# Review of Probability Theory and Linear Algebra

Mário A. T. Figueiredo

Instituto Superior Técnico & Instituto de Telecomunicações

Lisboa, **Portugal**

LxMLS: Lisbon Machine Learning School

July 19, 2012

# Outline

- Probability Theory

- Linear Algebra

# What is probability?

- Classical definition: $\mathbb{P}(A) = \dfrac{N_A}{N}$

...with $N$ mutually exclusive equally likely outcomes,
$N_A$ of which result in the occurrence of $A$. *Laplace, 1814*

**Example:** $\mathbb{P}(\text{randomly drawn card is } \clubsuit) = 13/52$.

**Example:** $\mathbb{P}(\text{getting 1 in throw a die}) = 1/6$.

# What is probability?

- Classical definition: $\mathbb{P}(A) = \dfrac{N_A}{N}$

  ...with $N$ mutually exclusive equally likely outcomes,
  $N_A$ of which result in the occurrence of $A$. *Laplace, 1814*

  **Example:** $\mathbb{P}(\text{randomly drawn card is } \clubsuit) = 13/52$.

  **Example:** $\mathbb{P}(\text{getting 1 in throw a die}) = 1/6$.

- Frequentist definition: $\mathbb{P}(A) = \lim_{N\rightarrow\infty} \dfrac{N_A}{N}$

  ...relative frequency of occurrence of $A$ in infinite number of trials.

# What is probability?

- Classical definition: $\mathbb{P}(A) = \dfrac{N_A}{N}$

  ...with $N$ mutually exclusive equally likely outcomes,
  $N_A$ of which result in the occurrence of $A$. *Laplace, 1814*

  **Example:** $\mathbb{P}(\text{randomly drawn card is } \clubsuit) = 13/52$.

  **Example:** $\mathbb{P}(\text{getting 1 in throw a die}) = 1/6$.

- Frequentist definition: $\mathbb{P}(A) = \lim_{N\to\infty} \dfrac{N_A}{N}$

  ...relative frequency of occurrence of $A$ in infinite number of trials.

- Subjective probability: $\mathbb{P}(A)$ is a degree of belief.

  ...gives meaning to $\mathbb{P}(\text{“tomorrow will rain”})$.

# Key concepts: Sample space and events

- Sample space $\mathcal{X}$ = set of possible outcomes of a random experiment.

  Examples:
  - Tossing two coins: $\mathcal{X} = \{HH, TH, HT, TT\}$
  - Roulette: $\mathcal{X} = \{1, 2, ..., 36\}$
  - Draw a card from a shuffled deck: $\mathcal{X} = \{A\clubsuit, 2\clubsuit, ..., Q\diamondsuit, K\diamondsuit\}$.

# Key concepts: Sample space and events

- Sample space $\mathcal{X}$ = set of possible outcomes of a random experiment.

  Examples:
  - Tossing two coins: $\mathcal{X} = \{HH, TH, HT, TT\}$
  - Roulette: $\mathcal{X} = \{1, 2, ..., 36\}$
  - Draw a card from a shuffled deck: $\mathcal{X} = \{A\clubsuit, 2\clubsuit, ..., Q\diamondsuit, K\diamondsuit\}$.

- An event is a subset of $\mathcal{X}$

  Examples:
  - "exactly one H in 2-coin toss": $A = \{TH, HT\} \subset \{HH, TH, HT, TT\}$.
  - "odd number in the roulette": $B = \{1, 3, ..., 35\} \subset \{1, 2, ..., 36\}$.
  - "drawn a $\heartsuit$ card": $C = \{A\heartsuit, 2\heartsuit, ..., K\heartsuit\} \subset \{A\clubsuit, ..., K\diamondsuit\}$

# Kolmogorov's Axioms for Probability

- Probability is a function that maps events $A$ into the interval $[0,1]$.

  Kolmogorov's axioms for probability (1933):

# Kolmogorov's Axioms for Probability

- Probability is a function that maps events $A$ into the interval $[0, 1]$.

  Kolmogorov's axioms for probability (1933):

  - For any $A \subseteq \mathcal{X}$, $\mathbb{P}(A) \geq 0$

# Kolmogorov's Axioms for Probability

- Probability is a function that maps events $A$ into the interval $[0, 1]$.

  Kolmogorov's axioms for probability (1933):

  - For any $A \subseteq \mathcal{X}$, $\mathbb{P}(A) \geq 0$
  - $\mathbb{P}(\mathcal{X}) = 1$

# Kolmogorov's Axioms for Probability

- Probability is a function that maps events $A$ into the interval $[0, 1]$.

  Kolmogorov's axioms for probability (1933):

  - For any $A \subseteq \mathcal{X}$, $\mathbb{P}(A) \geq 0$
  - $\mathbb{P}(\mathcal{X}) = 1$
  - If $A_1, A_2 \ldots \subseteq \mathcal{X}$ are disjoint events, then $\mathbb{P}\Big(\bigcup_i A_i\Big) = \sum_i \mathbb{P}(A_i)$

# Kolmogorov's Axioms for Probability

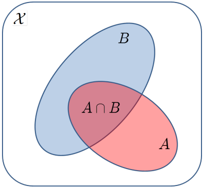
- Probability is a function that maps events $A$ into the interval $[0, 1]$.

  Kolmogorov's axioms for probability (1933):

  - For any $A \subseteq \mathcal{X}$, $\mathbb{P}(A) \geq 0$
  - $\mathbb{P}(\mathcal{X}) = 1$
  - If $A_1, A_2 \ldots \subseteq \mathcal{X}$ are disjoint events, then $\mathbb{P}\Big(\bigcup_i A_i\Big) = \sum_i \mathbb{P}(A_i)$

- From these axioms, many results can be derived. Examples:

  - $\mathbb{P}(\emptyset) = 0$
  - $C \subset D \;\Rightarrow\; \mathbb{P}(C) \leq \mathbb{P}(D)$
  - $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$



# Conditional Probability and Independence

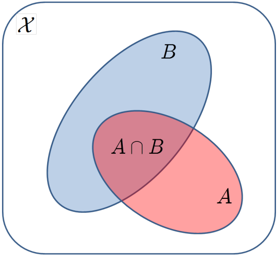
- If $\mathbb{P}(B) > 0$, $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ (conditional prob. of $A$ given $B$)

# Conditional Probability and Independence

- If $\mathbb{P}(B) > 0$, $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ (conditional prob. of $A$ given $B$)

- ...satisfies all Kolmogorov's axioms:

  - For any $A \subseteq \mathcal{X}$, $\mathbb{P}(A|B) \geq 0$
  - $\mathbb{P}(\mathcal{X}|B) = 1$
  - If $A_1, A_2 \ldots \subseteq \mathcal{X}$ are disjoint, then $\mathbb{P}\Big(\bigcup_i A_i \Big| B\Big) = \sum_i \mathbb{P}(A_i|B)$

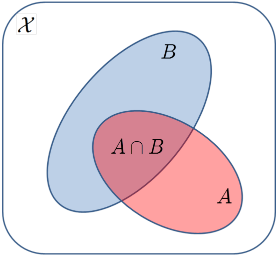


# Conditional Probability and Independence

- If $\mathbb{P}(B) > 0$, $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ (conditional prob. of $A$ given $B$)

- ...satisfies all Kolmogorov's axioms:

  - For any $A \subseteq \mathcal{X}$, $\mathbb{P}(A|B) \geq 0$
  - $\mathbb{P}(\mathcal{X}|B) = 1$
  - If $A_1, A_2 \ldots \subseteq \mathcal{X}$ are disjoint, then $\mathbb{P}\Big(\bigcup_i A_i \Big| B\Big) = \sum_i \mathbb{P}(A_i|B)$



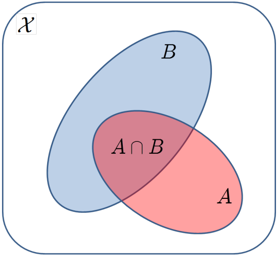

- Events $A$, $B$ are independent ($A \perp\!\!\!\perp B$) $\Leftrightarrow$ $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

# Conditional Probability and Independence

- If $\mathbb{P}(B) > 0$, $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ (conditional prob. of $A$ given $B$)

- ...satisfies all Kolmogorov's axioms:

  - For any $A \subseteq \mathcal{X}$, $\mathbb{P}(A|B) \geq 0$
  - $\mathbb{P}(\mathcal{X}|B) = 1$
  - If $A_1, A_2 \ldots \subseteq \mathcal{X}$ are disjoint, then $\mathbb{P}\Big(\bigcup_i A_i \Big| B\Big) = \sum_i \mathbb{P}(A_i|B)$




- Events $A$, $B$ are independent ($A \perp\!\!\!\perp B$) $\Leftrightarrow$ $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

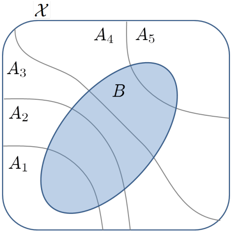
- Relationship with conditional probabilities:

$$A \perp\!\!\!\perp B \;\Leftrightarrow\; \mathbb{P}(A|B) = \mathbb{P}(A)$$

# Bayes Theorem

- Law of total probability: if $A_1, ..., A_n$ are a partition of $\mathcal{X}$

$$\mathbb{P}(B) = \sum_i \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

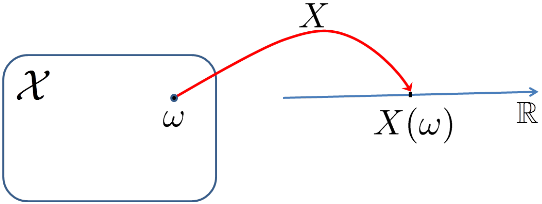



- Bayes' theorem: if $A_1, ..., A_n$ are a partition of $\mathcal{X}$

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B \cap A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_i)\,\mathbb{P}(A_i)}{\sum_i \mathbb{P}(B|A_i)\mathbb{P}(A_i)}$$

