$X \cdot Y$ ⇝ Dot product ⇝ It is a linear operation

$X \cdot Y > 0 \Rightarrow$ They are Positively correlated

$X \cdot Y < 0 \Rightarrow$ They are negatively correlated

$X \cdot Y = 0 \Rightarrow$ They are uncorrelated ⇝ They are orthogonal

If two vectors are linearly independent, can I say, they are uncorrelated ?

$$A = \begin{array}{c} \begin{array}{ccc} A_1 & A_2 & \phantom{13} \end{array} \\ \begin{bmatrix} 1 & 3 & 13 \\ 3 & 7 & 17 \\ 2 & 11 & 19 \end{bmatrix} \end{array}$$

Rank = 3 full rank

$A_1$ & $A_2$ are linearly Independence

$A_1 \cdot A_2 \neq 0$

$AX = X\Lambda$ ⇝ Eigen decomposition

$Ax = \delta x$ ⇝ eigen value & eigen vector

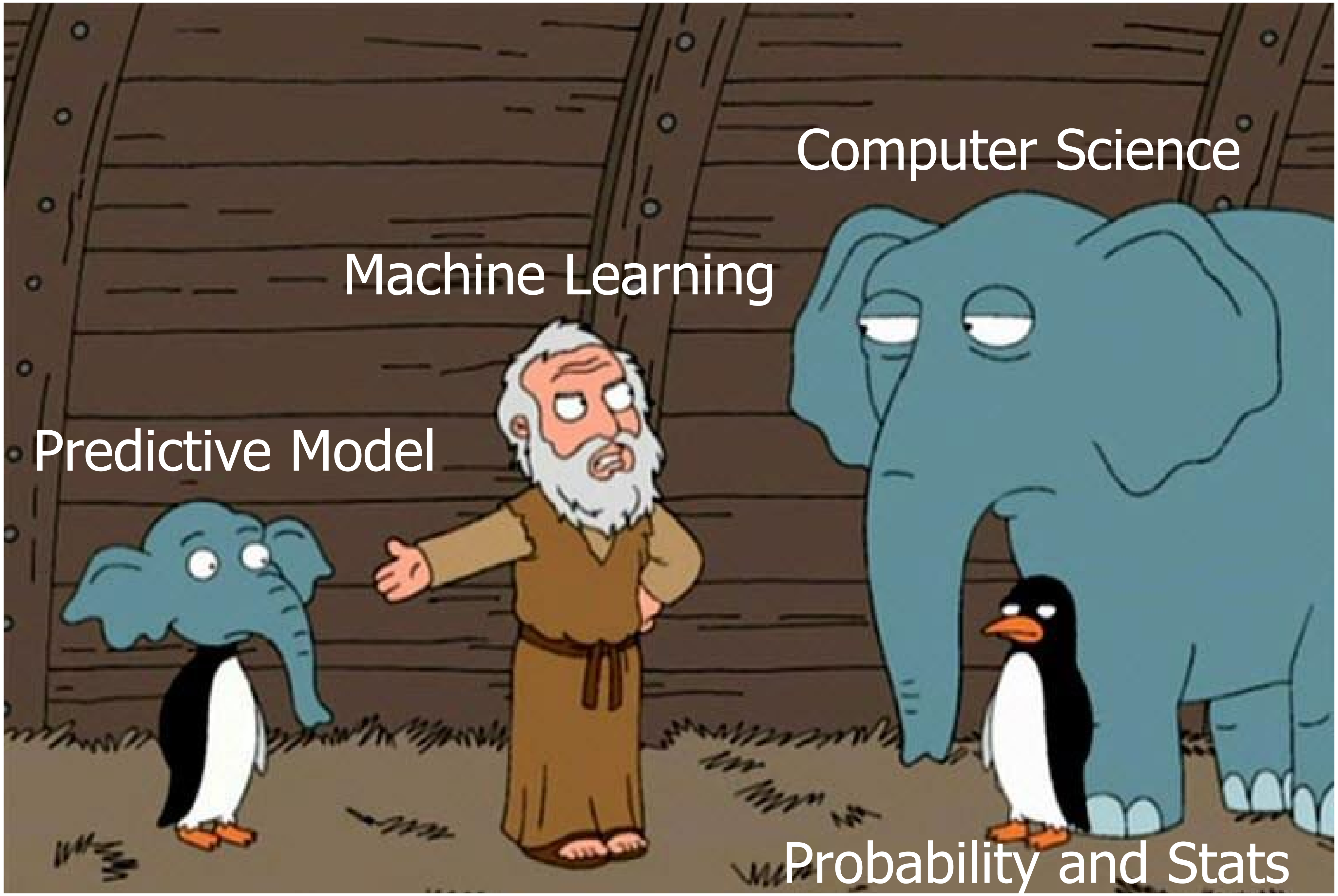
$$|A| = \prod_{i=1}^{d} \delta_i$$

what if one of the eigen values is equal to zero?

$|A| = 0$ ⇝ Singular matrix → Not invertible

→ non-full rank → some of columns or row are linearly dependent

$$A = \begin{array}{ccc} a_1 & a_2 & a_3 \\ \end{array}$$
$$A = \begin{bmatrix} \quad & \quad & \quad \\ & & \\ & & \end{bmatrix}$$


