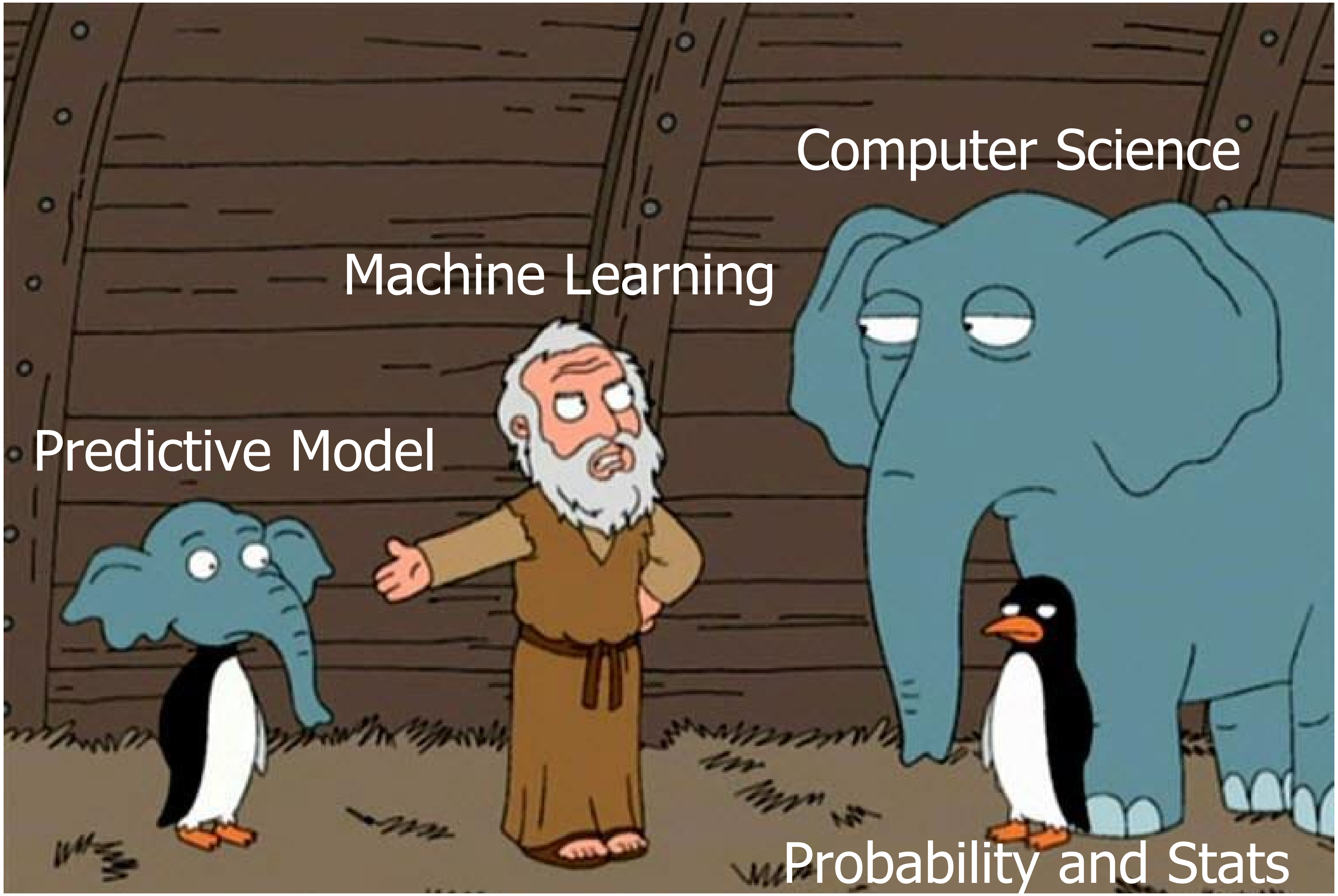

Computer Science
Machine Learning
Predictive Model
Probability and Stats

Probability and Statistics

Mahdi Roozbahani
Georgia Tech

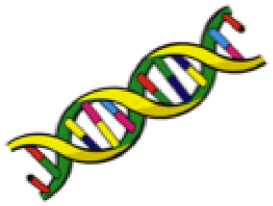
These slides are inspired based on slides from Le Song , Sam Roweis, and Chao Zhang.

Outline

- Probability Distributions
- Joint and Conditional Probability Distributions
- Bayes' Rule
- Mean and Variance
- Properties of Gaussian Distribution
- Maximum Likelihood Estimation

Probability

- A sample space S is the set of all possible outcomes of a conceptual or physical, repeatable experiment. (S can be finite or infinite.)
  - E.g., S may be the set of all possible outcomes of a dice roll: S $\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \end{pmatrix}$
  - E.g., S may be the set of all possible nucleotides of a DNA site: S $\begin{pmatrix} A & C & G & T \end{pmatrix}$



  - E.g., S may be the set of all possible time-space positions of an aircraft on a radar screen.
- An Event A is any subset of S
  - Seeing "1" or "6" in a dice roll; observing a "G" at a site; UA007 in space-time interval