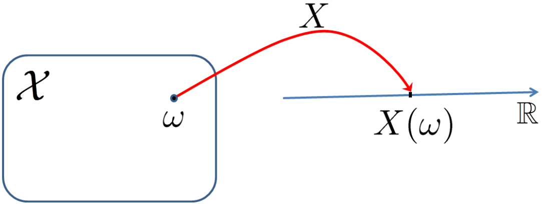
# Random Variables

- A (real) random variable (RV) is a function: $X : \mathcal{X} \to \mathbb{R}$



# Random Variables

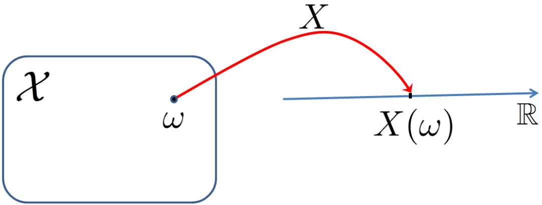
- A (real) random variable (RV) is a function: $X : \mathcal{X} \to \mathbb{R}$




  - Discrete RV: range of $X$ is countable (*e.g.*, $\mathbb{N}$ or $\{0, 1\}$)

# Random Variables

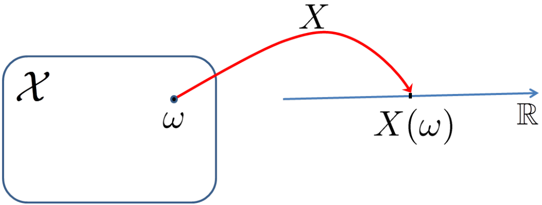
- A (real) random variable (RV) is a function: $X : \mathcal{X} \to \mathbb{R}$




  - Discrete RV: range of $X$ is countable (*e.g.*, $\mathbb{N}$ or $\{0, 1\}$)
  - Continuous RV: range of $X$ is uncountable (*e.g.*, $\mathbb{R}$ or [0, 1])

# Random Variables

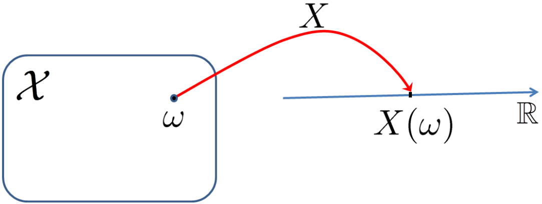
- A (real) random variable (RV) is a function: $X : \mathcal{X} \to \mathbb{R}$




  - Discrete RV: range of $X$ is countable (*e.g.*, $\mathbb{N}$ or $\{0, 1\}$)
  - Continuous RV: range of $X$ is uncountable (*e.g.*, $\mathbb{R}$ or $[0, 1]$)
  - Example: number of head in tossing two coins,
    $\mathcal{X} = \{HH, HT, TH, TT\}$,
    $X(HH) = 2$, $X(HT) = X(TH) = 1$, $X(TT) = 0$.
    Range of $X = \{0, 1, 2\}$.

# Random Variables

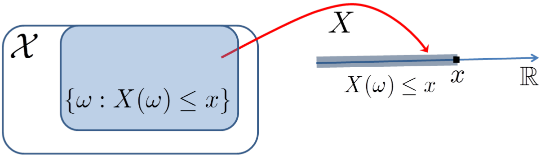
- A (real) random variable (RV) is a function: $X : \mathcal{X} \to \mathbb{R}$




  - Discrete RV: range of $X$ is countable (*e.g.*, $\mathbb{N}$ or $\{0, 1\}$)
  - Continuous RV: range of $X$ is uncountable (*e.g.*, $\mathbb{R}$ or [0, 1])
  - Example: number of head in tossing two coins,
    $\mathcal{X} = \{HH, HT, TH, TT\}$,
    $X(HH) = 2$, $X(HT) = X(TH) = 1$, $X(TT) = 0$.
    Range of $X = \{0, 1, 2\}$.
  - Example: distance traveled by a tossed coin; range of $X = \mathbb{R}_+$.

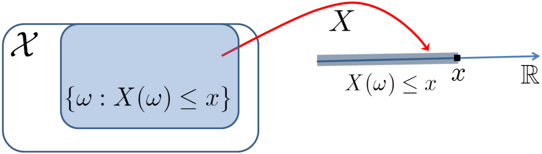
# Random Variables: Distribution Function

- Distribution function: $F_X(x) = \mathbb{P}(\{\omega \in \mathcal{X} : X(\omega) \leq x\})$

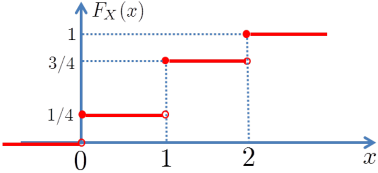


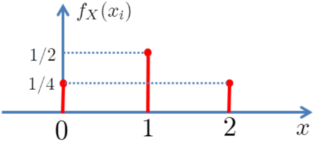
# Random Variables: Distribution Function

- Distribution function: $F_X(x) = \mathbb{P}(\{\omega \in \mathcal{X} : X(\omega) \leq x\})$



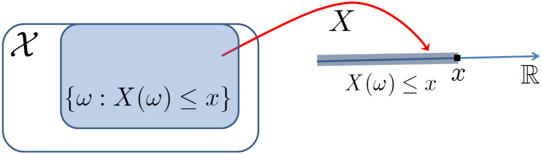

- Example: number of heads in tossing 2 coins; range$(X) = \{0, 1, 2\}$.



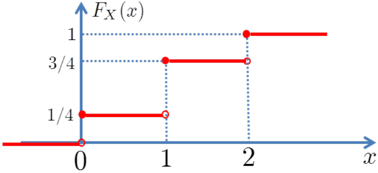



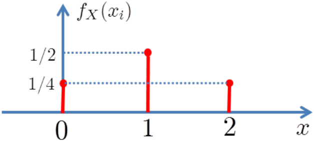
# Random Variables: Distribution Function

- Distribution function: $F_X(x) = \mathbb{P}(\{\omega \in \mathcal{X} : X(\omega) \leq x\})$




- Example: number of heads in tossing 2 coins; range$(X) = \{0, 1, 2\}$.




- Probability mass function (discrete RV): $f_X(x) = \mathbb{P}(X = x)$,

$$F_X(x) = \sum_{x_i \leq x} f_X(x_i).$$

# Important Discrete Random Variables

- Uniform: $X \in \{x_1, ..., x_K\}$, pmf $f_X(x_i) = 1/K$.

# Important Discrete Random Variables

- Uniform: $X \in \{x_1, ..., x_K\}$, pmf $f_X(x_i) = 1/K$.

- Bernoulli RV: $X \in \{0, 1\}$, pmf $f_X(x) = \begin{cases} p & \Leftarrow & x = 1 \\ 1-p & \Leftarrow & x = 0 \end{cases}$

  Can be written compactly as $f_X(x) = p^x(1-p)^{1-x}$.

# Important Discrete Random Variables

- Uniform: $X \in \{x_1, ..., x_K\}$, pmf $f_X(x_i) = 1/K$.
- Bernoulli RV: $X \in \{0, 1\}$, pmf $f_X(x) = \begin{cases} p & \Leftarrow & x = 1 \\ 1-p & \Leftarrow & x = 0 \end{cases}$

  Can be written compactly as $f_X(x) = p^x(1-p)^{1-x}$.
- Binomial RV: $X \in \{0, 1, ..., n\}$ (sum on $n$ Bernoulli RVs)

  $$f_X(x) = \text{Binomial}(x; n, p) = \binom{n}{x} p^x \, (1-p)^{(n-x)}$$

# Important Discrete Random Variables

- Uniform: $X \in \{x_1, ..., x_K\}$, pmf $f_X(x_i) = 1/K$.

- Bernoulli RV: $X \in \{0, 1\}$, pmf $f_X(x) = \begin{cases} p & \Leftarrow \quad x = 1 \\ 1 - p & \Leftarrow \quad x = 0 \end{cases}$

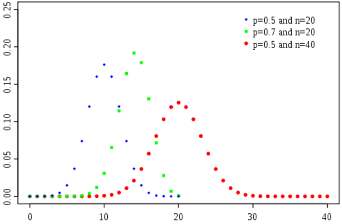
  Can be written compactly as $f_X(x) = p^x (1 - p)^{1-x}$.

