Probability Theory Review

CS221: Introduction to Artificial Intelligence
Naran Bayanbat
10/14/2011

Slides used material from CME106 course reader and CS229 handouts

Topics

- Axioms of Probability
- Product and chain rules
- Bayes Theorem
- Random variables
- PDFs and CDFs
- Expected value and variance

Introduction

- Sample space Ω - set of all possible outcomes of a random experiment
  - Dice roll: {1, 2, 3, 4, 5, 6}
  - Coin toss: {Tails, Heads}
- Event space $\mathcal{F}$ - subsets of elements in a sample space
  - Dice roll: {1, 2, 3} or {2, 4, 6}
  - Coin toss: {Tails}

Introduction

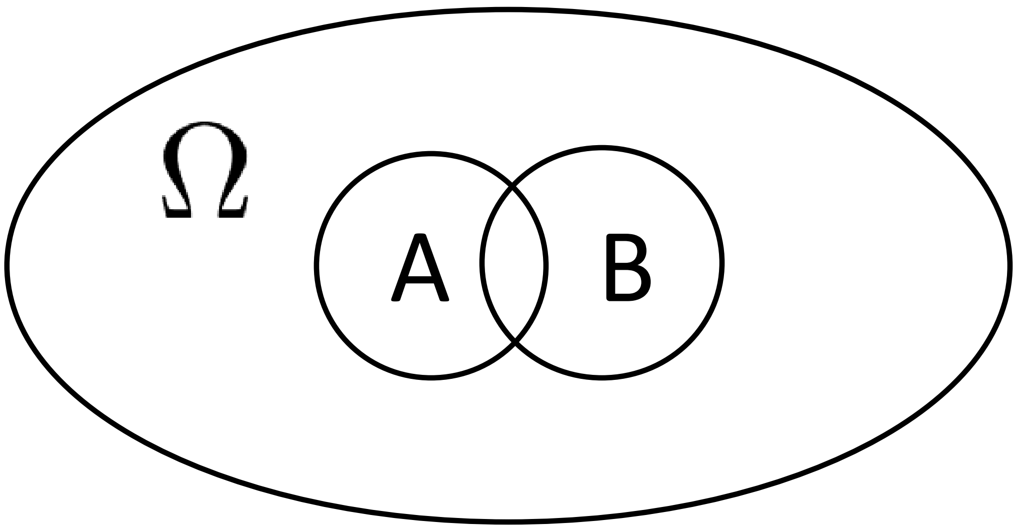
- $Probability\ measure\ P: \mathcal{F} \rightarrow \mathbb{R}$
- Axioms of Probability
  - $0 \le P(A) \le 1$, for all $A \in \mathcal{F}$
  - $P(\Omega) = 1$
  - for $A_1, A_2, \dots, A_n$ disjoint, $P(\bigcup A_i) = \sum P(A_i)$
- $P(A) = \lim_{n\to\infty} \frac{Number\ of\ outcomes \in A}{Total\ number\ of\ outcomes}$

Set operations

- Let $A = \{1, 2, 3\}$ and $\mathrm{B} = \{2, 4, 6\}$
  - $A \cap B = \{2\}$ and $A \cup B = \{1, 2, 3, 4, 6\}$
  - $A - B = \{1, 3\}$
- Properties:
  - $P(A \cap B) \leq \min\left(P(A), P(B)\right)$
  - $P(A \cup B) \leq P(A) + P(B)$
  - $P(\Omega - A) = 1 - P(A)$
  - If $A \subseteq B$ then $P(A) \leq P(B)$

