

# Probability Theory

- Experiments
  - Observations
  - Measurements
- Theory :
  - Hypotheses
  - Models
- Quantify Uncertainty

- Analysis:
- 1) Estimate (measure) parameter
  - 2) Quantify the uncertainty in  $\uparrow$
  - 3) Test extent prediction (test) agrees w/ data

### Sources of Uncertainty:

- 1) Stochastic: underlying randomness
  - e.g. QM
  - Random shuffling of cards
- 2) Incomplete Observations: Monty Hall Problem
- 3) Unknowns: incomplete modeling

Prob Theory  $\rightarrow$  analyze frequency of "events"

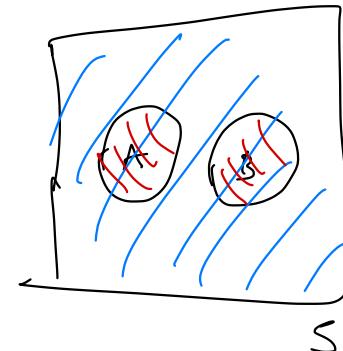
- 1. prob  $P$  of  $X$  event happening
  - $\Rightarrow$  repeatable observation of event
  - $\Rightarrow P \sim$  fraction that  $x$  would be outcome.
  - $\Rightarrow$  e.g. freq of a disease in a population  
 $\sim 1$  in 1000
  - $\Rightarrow$  "Frequentist"
- What if you got a test  $\rightarrow$  How do you interpret?
  - accuracy of test  $\rightarrow$  can't repeat
  - freq of disease  $\Rightarrow$  degree of belief

Same answers  
- computation  
are same



# Basic Definitions:

- Set  $S$  w/ subsets  $A, B$



$$P(S) = 1$$

$\Rightarrow$  for all  $\forall A \subset S : P(A) \geq 0$

$\uparrow$   
Subset

$$A \cap B = \emptyset$$

$\uparrow$  overlap     $\uparrow$  empty

$$P(A \cup B) = P(A) + P(B)$$

$\cap$  union

$$P(\bar{A}) = 1 - P(A)$$

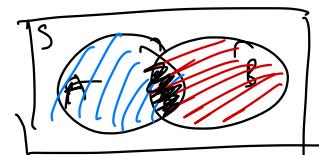
$$A \subset B \Rightarrow P(A) \leq P(B)$$

$\uparrow$   
Subset of  $B$

$$P(A \cup \bar{A}) = 1$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(\emptyset) = 0$$



$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

if  $A \cup B = \emptyset$

$$P(A, B) = P(A) P(B)$$

$$P(A|B) = \frac{P(A) P(B)}{P(B)} = P(A)$$

Conditional Prob:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Example:
- Die Roll (6-sided)
  - Jodi rolled 3 or less
  - What is the prob that you rolled a 1?

$$P(1) = \frac{P(\text{恰有 } n \text{ 事件})}{P(\text{events})} = \frac{\frac{1}{3}}{\frac{3}{6}} = \underline{\underline{\frac{1}{3}}}$$

• if  $\underline{A \cup B = \emptyset} \Rightarrow P(A; B) = \frac{P(A) P(B)}{P(A) P(B)} = P(A)$

// Interpretation : 1) Relative Freq:

$A, B, \dots$  outcomes of a repeatable experiment

$$P(A) = \lim_{n \rightarrow \infty} \frac{\text{times outcome is } A}{n}$$

2) Subjective prob

$A, B, \dots$  are hypothesis (true/false statements)

$P(A)$  is degree of belief that  $A$  is true

Bayes Jhm.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

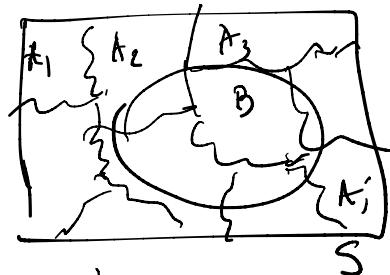
$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

$$\Rightarrow P(A|B) P(B) = P(B|A) P(A)$$

$$P(A \cap B) = P(B \cap A)$$

$$\Rightarrow P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

↑ ↗  
measurement Data



$$\bigcup_i A_i = S$$

$$\Rightarrow B = B \cap S = B \cap (\bigcup_i A_i) \Rightarrow \bigcup_i B \cap A_i$$

$$P(B) = \sum_i P(B|A_i) P(A_i)$$

posterior

$$\Rightarrow P(A|B) = \frac{P(B|A) P(A)}{\sum_i P(B|A_i) P(A_i)}$$

prior

$$\left. \begin{array}{l} P(\text{Sick}) = 0.001 \\ P(\text{no sick}) = 0.999 \end{array} \right\} \text{Prior Knowledge of the population}$$

$$P(+|\text{sick}) = 0.98 \quad \cdot \text{True Positve}$$

$$P(-|\text{sick}) = 0.02 \quad \left. \begin{array}{l} \text{Wrong} \\ \text{false neg} \end{array} \right\}$$

$$P(+|\text{not sick}) = 0.03 \quad \left. \begin{array}{l} \text{Wrong} \\ \text{false pos} \end{array} \right\}$$

$$P(-|\text{not sick}) = 0.97 \quad \cdot \text{True neg}$$

$$A_i = \{\text{sick}\} \cup \{\text{not sick}\}$$

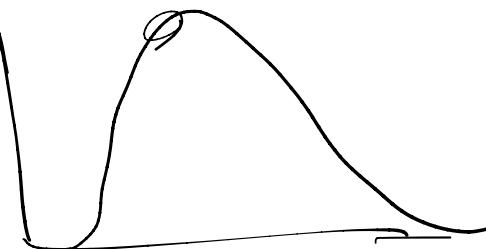
Test  $\Rightarrow$  get pos results  $\rightarrow$  what the prob that you are sick

$$P(\text{sick} | +) = \frac{P(+|\text{sick}) P(\text{sick})}{P(+|\text{sick}) P(\text{sick}) + P(+|\text{not sick}) P(\text{not sick})} = \underline{\underline{0.032}}$$