Three Key Ingredients in Probability Theory

A **sample space** is a collection of all possible **outcomes**

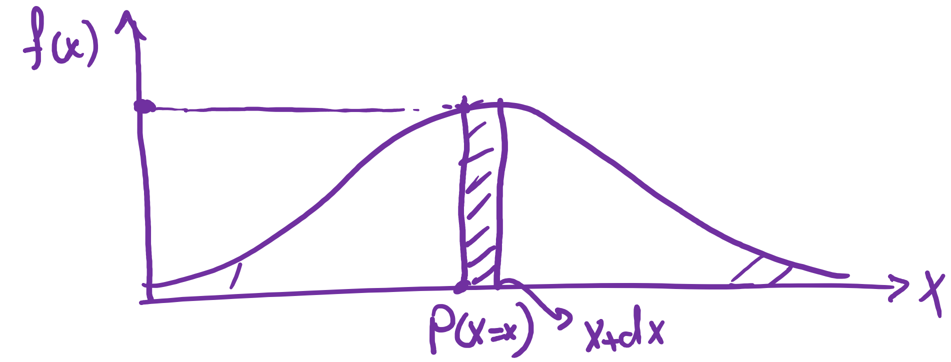
Random variables X represents **outcomes** in sample space

$P(X=2)=\frac{1}{6}$

Probability of a random variable to happen $p(x) = p(X = x)$

$p(x) \geq 0$

density or likelihood



Continuous variable

Continuous probability distribution

Probability density function

Density or likelihood value

Temperature (real number)

Gaussian Distribution

$$\int_x p(x)dx = 1$$

Discrete variable

Discrete probability distribution

Probability mass function

Probability value

Coin flip (integer)

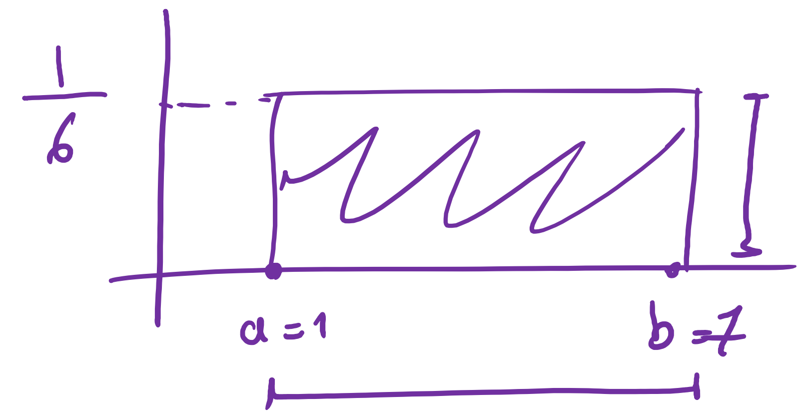
Bernoulli distribution

$$\sum_{x \epsilon A} p(x) = 1$$

Continuous Probability Functions

- Examples:
  - Uniform Density Function:

$$f_x(x) = \begin{cases} \frac{1}{b-a} & for\ a \le x \le b \\ 0 & otherwise \end{cases}$$



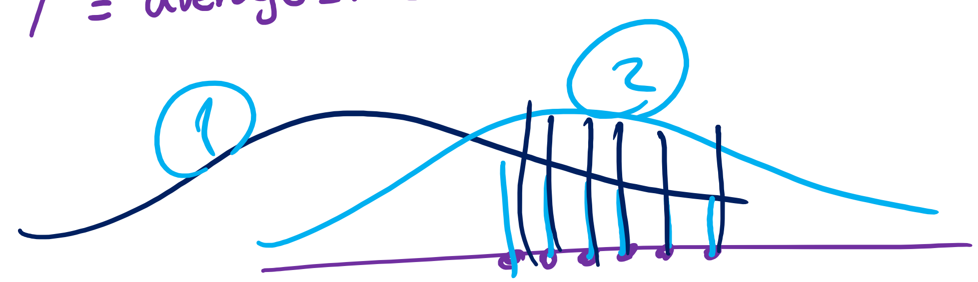
  - Exponential Density Function:

Parameter

$$f_x(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \qquad for\ x \ge 0$$

$$F_x(x) = 1 - e^{\frac{-x}{\mu}} \qquad for\ x \ge 0$$

μ = average = mean



  - Gaussian(Normal) Density Function

$$f_x(x) = \frac{1}{\sqrt{2\pi}\cancel{\sigma}\,b} e^{-\frac{(x-\cancel{\mu}\,a)^2}{2\cancel{\sigma^2}\,b^2}}$$

$\mu \rightarrow$ mean

$\sigma \rightarrow$ Standard deviation

$\sigma^2 \rightarrow$ Variance

Discrete Probability Functions

- Examples:
  - Bernoulli Distribution:
    - $\begin{cases} 1-p & for\ x=0 \\ p & for\ x=1 \end{cases}$

In Bernoulli, just a **single** trial is conducted

  - Binomial Distribution:
    - $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$

k is number of successes

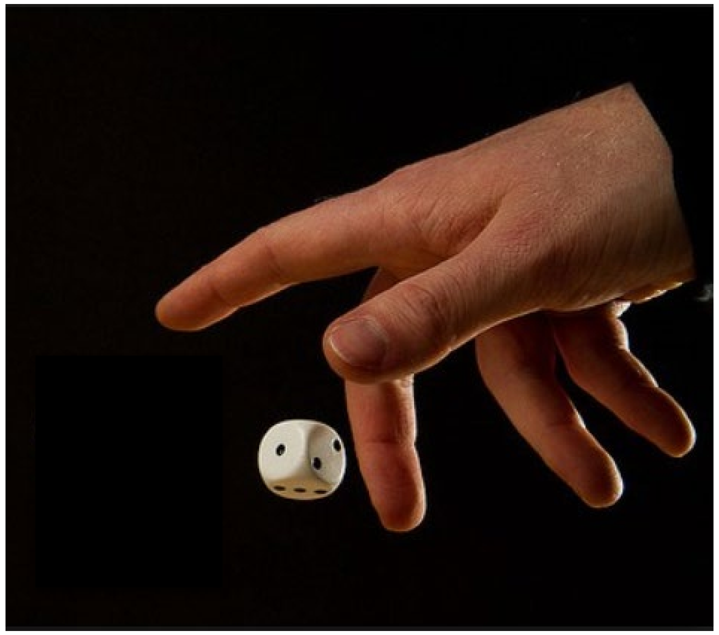
n-k is number of failures

$\binom{n}{k}$ The total number of ways of selection **k** distinct combinations of **n** trials, **irrespective of order**.