- Binomial RV: $X \in \{0, 1, ..., n\}$ (sum on $n$ Bernoulli RVs)

$$f_X(x) = \text{Binomial}(x; n, p) = \binom{n}{x} p^x \, (1 - p)^{(n-x)}$$

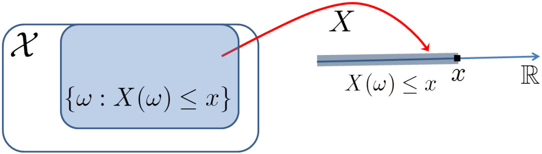
Binomial coefficients:

$$\binom{n}{x} = \frac{n!}{(n - x)! \, x!}$$



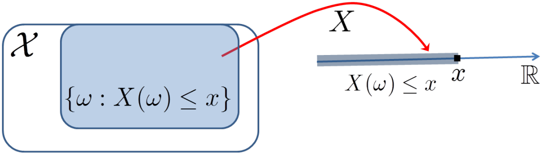
# Random Variables: Distribution Function

- Distribution function: $F_X(x) = \mathbb{P}(\{\omega \in \mathcal{X} : X(\omega) \leq x\})$

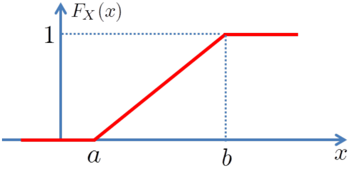


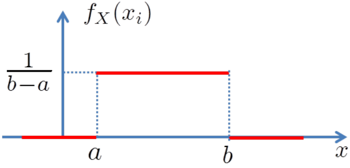
# Random Variables: Distribution Function

- Distribution function: $F_X(x) = \mathbb{P}(\{\omega \in \mathcal{X} : X(\omega) \leq x\})$



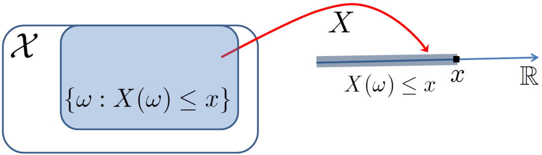

- Example: continuous RV with uniform distribution on $[a, b]$.



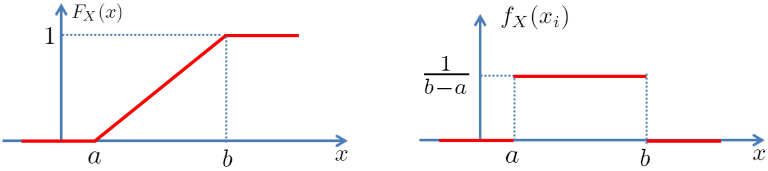



# Random Variables: Distribution Function

- Distribution function: $F_X(x) = \mathbb{P}(\{\omega \in \mathcal{X} : X(\omega) \leq x\})$



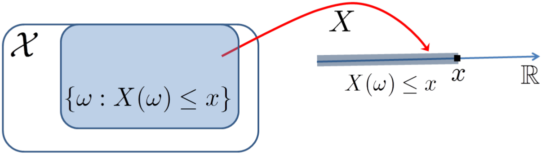

- Example: continuous RV with uniform distribution on $[a, b]$.



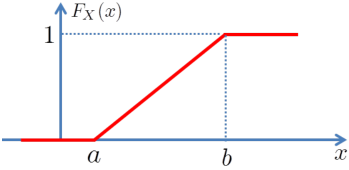

- Probability density function (pdf, continuous RV): $f_X(x)$

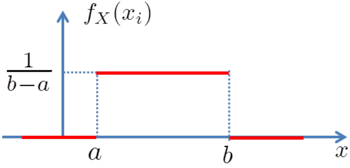
# Random Variables: Distribution Function

- Distribution function: $F_X(x) = \mathbb{P}(\{\omega \in \mathcal{X} : X(\omega) \leq x\})$




- Example: continuous RV with uniform distribution on $[a, b]$.



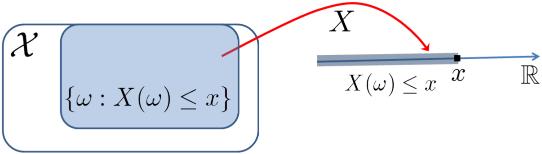




- Probability density function (pdf, continuous RV): $f_X(x)$

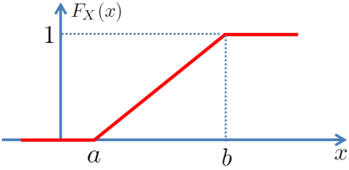
$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du,$$

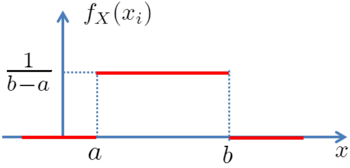
# Random Variables: Distribution Function

- Distribution function: $F_X(x) = \mathbb{P}(\{\omega \in \mathcal{X} : X(\omega) \leq x\})$




- Example: continuous RV with uniform distribution on $[a, b]$.



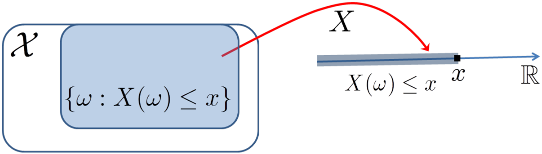




- Probability density function (pdf, continuous RV): $f_X(x)$

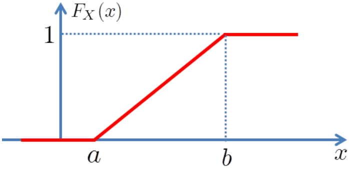
$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du, \quad \mathbb{P}(X \in [c, d]) = \int_{c}^{d} f_X(x)\, dx,$$

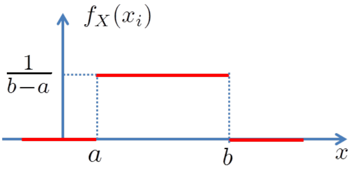
# Random Variables: Distribution Function

- Distribution function: $F_X(x) = \mathbb{P}(\{\omega \in \mathcal{X} : X(\omega) \leq x\})$




- Example: continuous RV with uniform distribution on $[a, b]$.




- Probability density function (pdf, continuous RV): $f_X(x)$

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du, \quad \mathbb{P}(X \in [c, d]) = \int_{c}^{d} f_X(x)\, dx, \quad \mathbb{P}(X = x) = 0$$

# Important Continuous Random Variables

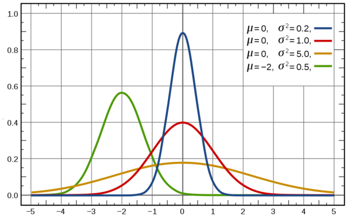
- Uniform: $f_X(x) = \text{Uniform}(x; a, b) = \begin{cases} \frac{1}{b-a} & \Leftarrow \quad x \in [a,\ b] \\ 0 & \Leftarrow \quad x \notin [a,\ b] \end{cases}$

  (see previous slide).

# Important Continuous Random Variables

- Uniform: $f_X(x) = \text{Uniform}(x; a, b) = \begin{cases} \frac{1}{b-a} & \Leftarrow \quad x \in [a,\ b] \\ 0 & \Leftarrow \quad x \notin [a,\ b] \end{cases}$ (see previous slide).

- Gaussian: $f_X(x) = \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}} e^{-\frac{(x-\mu)^2}{2\,\sigma^2}}$

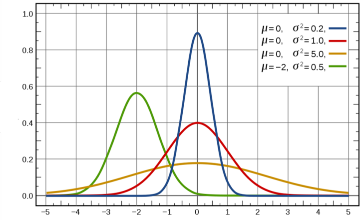


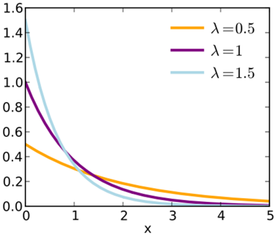
# Important Continuous Random Variables

- Uniform: $f_X(x) = \text{Uniform}(x; a, b) = \begin{cases} \frac{1}{b-a} & \Leftarrow \quad x \in [a, \ b] \\ 0 & \Leftarrow \quad x \notin [a, \ b] \end{cases}$

  (see previous slide).

- Gaussian: $f_X(x) = \mathcal{N}(x; \mu, \sigma^2) = \dfrac{1}{\sqrt{2\,\pi\,\sigma^2}} e^{-\frac{(x-\mu)^2}{2\,\sigma^2}}$




- Exponential: $f_X(x) = \text{Exp}(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & \Leftarrow \quad x \geq 0 \\ 0 & \Leftarrow \quad x < 0 \end{cases}$

# Expectation of Random Variables

- Expectation: $\mathbb{E}(X) = \begin{cases} \displaystyle\sum_i x_i f_X(x_i) & X \text{ discrete on } \{x_1, ...x_K\} \\ \displaystyle\int_{-\infty}^{\infty} x\, f_X(x)\, dx & X \text{ continuous} \end{cases}$

# Expectation of Random Variables

- Expectation: $\mathbb{E}(X) = \begin{cases} \displaystyle\sum_i x_i f_X(x_i) & X \text{ discrete on } \{x_1, ...x_K\} \\ \displaystyle\int_{-\infty}^{\infty} x\, f_X(x)\, dx & X \text{ continuous} \end{cases}$

- Example: Bernoulli, $f_X(x) = p^x\,(1-p)^{1-x}$, for $x \in \{0,\, 1\}$.

$$\mathbb{E}(X) = 0\,(1-p) + 1\,p = p.$$

# Expectation of Random Variables

- Expectation: $\mathbb{E}(X) = \begin{cases} \displaystyle\sum_i x_i f_X(x_i) & X \text{ discrete on } \{x_1, ...x_K\} \\ \displaystyle\int_{-\infty}^{\infty} x\, f_X(x)\, dx & X \text{ continuous} \end{cases}$

- Example: Bernoulli, $f_X(x) = p^x\,(1-p)^{1-x}$, for $x \in \{0,\, 1\}$.

  $\mathbb{E}(X) = 0\,(1-p) + 1\,p = p.$

- Example: Binomial, $f_X(x) = \binom{n}{x} p^x\,(1-p)^{n-x}$, for $x \in \{0, ..., n\}$.