	Age	Major	GPA	...
Student 1				
2				
3				
4				



$$\text{Age} = \{ 20, 21, 18, 25, 35, \dots \}$$

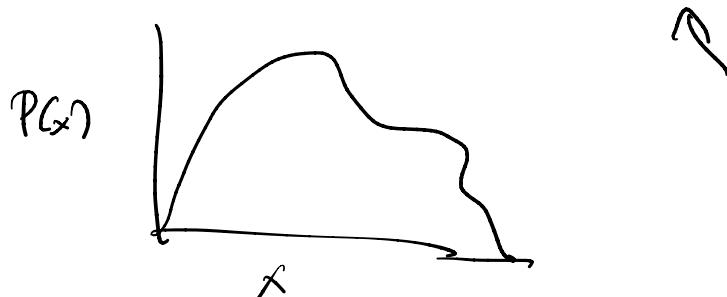


Age is a random variable

## Random variable

$X$  is a random variable  $\Rightarrow$  "draw" from  
prob mass distribution

$$X \sim P(x) \Rightarrow X = \{x_1, x_2, \dots, x_n\}$$



$P(x_i)$  = Prob of getting  $x_i$

· Freq: frac of times you get  $x_i$

· Bayes: how likely it is to measure  $x_i$

$X$  can be continuous  $\Rightarrow x = x_i$  is very unlikely

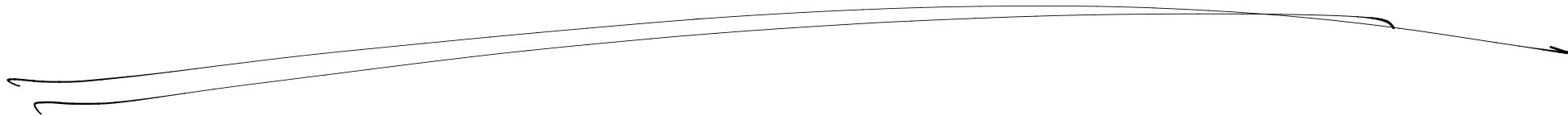
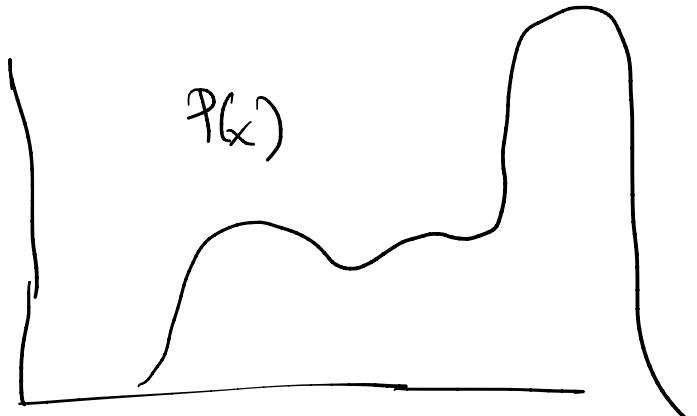
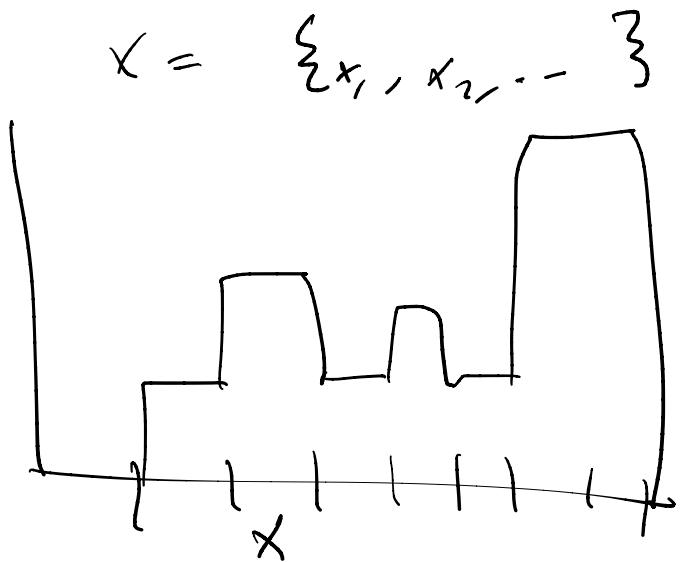
$$\lim_{\delta x \rightarrow 0} P(x \in [x_i, x_i + \delta x])$$

Prob density func

$$\lim_{\delta x \rightarrow 0} P(x) = P(x) \frac{\delta x}{\delta x}$$

$$X = \{x_1, x_2, \dots\}$$

Counting  
how often  
I see  $x$   
in specific  
range.



# Random Variable

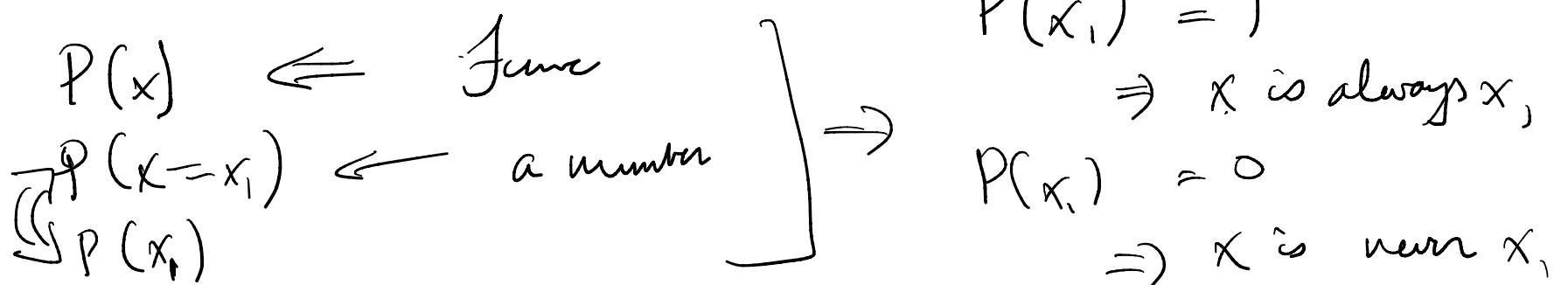
$$X \sim \{x_1, x_2, \dots, x_n\}$$

Discrete or continuous

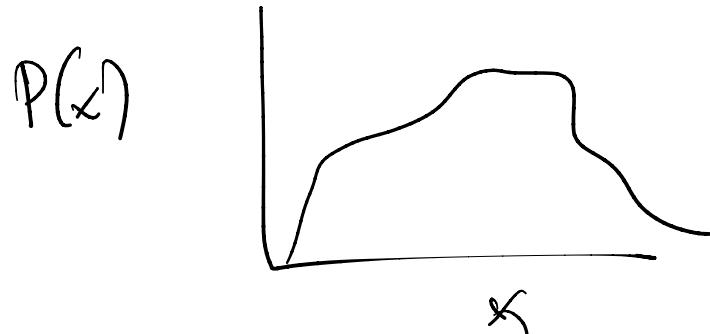
Prob Dist  $\Rightarrow$  How Prob is it for  $X$  to have a specific value.

Discrete  $\Rightarrow$  Prob Mass Distribution (PMF)

Continuous  $\Rightarrow$  Prob Density Distribution (PDF)



$x \sim P(x)$   
↳ "Drawn"



Joint Prob Dist  $P(x, y) \leftarrow$  multivariate

Properties of  $P(x)$

- Domain of  $P$  must be all possible values of  $x$
- $\forall x_i \in x \quad 0 \leq P(x_i) \leq 1$
- $\sum_{x_i \in x} P(x_i) = 1 \quad \leftarrow \text{normalization}$

if  $X$  is continuous  $\Rightarrow P(X = 0.5) = 0$

$$P(X \in [x_i, x_i + \delta x]) \rightarrow p(x) dx$$

capital P

lower case "p"

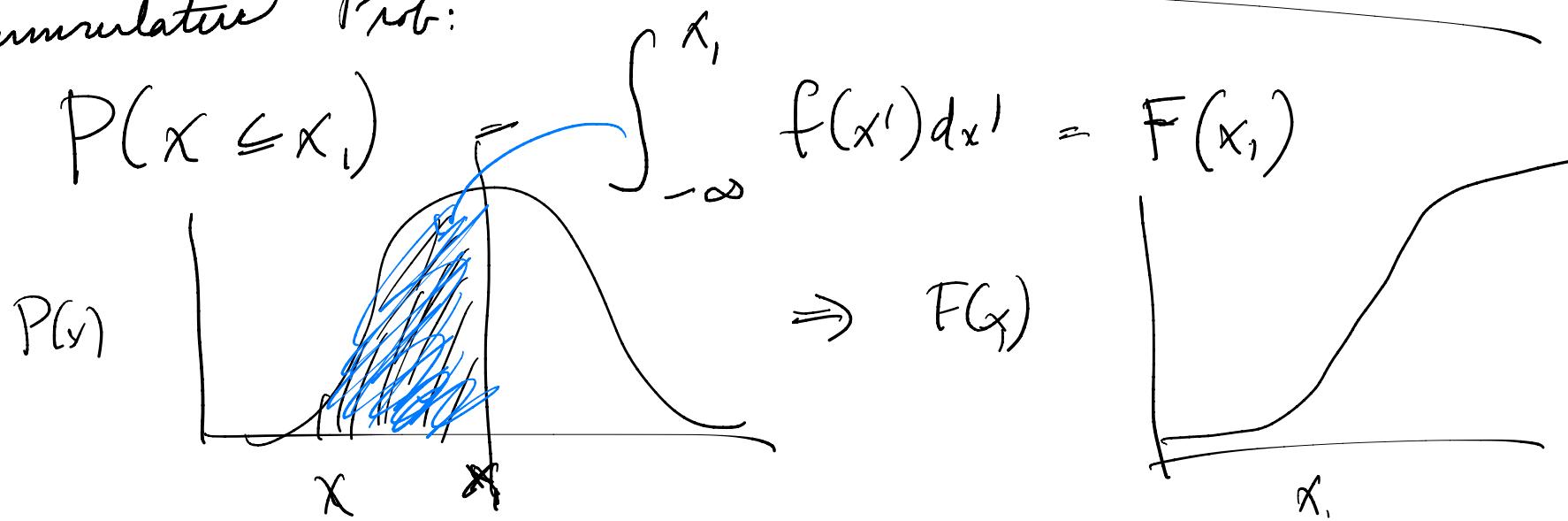
Prob Dist Func

$$\int_{-\infty}^{\infty} p(x) dx = 1$$

---

Cumulative Prob:

$$P(X \leq x_1) = \int_{-\infty}^{x_1} f(x') dx' = F(x_1)$$



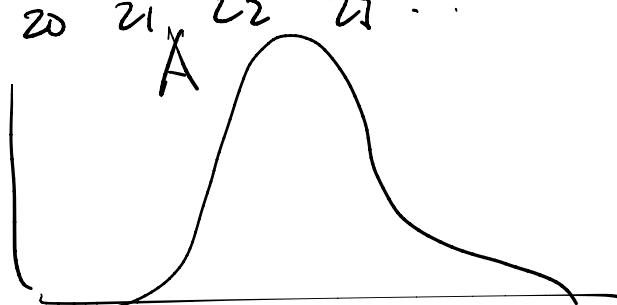
# Data

Name	(A) Age	(H) Height	(W). Weight
o	o	o	o
o	c	c	o
c	c	c	o

# Histogram



$$\Rightarrow P(A)$$

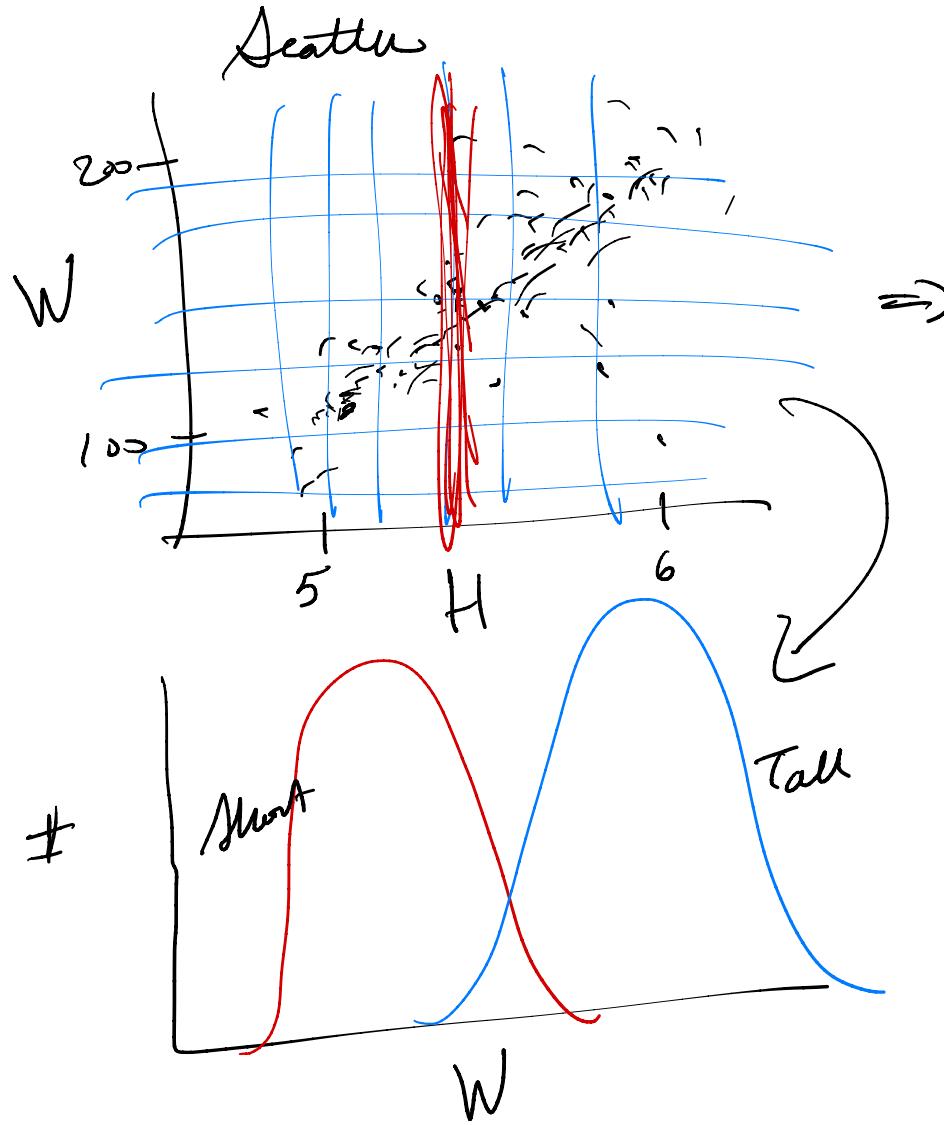


limit  
 $n \rightarrow \infty$   
 $\Delta x \rightarrow 0$

$$\frac{n(x)}{N \Delta x} = P(x)$$

$$P(x) \approx \frac{n(x)}{N \Delta x}$$

$\nearrow$  Total number of data points in histogram  
 $\nwarrow$  width of a bin

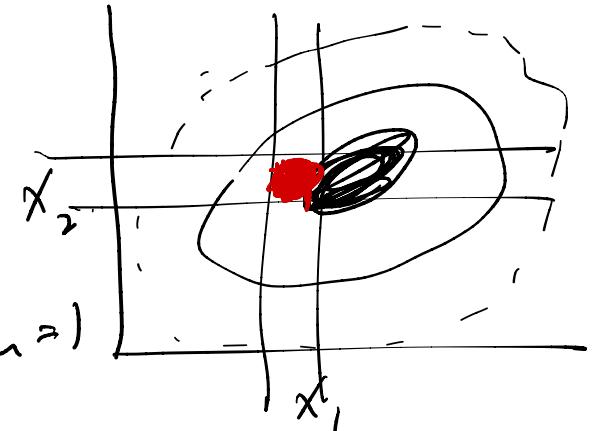


2D Histogram

- 
- . Freq  $P(x) dx$  = freq of measuring  $x: x \rightarrow x+dx$
  - . Bayesian  $P(x|dx)$  = if you measurement of  $x$   
how likely it is that  $x: x \rightarrow x+dx$

Multivariate:

$$P(x_1, x_2) dx_1 dx_2$$



normalization:

$$\int P(x_1, \dots, x_n) dx_1 \dots dx_n = 1$$

marginal prob:

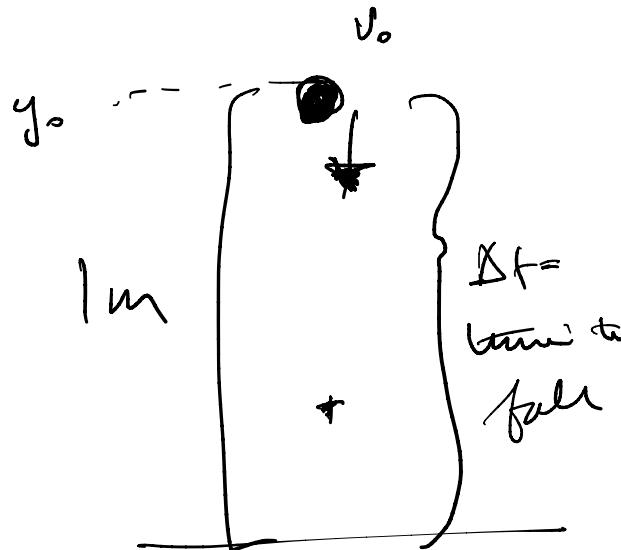
$$p(x_1) = \int P(x_1, \dots, x_n) dx_2 \dots dx_n =$$

independence:

$$\text{if } P(x_1, x_2) = P_1(x_1) P_2(x_2)$$

$\Rightarrow x_1 \& x_2$  are independent

# Simple Experiment



$$g = 9.8 \text{ m/s}^2$$

measure  $\Delta t \Rightarrow$  estimate  $g$

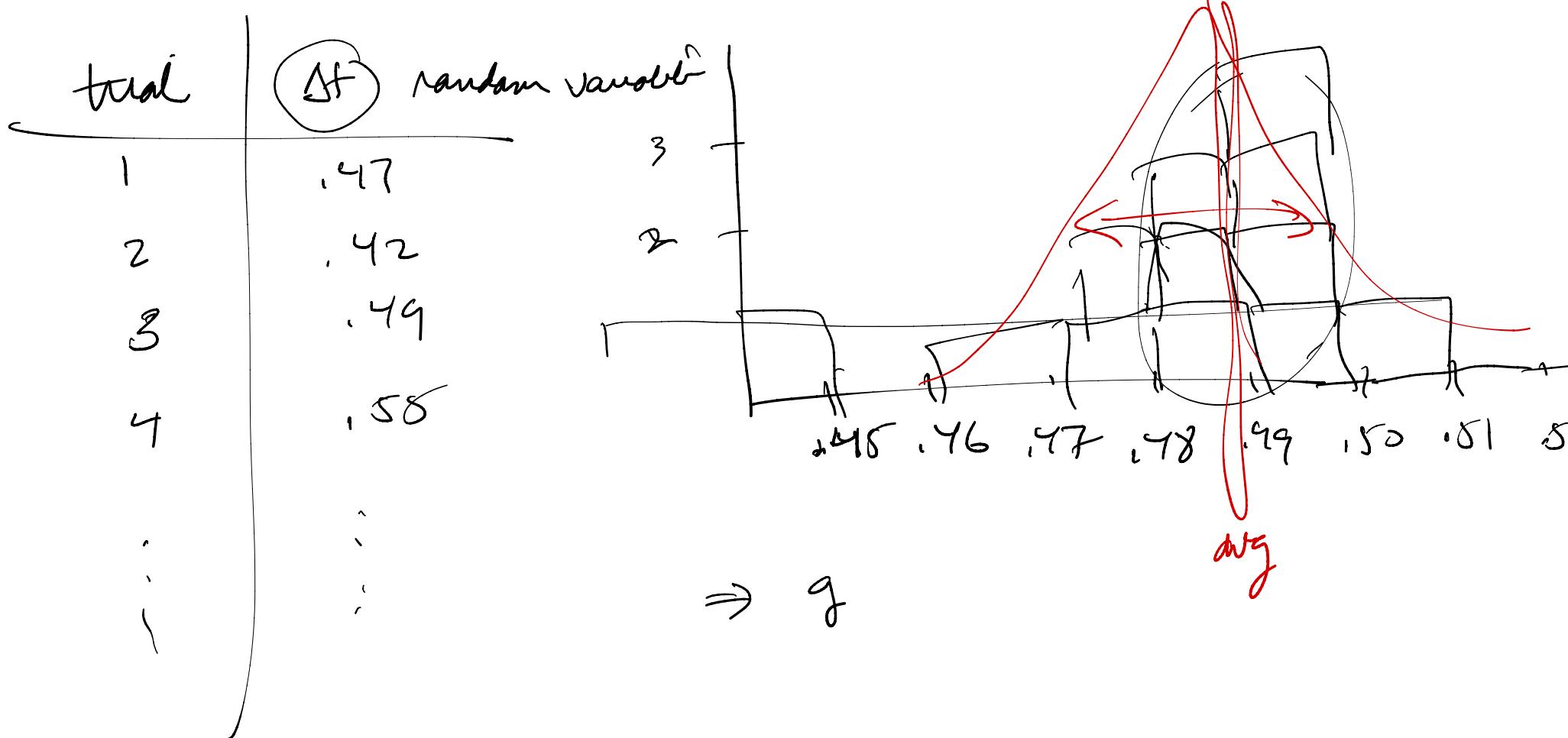
$$y = y_0 + v_0 t + \frac{1}{2} a t^2$$

~~$= g$~~

$$\Delta y = \frac{1}{2} g \Delta t^2$$

$$\Rightarrow g = \frac{2\Delta y}{\Delta t^2}$$

$$\Rightarrow \Delta t = \sqrt{\frac{2\Delta y}{g}} = \underline{\underline{.45 \text{ sec}}}$$



# Measurement

Define Expectation Value

$$\mathbb{E}_{x \sim p}[f(x)] = \int f(x) p(x) dx$$

random variable  
distr  $x$  is drawn from

fence

$$\mathbb{E}[x] = \int x p(x) dx \equiv \underline{\mu} \quad \text{"mu" in Greek}$$

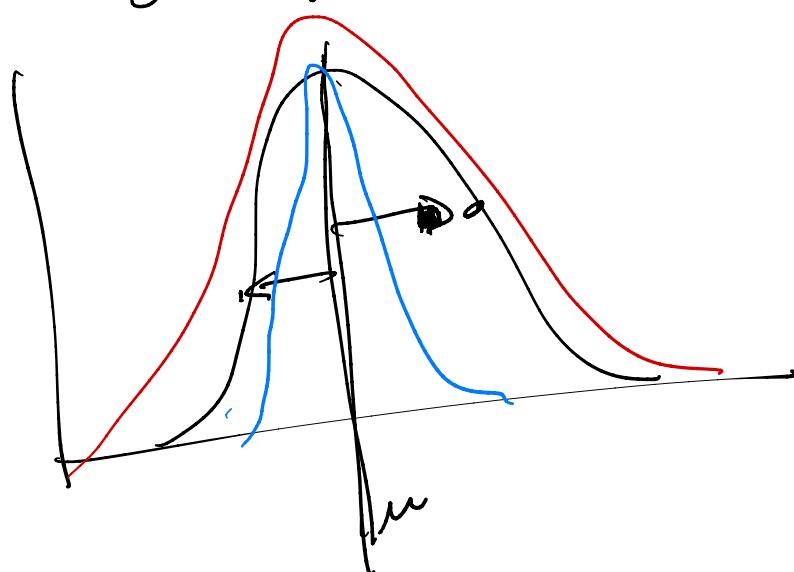
mean

$$= \sum_{x \in \Omega} x p(x)$$

Variance:

$$\mathbb{E}[x^2] - \mu^2 = \mathbb{E}[(x-\mu)^2] = \sigma^2$$

$$\sigma = \sqrt{\sigma^2} = \text{"standard deviation"} \quad \text{"signif."}$$



$P(x,y)$



covariance:

$$\begin{aligned}\text{cov}[x,y] &= \mathbb{E}[xy] - \mu_x \mu_y \\ &= \mathbb{E}[(x-\mu_x)(y-\mu_y)]\end{aligned}$$

correlation' coeff:

$$P_{xy} = \frac{\text{cov}[x,y]}{\sigma_x \sigma_y}$$

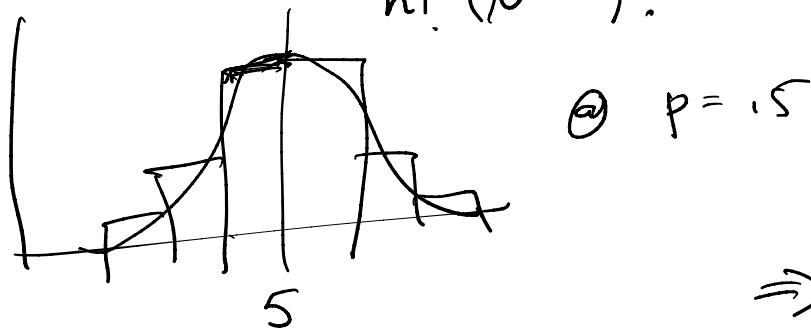
Binomial:  $N$  independent test  $\xrightarrow{\text{Success}}$   
 $\xrightarrow{\text{failure}}$

$N = \# \text{ of trials}$

$n = \# \text{ of successes}$

$p = \text{underlying prob for success}$

$$f(n|N,p) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$$



$$\Rightarrow 98\% = 98\%, = p$$

TPR, FPR, ...