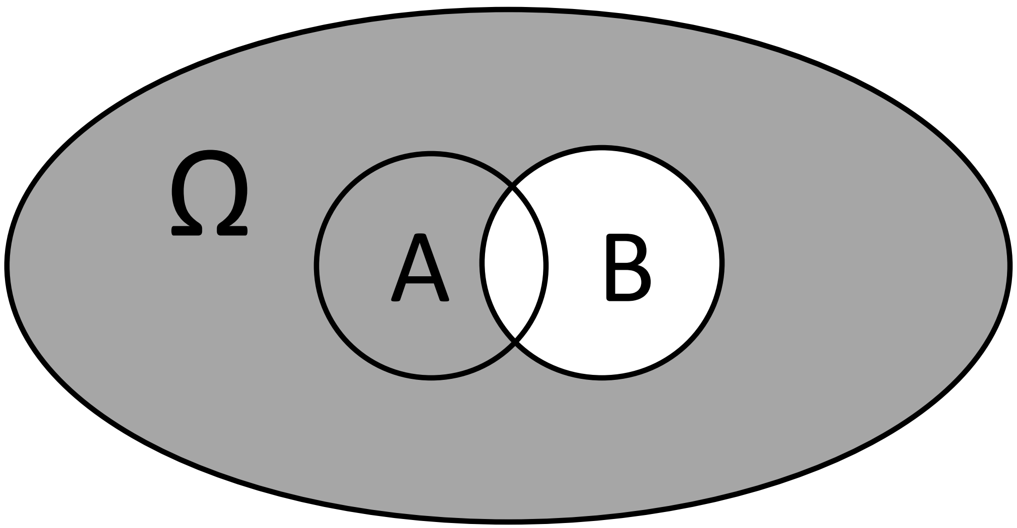
Conditional Probability

- $P(A|B)$ – probability of A given B



Conditional Probability

- $P(A|B)$ – probability of A given B



Conditional Probability

- $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A, B)}{P(B)}$
- A and B are independent if
  - $P(A|B) = P(A)$
- A and B are conditionally independent given C if
  - $P(A, B \mid C) = P(A \mid C)P(B \mid C)$

Conditional Probability

- Joint probability:
  - $P(A_1|A_2,\ldots,A_n) = \frac{P(A_1,\ldots,A_n)}{P}$
- Product rule:
  - $P(A \ldots)$

Conditional Probability

- $P(A|B)$ – probability of A given B
- Example:
  - $P(A)$ probability that school is closed
  - $P(B)$ probability that it snows
  - $P(A|B)$ probability of school closing if it snows
  - $P(B|A)$ probability of snowing if school is closed

Conditional Probability

- $P(A|B)$ – probability of A given B
- Example:
  - $P(A)$ probability that school is closed
  - $P(B)$ probability that it snows

P(A, B)	0.005
P(B)	0.02
P(A|B)	0.25

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

Bayes Theorem

- We can relate P(A|B) and P(B|A) through Bayes' rule:

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

Bayes Theorem

- We can relate $P(A|B)$ and $P(B|A)$ through

Bayes Theorem

- We can relate P(A|B) and P(B|A) through Bayes' rule:

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

- $P(B)$ can be eliminated with law of total probability:

$$P(B) = \sum_j P\left(B\middle|A_j\right)P(A_j)$$

Random Variables

- Random variable *X* is a function s.t.

$$X : \Omega \rightarrow \mathbb{R}$$

- Examples:
  - Russian roulette: *X* = 1 if gun fires and X = 0 otherwise

    $P(X = 1) = \frac{1}{6}$ and $P(X = 0) = \frac{5}{6}$
  - *X* = # of heads in 10 coin tosses

Cumulative Distribution Functions

- Defined as $F_X : \mathbb{R} \rightarrow [0,1]$ such that $F_X(x) = P(X \leq x)$
- Properties:
  - $0 \leq F_X(x) \leq 1$
  - $\lim_{x \to -\infty} F_X(x) = 0$
  - $\lim_{x \to \infty} F_X(x) = 1$
  - $x \leq y \Rightarrow F_X(x) \leq F_X(y)$

Probability Density Functions

- For discrete random variables, defined as:
  - $f(x_j) = P(x_{j-1} < X \leq x_j) = P(x_j)$
- Relates to CDFs:
  - $F_X(x) = \sum_{x_j \leq x} f(x_j)$
  - $f(x_j) = F_X(x_j) - F_X(x_{j-1})$