  $\mathbb{E}(X) = n\,p.$

# Expectation of Random Variables

- Expectation: $\mathbb{E}(X) = \begin{cases} \displaystyle\sum_i x_i f_X(x_i) & X \text{ discrete on } \{x_1, ... x_K\} \\ \displaystyle\int_{-\infty}^{\infty} x\, f_X(x)\, dx & X \text{ continuous} \end{cases}$

- Example: Bernoulli, $f_X(x) = p^x\,(1-p)^{1-x}$, for $x \in \{0,\,1\}$.

  $\mathbb{E}(X) = 0\,(1-p) + 1\,p = p.$

- Example: Binomial, $f_X(x) = \binom{n}{x} p^x\,(1-p)^{n-x}$, for $x \in \{0, ..., n\}$.

  $\mathbb{E}(X) = n\,p.$

- Example: Gaussian, $f_X(x) = \mathcal{N}(x; \mu, \sigma^2)$; $\mathbb{E}(X) = \mu$.

# Expectation of Random Variables

- Expectation: $\mathbb{E}(X) = \begin{cases} \displaystyle\sum_i x_i f_X(x_i) & X \text{ discrete on } \{x_1, ... x_K\} \\ \displaystyle\int_{-\infty}^{\infty} x\, f_X(x)\, dx & X \text{ continuous} \end{cases}$

- Example: Bernoulli, $f_X(x) = p^x\,(1-p)^{1-x}$, for $x \in \{0,\, 1\}$.

  $\mathbb{E}(X) = 0\,(1-p) + 1\,p = p.$

- Example: Binomial, $f_X(x) = \binom{n}{x} p^x\,(1-p)^{n-x}$, for $x \in \{0, ..., n\}$.

  $\mathbb{E}(X) = n\,p.$

- Example: Gaussian, $f_X(x) = \mathcal{N}(x; \mu, \sigma^2)$; $\mathbb{E}(X) = \mu$.

- Linearity of expectation: $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

# Expectation of Functions of Random Variables

- $$\mathbb{E}(g(X)) = \begin{cases} \displaystyle\sum_i g(x_i) f_X(x_i) & X \text{ discrete on } \{x_1, ...x_K\} \\ \displaystyle\int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx & X \text{ continuous} \end{cases}$$

# Expectation of Functions of Random Variables

- $$\mathbb{E}(g(X)) = \begin{cases} \displaystyle\sum_i g(x_i) f_X(x_i) & X \text{ discrete on } \{x_1, ...x_K\} \\ \displaystyle\int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx & X \text{ continuous} \end{cases}$$

- Example: variance, $\text{var}(X) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big)$

# Expectation of Functions of Random Variables

- $$\mathbb{E}(g(X)) = \begin{cases} \displaystyle\sum_i g(x_i) f_X(x_i) & X \text{ discrete on } \{x_1, ...x_K\} \\ \displaystyle\int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx & X \text{ continuous} \end{cases}$$

- Example: variance, $\text{var}(X) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

- Example: Bernoulli variance, $\mathbb{E}(X^2) = \mathbb{E}(X) = p$

# Expectation of Functions of Random Variables

- $$\mathbb{E}(g(X)) = \begin{cases} \displaystyle\sum_i g(x_i) f_X(x_i) & X \text{ discrete on } \{x_1, \ldots x_K\} \\ \displaystyle\int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx & X \text{ continuous} \end{cases}$$

- Example: variance, $\text{var}(X) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

- Example: Bernoulli variance, $\mathbb{E}(X^2) = \mathbb{E}(X) = p$ , thus $\text{var}(X) = p(1-p)$.

# Expectation of Functions of Random Variables

- $$\mathbb{E}(g(X)) = \begin{cases} \displaystyle\sum_i g(x_i) f_X(x_i) & X \text{ discrete on } \{x_1, \ldots x_K\} \\ \displaystyle\int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx & X \text{ continuous} \end{cases}$$

- Example: variance, $\mathrm{var}(X) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

- Example: Bernoulli variance, $\mathbb{E}(X^2) = \mathbb{E}(X) = p$ , thus $\mathrm{var}(X) = p(1-p)$.

- Example: Gaussian variance, $\mathbb{E}\big((X-\mu)^2\big) = \sigma^2$.

# Expectation of Functions of Random Variables

- $$\mathbb{E}(g(X)) = \begin{cases} \displaystyle\sum_i g(x_i) f_X(x_i) & X \text{ discrete on } \{x_1, ... x_K\} \\ \displaystyle\int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx & X \text{ continuous} \end{cases}$$

- Example: variance, $\text{var}(X) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

- Example: Bernoulli variance, $\mathbb{E}(X^2) = \mathbb{E}(X) = p$ , thus $\text{var}(X) = p(1-p)$.

- Example: Gaussian variance, $\mathbb{E}\big((X-\mu)^2\big) = \sigma^2$.

- Probability as expectation of indicator, $\mathbf{1}_A(x) = \begin{cases} 1 & \Leftarrow & x \in A \\ 0 & \Leftarrow & x \notin A \end{cases}$

$$\mathbb{P}(X \in A) = \int_A f_X(x)\, dx = \int \mathbf{1}_A(x)\, f_X(x)\, dx = \mathbb{E}(\mathbf{1}_A(X))$$

# Two (or More) Random Variables

- Joint pmf of two discrete RVs: $f_{X,Y}(x,y) = \mathbb{P}(X = x \wedge Y = y)$.

  Extends trivially to more than two RVs.

# Two (or More) Random Variables

- Joint pmf of two discrete RVs: $f_{X,Y}(x,y) = \mathbb{P}(X = x \wedge Y = y)$.

  Extends trivially to more than two RVs.

- Joint pdf of two continuous RVs: $f_{X,Y}(x,y)$, such that

$$\mathbb{P}(X \in A) = \int\int_A f_{X,Y}(x,y)\, dx\, dy, \qquad A \subset \mathbb{R}^2$$

Extends trivially to more than two RVs.

# Two (or More) Random Variables

- Joint pmf of two discrete RVs: $f_{X,Y}(x,y) = \mathbb{P}(X = x \wedge Y = y)$.

  Extends trivially to more than two RVs.

- Joint pdf of two continuous RVs: $f_{X,Y}(x,y)$, such that

$$\mathbb{P}(X \in A) = \iint_A f_{X,Y}(x,y)\, dx\, dy, \qquad A \subset \mathbb{R}^2$$

  Extends trivially to more than two RVs.

- Marginalization: $f_Y(y) = \begin{cases} \sum_x f_{X,Y}(x,y), & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx, & \text{if } X \text{ continuous} \end{cases}$

# Two (or More) Random Variables

- Joint pmf of two discrete RVs: $f_{X,Y}(x,y) = \mathbb{P}(X = x \wedge Y = y)$.

  Extends trivially to more than two RVs.

- Joint pdf of two continuous RVs: $f_{X,Y}(x,y)$, such that

$$\mathbb{P}(X \in A) = \iint_A f_{X,Y}(x,y)\,dx\,dy, \qquad A \subset \mathbb{R}^2$$

  Extends trivially to more than two RVs.

- Marginalization: $f_Y(y) = \begin{cases} \displaystyle\sum_x f_{X,Y}(x,y), & \text{if } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx, & \text{if } X \text{ continuous} \end{cases}$

- Independence: $X \perp\!\!\!\perp Y \;\Leftrightarrow\; f_{X,Y}(x,y) = f_X(x)\,f_Y(y)$.

# Conditionals and Bayes' Theorem

- Conditional pmf (discrete RVs):
  $$f_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x \wedge Y = y)}{\mathbb{P}(Y = y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

# Conditionals and Bayes' Theorem

- Conditional pmf (discrete RVs):
  $$f_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x \wedge Y = y)}{\mathbb{P}(Y = y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

- Conditional pdf (continuous RVs): $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$

  ...the meaning is technically delicate.

# Conditionals and Bayes' Theorem

- Conditional pmf (discrete RVs):
  $$f_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x \wedge Y = y)}{\mathbb{P}(Y = y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

- Conditional pdf (continuous RVs): $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$
  ...the meaning is technically delicate.

- Bayes' theorem: $f_{X|Y}(x|y) = \dfrac{f_{Y|X}(y|x)\, f_X(x)}{f_Y(y)}$ (pdf or pmf).

# Conditionals and Bayes' Theorem

- Conditional pmf (discrete RVs):
$$f_{X|Y}(x|y) = \mathbb{P}(X = x|Y = y) = \frac{\mathbb{P}(X = x \wedge Y = y)}{\mathbb{P}(Y = y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

- Conditional pdf (continuous RVs): $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$

  ...the meaning is technically delicate.

- Bayes' theorem: $f_{X|Y}(x|y) = \dfrac{f_{Y|X}(y|x)\, f_X(x)}{f_Y(y)}$ (pdf or pmf).

- Also valid in the mixed case (*e.g.*, $X$ continuous, $Y$ discrete).

# Joint, Marginal, and Conditional Probabilities: An Example

- A pair of binary variables $X, Y \in \{0, 1\}$, with joint pmf:

| $f_{X,Y}(x,y)$ | $Y = 0$ | $Y = 1$ |
|---|---|---|
| $X = 0$ | 1/5 | 2/5 |
| $X = 1$ | 1/10 | 3/10 |

# Joint, Marginal, and Conditional Probabilities: An Example

- A pair of binary variables $X, Y \in \{0, 1\}$, with joint pmf:

| $f_{X,Y}(x,y)$ | $Y = 0$ | $Y = 1$ |
|---|---|---|
| $X = 0$ | 1/5 | 2/5 |
| $X = 1$ | 1/10 | 3/10 |

- Marginals: $f_X(0) = \frac{1}{5} + \frac{2}{5} = \frac{3}{5}$, $\quad f_X(1) = \frac{1}{10} + \frac{3}{10} = \frac{4}{10}$,

  $f_Y(0) = \frac{1}{5} + \frac{1}{10} = \frac{3}{10}$, $\quad f_Y(1) = \frac{2}{5} + \frac{3}{10} = \frac{7}{10}$.