Computer Science
Machine Learning
Predictive Model
Probability and Stats

# Probability and Statistics

Mahdi Roozbahani
Georgia Tech

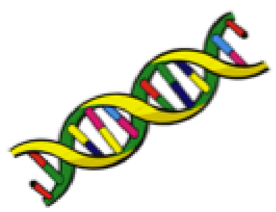
These slides are inspired based on slides from Le Song , Sam Roweis, and Chao Zhang.

# 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**

RV

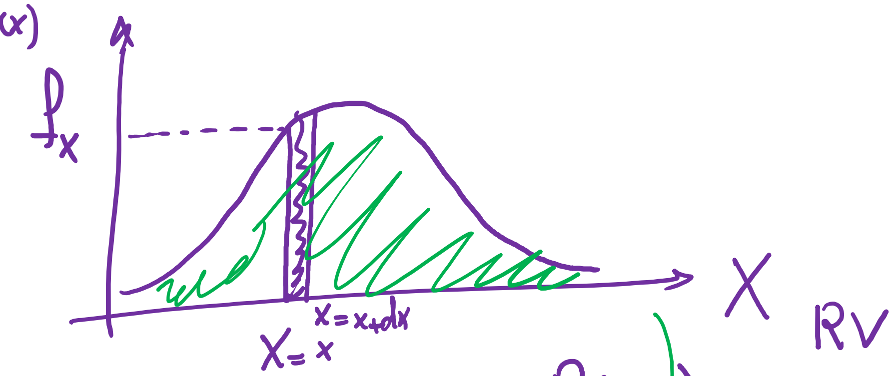
Random variables $X$ represents **outcomes** in sample space

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

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

$$p(x) \geq 0$$

density = Likelihood = $P(x) = f(x)$




$P(X=x) =$

## Continuous variable

Continuous probability distribution

pdf ← Probability density function

Density or likelihood value

Temperature (real number)

Gaussian Distribution

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

$$\int_x f(x)\, dx = 1$$

## Discrete variable

Discrete probability distribution

pmf Probability mass function

Probability value

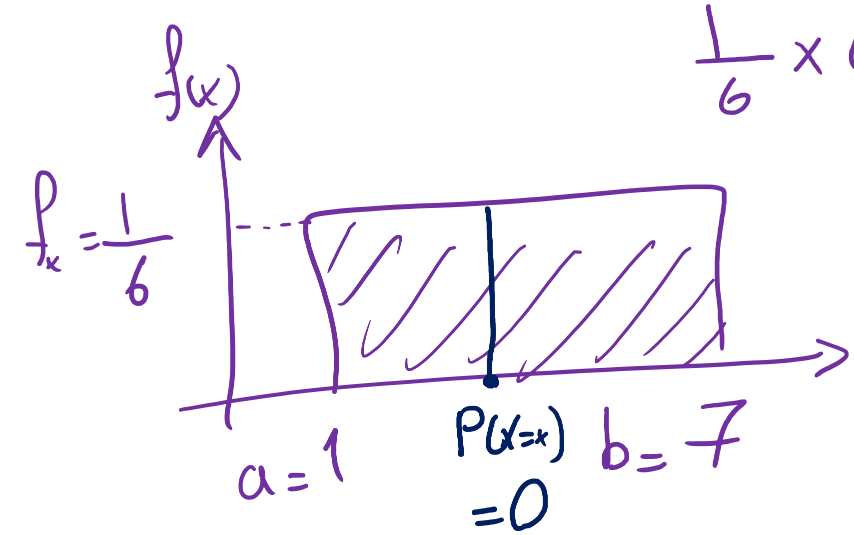
Coin flip (integer)