Probability Density Functions

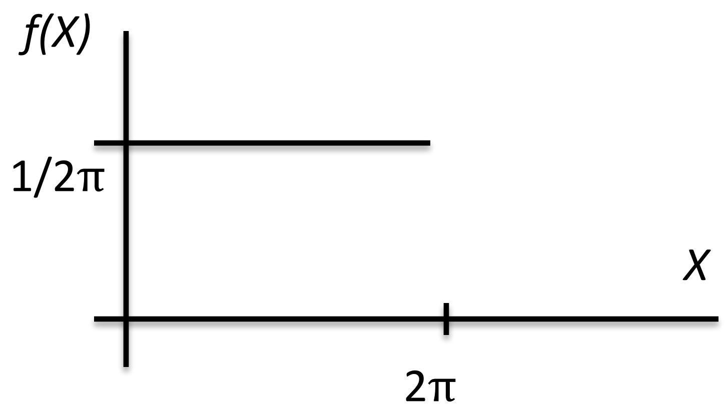
- For continuous random variables, defined as:
  - $f\left(x_j\right) \Delta\, x \approx P(x < X \leq x + \Delta\, x)$
- Relates to CDFs:
  - $F_X(x) = \int_{-\infty}^{x} f_X(x)dx$
  - $f_X(x) = \frac{dF_X(x)}{dx}$

Probability Density Functions

- Example:
  - Let X be the angular position of freely spinning pointer on a circle
  - $f(x)$

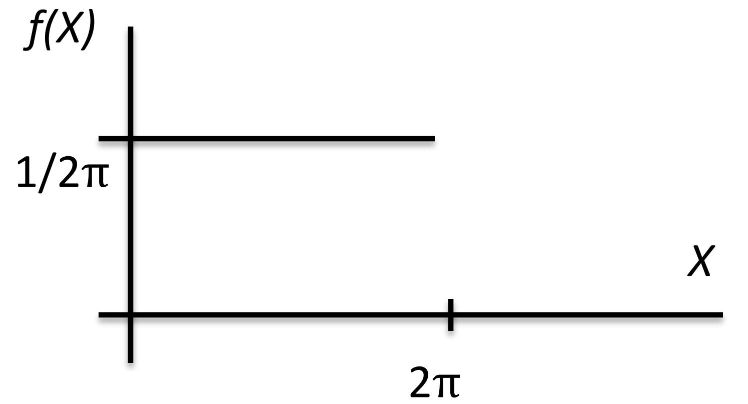
Probability Density Functions

- Example:
  - Let *X* be the angular position of freely spinning pointer on a circle
  - $f(x) = \frac{1}{2\pi}$



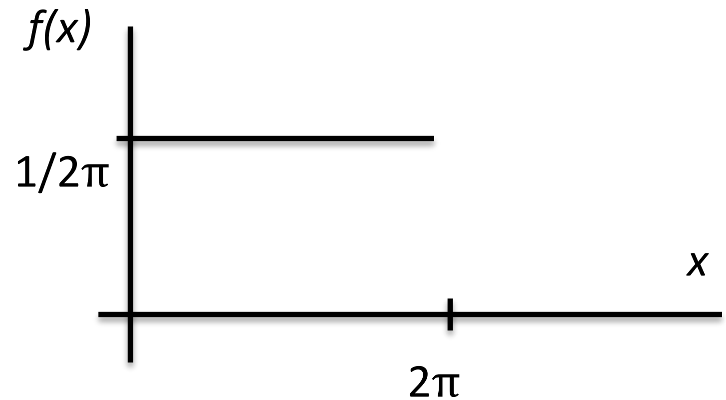
Probability Density Functions

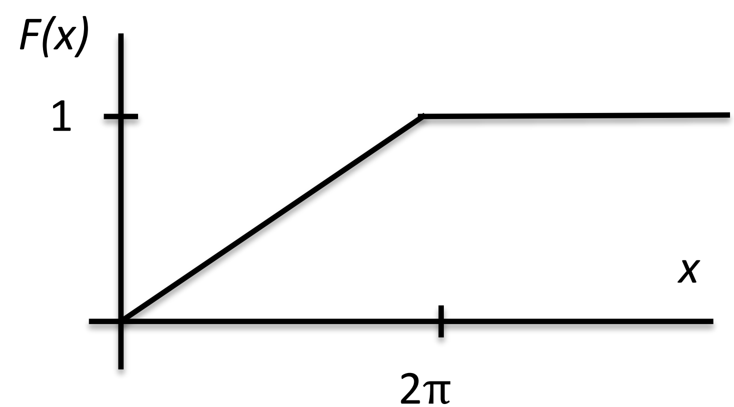
- Example:
  - Let X be the angular position of freely spinning pointer on a circle
  - $f(x) = \frac{1}{2\pi}$
  - $F(x)$



