Probability Theory

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

Analysis:

1) Estimate (measure) parameter
2) Quantify the uncertainty in
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

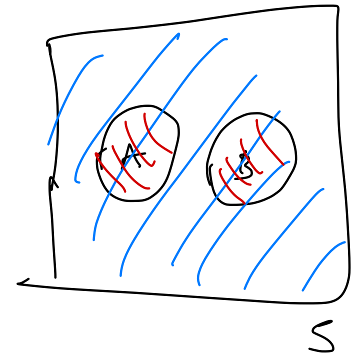
Prob Theory → analyze frequency of "events"

- Same answers
- computation are same

1. prob P of X event happening
   ⇒ repeatable observation of event
      ⇒ P ~ fraction that X would be outcome.
   ⇒ e.g freq of a disease in a population
      ~ 1 in 1000
         ⇒ "Frequentist"

- What if you got a test → How do you interpret?
   - accuracy of test ⇒ can't repeat
   - freq of disease
      ⇒ degree of belief
         ⇒ Bayesian

Basic Definitions:



- Set S w/ subsets A, B
- $P(S) = 1$

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

  subset

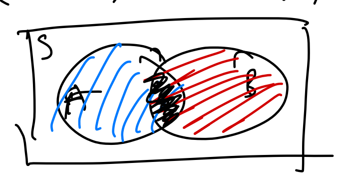
- $A \cap B = \phi$; $P(A \cup B) = P(A) + P(B)$

  overlap ; empty ; union

- $P(\bar{A}) = 1 - P(A)$
- $A \subset B \Rightarrow P(A) \leq P(B)$

  subset of B

- $P(A \cup \bar{A}) = 1$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



- $P(\phi) = 0$

- Conditional Prob:

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

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

$$P(1) = \frac{P(<3 \text{ in } n \text{ events})}{P(\text{events})} = \frac{1/3}{3/6} = \underline{\underline{1/3}}$$

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

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

// Interpretation: i) Relative Freq:

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

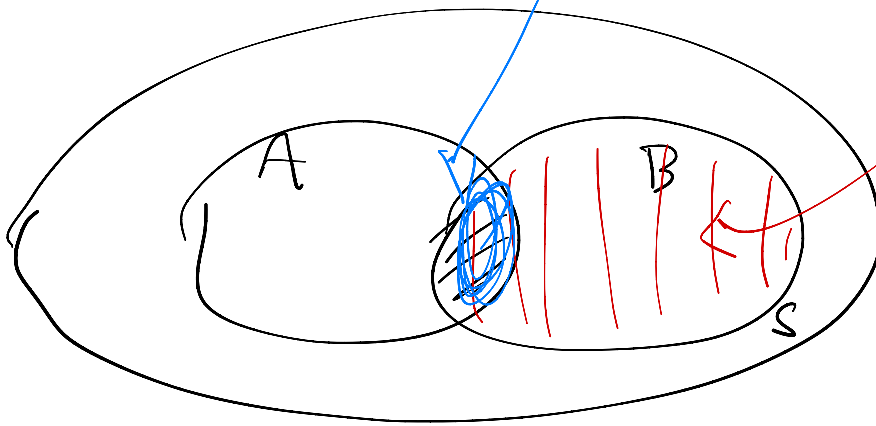
$$P(A) = \lim_{n \to \infty} \frac{\text{times outcome is } A}{n}$$

2) Subjective prob

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

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

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



if $\quad A \cup B = \phi$

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

$$P(A \mid B) = \frac{P(A)\, P(B)}{P(B)} = P(A)$$

Bayes Thm.:

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

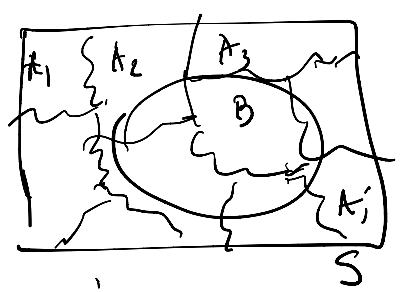
$$P(B|A) = \frac{P(B\cap A)}{P(A)}$$

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

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

measurement

Data



$$\bigcup_i A_i = S$$

$$\Rightarrow B = B\cap S = B\cap\left(\bigcup_i A_i\right) = \underline{\bigcup_i B\cap A_i}$$

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

posterior

prior

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

$$\left.\begin{aligned} P(\text{Sick}) &= 0.001 \\ P(\text{no sick}) &= 0.999 \end{aligned}\right\}$$ Prior knowledge of the population

$P(+ \mid \text{sick}) = 0.98$. True Positive

$P(- \mid \text{sick}) = 0.02$. false neg

$P(+ \mid \text{not sick}) = 0.03$. false pos

$P(- \mid \text{not sick}) = 0.97$. True neg

wrong

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

Test ⇒ get pos results ⇒ what the prob that you are sick

$$P(\text{sick} \mid +) = \frac{\overset{.98}{P(+ \mid \text{sick})}\ \overset{.001}{P(\text{sick})}}{\underset{.98}{P(+ \mid \text{sick})}\ \underset{.001}{P(\text{sick})} + \underset{.03}{P(+ \mid \text{not sick})}\ \underset{.999}{P(\text{no sick})}} = \underline{0.032}$$

	Age	Major	GPA	...
Student 1				
2				
3				
4				

Age = { 20, 21, 18, 25, 35, ... }

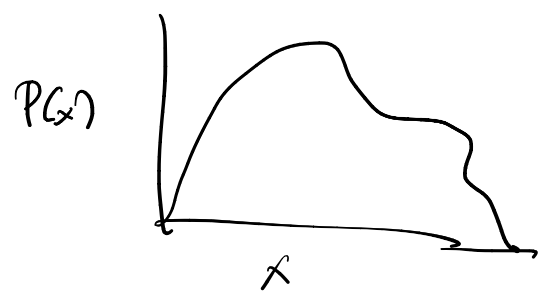
Age is a random variable

Random variable

X is a random variable $\Rightarrow$ "drawn" from

Prob mass distribution

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



$P(x_1)$ = Prob of getting x_1

- Freq: frac of times you get x_1
- Bayes: how likely it is to measure x_1

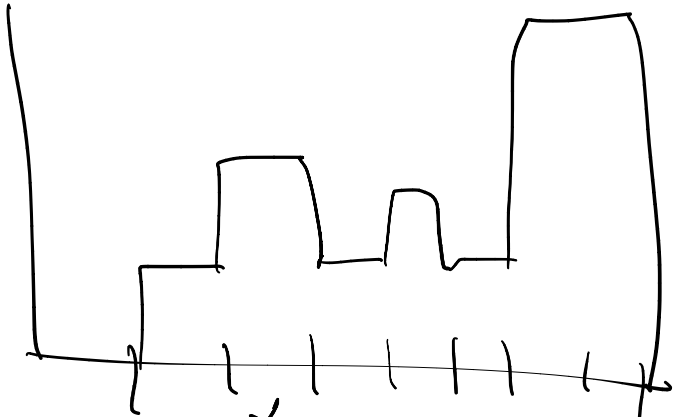
X can be continuous $\Rightarrow$ $X = x_1$ is very unlikely