Bernoulli distribution

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

# Continuous Probability Functions

$\frac{1}{6} \times 6 = 1$




- 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:

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

$\mu$ as a Parameter

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

  - Gaussian(Normal) Density Function

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

$\mu$

$\sigma$ we have two Parameters

# 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

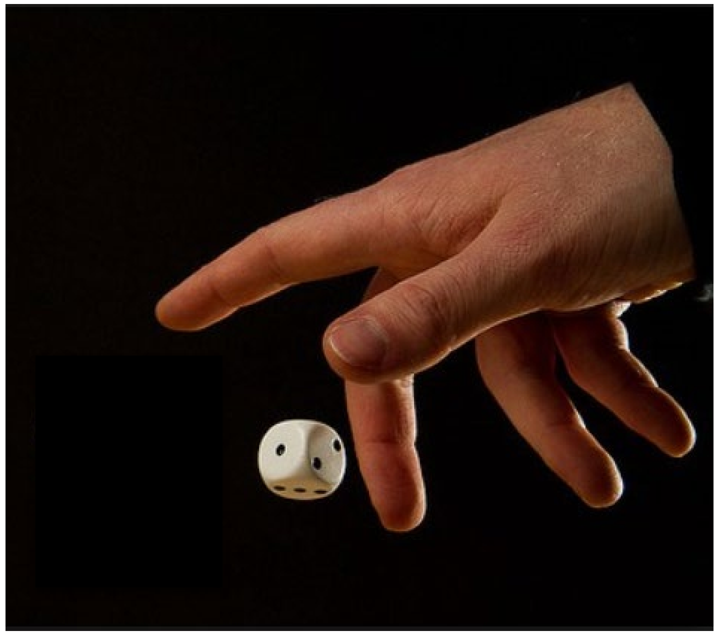
**n-k** is number of failures

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

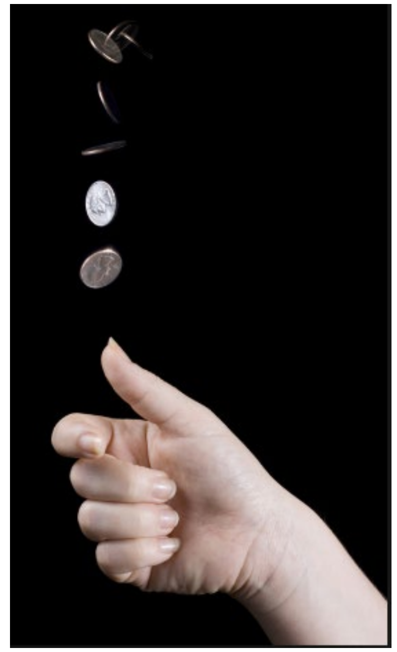
# Outline

- 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**

| **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 |

**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 |

$$P(y=h, X=2) = \frac{2}{35} = \frac{n_{ij}}{N}$$

$$P(y=t) = \frac{20}{35} = \frac{C_j}{N} \qquad P(x=5) = \frac{5}{35} = \frac{C_i}{N}$$

$$P(y=t \mid x=1) = \frac{3}{5} = \frac{n_{ij}}{C_i} \qquad P(x=1 \mid y=t) = \frac{3}{20} = \frac{n_{ij}}{C_j}$$

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

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

**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**

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

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

# Conditional Independence

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

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

$$= P(H|F,D)\ P(F|V)\ \underbrace{P(V|D)\ P(D)}_{P(V)}$$

- Examples:

  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

two important rules:

① Sum rule $P(x) = \sum_{y} P(X, Y=y)$

② Product rule $P(x,y) = P(x|\underset{=}{y})\, P(y) = P(y|x)\, P(x)$

# Bayes' Rule

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

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

$P(y|\underline{\underline{x}}) = \frac{P(x,y)}{P(x)} = \frac{P(x|y)\ P(y)}{P(x)}$

- 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?