Probability Density Functions

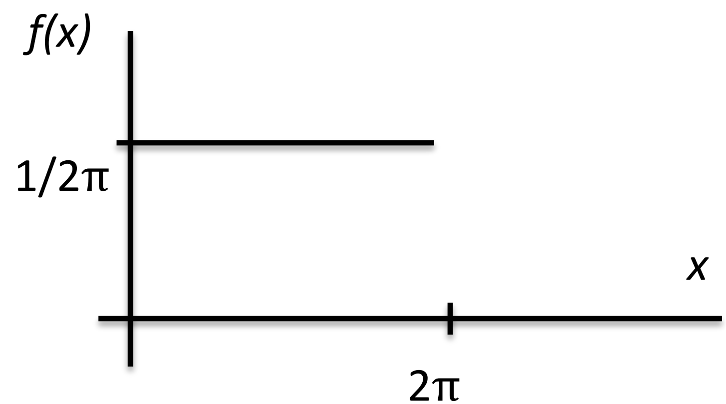
- Example:
  - Let *X* be the angular position of freely spinning pointer on a circle
  - $f(x) = \frac{1}{2\pi}$
  - $F(x) = \frac{1}{2\pi} x$ for $0 \leq x \leq 2\pi$
  - $P(0 \leq x \leq \pi/3\ )$

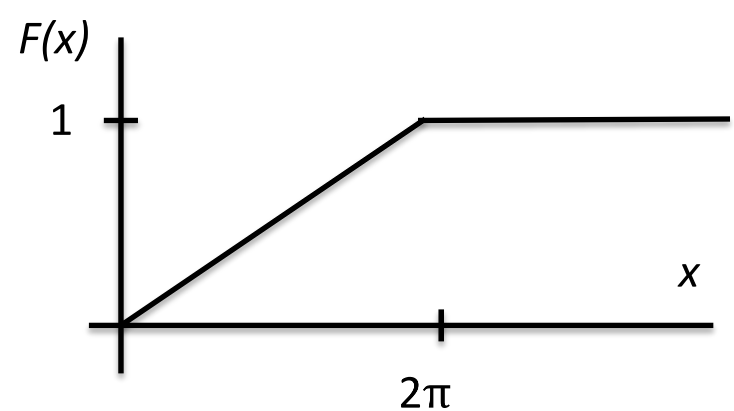




Probability Density Functions

- Example:
  - Let X be the angular position of freely spinning pointer on a circle
  - $f(x) = \frac{1}{2\pi}$
  - $F(x) = \frac{1}{2\pi} x$ for $0 \leq x \leq 2\pi$
  - $P(0 \leq x \leq \pi/3\ ) = \frac{1}{6}$





Expectation

- Given a discrete r.v. X and a function $g : \mathbb{R} \to \mathbb{R}$,

$$E[g(X)] = \sum_{x \in Val(X)} g(x) f_X(x)$$

- For a continuous r.v. X and a function $g : \mathbb{R} \to \mathbb{R}$,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

Expectation

- Expectation of a r.v. X is known as the first moment, or the mean value, of that variable -> $\mathrm{X} = g(X)$
- Properties:
  - $E[a] = a$ for any constant $a \in \mathbb{R}$
  - $E[af(X)] = aE[f(X)]$ for any constant $a \in \mathbb{R}$
  - $E[f(X) + g(X)] = E[f(X)] + E[g(X)]$

Variance

- For a random variable X, define:

$$Var[X] = E\left[\left(X - E(X)\right)^2\right]$$

- Another common form:

$$Var[X] = E[X^2] - E[X]^2$$

- Properties:
  - $Var[a] = 0$ for any constant $a \in \mathbb{R}$
  - $Var[af(X)] = a^2 Var[f(X)]$ for any constant $a \in \mathbb{R}$

Gaussian Distributions

- Let $X \sim Normal(\mu, \sigma^2)$, then

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

where μ represents the mean and σ^2 the variance

- Occurs naturally in many phenomena
  - Noise/error
  - Central limit theorem

Questions?