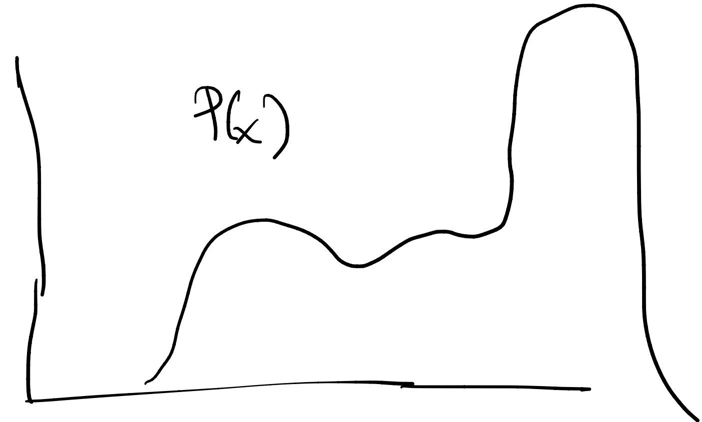
$P(X \in [x_1, x_1 + \delta x])$ Prob density func

$\lim_{\delta x \to 0} P(x) = p(x)\, \underline{dx}$

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

Counting how often we see X in specific range.





Random Variables

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

Discrete or continuous

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

Discrete $\Rightarrow$ Prob Mass Distribution (PMD)

Continuous $\Rightarrow$ Prob Density Distribution (PDF)

$P(x) \Leftarrow$ func

$P(x = x_1) \leftarrow$ a number

$P(x_1)$

$\Rightarrow$

$P(x_1) = 1$

$\Rightarrow$ x is always x_1

$P(x_1) = 0$

$\Rightarrow$ x is never x_1

$x \sim P(x)$

"Drawn"

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

Properties of $P(x)$

- Domain of P must be all possible values of x
- $\forall x_1 \in x \qquad 0 \leq P(x_1) \leq 1$
- $$\sum_{x_1 \in x} P(x) = 1$$ $\leftarrow$ normalization

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

$$P(x \in [x_1, x_1 + \delta x]) = p(x)\, dx$$

capital P

lower case "p"

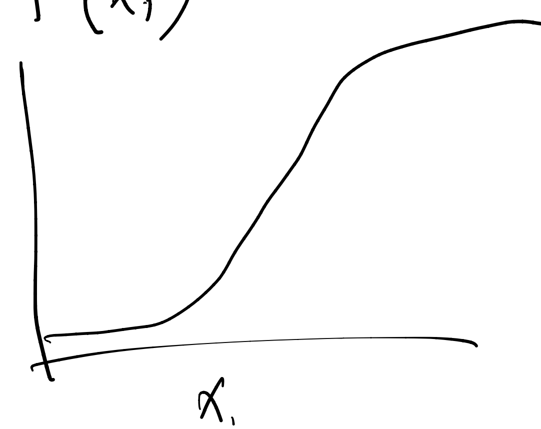
Prob Dist Func

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

Cummulative Prob:

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

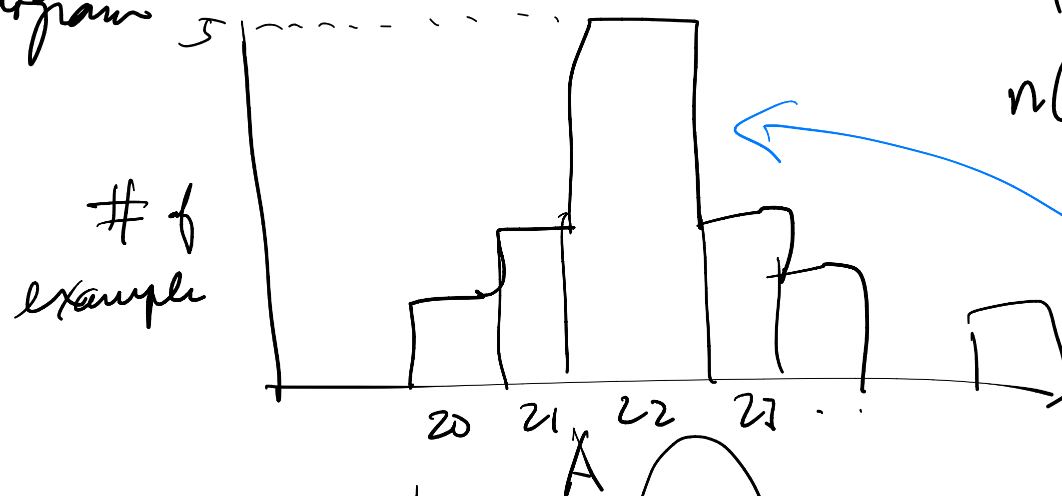
$\Rightarrow$ $F(x_1)$



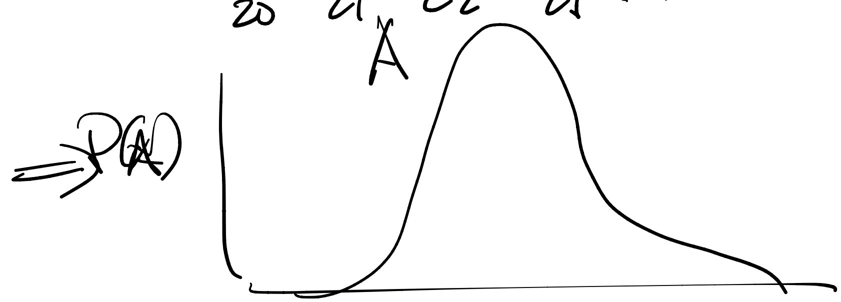
Data

name	(A) Age	(H) Height	(W) Weight

Histogram



$n(x)$

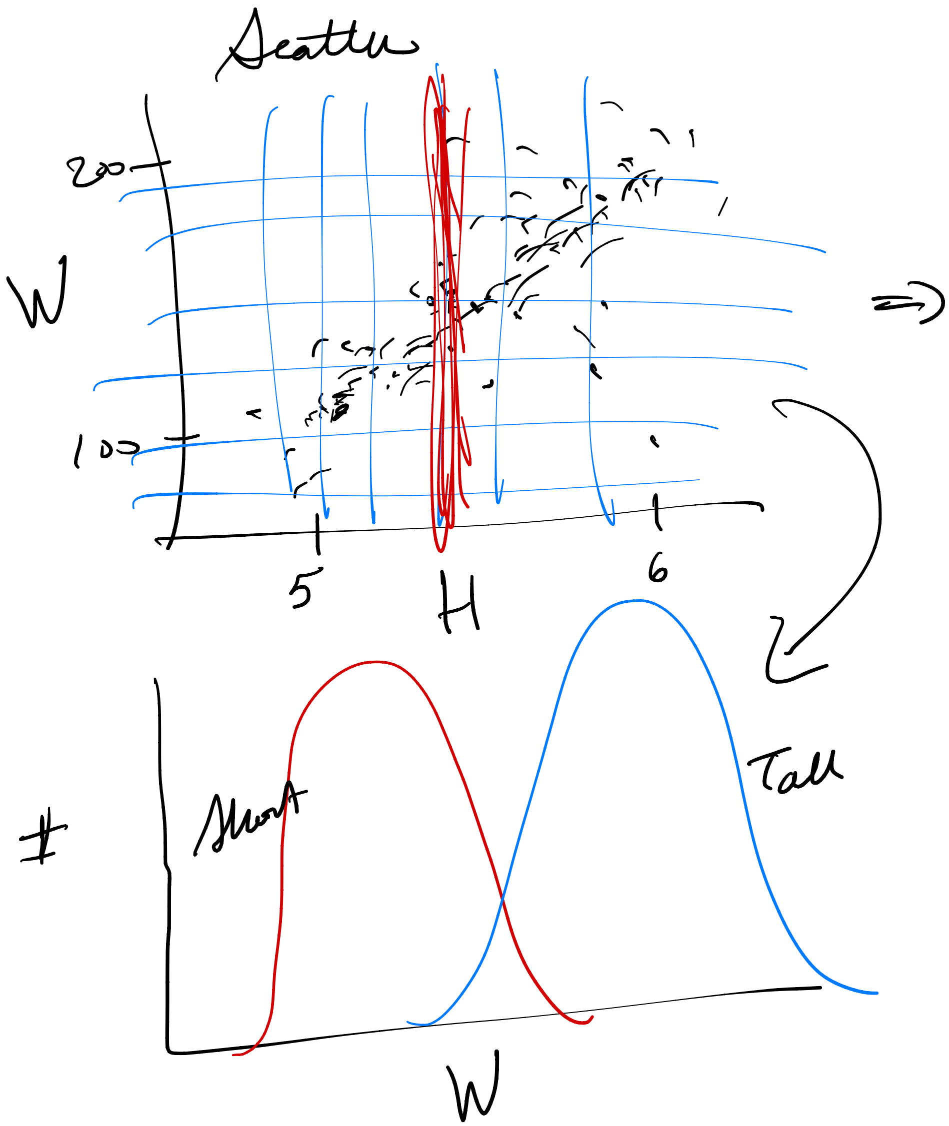


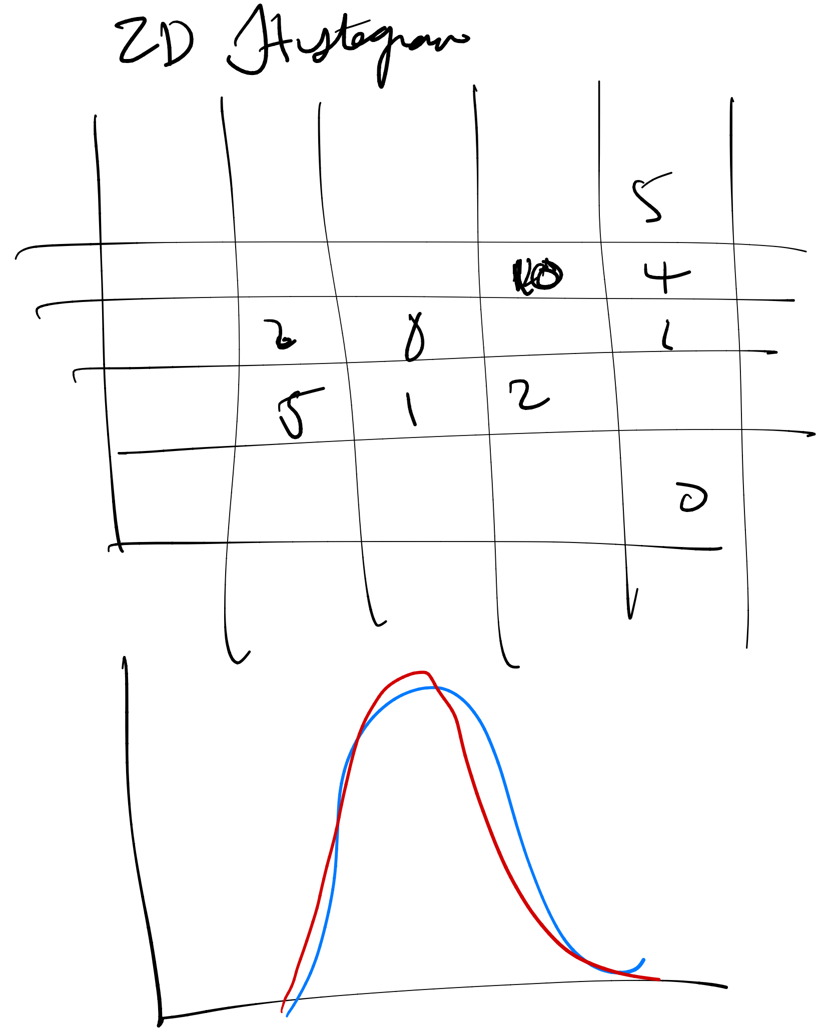
$$P(x) \approx \frac{n(x)}{N\,\Delta x}$$

Total number of data points in histogram

width of a bin

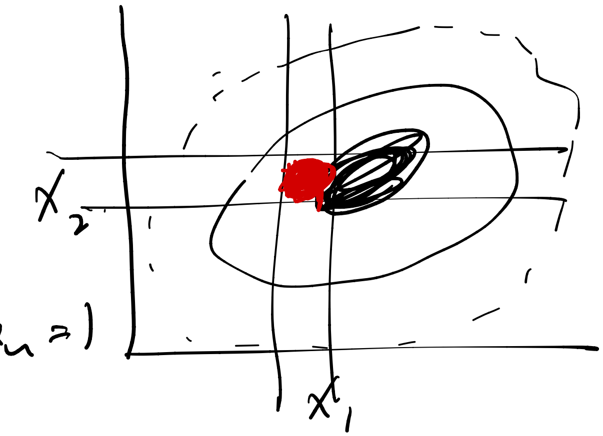
$$\lim_{\substack{N \to \infty \\ \Delta x \to 0}} \frac{n(x)}{N\,\Delta x} = P(x)$$