Efficiency

What fraction of ...?

$$\Rightarrow E[n] = Np$$

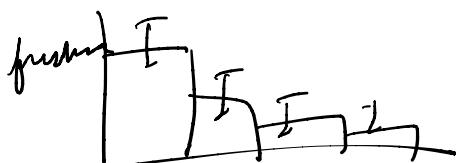
$$\sqrt{[n]} = \sqrt{Np(1-p)}$$

$$100(.98)(1-.98) =$$

$$p = \frac{n}{N}$$

$$\sqrt{[p]} = \sqrt{N \left[ \frac{p}{N} \right]} =$$

$$\text{std}[p] = \frac{1}{\sqrt{N}} \sqrt{p(1-p)}$$



A

Start of Binomial  $N \rightarrow \infty$   $p \rightarrow 0$   $\Rightarrow E[n] = Np \rightarrow 0$   
 $\uparrow$   
 $n \text{ per } \Delta t$

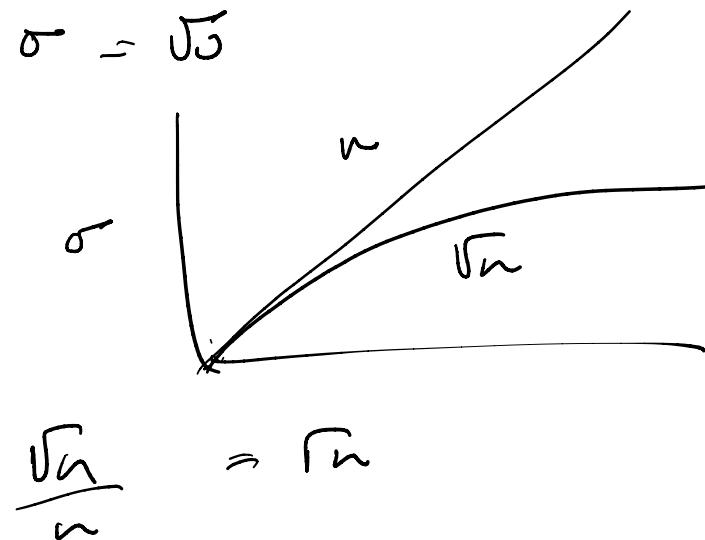
$$f(n; \sigma) = \frac{\sigma^n}{n!} e^{-\sigma} \leftarrow \text{"Poisson"}$$

$$E[n] = \sigma$$

$$V[n] = \sigma = \sigma = \sqrt{\sigma}$$

$$n \pm \sqrt{n}$$

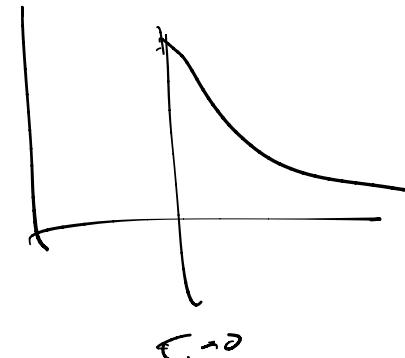
How many  $\rightarrow$  rate



$$\frac{\sqrt{n}}{n} = r_n$$

Exponentiated Dist

$$f(t; \tilde{\sigma}) = \begin{cases} \pi e^{-t/\tilde{\sigma}} & t \geq 0 \\ 0 & t < 0 \end{cases}$$



$$\tilde{\sigma} = \frac{1}{\sqrt{2}}$$

$$\mathbb{E}[t] = \tilde{\sigma}$$

$$\text{Var}[t] = \tilde{\sigma}^2 \Rightarrow \sigma = \tilde{\sigma}$$

How long?

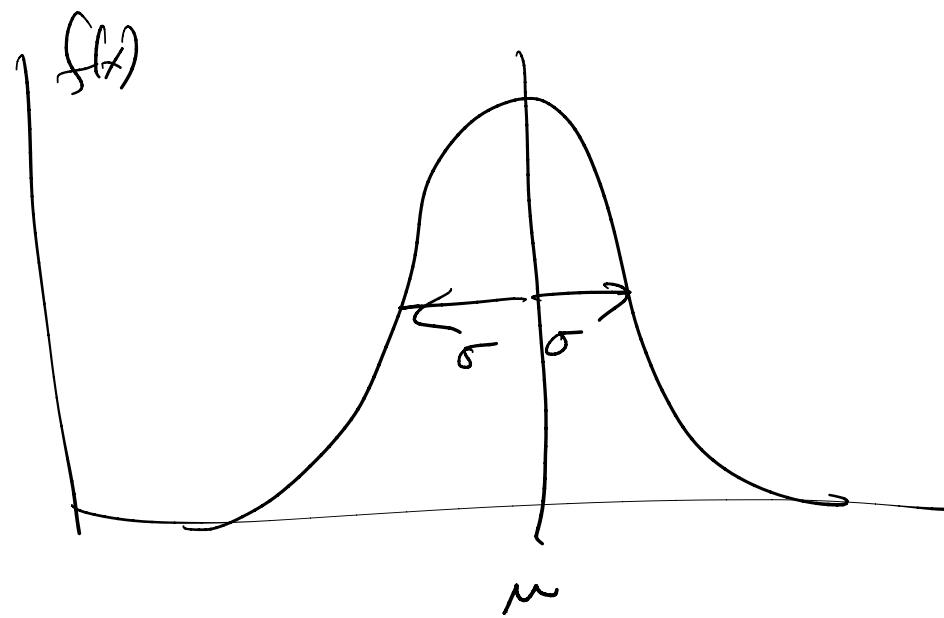
Central Limit Thm: add (independent) random variables

$x, y \Rightarrow x+y \rightarrow$  normal distribution

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

$$\mathbb{E}[x] = \mu$$

$$\text{Var}[x] = \sigma^2 = \sigma = \sigma$$



How to Compute ...

$$= \sum_i x_i f(x_i) \Delta x$$

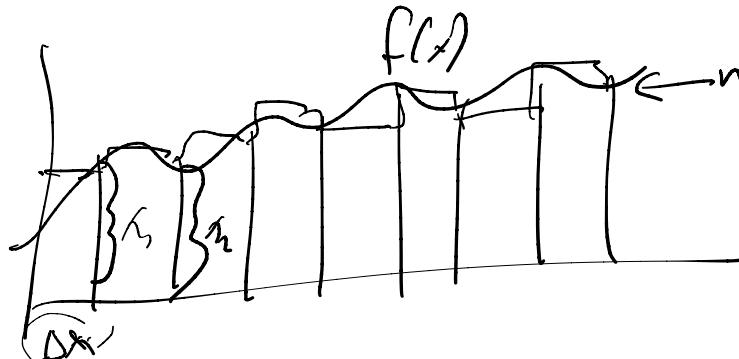
mean:  $E[x] \quad (x \sim P(x)) : \mu = \int x f(x) dx$

possible values set  $x = \{x_1, x_2, \dots\}$

$$= \sum_i x_i P(x_i)$$

Dataset of  $x \quad x = \{x_1, x_2, \dots\} \leftarrow$  random variable

over all possible values of  $x$



$$P(x) \approx \frac{n(x)}{N \Delta x}$$

$$\mu = \frac{1}{N} \sum_{x \in \text{Dataset}} x_i$$

$$\mu = \sum_i x_i f(x_i) \Delta x$$

$$= \sum_i x_i \cdot \frac{n(x_i)}{N \Delta x} \Delta x$$

$$\sum_{x \in \text{Dataset}} x_i = \frac{1}{N} \sum_i x_i \cdot n(x_i)$$

## Variance

$$V[x] = E[x^2] - \mu^2 = \sum_i x_i^2 p(x) - \mu^2$$
$$= E[(x-\mu)^2]$$

$\frac{1}{N} \sum_{x \in \text{Dataset}} x_i^2 - \mu^2$

$$\Rightarrow \sigma = \sqrt{\frac{1}{N} \sum_{x_i \in \text{Dataset}} (x_i - \mu)^2}$$

$$\sigma = \sqrt{\frac{1}{N-1} \sum_i (x_i - \mu)^2}$$

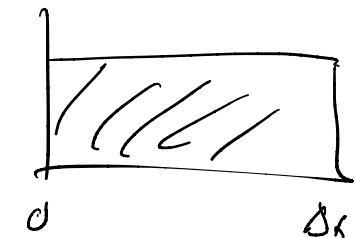


Bessel Correction

# Generate Distributions

## Uniform Dist

$$P(x) = \frac{1}{\Delta x}$$



• How do you do this in a computer?

• rely on natural random processes

e.g. noise, quantum phenomena...