# Joint, Marginal, and Conditional Probabilities: An Example

- A pair of binary variables $X, Y \in \{0, 1\}$, with joint pmf:

| $f_{X,Y}(x,y)$ | $Y=0$ | $Y=1$ |
|---|---|---|
| $X=0$ | 1/5 | 2/5 |
| $X=1$ | 1/10 | 3/10 |

- Marginals: $f_X(0) = \frac{1}{5} + \frac{2}{5} = \frac{3}{5}$, $\quad f_X(1) = \frac{1}{10} + \frac{3}{10} = \frac{4}{10}$,

$$f_Y(0) = \frac{1}{5} + \frac{1}{10} = \frac{3}{10}, \quad f_Y(1) = \frac{2}{5} + \frac{3}{10} = \frac{7}{10}.$$

- Conditional probabilities:

| $f_{X\|Y}(x\|y)$ | $Y=0$ | $Y=1$ |
|---|---|---|
| $X=0$ | 2/3 | 4/7 |
| $X=1$ | 1/3 | 3/7 |

| $f_{Y\|X}(y\|x)$ | $Y=0$ | $Y=1$ |
|---|---|---|
| $X=0$ | 1/3 | 2/3 |
| $X=1$ | 1/4 | 3/4 |

# An Important Multivariate RV: Multinomial

- Multinomial: $X = (X_1, ..., X_K)$, $X_i \in \{0, ..., n\}$, such that $\sum_i X_i = n$,

$$f_X(x_1, ..., x_K) = \begin{cases} \binom{n}{x_1\ x_2\ \cdots\ x_K} p_1^{x_1}\, p_2^{x_2} \cdots p_k^{x_K} & \Leftarrow \quad \sum_i x_i = n \\ 0 & \Leftarrow \quad \sum_i x_i \neq n \end{cases}$$

$$\binom{n}{x_1\ x_2\ \cdots\ x_K} = \frac{n!}{x_1!\, x_2!\, \cdots\, x_K!}$$

Parameters: $p_1, ..., p_K \geq 0$, such that $\sum_i p_i = 1$.

# An Important Multivariate RV: Multinomial

- Multinomial: $X = (X_1, ..., X_K)$, $X_i \in \{0, ..., n\}$, such that $\sum_i X_i = n$,

$$f_X(x_1, ..., x_K) = \begin{cases} \binom{n}{x_1\ x_2\ \cdots\ x_K} p_1^{x_1}\, p_2^{x_2} \cdots p_k^{x_K} & \Leftarrow \quad \sum_i x_i = n \\ 0 & \Leftarrow \quad \sum_i x_i \neq n \end{cases}$$

$$\binom{n}{x_1\ x_2\ \cdots\ x_K} = \frac{n!}{x_1!\, x_2! \cdots x_K!}$$

Parameters: $p_1, ..., p_K \geq 0$, such that $\sum_i p_i = 1$.

- Generalizes the binomial from binary to $K$-classes.

# An Important Multivariate RV: Multinomial

- Multinomial: $X = (X_1, ..., X_K)$, $X_i \in \{0, ..., n\}$, such that $\sum_i X_i = n$,

$$f_X(x_1, ..., x_K) = \begin{cases} \binom{n}{x_1\ x_2\ \cdots\ x_K} p_1^{x_1}\, p_2^{x_2} \cdots p_k^{x_K} & \Leftarrow \quad \sum_i x_i = n \\ 0 & \Leftarrow \quad \sum_i x_i \neq n \end{cases}$$

$$\binom{n}{x_1\ x_2\ \cdots\ x_K} = \frac{n!}{x_1!\, x_2! \cdots x_K!}$$

Parameters: $p_1, ..., p_K \geq 0$, such that $\sum_i p_i = 1$.

- Generalizes the binomial from binary to $K$-classes.

- Example: tossing $n$ independent fair dice, $p_1 = \cdots = p_6 = 1/6$.
  $x_i$ = number of outcomes with $i$ dots. Of course, $\sum_i x_i = n$.

# An Important Multivariate RV: Gaussian

- Multivariate Gaussian: $X \in \mathbb{R}^n$,

$$f_X(x) = \mathcal{N}(x;\mu,C) = \frac{1}{\sqrt{\det(2\,\pi\,C)}}\exp\left(-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right)$$

# An Important Multivariate RV: Gaussian

- Multivariate Gaussian: $X \in \mathbb{R}^n$,

$$f_X(x) = \mathcal{N}(x; \mu, C) = \frac{1}{\sqrt{\det(2\,\pi\,C)}} \exp\left(-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right)$$

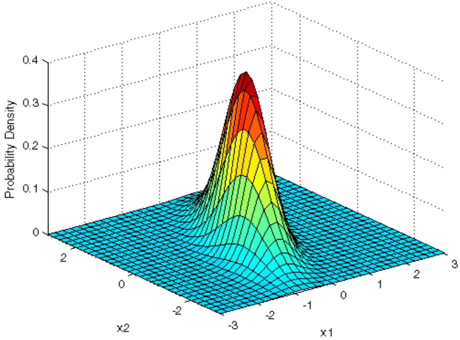
- Parameters: mean $\mu \in \mathbb{R}^n$ and covariance matrix $C \in \mathbb{R}^{n \times n}$.

# An Important Multivariate RV: Gaussian

- Multivariate Gaussian: $X \in \mathbb{R}^n$,

$$f_X(x) = \mathcal{N}(x; \mu, C) = \frac{1}{\sqrt{\det(2\,\pi\,C)}} \exp\left( -\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu) \right)$$

- Parameters: mean $\mu \in \mathbb{R}^n$ and covariance matrix $C \in \mathbb{R}^{n \times n}$.



# Statistical Inference 101 (A Lot More Tomorrow)

- Two RVs, with joint pdf or pmf $f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x)$.

  $f_{Y|X}(y|x)$ is often called likelihood function; $f_X(x)$ is called prior.

# Statistical Inference 101 (A Lot More Tomorrow)

- Two RVs, with joint pdf or pmf $f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x)$.

  $f_{Y|X}(y|x)$ is often called likelihood function; $f_X(x)$ is called prior.

- One RV is observed, say $Y = y$; goal is to infer the other RV $X$.

# Statistical Inference 101 (A Lot More Tomorrow)

- Two RVs, with joint pdf or pmf $f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x)$.

  $f_{Y|X}(y|x)$ is often called likelihood function; $f_X(x)$ is called prior.

- One RV is observed, say $Y = y$; goal is to infer the other RV $X$.

- Maximum likelihood (ML) criterion: $\widehat{x}_{\text{ML}} = \arg\max_x f_{Y|X}(y|x)$

# Statistical Inference 101 (A Lot More Tomorrow)

- Two RVs, with joint pdf or pmf $f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x)$.

  $f_{Y|X}(y|x)$ is often called likelihood function; $f_X(x)$ is called prior.

- One RV is observed, say $Y = y$; goal is to infer the other RV $X$.

- Maximum likelihood (ML) criterion: $\widehat{x}_{\text{ML}} = \arg\max_x f_{Y|X}(y|x)$

- Maximum *a posteriori* (MAP) criterion:

$$\widehat{x}_{\text{MAP}} = \arg\max_x f_{X|Y}(x|y) = \arg\max_x \frac{f_{Y|X}(y|x)\, f_X(x)}{f_Y(y)} = \arg\max_x f_{Y|X}(y|x)\, f_X(x)$$

# Statistical Inference 101 (A Lot More Tomorrow)

- Two RVs, with joint pdf or pmf $f_{X,Y}(x,y) = f_{Y|X}(y|x)\, f_X(x)$.

  $f_{Y|X}(y|x)$ is often called likelihood function; $f_X(x)$ is called prior.

- One RV is observed, say $Y = y$; goal is to infer the other RV $X$.

- Maximum likelihood (ML) criterion: $\widehat{x}_{\text{ML}} = \arg\max_x f_{Y|X}(y|x)$

- Maximum *a posteriori* (MAP) criterion:

$$\widehat{x}_{\text{MAP}} = \arg\max_x f_{X|Y}(x|y) = \arg\max_x \frac{f_{Y|X}(y|x)\, f_X(x)}{f_Y(y)} = \arg\max_x f_{Y|X}(y|x)\, f_X(x)$$

- Posterior mean (PM) criterion (for continuous RVs):

$$\widehat{x}_{\text{PM}} = \mathbb{E}(X|Y=y) = \int x\, f_{X|Y}(x|y)\, dx$$

# Statistical Inference: Example

- Observed $n$ i.i.d. (independent identically distributed) Bernoulli RVs: $Y = (Y_1, ..., Y_n)$, with $Y_i \in \{0, 1\}$.
  Common pmf $f_{Y_i|X}(y|x) = x^y(1-x)^{1-y}$, where $x \in [0,\ 1]$.

# Statistical Inference: Example

- Observed $n$ i.i.d. (independent identically distributed) Bernoulli RVs: $Y = (Y_1, ..., Y_n)$, with $Y_i \in \{0, 1\}$.
  Common pmf $f_{Y_i|X}(y|x) = x^y (1-x)^{1-y}$, where $x \in [0,\, 1]$.