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

Multivariate: $P(x_1, x_2)\, dx_1\, dx_2$



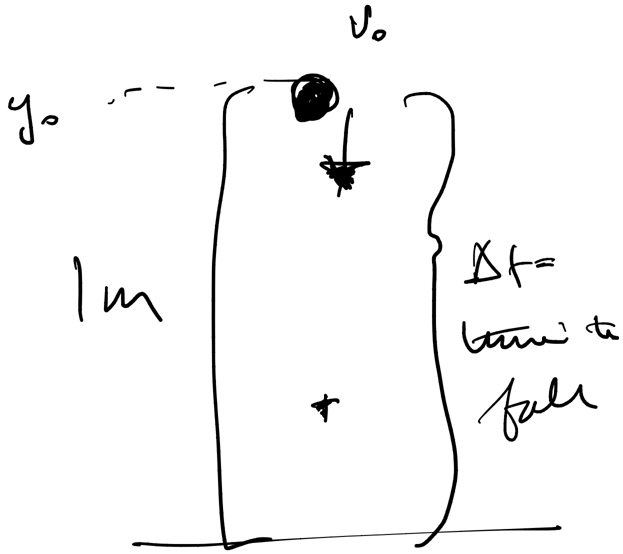
normalization: $\int P(x_1, \dots x_n)\, dx_1 \dots dx_n = 1$

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

independence: if $P(x_1, x_2) = P_1(x_1)\, P_2(x_2)$

$\Rightarrow$ x_1 & x_2 are independent

Simple Experiment



$g = 9.8\ m/s^2$

measure Δt $\Rightarrow$ estimate g

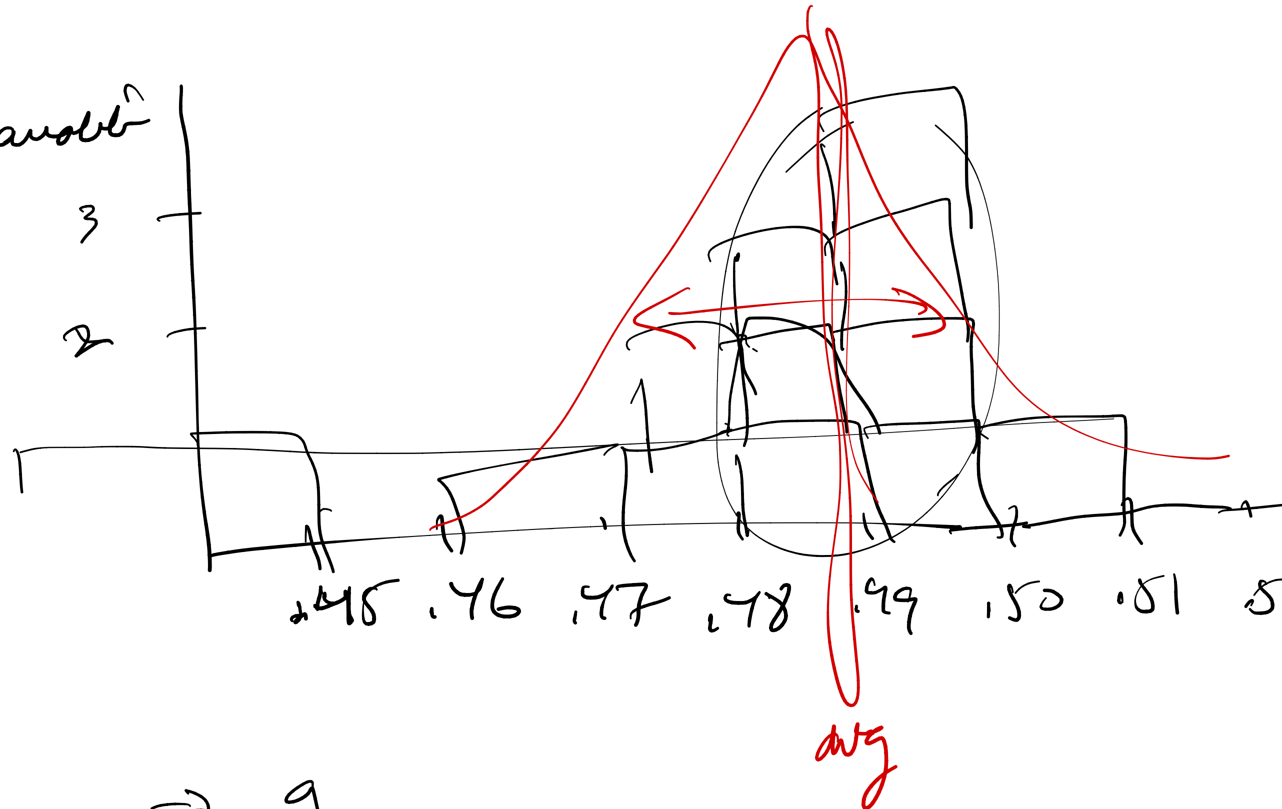
$$y = y_0 + \cancel{v_0 t}^{\,0} + \frac{1}{2}\underbrace{a}_{=-g} t^2$$

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

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

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

trial	Δf random variable
1	.47
2	.42
3	.49
4	.58
⋮	⋮



$\Rightarrow g$

Measurement

Define Expectation Value

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

random variable

dist x is drawn from

func

$$\mathbb{E}[x] = \int x\, P(x)\, dx \equiv \mu$$

"mu" in Greek

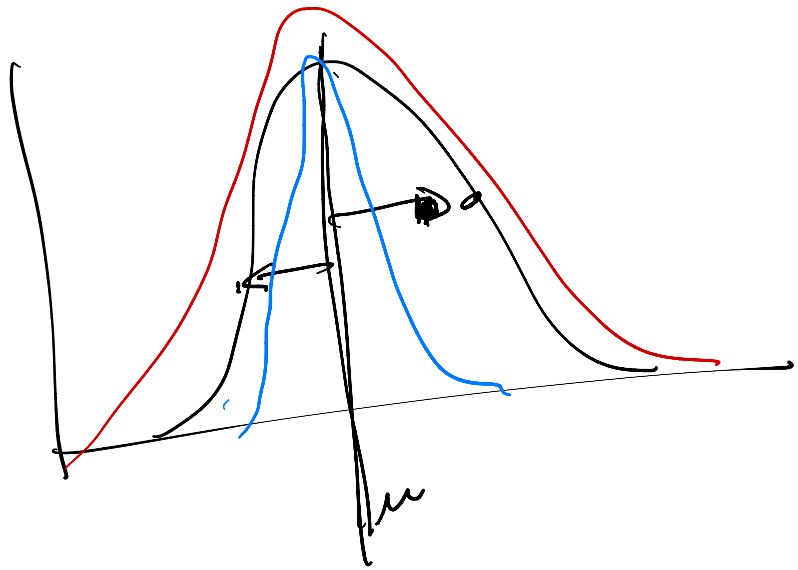
mean

$$= \sum_{x=\{\}} x\, P(x)$$

Variance: $E[x^2] - \mu^2 = E[(x-\mu)^2] = \sigma^2$

"sigma"

$\sigma = \sqrt{\sigma^2}$ = "standard deviation"



$P(x,y) \Rightarrow$ covariance: $\mathrm{cov}[x,y] = \mathbb{E}[xy] - \mu_x \mu_y$

$= \mathbb{E}[(x-\mu_x)(y-\mu_y)]$

correlation coef: $\rho_{xy} = \dfrac{\mathrm{cov}[x,y]}{\sigma_x \sigma_y}$

Binomial: N independent test $\rightarrow$ success
$\rightarrow$ failure

N = # of trials

n = # of successes

p = underlying prob for success

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

f(n)

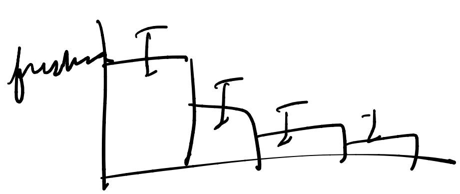
5

@ $p = .5$

· TPR, FPR, ...

