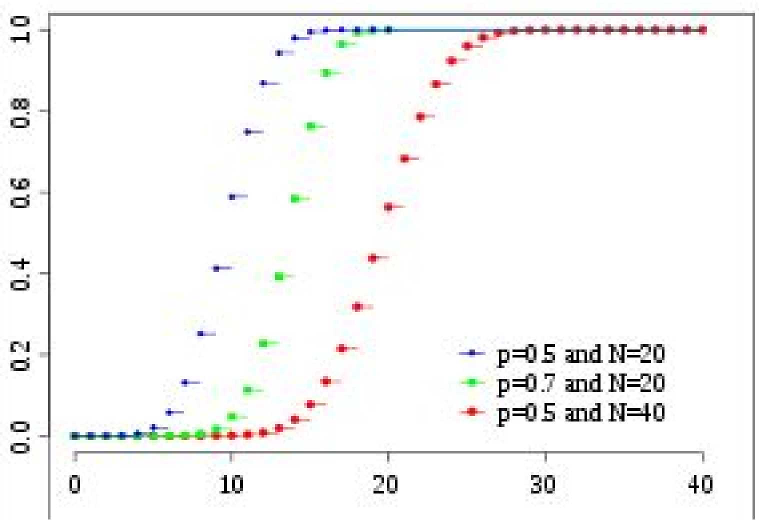
For a given random variable X, the *cumulative distribution function* (CDF), *Fx:* $\mathbb{R} \to [0, 1]$, is defined by:

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

Discrete Uniform



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



Binomial (n, p)

(like normal)

Discrete Random Variables

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.

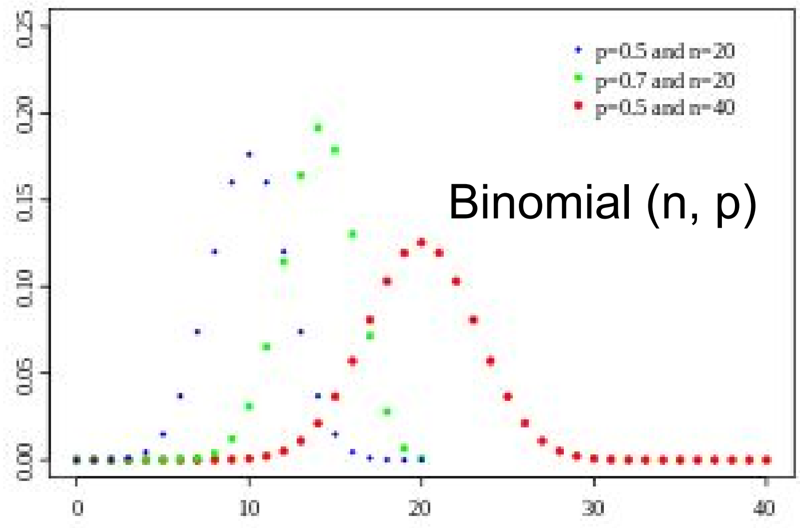
Discrete Random Variables

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

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

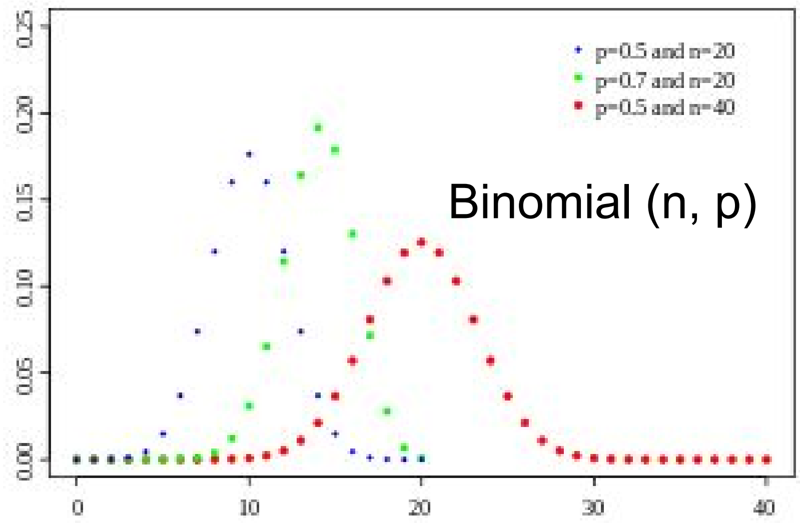
For a given discrete random variable X, *probability mass function* (*pmf*), *fx:* $\mathbb{R} \rightarrow [0, 1]$, is defined by:

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



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

Discrete Random Variables



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:

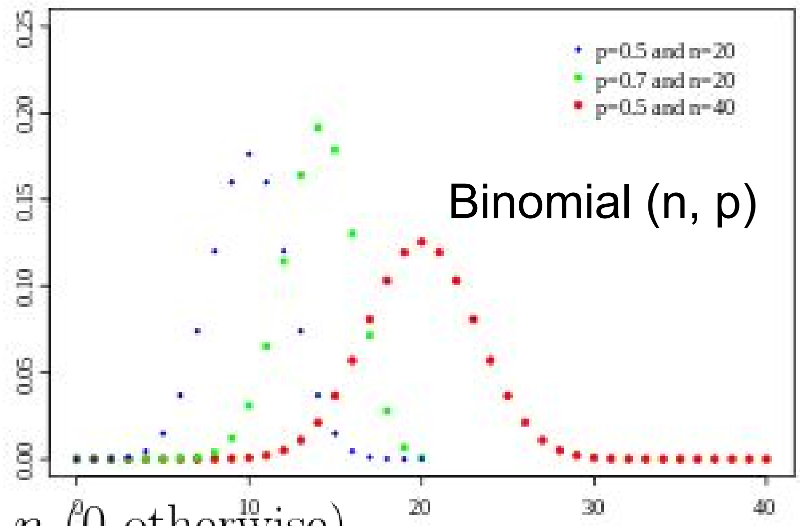
$$f_X(x) = \mathrm{P}(X = x)$$

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 Random Variables



Common Discrete Random Variables

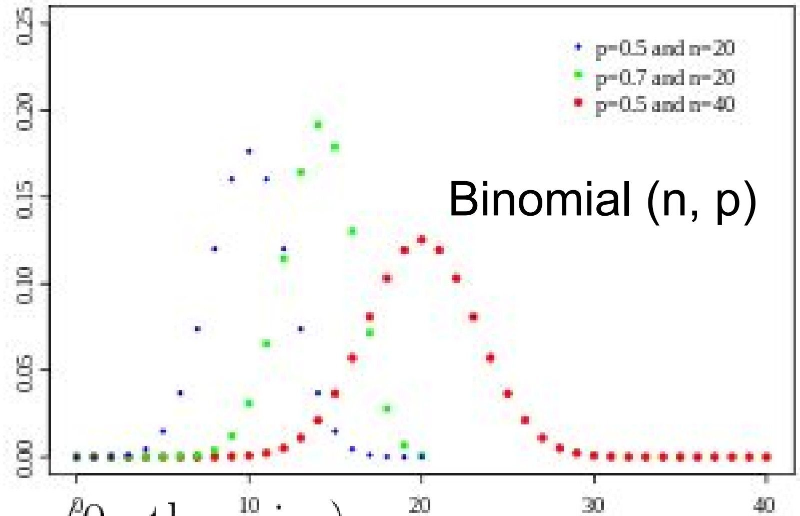
- Binomial(n, p)

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

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

  example: one trial of success or failure

Discrete Random Variables



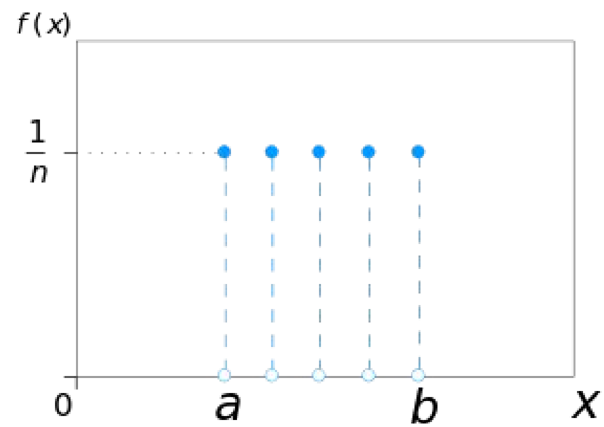
Common Discrete Random Variables

- Binomial(n, p)