    ↳ source of natural entropy  
    ⇒ hard to be blocking ⇒ rate-limited

• "pseudo-random"

    • key/seed → start square

    • deterministic

    • repeats eventually

# Linear Congruent Generator

$$x_{n+1} = (ax_n + b) \bmod m \leftarrow \text{prime}$$

large number

seed

Generate Other Distributions ?

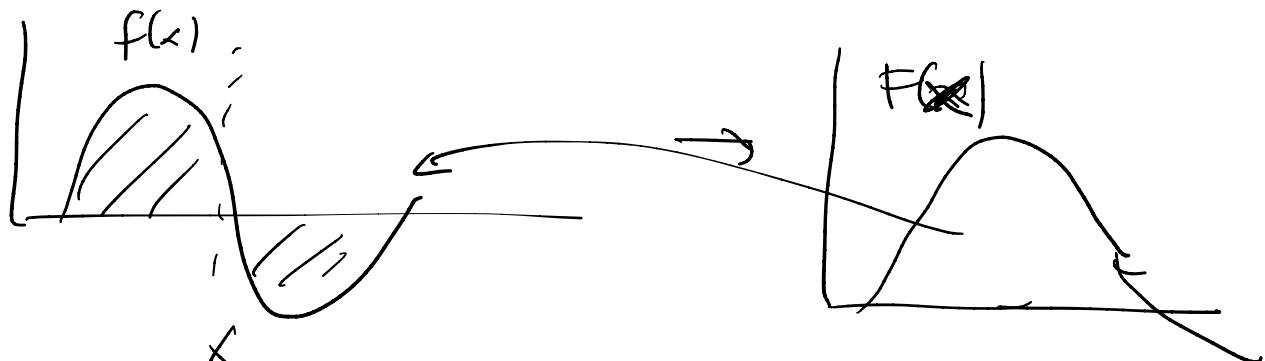
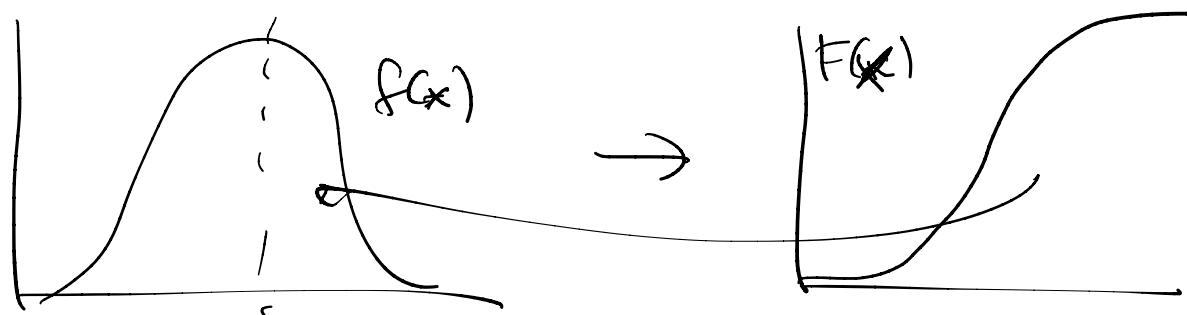
$$y \sim \text{uniform} \Rightarrow x \sim P(y)$$

desired dist

Inversen):  $x \sim f(x)$   $x = \{-\}$

$$F(x) = \int_{-\infty}^x f(x) dx \quad \leftarrow \text{cumulative func}$$

If  $y = F(x)$  is continuous and increasing



$$\Rightarrow x = F^{-1}(y) \quad y \sim \text{unif} \\ x \sim f(x)$$

Example:  $P(t) = \begin{cases} \frac{e^{-t/\tau}}{\tau} & t \geq 0 \\ 0 & t < 0 \end{cases}$

Inverse:  $F(T) = \dots = u \Rightarrow F^{-1}(u) \rightarrow t$   
 $u \sim \text{unif}$   
 $t \sim P(t)$

$$F(T) = \int_0^T \frac{e^{-t/\tau}}{\tau} dt = e^{-t/\tau} \Big|_0^T$$

$$= 1 - e^{-T/\tau} = u$$

$$F^{-1}(u) = -\tau e^{-T/\tau} = u - 1$$

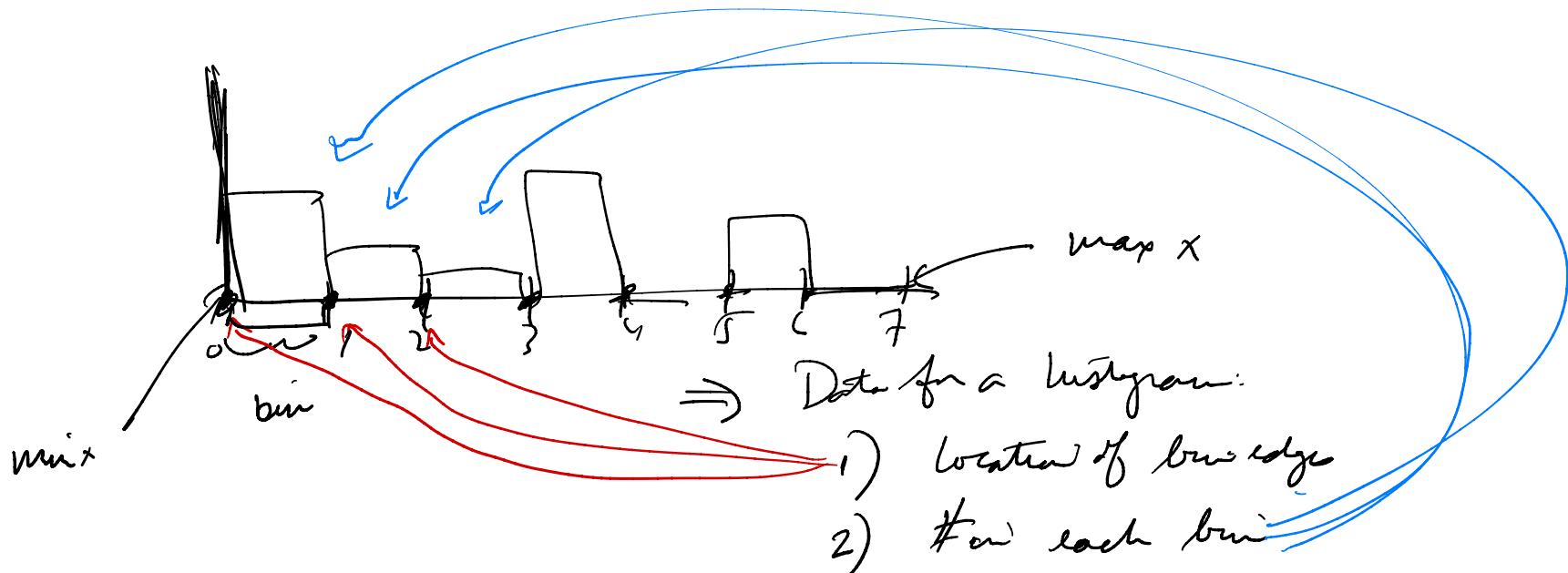
$$e^{-T/\tau} = 1 - u$$

$$-\tau/\tau = \ln(1-u)$$

$$F^{-1}(u) = T = -\tau \ln(1-u)$$

# Histogramming

$x = [1, 1.5, 2, -1, \dots]$  ← want to histogram



Input :  $x$  a list

min  $x$

or bins.

max  $x$

- n-bins defaults to 10
- 1) find min & max  $x$   
(if not provided)
- 2) bin size  $\frac{x_{\text{max}} - x_{\text{min}}}{n\text{-bins}}$

3) Create a list of size n-knis w/ all values = 0.