· Efficiency

· What fraction of ...?



A

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

$V[n] = Np(1-p)$

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

$\Rightarrow$ 98 = 98% = p

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

$V[p] = V[\frac{n}{N}] =$

$p(1-p)$

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

Start of Binomial $\quad N \to \infty \quad p \to 0 \quad \Rightarrow \quad E[n] = Np \to \nu$

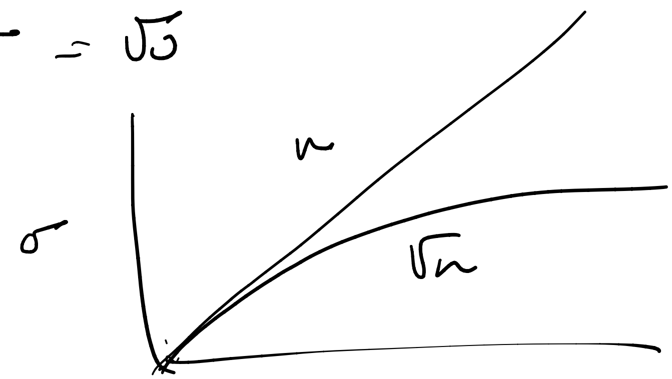
"nu"

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

$$\mathbb{E}[n] = \nu$$

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

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

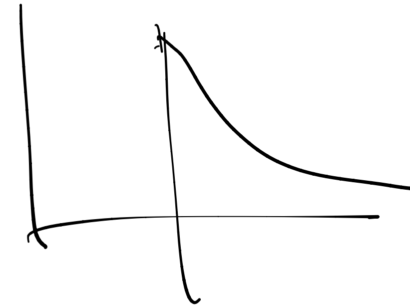


How many $\to$ rate

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

Exponential Dist

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



$\tau \to 0$

$\tau = \frac{1}{\nu}$

$E[t] = \tau$

$V[t] = \tau^2 \Rightarrow \sigma = \tau$

How long?

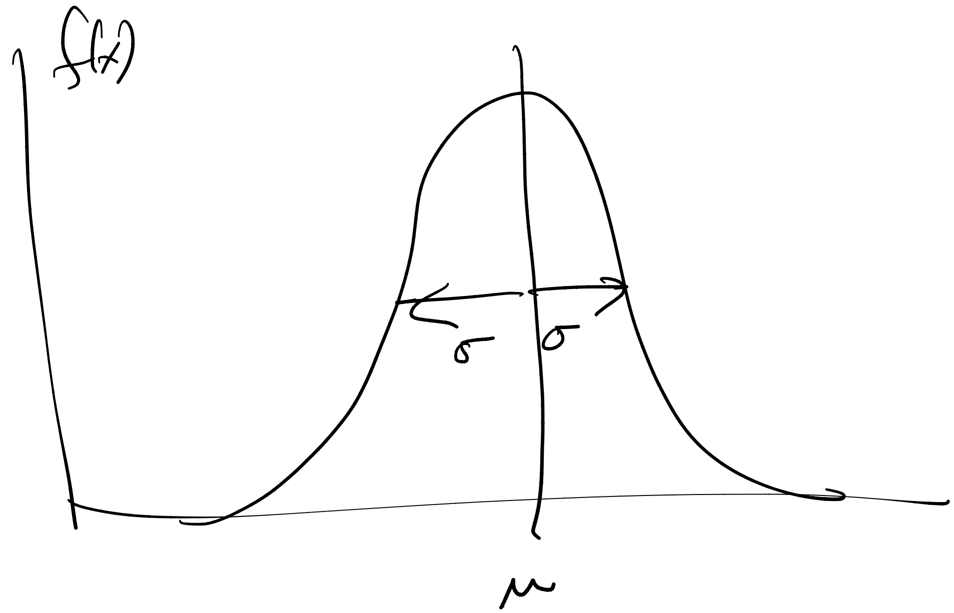
Central Limit Thm: add (combine) random variables

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

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

$E[x] = \mu$

$V[x] = \sigma^2 = \sigma \to \sigma$



f(x)
σ
σ
μ

How to Compute ...

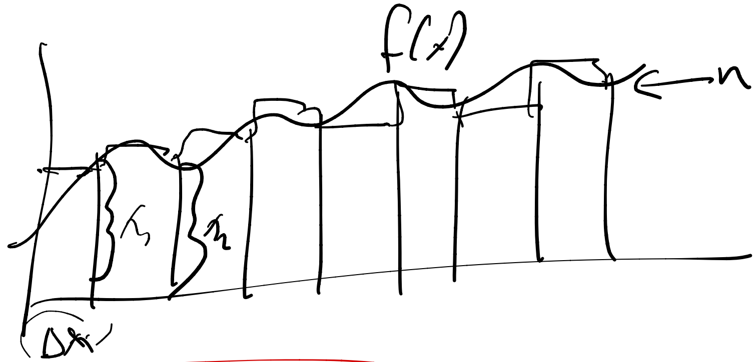
mean: $\mathbb{E}[x]$ $\quad x \sim P(x)$: $\mu \equiv \int x\, f(x)\, dx \approx \sum_i x_i f(x_i) \Delta x$

$\equiv \sum_i x_i P(x_i)$

over all possible values of x

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

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



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

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

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

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

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

Variance

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

$$= E[(x-\mu)^2] \qquad = \frac{1}{N} \sum_{x \in \text{Dataset}} x_i^2 - \mu^2$$

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

x_i ← Dataset

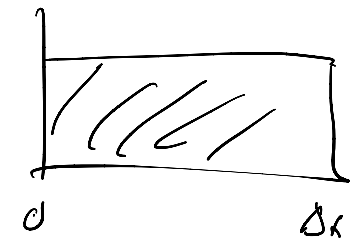
$$\sigma = \sqrt{\frac{1}{N-1} \sum_i (x_i - \mu)^2}$$

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
    ⇒ tend to be blocking ⇒ rate-limited
  - "pseudo-random"
    - key/seed → start sequence
    - deterministic
    - repeats eventually

Linear Congruent Generator

$$x_{n+1} = (a x_n + b) \bmod m$$

large number

prime

"%"

seed

Generate Other Distributions?