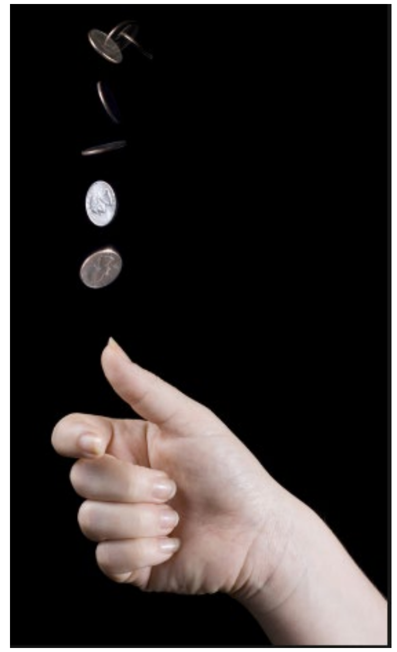
Outline

- Probability Distributions
- Joint and Conditional Probability Distributions
- Bayes' Rule
- Mean and Variance
- Properties of Gaussian Distribution
- Maximum Likelihood Estimation

Example



X = Throw a dice



Y = Flip a coin

X and **Y** are random variables

N = total number of trials

$\boldsymbol{n_{ij}}$ = Number of occurrence

X

		$x_{i=1} = 1$	$x_{i=2} = 2$	$x_{i=3} = 3$	$x_{i=4} = 4$	$x_{i=5} = 5$	$x_{i=6} = 6$	C_j
Y	$y_{j=2} = tail$	$n_{ij} = 3$	$n_{ij} = 4$	$n_{ij} = 2$	$n_{ij} = 5$	$n_{ij} = 1$	$n_{ij} = 5$	20
	$y_{j=1} = head$	$n_{ij} = 2$	$n_{ij} = 2$	$n_{ij} = 4$	$n_{ij} = 2$	$n_{ij} = 4$	$n_{ij} = 1$	15
	C_i	5	6	6	7	5	6	N=35

X

Y		$x_{i=1}=1$	$x_{i=2}=2$	$x_{i=3}=3$	$x_{i=4}=4$	$x_{i=5}=5$	$x_{i=6}=6$	C_j
	$y_{j=2}=tail$	$n_{ij}=3$	$n_{ij}=4$	$n_{ij}=2$	$n_{ij}=5$	$n_{ij}=1$	$n_{ij}=5$	20
	$y_{j=1}=head$	$n_{ij}=2$	$n_{ij}=2$	$n_{ij}=4$	$n_{ij}=2$	$n_{ij}=4$	$n_{ij}=1$	15
	C_i	5	6	6	7	5	6	N=35

$$P(x=2, y=\text{tail}) = \frac{4}{35} = \frac{n_{ij}}{N}$$

$$P(y=\text{head}) = \frac{15}{35} = \frac{C_j}{N}$$

$$P(x=3) = \frac{6}{35} = \frac{C_i}{N} = \sum_{y} P(x=3, y=y) \rightarrow \text{Sum Rule}$$

$$P(y=\text{tail} \mid x=4) = \frac{5}{7} = \frac{n_{ij}}{C_i}$$

$$P(x=4 \mid y=\text{tail}) = \frac{5}{20} = \frac{n_{ij}}{C_j}$$

$$P(x,y) = \frac{n_{ij}}{N} = \frac{n_{ij}}{C_i}\frac{C_i}{N} = \frac{n_{ij}}{C_j}\frac{C_j}{N}$$

$$P(x,y) = P(y|x)\,P(x) = P(x|y)\,P(y)$$

↳ product rule

Probability:

$$p(X = x_i) = \frac{c_i}{N}$$

Joint probability:

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

Conditional probability:

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Sum rule

$$p(X = x_i) = \sum_{j=1}^{L} p\left(X = x_i, Y = y_j\right) \Rightarrow \; p(X) = \sum_{Y} P(X, Y)$$

Product rule

$$p\left(X = x_i, Y = y_j\right) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i}\frac{c_i}{N} = p\left(Y = y_j \middle| X = x_i\right)p(X = x_i)$$

$$p(X, Y) = p(Y|X)p(X)$$

$P(a,b) = P(a|b)\,P(b)$ $P(a,b,c) = P(a|\underline{b,c})\,P(b,c)$

Conditional Independence

- Examples:

$P(H,F,V,D) = P(H|F,V,D) \times P(F,V,D)$

$= P(H|F,D) \times P(F|V,D) \times P(V,D)$

$= P(H|F,D) \times P(F|V) \times P(V|D) \times P(D)$

$= P(H|F,D) \times P(F|V) \times P(V) \times P(D)$

P(Virus | Drink Beer) = P(Virus)

iff Virus is independent of Drink Beer

P(Flu | Virus,DrinkBeer) = P(Flu|Virus)

iff Flu is independent of Drink Beer, given Virus

P(Headache | Flu,Virus,DrinkBeer) =

P(Headache|Flu,DrinkBeer)

iff Headache is independent of Virus, given Flu and Drink Beer

Assume the above independence, we obtain:

P(Headache;Flu;Virus;DrinkBeer)

=P(Headache | Flu;Virus;DrinkBeer) P(Flu | Virus;DrinkBeer)

P(Virus | Drink Beer) P(DrinkBeer)

=P(Headache|Flu;DrinkBeer) P(Flu|Virus) P(Virus) P(DrinkBeer)

Outline

- Probability Distributions
- Joint and Conditional Probability Distributions
- Bayes' Rule
- Mean and Variance
- Properties of Gaussian Distribution
- Maximum Likelihood Estimation

Bayes' Rule

- P(X|Y)= Fraction of the worlds in which X is true given that Y is also true.