- Likelihood function: $f_{Y|X}\big((y_1, ..., y_n)|x\big) = \prod_{i=1}^{n} x^{y_i}(1-x)^{1-y_i}$

  Log-likelihood function:

$$\log f_{Y|X}\big((y_1, ..., y_n)|x\big) = n \log(1-x) + \log \frac{x}{1-x} \sum_{i=1}^{n} y_i$$

# Statistical Inference: Example

- Observed $n$ i.i.d. (independent identically distributed) Bernoulli RVs: $Y = (Y_1, ..., Y_n)$, with $Y_i \in \{0, 1\}$.
  Common pmf $f_{Y_i|X}(y|x) = x^y (1-x)^{1-y}$, where $x \in [0,\, 1]$.

- Likelihood function: $f_{Y|X}\big((y_1, ..., y_n)|x\big) = \prod_{i=1}^{n} x^{y_i}(1-x)^{1-y_i}$

  Log-likelihood function:

$$\log f_{Y|X}\big((y_1, ..., y_n)|x\big) = n \log(1-x) + \log \frac{x}{1-x} \sum_{i=1}^{n} y_i$$

- Maximum likelihood: $\widehat{x}_{\text{ML}} = \arg\max_x f_{Y|X}(y|x) = \frac{1}{n} \sum_{i=1}^{n} y_i$

# Statistical Inference: Example

- Observed $n$ i.i.d. (independent identically distributed) Bernoulli RVs: $Y = (Y_1, ..., Y_n)$, with $Y_i \in \{0, 1\}$.
  Common pmf $f_{Y_i|X}(y|x) = x^y (1-x)^{1-y}$, where $x \in [0,\ 1]$.

- Likelihood function: $f_{Y|X}\big((y_1, ..., y_n)|x\big) = \prod_{i=1}^{n} x^{y_i}(1-x)^{1-y_i}$

  Log-likelihood function:

$$\log f_{Y|X}\big((y_1, ..., y_n)|x\big) = n \log(1-x) + \log \frac{x}{1-x} \sum_{i=1}^{n} y_i$$

- Maximum likelihood: $\widehat{x}_{\mathrm{ML}} = \arg\max_x f_{Y|X}(y|x) = \frac{1}{n} \sum_{i=1}^{n} y_i$

- Example: $n = 10$, observed $y = (1, 1, 1, 0, 1, 0, 0, 1, 1, 1)$, $\widehat{x}_{\mathrm{ML}} = 7/10$.

# Statistical Inference: Example (Continuation)

- Observed $n$ i.i.d. (independent identically distributed) Bernoulli RVs.

# Statistical Inference: Example (Continuation)

- Observed $n$ i.i.d. (independent identically distributed) Bernoulli RVs.

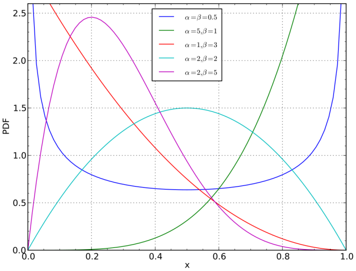
- Likelihood: $f_{Y|X}\big((y_1, ..., y_n)|x\big) = \prod_{i=1}^{n} x^{y_i}(1-x)^{1-y_i} = x^{\sum_i y_i}(1-x)^{n-\sum_i y_i}$

# Statistical Inference: Example (Continuation)

- Observed $n$ i.i.d. (independent identically distributed) Bernoulli RVs.

- Likelihood: $f_{Y|X}\big((y_1, ..., y_n)|x\big) = \prod_{i=1}^{n} x^{y_i}(1-x)^{1-y_i} = x^{\sum_i y_i}(1-x)^{n-\sum_i y_i}$

- Beta (conjugate) prior: $f_X(x) \propto x^{\alpha-1}(1-x)^{\beta-1}$, $\alpha, \beta > 0$



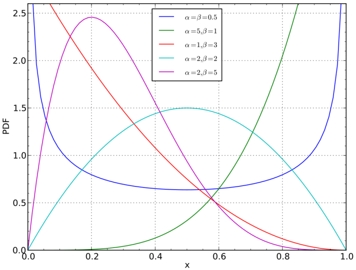
# Statistical Inference: Example (Continuation)

- Observed $n$ i.i.d. (independent identically distributed) Bernoulli RVs.

- Likelihood: $f_{Y|X}\big((y_1,...,y_n)|x\big) = \prod_{i=1}^{n} x^{y_i}(1-x)^{1-y_i} = x^{\sum_i y_i}(1-x)^{n-\sum_i y_i}$

- Beta (conjugate) prior: $f_X(x) \propto x^{\alpha-1}(1-x)^{\beta-1}$, $\alpha,\beta > 0$

  - MAP estimate: $\widehat{x}_{\text{MAP}} = \frac{\alpha+\sum_i y_i - 1}{\alpha+\beta+n-2}$



# Statistical Inference: Example (Continuation)

- Observed $n$ i.i.d. (independent identically distributed) Bernoulli RVs.

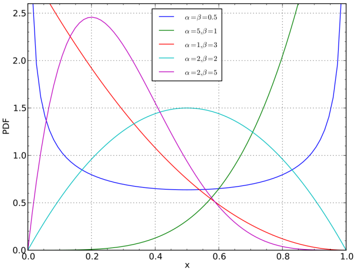
- Likelihood: $f_{Y|X}\left((y_1,...,y_n)|x\right) = \prod_{i=1}^{n} x^{y_i}(1-x)^{1-y_i} = x^{\sum_i y_i}(1-x)^{n-\sum_i y_i}$

- Beta (conjugate) prior: $f_X(x) \propto x^{\alpha-1}(1-x)^{\beta-1}$, $\alpha, \beta > 0$

  - MAP estimate: $\widehat{x}_{\text{MAP}} = \frac{\alpha+\sum_i y_i - 1}{\alpha+\beta+n-2}$

  - Example: $\alpha = 2$, $\beta = 5$, $n = 10$,
    $y = (1,1,1,0,1,0,0,1,1,1)$,

    $$\widehat{x}_{\text{MAP}} = \frac{8}{15} \quad \left(\text{recall } \widehat{x}_{\text{ML}} = 7/10\right)$$



# Agenda

- ~~Probability Theory~~ ✓
- Linear Algebra

# Linear Algebra

- Linear algebra provides (among many other things) a compact way of representing, studying, and solving linear systems of equations

# Linear Algebra

- Linear algebra provides (among many other things) a compact way of representing, studying, and solving linear systems of equations
- Example: the system

$$\begin{aligned} 4\,x_1 - 5\,x_2 &= -13 \\ -2\,x_1 + 3\,x_2 &= 9 \end{aligned}$$

can be written compactly as $Ax = b$, where

$$A = \left[\begin{array}{cc} 4 & -5 \\ -2 & 3 \end{array}\right], \; b = \left[\begin{array}{c} -13 \\ 9 \end{array}\right]$$

# Linear Algebra

- Linear algebra provides (among many other things) a compact way of representing, studying, and solving linear systems of equations
- Example: the system

$$\begin{aligned} 4\,x_1 - 5\,x_2 &= -13 \\ -2\,x_1 + 3\,x_2 &= 9 \end{aligned}$$

can be written compactly as $Ax = b$, where

$$A = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix}, \; b = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

- It can be solved as $x = A^{-1}b$ (if $A^{-1}$ exists).

# Notation: Matrices and Vectors

- $A \in \mathbb{R}^{m \times n}$ is a matrix with $m$ rows and $n$ columns.

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{bmatrix}.$$

# Notation: Matrices and Vectors

- $A \in \mathbb{R}^{m \times n}$ is a matrix with $m$ rows and $n$ columns.

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{bmatrix}.$$

- $x \in \mathbb{R}^n$ is a vector with $n$ components,

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

# Notation: Matrices and Vectors

- $A \in \mathbb{R}^{m\times n}$ is a matrix with $m$ rows and $n$ columns.

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{bmatrix}.$$

- $x \in \mathbb{R}^n$ is a vector with $n$ components,

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

- A (column) vector is a matrix with $n$ rows and 1 column.

# Notation: Matrices and Vectors

- $A \in \mathbb{R}^{m\times n}$ is a matrix with $m$ rows and $n$ columns.

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{bmatrix}.$$

- $x \in \mathbb{R}^n$ is a vector with $n$ components,

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

- A (column) vector is a matrix with $n$ rows and 1 column.

- A matrix with 1 row and $n$ columns is called a row vector.

# Matrix Transpose and Products

- Given matrix $A \in \mathbb{R}^{m \times n}$, its transpose $A^T$ is such that $(A^T)_{i,j} = A_{j,i}$.

# Matrix Transpose and Products

- Given matrix $A \in \mathbb{R}^{m \times n}$, its transpose $A^T$ is such that $(A^T)_{i,j} = A_{j,i}$.
- A matrix $A$ is symmetric if $A^T = A$.