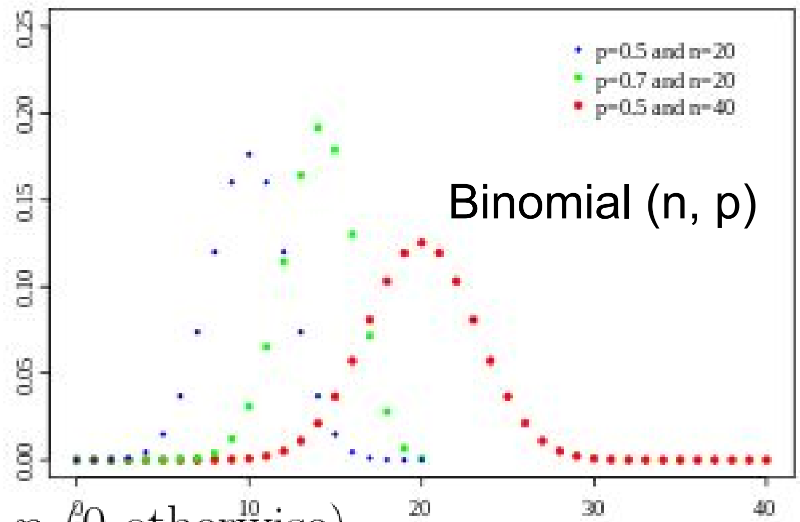
  $$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ if } 0 \le x \le n \text{ (0 otherwise)}$$

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

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



Discrete Random Variables



Common Discrete Random Variables

- Binomial(n, p)

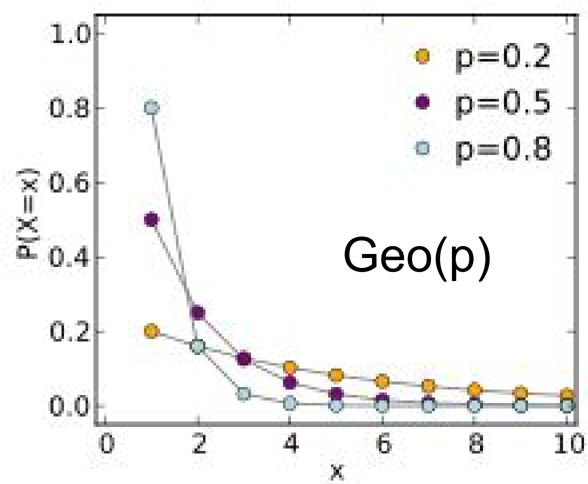
$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ if } 0 \leq x \leq n \text{ (0 otherwise)}$$

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

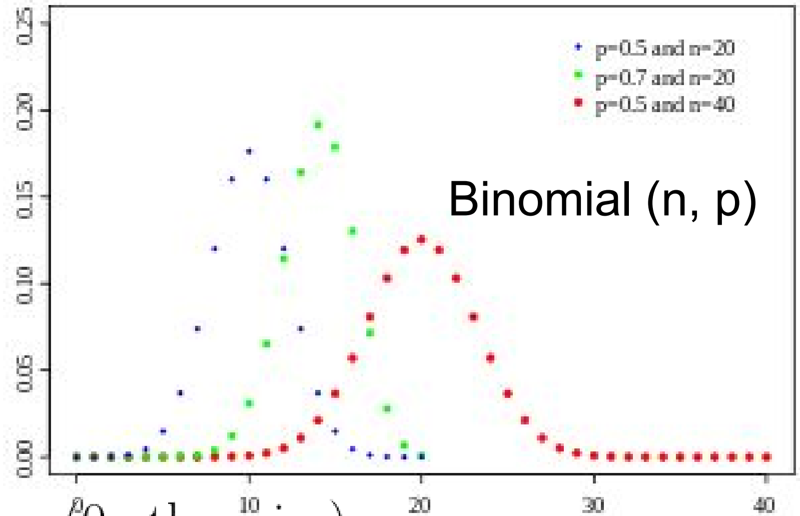
  example: one trial of success or failure
- Discrete Uniform(a, b)
- Geometric(p)

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

  example: coin flips until first head



Discrete Random Variables



Common Discrete Random Variables

- Binomial(n, p)