$P(X,Y) = P(X|Y)P(Y) \Rightarrow P(X|Y) = \frac{P(X,Y)}{P(Y)} = \frac{P(Y|X)P(X)}{P(Y)}$

- For example:
  - H="Having a headache"
  - F="Coming down with flu"
  - $P(Headche|Flu)$ = fraction of flu-inflicted worlds in which you have a headache. How to calculate?

$P(Y) = \sum_{X} P(Y, X=x) = \sum P(Y|X)P(X)$

- Definition:

$$P(X|Y) = \frac{P(X,Y)}{P(Y)} = \frac{\mathrm{P(Y|X)P(X)}}{\mathrm{P(Y)}}$$

Corollary:

$$P(X,Y) = P(Y|X)P(X)$$

This is called Bayes Rule

Bayes' Rule

- $P(Headache|Flu) = \frac{P(Headache,Flu)}{P(Flu)}$

$$= \frac{P(Flu|Headache)P(Headache)}{P(Flu)}$$

Other cases:

- $P(Y|X) = \frac{P(X|Y)P(Y)}{P(X|Y)P(Y)+P(X|\neg Y)P(\neg Y)}$
- $P(Y = y_i|X) = \frac{P(X|Y)P(Y)}{\sum_{i \in S} P(X|Y = y_i)P(Y=y_i)}$
- $P(Y|X,Z) = \frac{\boxed{P(X|Y,Z)P(Y,Z)}}{P(X,Z)} =$ $P(X,Y,Z) = P(X|Y,Z)P(Y,Z)$

$$\frac{P(X|Y,Z)P(Y,Z)}{P(X|Y,Z)P(Y,Z)+P(X|\neg Y,Z)P(\neg Y,Z)}$$

Outline

- Probability Distributions
- Joint and Conditional Probability Distributions
- Bayes' Rule
- Mean and Variance
- Properties of Gaussian Distribution
- Maximum Likelihood Estimation

Mean and Variance

$E[g(x)] = \sum g(x)\, p(x)$

- Expectation: The mean value, center of mass, first moment:

$$E_X[g(X)] = \int_{-\infty}^{\infty} g(x) p_X(x) dx = \mu$$

- N-th moment: $g(x) = x^n$
- N-th central moment: $g(x) = (x - \mu)^n$
- Mean: $E_X[X] = \int_{-\infty}^{\infty} x p_X(x) dx$
  - $E[\alpha X] = \alpha E[X]$
  - $E[\alpha + X] = \alpha + E[X]$
- Variance(Second central moment): $Var(x) =$

$Var(x) = \boxed{E_X[(X - E_X[X])^2]} = E_X[X^2] - E_X[X]^2$ $\quad E[x^2] = Var(x) + E[x]^2$

  - $Var(\alpha X) = \alpha^2 Var(X)$
  - $Var(\alpha + X) = Var(X)$

$g(x) = x \qquad x = [1, 2, 3]$

$P(x=1) = \frac{1}{5} \qquad P(x=2) = \frac{2}{5} \qquad P(x=3) = \frac{2}{5}$

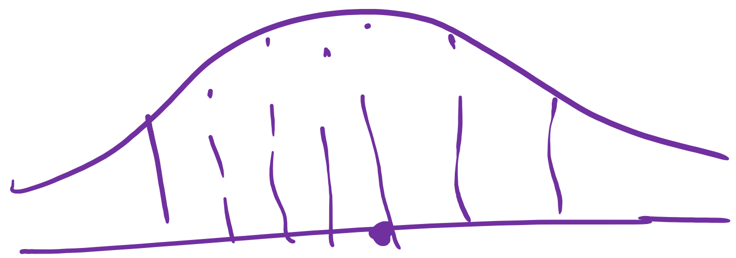
$$E[g(x)] = \sum_{i=1}^{N} g(x)\,p(x)$$

$$E[g(x)] = 1 \times \frac{1}{5} + 2 \times \frac{2}{5} + 3 \times \frac{2}{5} = \frac{11}{5}$$

$$\mu = \text{mean} = \frac{1+2+3}{3} = 2$$

$x = [1, 2, 2, 3, 3]$

$$\mu = \frac{1+2+2+3+3}{5} = \frac{11}{5} = E[g(x)]$$



$$X = \begin{array}{c} \text{height} = h \\ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_{n \times d} \end{array} \qquad \mu_h = \frac{1+2+3}{3} = 2 \qquad \bar{X} = \begin{array}{c} \bar{h} \\ \begin{bmatrix} 1-\mu_h \\ 2-\mu_h \\ 3-\mu_h \end{bmatrix} \end{array} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