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

- 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:

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

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

$$P(y|\underline{\underline{x,z}}) = \frac{P(y,x,z)}{P(x,z)}$$

$$= \frac{P(x|y,z)\,P(y,z)}{\underline{P(x,z)}}$$

$$= \sum_{y} P(x,z,Y=y)$$

- **Course** ML-7641-Spring23
- **Session ID** 222937

# Outline

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

# Mean and Variance

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

$Var(x) = E[x^2] - (E[x])^2$

$E[x^2] = Var(x) + (E[x])^2$

$g(x) = x$

$$\begin{array}{cccc} g(x) & [1, & 2, & 3] \\ p(x) & \frac{1}{6} & \frac{3}{6} & \frac{2}{6} \end{array}$$

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

$$E[g(x)] = 1 \times \frac{1}{6} + 2 \times \frac{3}{6} + 3 \times \frac{2}{6} =$$

$$E[g(x)] = \frac{13}{6}$$

---

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

---

$g(x)$ $[1, 2, 2, 2, 3, 3]$

$$\mu = \frac{1+2+2+2+3+3}{6} = \frac{13}{6}$$

$$\text{height} = h$$

$$\underset{n \times d}{X} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$n = 3$ & $d = 1$

$$\mu_h = \frac{1+2+3}{3} = 2$$

$$\sigma_h^2 = Var_h = \frac{\sum_{i=1}^{N} (x - \mu)^2}{N} = \frac{\sum (x - E_h[x])^2}{N}$$

$$= \frac{1}{N} \sum_{i=1}^{N} (x - E_h[x])^2$$

$$= E_h\left[ (x - E_h[x])^2 \right]$$

$$\sigma_h^2 = Var_h = \frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3}$$

$$X = \begin{array}{c} h \\ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}_{3\times 1} \end{array}$$

$$\sigma_h^2 = \frac{\sum_{i=1}^{3} (x - \mu_h)^2}{n}$$

$$\mu_h = 2$$

$$\bar{X} = \begin{array}{c} \bar{h} \\ \begin{bmatrix} 1 - \mu_h \\ 2 - \mu_h \\ 3 - \mu_h \end{bmatrix} \end{array}$$

$\downarrow$

centered matrix

$$\mu_{\bar{X}_h} = 0$$

$$Var_h = \frac{\bar{X}^T \bar{X}}{n} = \frac{\begin{bmatrix} 1 - \mu_h & 2 - \mu_h & 3 - \mu_h \end{bmatrix} \begin{bmatrix} 1 - \mu_h \\ 2 - \mu_h \\ 3 \ \ \mu_h \end{bmatrix}}{n} = \frac{(1 - \mu_h)^2 + \cdots + (3 - \mu_h)^2}{3}$$

$$X = \begin{array}{c} \begin{array}{cc} h=\text{height} & w=\text{weight} \end{array} \\ \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}_{3\times 2} \end{array}$$

$$\mu_h = 2$$

$$\mu_w = 5$$

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

$$\text{Covariance} = \frac{\bar{X}^T_{2\times 3}\,\bar{X}_{3\times 2}}{n} = \frac{1}{n}\begin{bmatrix} 1-\mu_h & \cdots & 3-\mu_h \\ 4-\mu_w & \cdots & 6-\mu_w \end{bmatrix}\begin{bmatrix} 1-\mu_h & 4-\mu_w \\ \vdots & \vdots \\ 3-\mu_h & 6-\mu_w \end{bmatrix} =$$

$$= \begin{array}{c|c} & \begin{array}{cc} h & w \end{array} \\ \begin{array}{c} h \\ w \end{array} & \begin{bmatrix} \sigma_h^2 = \sigma_{hh} & \sigma_{hw} \\ \sigma_{wh} & \sigma_w^2 = \sigma_{ww} \end{bmatrix}_{d\times d} \end{array}$$

Symmetrical

Standardization

$\sigma \rightarrow$ standard deviation $\quad \sigma^2 \rightsquigarrow$ variance