4) Loop over all values of  $x$  provide

    Loop over all bins(index i)

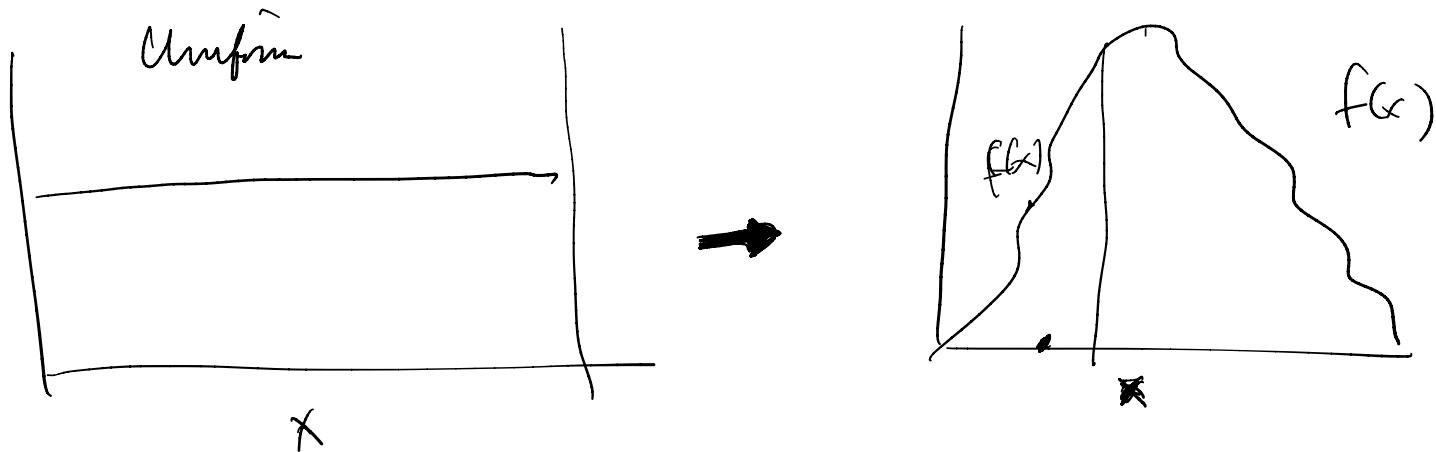
    if      $x_{\text{min}} + i \leq \text{bin-size} < x$  ← lower edge  
           $< x_{\text{min}} + (i+1) \leq \text{bin-size}$

        ⇒ increment bin<sup>i</sup>

    break

    [return]

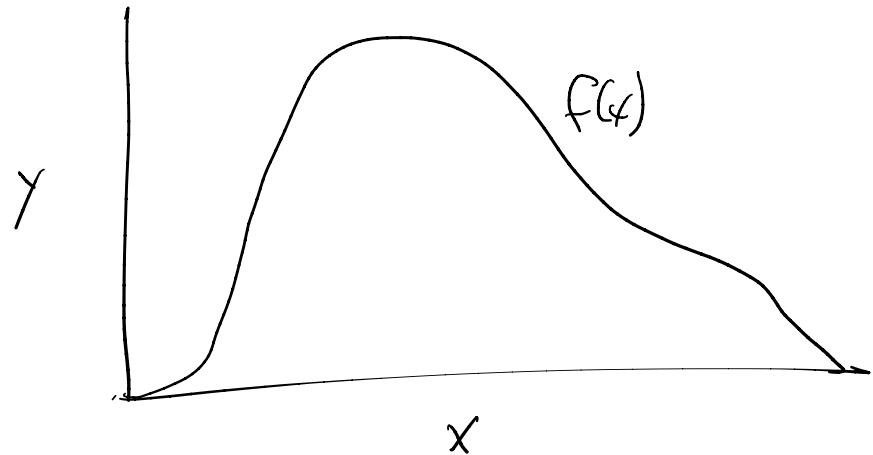
# Accept Reject



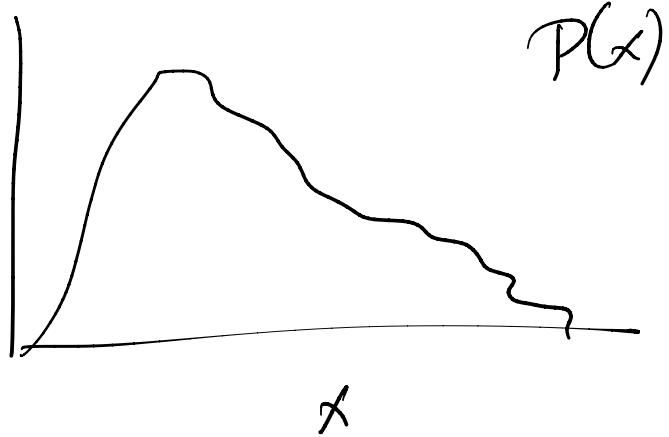
- Draw  $x$  from uniform dist
- Prob of getting  $x$  is  $f(x)$   $\Rightarrow$  draw  $y$  from uniform dist  
if  $y < f(x)$   $\Rightarrow$  accept  $x \Rightarrow$  add to output list  
else  $\Rightarrow$  reject  $x \Rightarrow$  try again

1. Find  $x_{\text{min}}$ ,  $x_{\text{max}}$   
 $y_{\text{min}}$ ,  $y_{\text{max}}$  → assume 0

2.