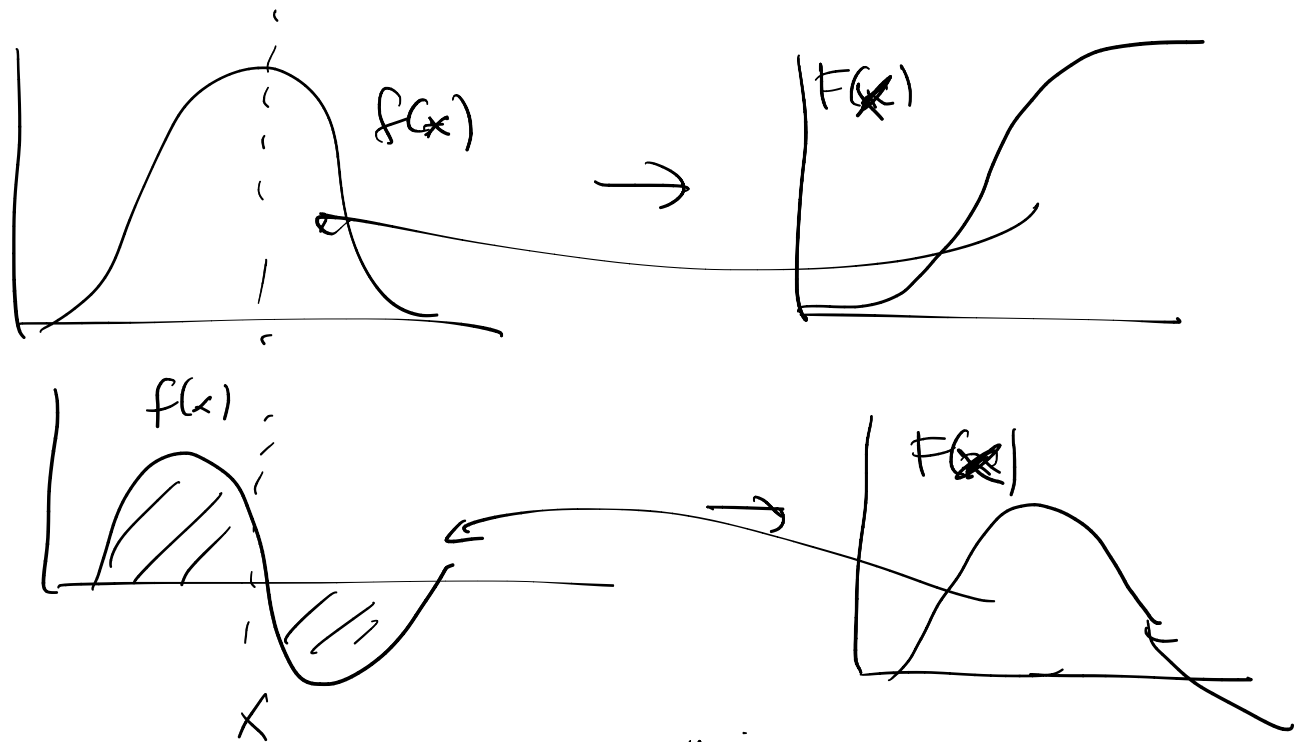
$$y \sim \text{uniform} \Rightarrow x \sim P(y)$$

desired dist

Inversion: $x \sim f(x)$ $\quad x = \{ \cdot - \}$

$$F(X) = \int_{-\infty}^{X} f(x)\,dx$$ ← cumulative func

if $y = F(\cancel{x})$ is continuous and increasing



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

$x \sim f(x)$

Example:

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

Inverse: $F(T) = \cdots = u \Rightarrow F^{-1}(u) \to t$

$u \sim$ uniform

$t \sim P(t)$

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

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

$F^{-1}(u)$

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

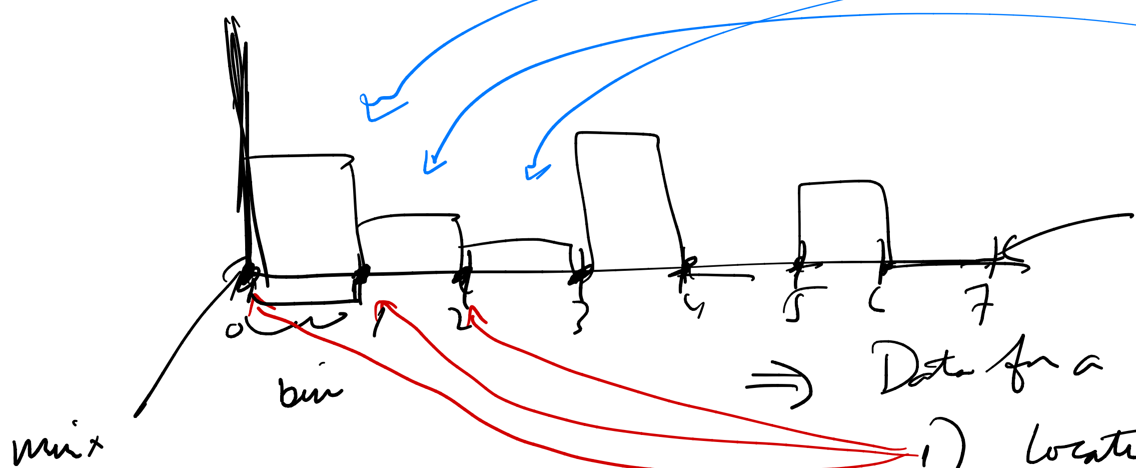
$$e^{-T/\tau} = 1 - u$$

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

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

Histogramming

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



⟹ Data for a histogram:

1) location of bin edges
2) # in each bin

Input: x a list

min x   n-bins.

max x

- n_bins defaults to 10

1) find min x & max x (if not provided)
2) bin size $\dfrac{x_{max} - x_{min}}{n\_bins}$

3) Create a list of size n_bins w/ all values = 0.

4) Loop over all values of x provided

Loop over all bins (index i)

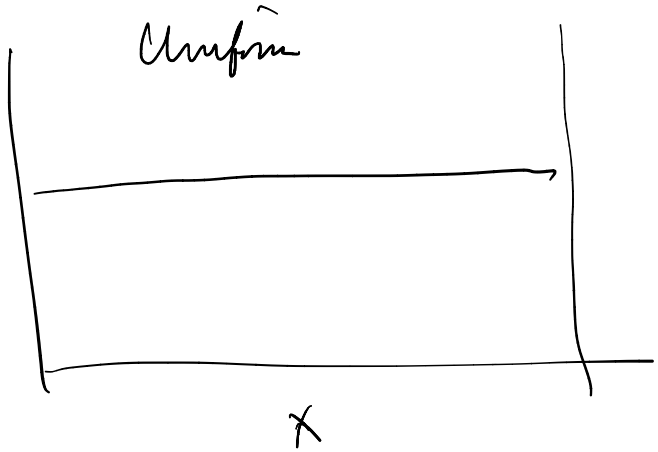
if $x\_min + i * bin\_size < x < x\_min + (i+1) * bin\_size$ ← lower edge