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

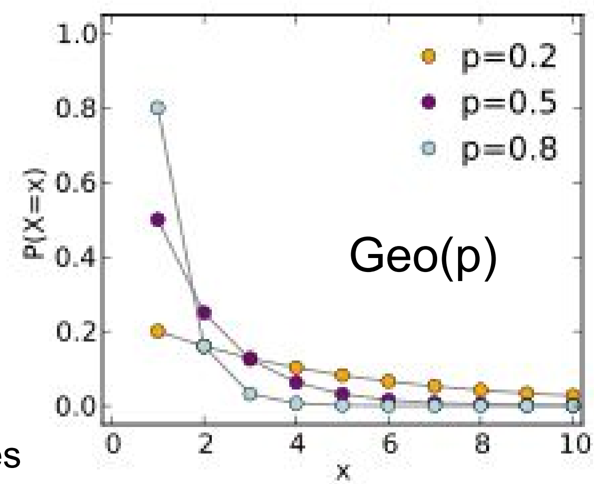
  example: number of heads after n coin flips (p, probability of heads)
- Bernoulli(p) = Binomial(p, 1)

  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

discrete random variables



Maximum Likelihood Estimation (parameter estimation)

Given data and a distribution, how does one choose the parameters?

Maximum Likelihood Estimation (parameter estimation)

Given data and a distribution, how does one choose the parameters?

likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

maximum likelihood estimation: What is the θ that maximizes *L*?

Maximum Likelihood Estimation (parameter estimation)

Given data and a distribution, how does one choose the parameters?

likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

log-likelihood function:

$$l(\theta) = log \sum_{i=1}^{n} f(X_i; \theta)$$

maximum likelihood estimation: What is the θ that maximizes *L*?

Maximum Likelihood Estimation (parameter estimation)

Given data and a distribution, how does one choose the parameters?

likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

log-likelihood function:

$$l(\theta) = log \sum_{i=1}^{n} f(X_i; \theta)$$

maximum likelihood estimation: What is the θ that maximizes *L*?

Example: $X_1, X_2, \ldots, X_n \sim$ Bernoulli(p), then $f(x;p) = p^x(1 - p)^{1-x}$, *for* $x = 0, 1$.

Maximum Likelihood Estimation (parameter estimation)

Given data and a distribution, how does one choose the parameters?

likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

log-likelihood function:

$$l(\theta) = log \sum_{i=1}^{n} f(X_i; \theta)$$

maximum likelihood estimation: What is the θ that maximizes *L*?

Example: $X_1, X_2, \ldots, X_n \sim$ Bernoulli(p), then $f(x;p) = p^x(1 - p)^{1-x}$, *for* $x = 0, 1$.

$$L_n(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-X_i} = p^S(1-p)^{n-S}, \text{where } S = \sum_i X_i$$

Maximum Likelihood Estimation (parameter estimation)

Given data and a distribution, how does one choose the parameters?

likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

log-likelihood function:

$$l(\theta) = log \sum_{i=1}^{n} f(X_i; \theta)$$

maximum likelihood estimation: What is the θ that maximizes *L*?

Example: $X_1, X_2, \ldots, X_n$ ~ Bernoulli(p), then $f(x;p) = p^x(1 - p)^{1-x}$, *for* $x = 0, 1$.

$$L_n(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-X_i} = p^S(1-p)^{n-S}, \text{where } S = \sum_i X_i$$

$$l_n(p) = S\, log\, p + (n - s)\, log(1 - p)$$

Maximum Likelihood Estimation (parameter estimation)

Given data and a distribution, how does one choose the parameters?

likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

log-likelihood function:

$$l(\theta) = log \sum_{i=1}^{n} f(X_i; \theta)$$

maximum likelihood estimation: What is the θ that maximizes *L*?

Example: $X_1, X_2, \ldots, X_n \sim$ Bernoulli(p), then $f(x;p) = p^x(1 - p)^{1-x}$*, for* $x = 0, 1$*.*

$$L_n(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-X_i} = p^S(1-p)^{n-S}, \text{where } S = \sum_i X_i$$

$$l_n(p) = S\, log\, p + (n - s)\, log(1 - p)$$

take the derivative and set to 0 to find: $\hat{p} = \dfrac{S}{n}$