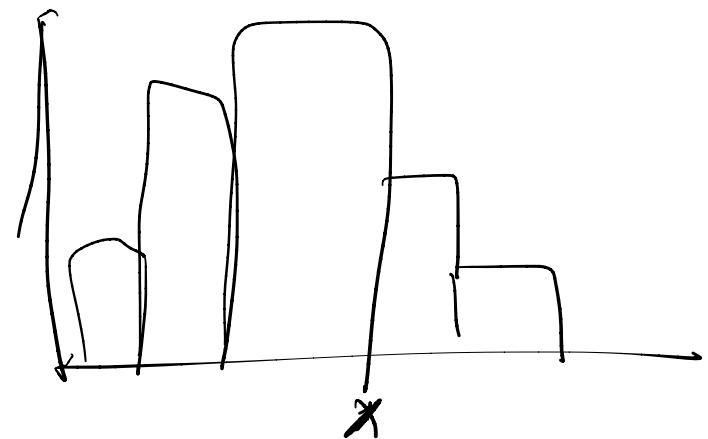
Accept Reject

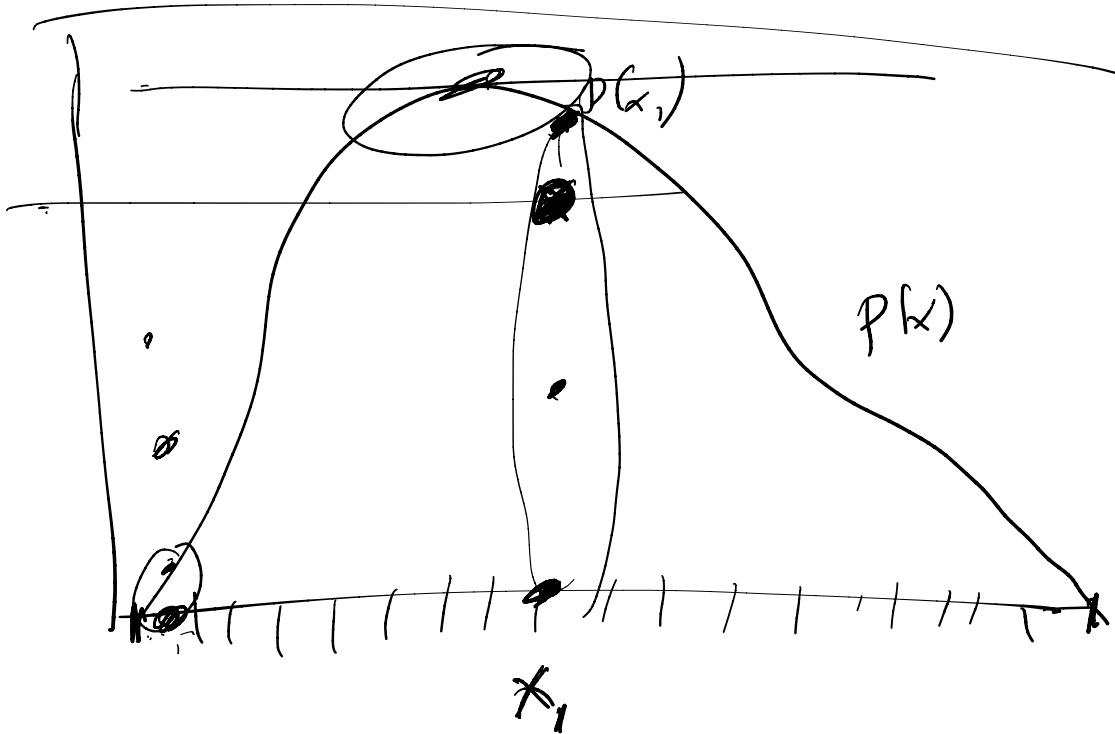


$P(x)$

$x \sim P(x)$

$\Rightarrow x = \{x_1, \dots, x_n\}$





1) Range of  $x \Rightarrow x_{\min} \rightarrow x_{\max}$

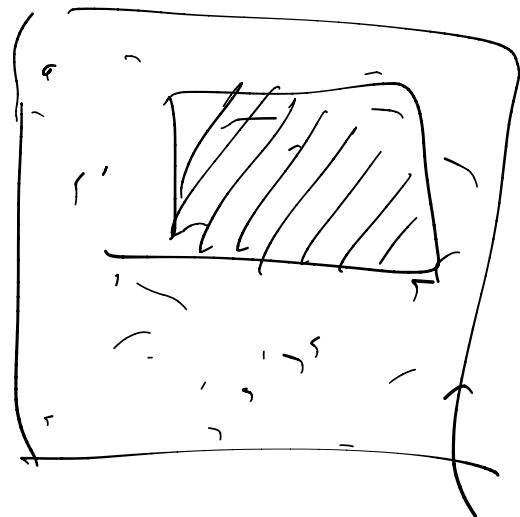
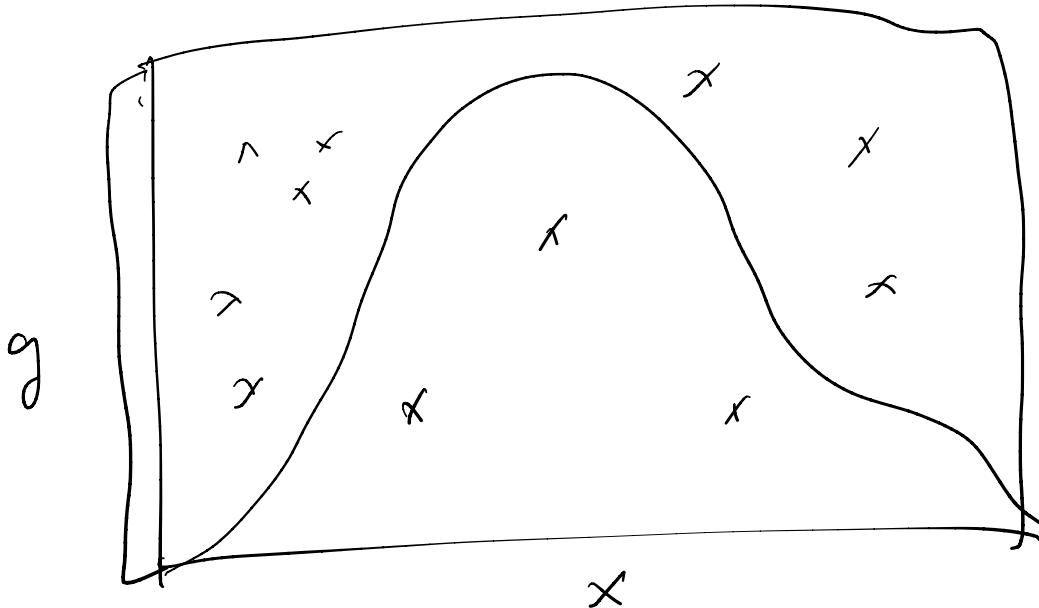
2)  $x \sim \text{uniform } (x_{\min} \rightarrow x_{\max})$  ~~given~~ (provided a given input)

\* 3) Range of  $y \Rightarrow y_{\min} \rightarrow y_{\max}$

do until  
Non output  
list

4)  $y \sim \text{uniform } (y_{\min} \rightarrow y_{\max})$  ~~given~~

5) if  $y < P(x) \Rightarrow$  accept  $x \rightarrow$  put in output list  
else reject



$$\text{Area} = A \quad \frac{\# \text{ accepts}}{\# \text{ trials}}$$

$$A = (x_{\max} - x_{\min})(y_{\max} - y_{\min})$$

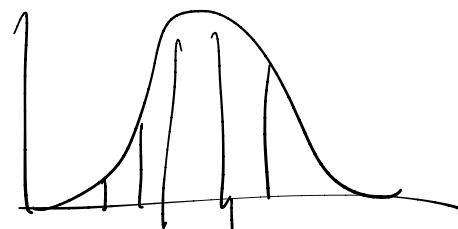
Monte Carlo Integration

10<sup>4</sup>0

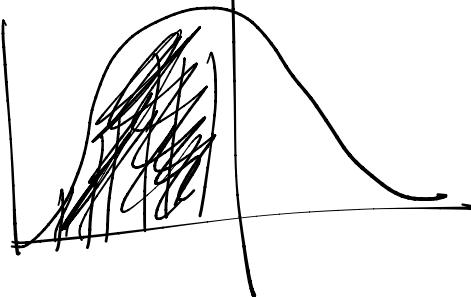
1) generate

$X \sim N(\mu)$   $\Rightarrow$  list of  $X$  values  
↑ normal dist

2) histogram



$\Rightarrow$  histo a list



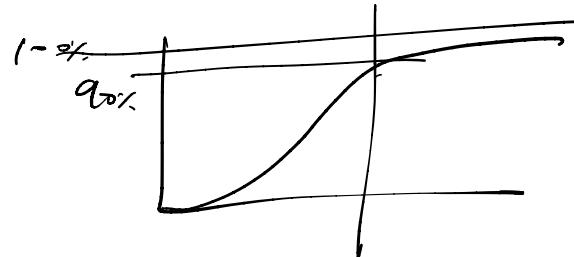
$\rightarrow$  cum sum list

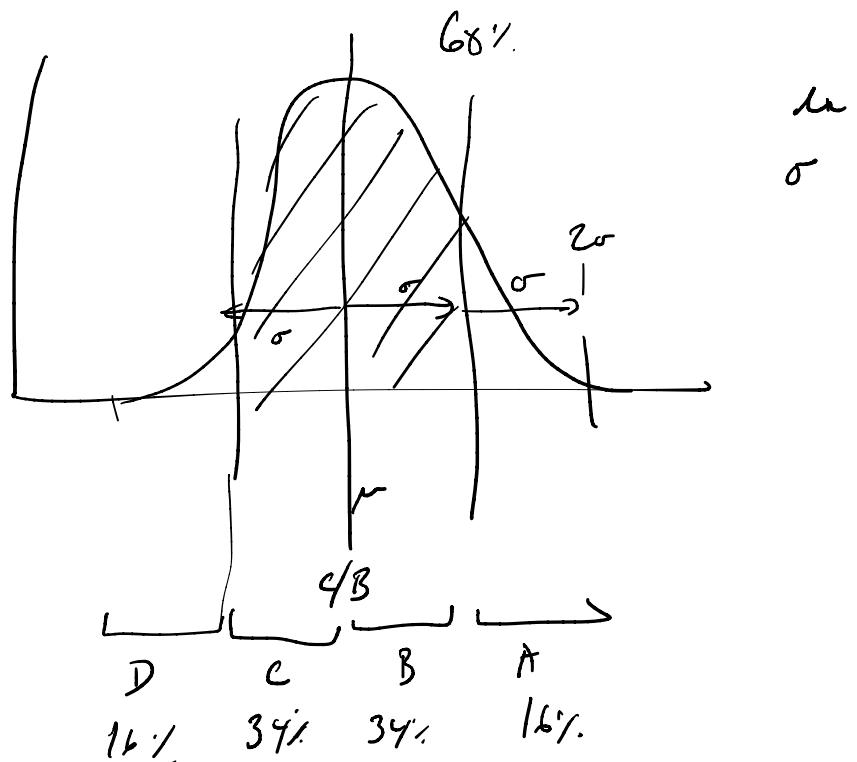
3) cum sum

4) What point is the cum sum 90% of total sum?

900 - 90

5) Get bin edge  $\Rightarrow x_{q_0}$





$\mu$   
 $\sigma$