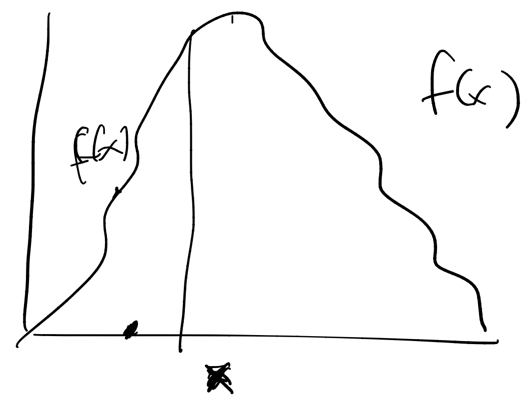
⇒ increment bin i

break

continue

Accept Reject

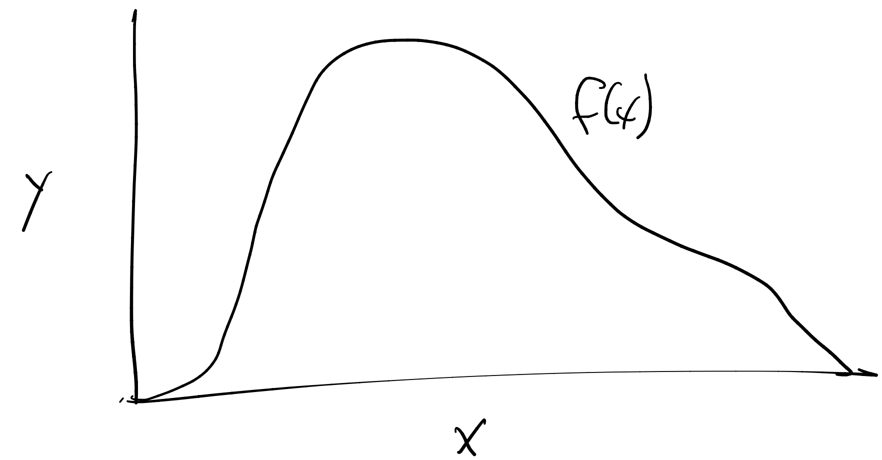




- Draw x from uniform dist
- Prob of getting x is f(x) ⇒ draw y from uniform dist

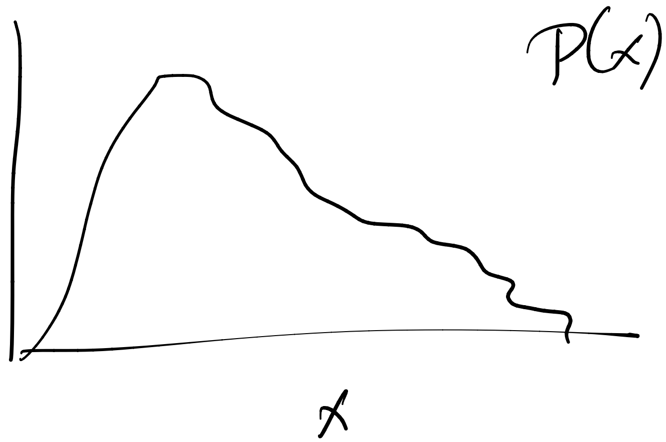
  if $y < f(x)$ ⇒ accept x ⇒ add to output list
  else ⇒ reject x ⇒ try again

1. Find x_min, x_max
y_min, y_max ← assume 0



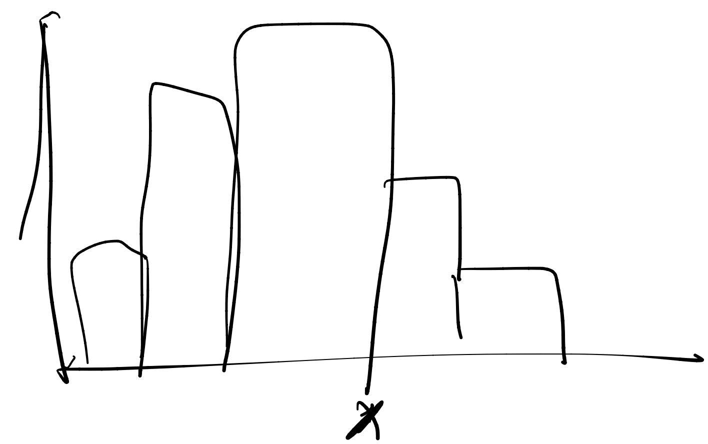
2.

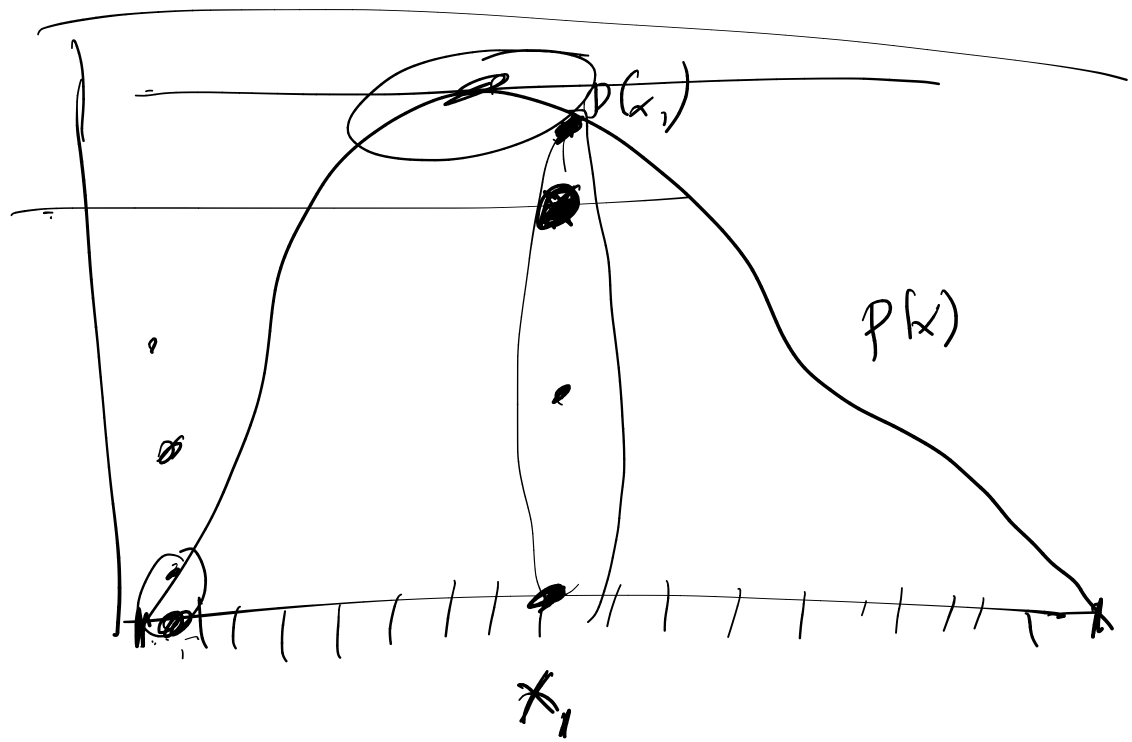
Accept Reject



$x \sim P(x)$

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





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

2) $x \sim$ uniform $(x_{min} \rightarrow x_{max})$ ← given (provided as input)