$d=1 \quad n=3$

$$\sigma_h^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_h)^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - E[h])^2$$

$$= E\left[(x_i - E[h])^2\right]$$

$$\sigma_h^2 = \frac{\bar{X}^T \bar{X}}{N} = \frac{1}{N} \begin{bmatrix} 1-\mu_h & 2-\mu_h & 3-\mu_h \end{bmatrix} \begin{bmatrix} 1-\mu_h \\ 2-\mu_h \\ 3-\mu_h \end{bmatrix}$$

$$= \frac{1}{N} \left[(1-\mu_h)^2 + (2-\mu_h)^2 + (3-\mu_h)^2 \right] = \frac{\sum_{i=1}^{N} (x-\mu_h)^2}{N}$$

$$X = \begin{array}{cc} h & \text{weight} = w \\ \left[\begin{array}{c} 1 \\ 2 \\ 3 \end{array}\right. & \left.\begin{array}{c} 4 \\ 5 \\ 6 \end{array}\right] \end{array}_{\substack{3\times2 \\ n\times d}}$$

$$\mu_h = 2 \qquad \mu_w = 5$$

$$\bar{X} = \begin{array}{cc} \bar{h} & \bar{w} \\ \left[\begin{array}{c} 1-\mu_h \\ 2-\mu_h \\ 3-\mu_h \end{array}\right. & \left.\begin{array}{c} 4-\mu_w \\ 5-\mu_w \\ 6-\mu_w \end{array}\right] \end{array}$$

$$\mathrm{Cov} = \frac{\bar{X}^T \bar{X}}{N} = \frac{1}{N} \begin{bmatrix} 1-\mu_h & 2-\mu_h & 3-\mu_h \\ 4-\mu_w & 5-\mu_w & 6-\mu_w \end{bmatrix} \begin{bmatrix} 1-\mu_h & 4-\mu_w \\ 2-\mu_h & 5-\mu_w \\ 3-\mu_h & 6-\mu_w \end{bmatrix}$$

$$\mathrm{Cov} = \begin{bmatrix} \sigma_h^2 = \sigma_{hh} & \sigma_{hw} \\ \sigma_{wh} & \sigma_w^2 = \sigma_{ww} \end{bmatrix}_{\substack{2\times2 \\ \underline{d\times d}}}$$

$$X - \mu_h \perp X - \mu_w$$

$$\begin{bmatrix} 1-\mu_h \\ 2-\mu_h \\ 3-\mu_h \end{bmatrix} \qquad \begin{bmatrix} 4-\mu_w \\ 5-\mu_w \\ 6-\mu_w \end{bmatrix}$$

$$\bar{X} = \begin{array}{c} \begin{array}{cc} \bar{h} & \bar{w} \end{array} \\ \begin{bmatrix} 1-\mu_h & 4-\mu_w \\ 2-\mu_h & 5-\mu_w \\ 3-\mu_h & 6-\mu_w \end{bmatrix} \end{array}$$

Standardization

$$\bar{X}^* = \begin{array}{c} \begin{array}{cc} \bar{h}^* & \bar{w}^* \end{array} \\ \begin{bmatrix} \frac{1-\mu_h}{\sigma_h} & \frac{4-\mu_w}{\sigma_w} \\ \frac{2-\mu_h}{\sigma_h} & \frac{5-\mu_w}{\sigma_w} \\ \frac{3-\mu_h}{\sigma_h} & \frac{6-\mu_w}{\sigma_w} \end{bmatrix} \end{array}$$

$$\text{Correlation} = \frac{\bar{X}^{*T} \bar{X}^*}{n} = \frac{1}{N} \begin{bmatrix} \frac{1-\mu_h}{\sigma_h} & \frac{2-\mu_h}{\sigma_h} & \frac{3-\mu_h}{\sigma_h} \\ . & . & . \end{bmatrix} \begin{bmatrix} \frac{1-\mu_h}{\sigma_h} & . \\ \frac{2-\mu_h}{\sigma_h} & . \\ \frac{3-\mu_h}{\sigma_h} & . \end{bmatrix}$$

Cor

$$\text{Cor} = \begin{bmatrix} \frac{\sigma_h^2}{\sigma_h^2} = 1 & -1 \leq \sigma_{hw} \leq 1 \\ -1 \leq \sigma_{wh} \leq 1 & \frac{\sigma_w^2}{\sigma_w^2} = 1 \end{bmatrix}$$

For Joint Distributions

- Expectation and Covariance:
  - $E[X+Y] = E[X] + E[Y]$
  - $cov(X,Y) = E[(X - E_X[X])(Y - E_Y(Y)] = E[XY] - E[X]E[Y]$
  - $Var(X+Y) = Var(X) + 2cov(X,Y) + Var(Y)$

$X = Z$ $\quad Y = Z^2$ $\quad E[X^2] = Var(X) + E[X]^2$

$\mu = 0 \quad \sigma = 1$ $\quad \mu = ? \quad \sigma = ?$ $\quad \mu_Y = E[Z^2] = Var(Z) + E[Z]^2 = 1 + 0 = 1$

$Cov(X,Y) = E[XY] - \underbrace{E[X]}\,E[Y] = \underbrace{E[Z^3]}_{0} - 0 \times 1 = 0$