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

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

$$\text{Correlation} = \frac{\bar{X}^{*T} \bar{X}^*}{n} = \begin{array}{c|cc} & h & w \\ \hline h & 1 & -1 \le \sigma_{hw} \le 1 \\ w & -1 \le \sigma_{wh} \le 1 & 1 \end{array}$$

# For Joint Distributions

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

$X = Z$ $\quad$ $Y = Z^2$

$\mu = 0 \quad \sigma = 1$ $\quad$ $\mu = ? \quad \sigma = ?$

$E[Z^2] = \underbrace{Var(Z)}_{1} + (\underbrace{E[Z]}_{0})^2 = 1$

$Cov(X,Y) = E[\underbrace{Z^3}_{0}] - E[\underbrace{Z}_{0}]\,E[\underbrace{Z^2}_{1}] = 0$

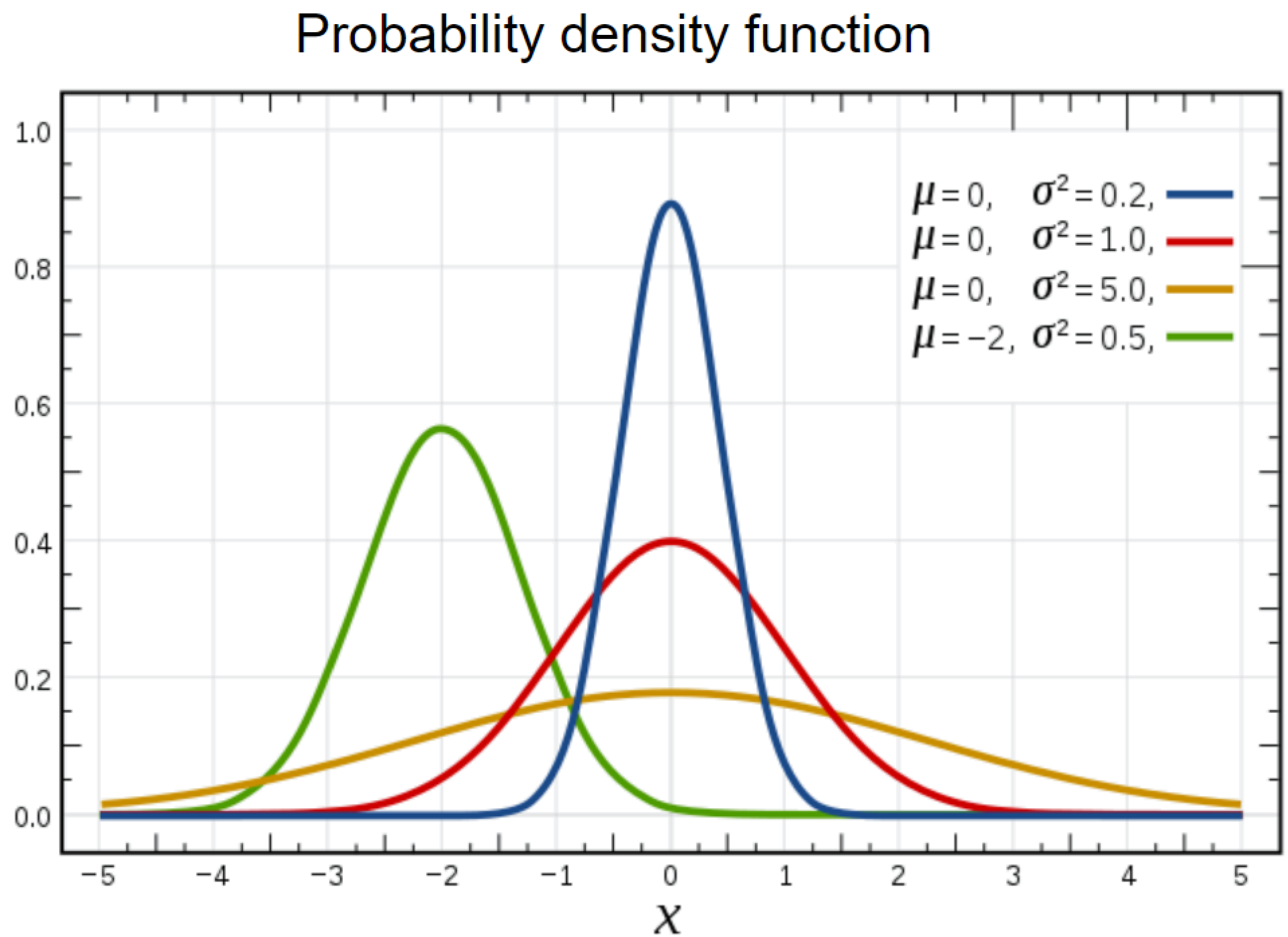
[Uncorrelated vs Independent RV](#)