# Matrix Transpose and Products

- Given matrix $A \in \mathbb{R}^{m\times n}$, its transpose $A^T$ is such that $(A^T)_{i,j} = A_{j,i}$.
- A matrix $A$ is symmetric if $A^T = A$.
- Given matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, their product is

$$C = A\,B \; \in \mathbb{R}^{m\times p} \quad \text{where} \quad C_{i,j} = \sum_{k=1}^{n} A_{i,k}\,B_{k,j}$$

# Matrix Transpose and Products

- Given matrix $A \in \mathbb{R}^{m\times n}$, its transpose $A^T$ is such that $(A^T)_{i,j} = A_{j,i}$.
- A matrix $A$ is symmetric if $A^T = A$.
- Given matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, their product is

$$C = A\,B \;\in \mathbb{R}^{m\times p} \quad \text{where} \quad C_{i,j} = \sum_{k=1}^{n} A_{i,k}\, B_{k,j}$$

- Inner product between vectors $x, y \in \mathbb{R}^n$:

$$\langle x, y\rangle = x^T y = y^T x = \sum_{i=1}^{n} x_i y_i \in \mathbb{R}.$$

# Matrix Transpose and Products

- Given matrix $A \in \mathbb{R}^{m \times n}$, its transpose $A^T$ is such that $(A^T)_{i,j} = A_{j,i}$.
- A matrix $A$ is symmetric if $A^T = A$.
- Given matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$, their product is

$$C = A\,B \ \in \mathbb{R}^{m \times p} \quad \text{where} \quad C_{i,j} = \sum_{k=1}^{n} A_{i,k}\, B_{k,j}$$

- Inner product between vectors $x, y \in \mathbb{R}^n$:

$$\langle x, y \rangle = x^T y = y^T x = \sum_{i=1}^{n} x_i y_i \in \mathbb{R}.$$

- Outer product between vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$: $x\,y^T \in \mathbb{R}^{n \times m}$, where $(x\,y^T)_{i,j} = x_i\, y_j$.

# Properties of Matrix Products and Transposes

- Given matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, their product is

$$C = A\,B \;\in \mathbb{R}^{m\times p} \quad \text{where} \;\; C_{i,j} = \sum_{k=1}^{n} A_{i,k}\, B_{k,j}$$

# Properties of Matrix Products and Transposes

- Given matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, their product is

$$C = A\,B \;\in \mathbb{R}^{m\times p} \quad \text{where} \quad C_{i,j} = \sum_{k=1}^{n} A_{i,k}\, B_{k,j}$$

- Matrix product is associative: $(AB)C = A(BC)$.

# Properties of Matrix Products and Transposes

- Given matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, their product is

$$C = A\,B \;\in \mathbb{R}^{m\times p} \quad \text{where} \;\; C_{i,j} = \sum_{k=1}^{n} A_{i,k}\, B_{k,j}$$

- Matrix product is associative: $(AB)C = A(BC)$.
- In general, matrix product is not commutative: $AB \neq BA$.

# Properties of Matrix Products and Transposes

- Given matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, their product is

$$C = A\,B \;\in \mathbb{R}^{m\times p} \quad \text{where} \;\; C_{i,j} = \sum_{k=1}^{n} A_{i,k}\, B_{k,j}$$

- Matrix product is associative: $(AB)C = A(BC)$.
- In general, matrix product is not commutative: $AB \neq BA$.
- Transpose of product: $(A\,B)^T = B^T A^T$.

# Properties of Matrix Products and Transposes

- Given matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$, their product is

$$C = A\,B \;\in \mathbb{R}^{m\times p} \quad \text{where} \;\; C_{i,j} = \sum_{k=1}^{n} A_{i,k}\, B_{k,j}$$

- Matrix product is associative: $(AB)C = A(BC)$.
- In general, matrix product is not commutative: $AB \neq BA$.
- Transpose of product: $(A\,B)^T = B^T A^T$.
- Transpose of sum: $(A+B)^T = A^T + B^T$.

# Norms

- The norm of a vector is (informally) its "length". Euclidean norm:

$$\|x\|_2 = \sqrt{\langle x, x\rangle} = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

# Norms

- The norm of a vector is (informally) its "length". Euclidean norm:

$$\|x\|_2 = \sqrt{\langle x, x\rangle} = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

- More generally, the $\ell_p$ norm of a vector $x \in \mathbb{R}^n$, where $p \geq 1$,

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

# Norms

- The norm of a vector is (informally) its "length". Euclidean norm:

$$\|x\|_2 = \sqrt{\langle x, x\rangle} = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

- More generally, the $\ell_p$ norm of a vector $x \in \mathbb{R}^n$, where $p \geq 1$,

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

- Notable case: the $\ell_1$ norm, $\|x\|_1 = \sum_i |x_i|$.

# Norms

- The norm of a vector is (informally) its "length". Euclidean norm:

$$\|x\|_2 = \sqrt{\langle x, x\rangle} = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

- More generally, the $\ell_p$ norm of a vector $x \in \mathbb{R}^n$, where $p \geq 1$,

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

- Notable case: the $\ell_1$ norm, $\|x\|_1 = \sum_i |x_i|$.

- Notable case: the $\ell_\infty$ norm, $\|x\|_\infty = \max\{|x_1|, ..., |x_n|\}$.

# Norms

- The norm of a vector is (informally) its "length". Euclidean norm:

$$\|x\|_2 = \sqrt{\langle x, x\rangle} = \sqrt{x^T x} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

- More generally, the $\ell_p$ norm of a vector $x \in \mathbb{R}^n$, where $p \geq 1$,

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

- Notable case: the $\ell_1$ norm, $\|x\|_1 = \sum_i |x_i|$.
- Notable case: the $\ell_\infty$ norm, $\|x\|_\infty = \max\{|x_1|, ..., |x_n|\}$.
- Notable case: the $\ell_0$ "norm" (not): $\|x\|_0 = |\{i : x_i \neq 0\}|$.

# Special Matrices

- The identity matrix $I \in \mathbb{R}^{n \times n}$ is a square matrix such that

$$I_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \qquad I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

# Special Matrices

- The identity matrix $I \in \mathbb{R}^{n\times n}$ is a square matrix such that

$$I_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \qquad I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

- Neutral element of matrix product: $A\,I = I\,A = A$.

# Special Matrices

- The identity matrix $I \in \mathbb{R}^{n\times n}$ is a square matrix such that

$$I_{ij} = \left\{ \begin{array}{ll} 1 & i = j \\ 0 & i \neq j \end{array} \right. \qquad I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

- Neutral element of matrix product: $A\, I = I\, A = A$.
- Diagonal matrix: $A \in \mathbb{R}^{n\times n}$ is diagonal if $(i \neq j) \Rightarrow A_{i,j} = 0$.

# Special Matrices

- The identity matrix $I \in \mathbb{R}^{n\times n}$ is a square matrix such that

$$I_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \qquad I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

- Neutral element of matrix product: $A\,I = I\,A = A$.
- Diagonal matrix: $A \in \mathbb{R}^{n\times n}$ is diagonal if $(i \neq j) \Rightarrow A_{i,j} = 0$.
- Upper triangular matrix: $(j < i) \Rightarrow A_{i,j} = 0$.

# Special Matrices

- The identity matrix $I \in \mathbb{R}^{n\times n}$ is a square matrix such that

$$I_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \qquad I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

- Neutral element of matrix product: $A\,I = I\,A = A$.
- Diagonal matrix: $A \in \mathbb{R}^{n\times n}$ is diagonal if $(i \neq j) \Rightarrow A_{i,j} = 0$.
- Upper triangular matrix: $(j < i) \Rightarrow A_{i,j} = 0$.
- Lower triangular matrix: $(j > i) \Rightarrow A_{i,j} = 0$.

# Eigenvalues, eigenvectors, determinant, trace

- A vector $x \in \mathbb{R}^n$ is an eigenvector of matrix $A \in \mathbb{R}^{n \times n}$ if

$$A\,x = \lambda\,x,$$

where $\lambda \in \mathbb{R}$ is the corresponding eigenvalue.

# Eigenvalues, eigenvectors, determinant, trace

- A vector $x \in \mathbb{R}^n$ is an eigenvector of matrix $A \in \mathbb{R}^{n \times n}$ if

$$A\,x = \lambda\,x,$$

  where $\lambda \in \mathbb{R}$ is the corresponding eigenvalue.

- Eigenvalues of diagonal matrix are the elements in the diagonal.

# Eigenvalues, eigenvectors, determinant, trace

- A vector $x \in \mathbb{R}^n$ is an eigenvector of matrix $A \in \mathbb{R}^{n \times n}$ if

$$A\,x = \lambda\,x,$$

  where $\lambda \in \mathbb{R}$ is the corresponding eigenvalue.

- Eigenvalues of diagonal matrix are the elements in the diagonal.

- Matrix trace:

$$\text{trace}(A) = \sum_i A_{i,i} = \sum_i \lambda_i$$

# Eigenvalues, eigenvectors, determinant, trace

- A vector $x \in \mathbb{R}^n$ is an eigenvector of matrix $A \in \mathbb{R}^{n \times n}$ if
$$A\,x = \lambda\,x,$$
where $\lambda \in \mathbb{R}$ is the corresponding eigenvalue.

- Eigenvalues of diagonal matrix are the elements in the diagonal.

- Matrix trace:
$$\text{trace}(A) = \sum_i A_{i,i} = \sum_i \lambda_i$$

- Matrix determinant:
$$|A| = \det(A) = \prod_i \lambda_i$$

# Eigenvalues, eigenvectors, determinant, trace

- A vector $x \in \mathbb{R}^n$ is an eigenvector of matrix $A \in \mathbb{R}^{n\times n}$ if
$$A\,x = \lambda\,x,$$
where $\lambda \in \mathbb{R}$ is the corresponding eigenvalue.
- Eigenvalues of diagonal matrix are the elements in the diagonal.
- Matrix trace:
$$\text{trace}(A) = \sum_i A_{i,i} = \sum_i \lambda_i$$
- Matrix determinant:
$$|A| = \det(A) = \prod_i \lambda_i$$
- Properties: $|AB| = |A||B|$,

# Eigenvalues, eigenvectors, determinant, trace

- A vector $x \in \mathbb{R}^n$ is an eigenvector of matrix $A \in \mathbb{R}^{n \times n}$ if
$$A\,x = \lambda\,x,$$
where $\lambda \in \mathbb{R}$ is the corresponding eigenvalue.