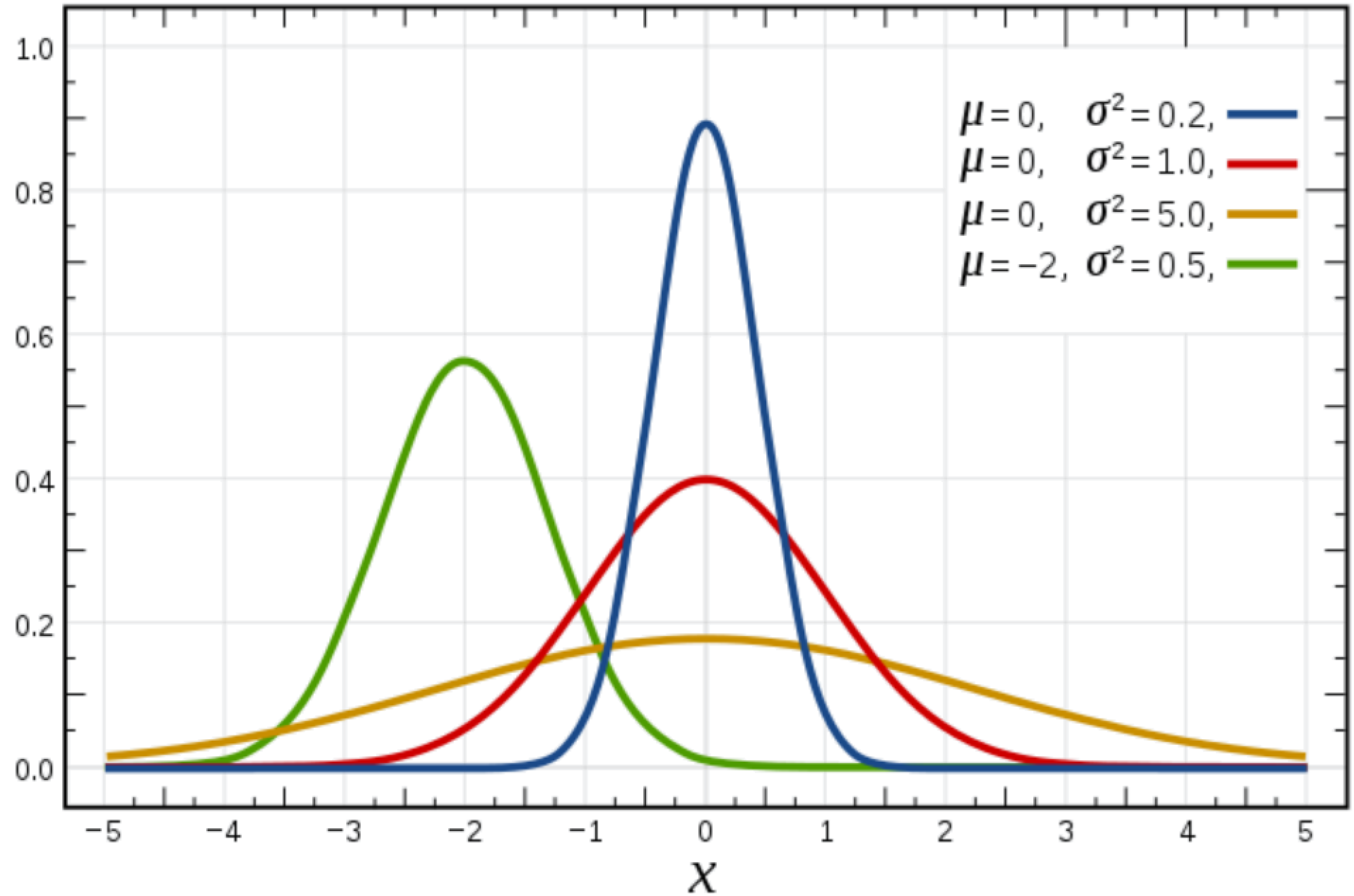
Outline

- Probability Distributions
- Joint and Conditional Probability Distributions
- Bayes' Rule
- Mean and Variance
- Properties of Gaussian Distribution
- Maximum Likelihood Estimation

Gaussian Distribution

- Gaussian Distribution: $$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Probability density function



Probability versus likelihood

Probability is about fact $P(x=H) = 0.5$

H, H, H, T $P(x=H) = \frac{3}{4}$

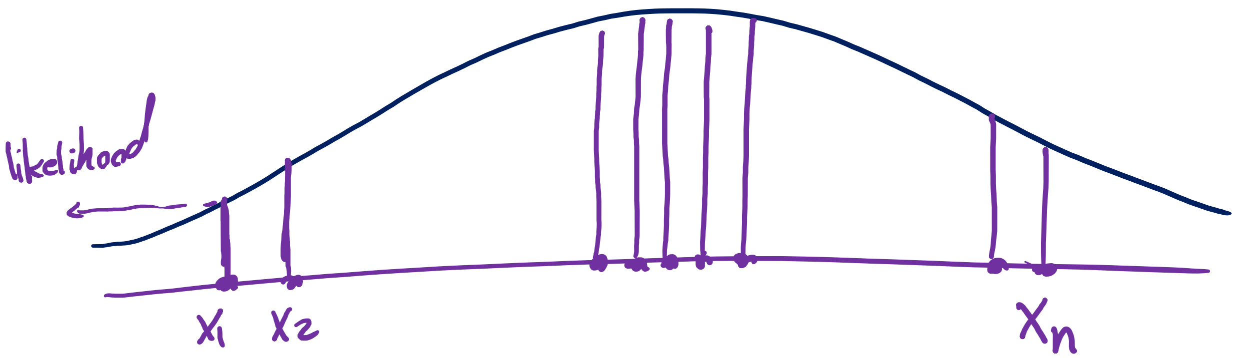
$$P(x \mid \mu, \sigma) = f(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \qquad \sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$

$$P(x \mid a, b) = f(x \mid a, b) = \frac{1}{\sqrt{2\pi b^2}} e^{-\frac{(x-a)^2}{2b^2}}$$

$L(a,b \mid x)$?