# 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 versus likelihood]

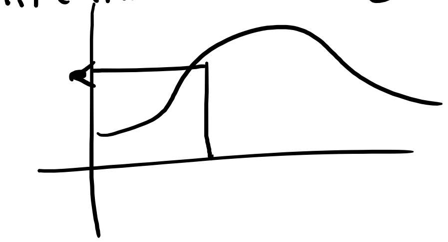
$P(\text{coin}=T)=\frac{1}{2}$    T, T, T, H, H    $P(\text{coin}=T)=\frac{3}{5}$

---

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

$$\boxed{\int f(x|a,b)\,dx = 1}$$

likelihood or density



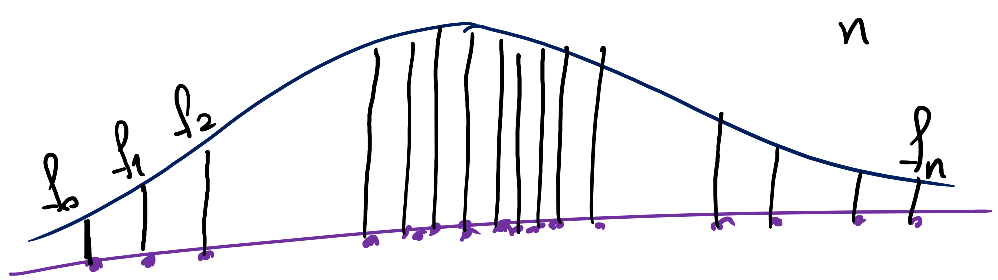
$$L(a,b\,|\,X)=\,?$$

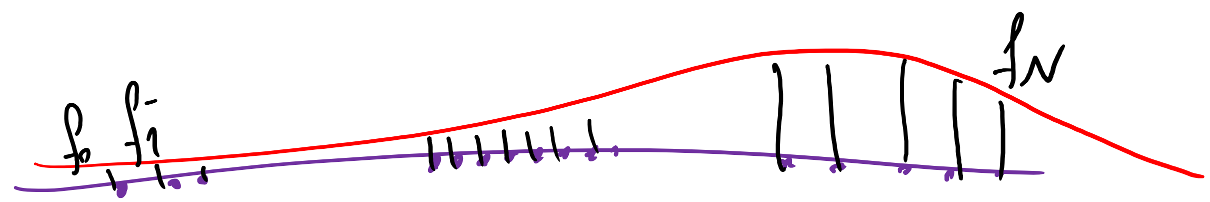
①

②

$$a=\frac{\sum(x-\mu)^2}{n}$$

$$b=\frac{\sum x_i}{n}$$




$$L_0 = f_0 \times f_1 \times \cdots f_n$$

$$L_1 = f_0 \times f_1 \cdots f_N$$

$$L_0 > L_1$$

$$f(x_1, x_2, \ldots x_n) = f(x_1)\, f(x_2) \cdots f(x_n)$$

# Multivariate Gaussian Distribution

$$p(x|\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(x-\mu)^{\top}\Sigma^{-1}(x-\mu)\}$$