$y_{min} = 0$

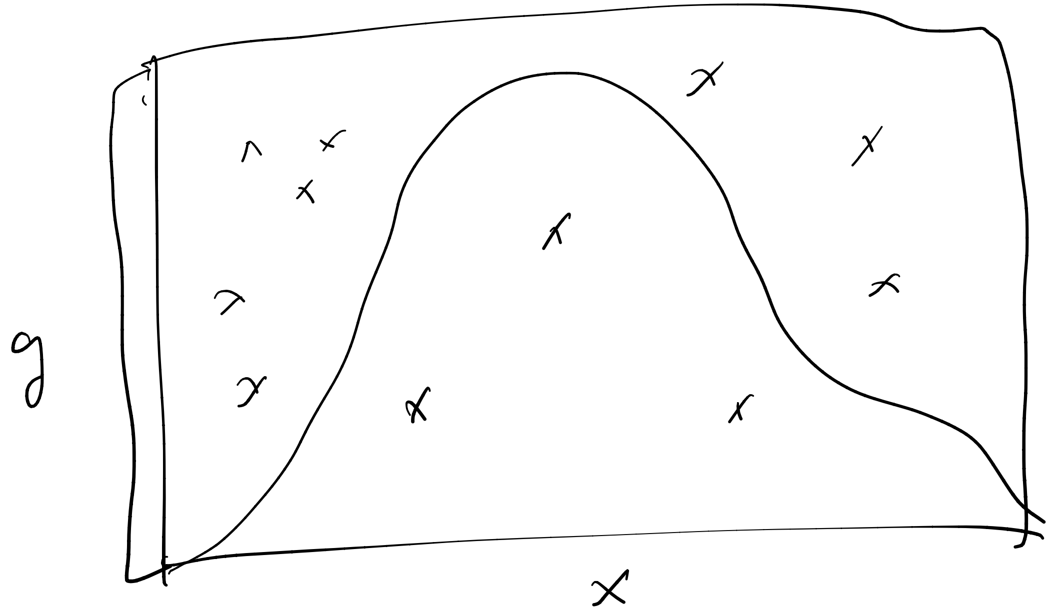
3) Range of y $\Rightarrow$ $y_{min} \rightarrow y_{max}$

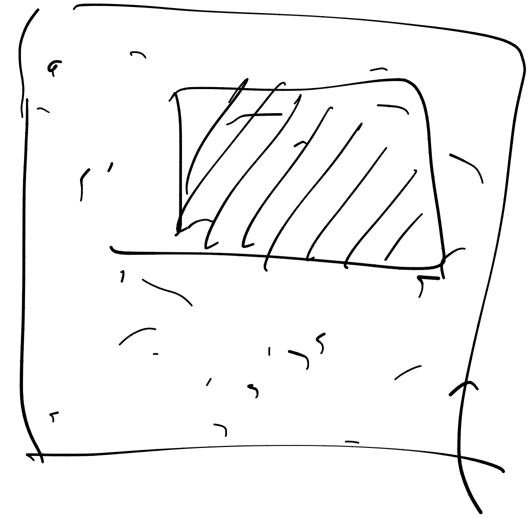
4) $y \sim$ uniform $(y_{min} \rightarrow y_{max})$ ←

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

else reject

do until N in output list





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

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

<u>Monte Carlo</u> Integration

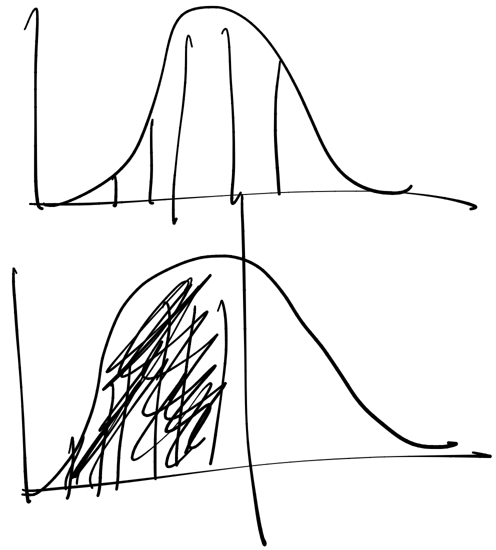
1080

1) Generate

$X \sim N(x)$ ⇒ list of x values

normal dist

2) histogram



⇒ histo a list

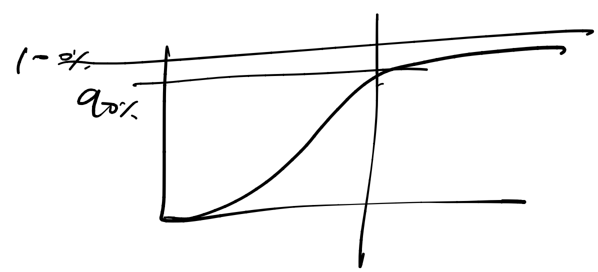
3) cum sum

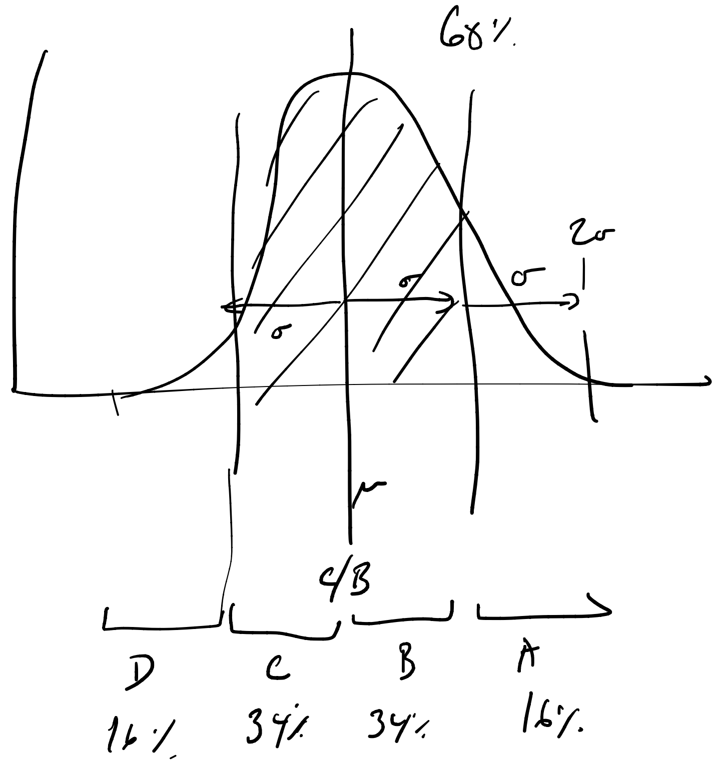
→ cum sum list

900 - 48

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

5) Get bin edge ⇒ x_{90}





μ

σ