(1)



$$f(x_1, x_2, \dots, x_n \mid \overbrace{a,b}^{\theta})$$

iid $f(x_1 \mid a,b)\, f(x_2 \mid a,b) \cdots f(x_n \mid a,b)$

$$\log\left(f(x_1 \mid a,b) \cdots f(x_n \mid a,b)\right)$$

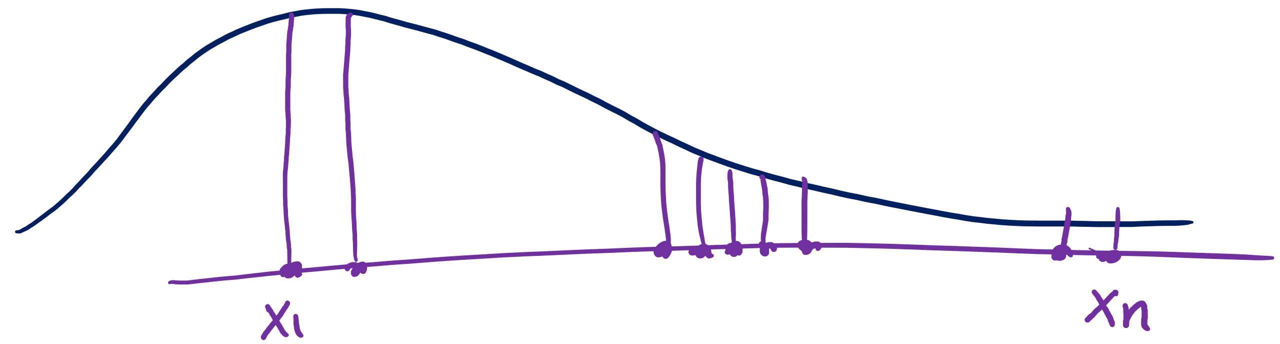
$$\log f(x_1 \mid a,b) + \dots + \log f(x_n \mid a,b)$$

$$f(x_1 \mid \theta)$$

$\hookrightarrow \theta \in \{a,b\}$

$L(a,b \mid x)$?

(2)



$$f(x_1, x_2, \dots, x_n \mid a,b)$$

$$f(x_1 \mid a,b)\, f(x_2 \mid a,b) \cdots f(x_n \mid a,b)$$

Multivariate Gaussian Distribution

$h \quad w$

$X_{n\times d} = [\;]$

$x^{\{\}}_{1\times d}$

1×2

$X_{d\times n} = \begin{array}{c} h \\ w \end{array}\begin{bmatrix} - & - & - & - \\ - & - & - & - \end{bmatrix}$

$x_{d\times 1}$

$\mu = [\mu_h \quad \mu_w]_{1\times d}$

$$\underbrace{p(x|\mu,\Sigma)}_{\mathbb{R}} = \frac{1}{(2\pi)^{n/2}\underbrace{|\Sigma|}_{1\times 1}^{1/2}}\exp\{\underbrace{-\frac{1}{2}\underbrace{(\underset{d\times 1}{x}-\underset{d\times 1}{\mu})^{\mathsf{T}}}_{1\times d}\underset{d\times d}{\Sigma^{-1}}\underbrace{(x-\mu)}_{d\times 1}}_{1\times 1}\}$$

covariance $\Sigma_{d\times d}$

- Moment Parameterization $\mu = E(X)$

$$\Sigma = Cov(X) = E[(X-\mu)(X-\mu)^{\mathsf{T}}]$$

- Mahalanobis Distance $\Delta^2 = (x - \mu)^{\mathsf{T}}\Sigma^{-1}(x-\mu)$

- Tons of applications (MoG, FA, PPCA, Kalman filter,...)

Properties of Gaussian Distribution

- The linear transform of a Gaussian r.v. is a Gaussian. Remember that no matter how x is distributed

$$E(AX+b) = AE(X)+b$$
$$Cov(AX+b) = ACov(X)A^{\top}$$

this means that for Gaussian distributed quantities:

$$X \sim N(\mu,\Sigma) \;\; \rightarrow \;\; AX+b \sim N(A\mu+b, A\Sigma A^{\top})$$

- The sum of two independent Gaussian r.v. is a Gaussian

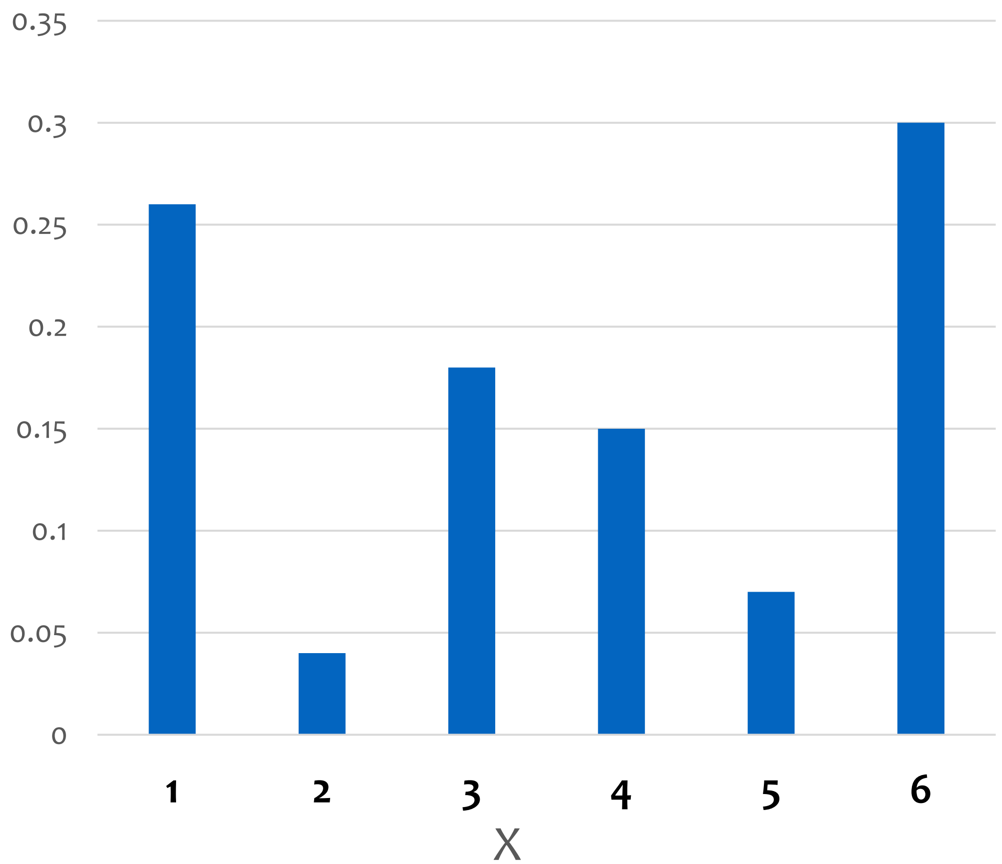
$$\mathrm{Y} = \mathrm{X}_1 + \mathrm{X}_2,\;\; \mathrm{X}_1 \perp X_2 \;\; \rightarrow \mu_y = \mu_1 + \mu_2, \Sigma_y = \Sigma_1 + \Sigma_2$$