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

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

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

- 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

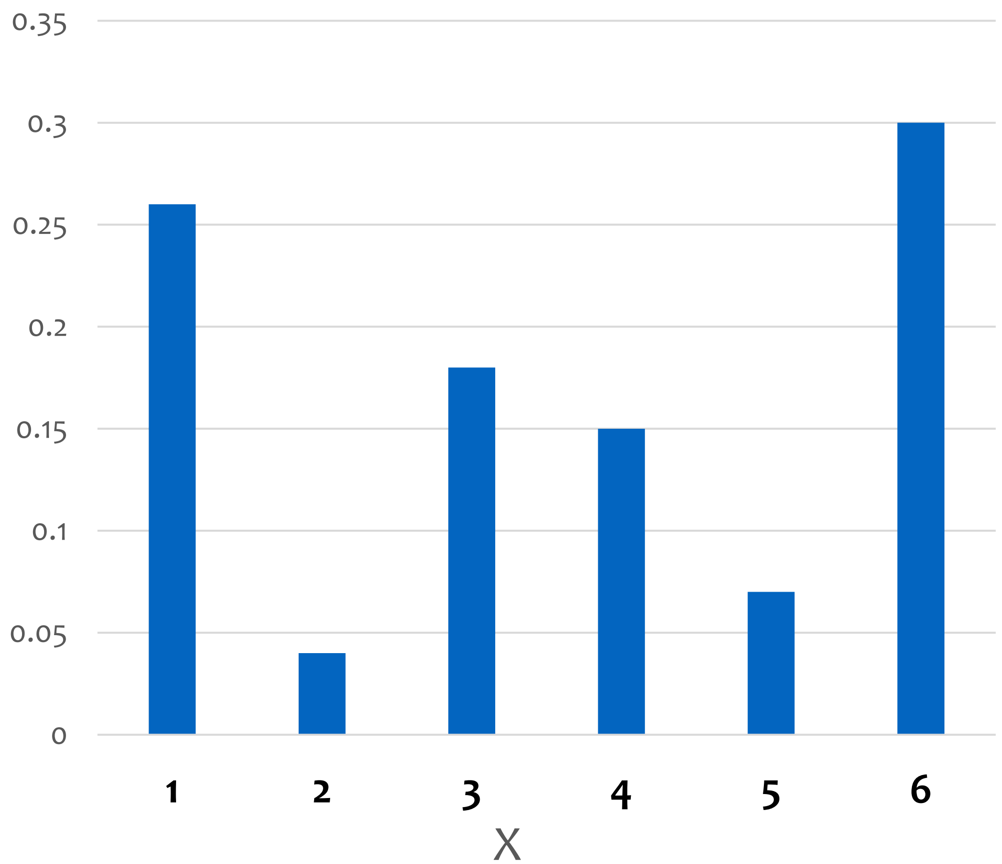
$$\mathrm{Y} = \mathrm{X}_1 + \mathrm{X}_2, \ \mathrm{X}_1 \perp X_2 \ \rightarrow \mu_y = \mu_1 + \mu_2, \Sigma_y = \Sigma_1 + \Sigma_2$$

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

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

# Central Limit Theorem

Probability mass function of a **biased** dice



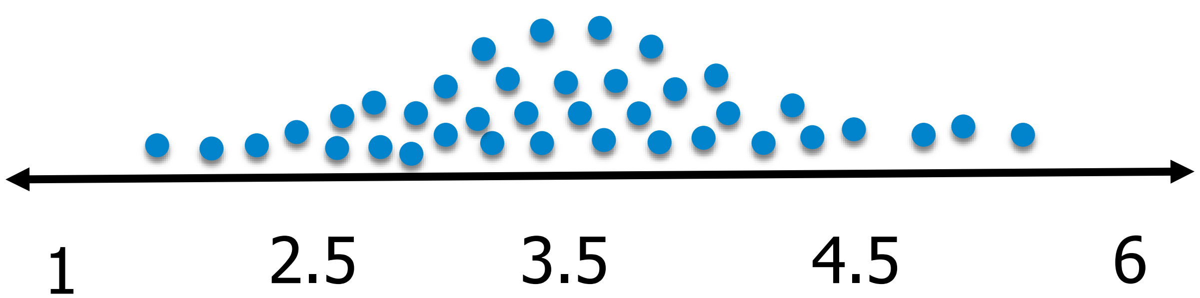

Let's say, I am going to get a sample from this pmf having a size of $n = 4$

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

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

$$\vdots$$

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




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

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

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

$f(x) = x^2$

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

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

How to optimize $\theta$?

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

60 T 40 H

$$\theta = \frac{1}{n}\sum_{i=1}^{n} x_i \rightarrow \boxed{\frac{60}{100}}$$