- Eigenvalues of diagonal matrix are the elements in the diagonal.

- Matrix trace:
$$\text{trace}(A) = \sum_i A_{i,i} = \sum_i \lambda_i$$

- Matrix determinant:
$$|A| = \det(A) = \prod_i \lambda_i$$

- Properties: $|AB| = |A||B|$, $\;|A^T| = |A|$,

# Eigenvalues, eigenvectors, determinant, trace

- A vector $x \in \mathbb{R}^n$ is an eigenvector of matrix $A \in \mathbb{R}^{n\times n}$ if
$$A\,x = \lambda\,x,$$
where $\lambda \in \mathbb{R}$ is the corresponding eigenvalue.

- Eigenvalues of diagonal matrix are the elements in the diagonal.

- Matrix trace:
$$\text{trace}(A) = \sum_i A_{i,i} = \sum_i \lambda_i$$

- Matrix determinant:
$$|A| = \det(A) = \prod_i \lambda_i$$

- Properties: $|AB| = |A||B|, \quad |A^T| = |A|, \quad |\alpha\,A| = \alpha^n |A|$

# Matrix Inverse

- Matrix $A \in \mathbb{R}^{n\times n}$ in invertible if there is $B \in \mathbb{R}^{n\times n}$ s.t. $AB = BA = I$.

# Matrix Inverse

- Matrix $A \in \mathbb{R}^{n\times n}$ in invertible if there is $B \in \mathbb{R}^{n\times n}$ s.t. $AB = BA = I$.
- ...matrix $B$ such that $AB = BA = I$ denoted $B = A^{-1}$ (inverse of $A$).

# Matrix Inverse

- Matrix $A \in \mathbb{R}^{n \times n}$ in invertible if there is $B \in \mathbb{R}^{n \times n}$ s.t. $AB = BA = I$.
- ...matrix $B$ such that $AB = BA = I$ denoted $B = A^{-1}$ (inverse of $A$).
- Matrix $A \in \mathbb{R}^{n \times n}$ is invertible $\Leftrightarrow \det(A) \neq 0$.

# Matrix Inverse

- Matrix $A \in \mathbb{R}^{n\times n}$ in invertible if there is $B \in \mathbb{R}^{n\times n}$ s.t. $AB = BA = I$.
- ...matrix $B$ such that $AB = BA = I$ denoted $B = A^{-1}$ (inverse of $A$).
- Matrix $A \in \mathbb{R}^{n\times n}$ is invertible $\Leftrightarrow \det(A) \neq 0$.
- Determinant of inverse: $\det(A^{-1}) = \dfrac{1}{\det(A)}$.

# Matrix Inverse

- Matrix $A \in \mathbb{R}^{n\times n}$ in invertible if there is $B \in \mathbb{R}^{n\times n}$ s.t. $AB = BA = I$.
- ...matrix $B$ such that $AB = BA = I$ denoted $B = A^{-1}$ (inverse of $A$).
- Matrix $A \in \mathbb{R}^{n\times n}$ is invertible $\Leftrightarrow \det(A) \neq 0$.
- Determinant of inverse: $\det(A^{-1}) = \dfrac{1}{\det(A)}$.
- Solving system $Ax = b$, if $A$ is invertible: $x = A^{-1}b$.

# Matrix Inverse

- Matrix $A \in \mathbb{R}^{n\times n}$ in invertible if there is $B \in \mathbb{R}^{n\times n}$ s.t. $AB = BA = I$.
- ...matrix $B$ such that $AB = BA = I$ denoted $B = A^{-1}$ (inverse of $A$).
- Matrix $A \in \mathbb{R}^{n\times n}$ is invertible $\Leftrightarrow$ $\det(A) \neq 0$.
- Determinant of inverse: $\det(A^{-1}) = \dfrac{1}{\det(A)}$.
- Solving system $Ax = b$, if $A$ is invertible: $x = A^{-1}b$.
- Properties: $(A^{-1})^{-1} = A$,

# Matrix Inverse

- Matrix $A \in \mathbb{R}^{n\times n}$ in invertible if there is $B \in \mathbb{R}^{n\times n}$ s.t. $AB = BA = I$.
- ...matrix $B$ such that $AB = BA = I$ denoted $B = A^{-1}$ (inverse of $A$).
- Matrix $A \in \mathbb{R}^{n\times n}$ is invertible $\Leftrightarrow \det(A) \neq 0$.
- Determinant of inverse: $\det(A^{-1}) = \dfrac{1}{\det(A)}$.
- Solving system $Ax = b$, if $A$ is invertible: $x = A^{-1}b$.
- Properties: $(A^{-1})^{-1} = A$, $(A^{-1})^T = (A^T)^{-1}$,

# Matrix Inverse

- Matrix $A \in \mathbb{R}^{n\times n}$ in invertible if there is $B \in \mathbb{R}^{n\times n}$ s.t. $AB = BA = I$.
- ...matrix $B$ such that $AB = BA = I$ denoted $B = A^{-1}$ (inverse of $A$).
- Matrix $A \in \mathbb{R}^{n\times n}$ is invertible $\Leftrightarrow \det(A) \neq 0$.
- Determinant of inverse: $\det(A^{-1}) = \dfrac{1}{\det(A)}$.
- Solving system $Ax = b$, if $A$ is invertible: $x = A^{-1}b$.
- Properties: $(A^{-1})^{-1} = A$, $(A^{-1})^T = (A^T)^{-1}$, $(A\,B)^{-1} = B^{-1}A^{-1}$

# Matrix Inverse

- Matrix $A \in \mathbb{R}^{n\times n}$ in invertible if there is $B \in \mathbb{R}^{n\times n}$ s.t. $AB = BA = I$.
- ...matrix $B$ such that $AB = BA = I$ denoted $B = A^{-1}$ (inverse of $A$).
- Matrix $A \in \mathbb{R}^{n\times n}$ is invertible $\Leftrightarrow \det(A) \neq 0$.
- Determinant of inverse: $\det(A^{-1}) = \dfrac{1}{\det(A)}$.
- Solving system $Ax = b$, if $A$ is invertible: $x = A^{-1}b$.
- Properties: $(A^{-1})^{-1} = A$, $(A^{-1})^T = (A^T)^{-1}$, $(A\,B)^{-1} = B^{-1}A^{-1}$
- There are many algorithms to compute $A^{-1}$; general case, computational cost $O(n^3)$.

# Quadratic Forms and Positive (Semi-)Definite Matrices

- Given matrix $A \in \mathbb{R}^{n\times n}$ and vector $x \in \mathbb{R}^n$,

$$x^T A x = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j}\, x_i\, x_j \;\in\; \mathbb{R}$$

is called a quadratic form.

# Quadratic Forms and Positive (Semi-)Definite Matrices

- Given matrix $A \in \mathbb{R}^{n\times n}$ and vector $x \in \mathbb{R}^n$,

$$x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j}\, x_i\, x_j \ \in\ \mathbb{R}$$

  is called a quadratic form.

- A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is positive semi-definite (PSD) if, for any $x \in \mathbb{R}^n$, $x^T A x \geq 0$.

# Quadratic Forms and Positive (Semi-)Definite Matrices

- Given matrix $A \in \mathbb{R}^{n \times n}$ and vector $x \in \mathbb{R}^n$,

$$x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j}\, x_i\, x_j \ \in \ \mathbb{R}$$

  is called a quadratic form.

- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive semi-definite (PSD) if, for any $x \in \mathbb{R}^n$, $x^T A x \geq 0$.

- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive definite (PD) if, for any $x \in \mathbb{R}^n$, $(x \neq 0) \Rightarrow x^T A x > 0$.

# Quadratic Forms and Positive (Semi-)Definite Matrices

- Given matrix $A \in \mathbb{R}^{n \times n}$ and vector $x \in \mathbb{R}^n$,

$$x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j}\, x_i\, x_j \;\in\; \mathbb{R}$$

  is called a quadratic form.

- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive semi-definite (PSD) if, for any $x \in \mathbb{R}^n$, $x^T A x \geq 0$.

- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive definite (PD) if, for any $x \in \mathbb{R}^n$, $(x \neq 0) \Rightarrow x^T A x > 0$.

- Matrix $A \in \mathbb{R}^{n \times n}$ is PSD $\Leftrightarrow$ all $\lambda_i(A) \geq 0$.

# Quadratic Forms and Positive (Semi-)Definite Matrices

- Given matrix $A \in \mathbb{R}^{n\times n}$ and vector $x \in \mathbb{R}^n$,

$$x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j}\, x_i\, x_j \;\in\; \mathbb{R}$$

  is called a quadratic form.

- A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is positive semi-definite (PSD) if, for any $x \in \mathbb{R}^n$, $x^T A x \geq 0$.

- A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is positive definite (PD) if, for any $x \in \mathbb{R}^n$, $(x \neq 0) \Rightarrow x^T A x > 0$.

- Matrix $A \in \mathbb{R}^{n\times n}$ is PSD $\Leftrightarrow$ all $\lambda_i(A) \geq 0$.

- Matrix $A \in \mathbb{R}^{n\times n}$ is PD $\Leftrightarrow$ all $\lambda_i(A) > 0$.

# Agenda

- ~~Probability Theory~~ ✓
- ~~Linear Algebra~~ ✓

Enjoy LxMLS!