- The multiplication of two Gaussian functions is another Gaussian function (although no longer normalized)

$$N(a,A)N(b,B) \propto N(c,C),$$
$$where\; C = (A^{-1}+B^{-1})^{-1}, c = \mathrm{C}\mathrm{A}^{-1}a + CB^{-1}b$$

Central Limit Theorem

Probability mass function of a **biased** dice



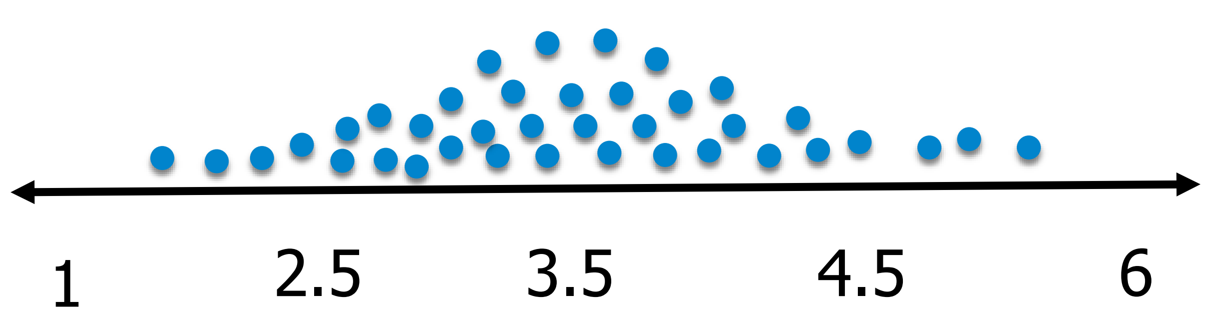
Let's say, I am going to get a sample from this pmf having a size of $n = 4$

$$S_1 = \{1,1,1,6\} \Rightarrow E(S_1) = 2.25$$

$$S_2 = \{1,1,3,6\} \Rightarrow E(S_2) = 2.75$$

$$\vdots$$

$$S_m = \{1,4,6,6\} \Rightarrow E(S_m) = 4.25$$



According to CLT, it will follow a bell curve distribution (normal distribution)

Outline

- Probability Distributions
- Joint and Conditional Probability Distributions
- Bayes’ Rule
- Mean and Variance
- Properties of Gaussian Distribution
- Maximum Likelihood Estimation

Maximum Likelihood Estimation

- Probability: inferring probabilistic quantities for data given fixed models (e.g. prob. of events, marginals, conditionals, etc).
- Statistics: inferring a model given fixed data observations (e.g. clustering, classification, regression).

Main assumption:

Independent and identically distributed random variables

i.i.d

Maximum Likelihood Estimation

For Bernoulli (i.e. flip a coin):

Objective function: $P(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}$ $\quad x_i \in \{0,1\}\ or\ \{head, tail\}$

$$L(\theta|X) = L(\theta|X = x_1, X = x_2, X = x_3, \dots, X = x_n)$$

$P(x_1, x_2, \dots, x_n | \theta)$

$P(x_1|\theta) \dots P(x_n|\theta)$

i.i.d assumption

$$L(\theta|X) = \prod_{i=1}^{n} P(x_i|\theta)$$

$$L(\theta|X) = \prod_{i=1}^{n} P(x_i|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$

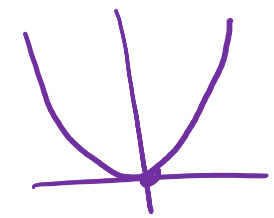
$$L(\theta|X) = \theta^{x_1}(1-\theta)^{1-x_1} \times \theta^{x_2}(1-\theta)^{1-x_2} \dots \times \theta^{x_n}(1-\theta)^{1-x_n} =$$
$$= \theta^{\sum x_i}(1-\theta)^{\sum(1-x_i)}$$

We don't like multiplication, let's convert it into summation

What's the trick? Take the log

$\log_a b$

$y = x^2$



$$L(\theta|X) = \theta^{\sum x_i}(1-\theta)^{\sum(1-x_i)}$$

$$logL(\theta|X) = l(\theta|X) = \log(\theta)\sum_{i=1}^{n} x_i + \log(1-\theta)\sum_{i=1}^{n}(1-x_i)$$

How to optimize θ?

$$\frac{\partial l(\theta|X)}{\partial\theta} = 0 \qquad \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{\sum_{i=1}^{n}(1-x_i)}{1-\theta} = 0$$

$x_i \in \{1,0\}$

$$\theta = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\overbrace{1+1+1+1\ldots}^{8\times 1} + \overbrace{0+0}^{2\times 0}}{10} = 0.8$$