Probability

CS 470 Introduction To Artificial Intelligence

Daqing Yi

Department of Computer Science
Brigham Young University

Outline

Outline

Probability space

Probability space

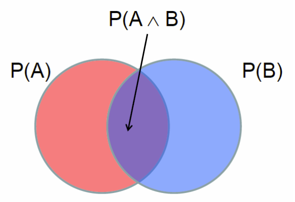
- Ω - world physical states
- defining P
  - $P : 2^{\Omega} \Longrightarrow [0, 1]$
  - for all $A \subseteq \Omega, P(A) \in [0, 1]$
  - $P(\Omega) = 1$

Probability space

Kolmogorvo's axioms

For any set of propositions (events) A, B

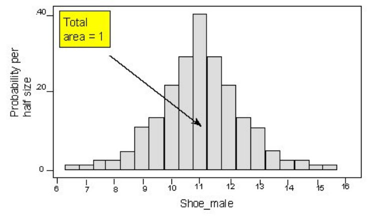
- $0 \leq P(A) \leq 1$
- $P(True) = 1$ and $P(False) = 0$
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$



Type

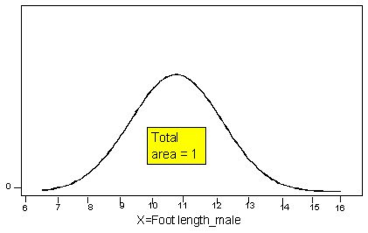
Discrete

Discrete random variable



Continuous

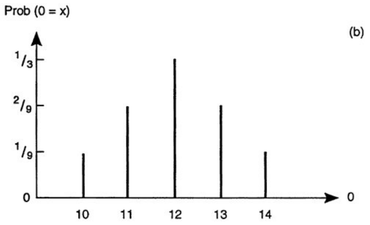
Continuous random variable



Discrete random variable

Probability mass function

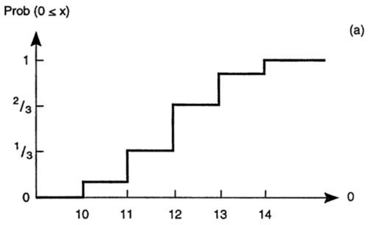
- $\forall x \in S, P(X = x) = f(x) \leq 0$
- $\sum_{x \in S} f(x) = 1$
- $P(A) = \sum_{x \in A} f(x)$



Discrete random variable

Cumulative mass function

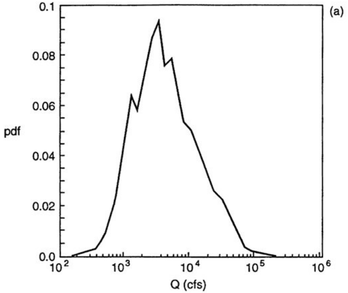
- $F(x) = P(X \leq x)$
- $P(a < X \leq b) = F(b) - F(a)$



Continuous random variable

Probability density function

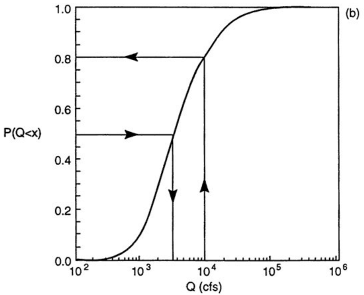
- $\forall x \in S, P(X = x) = f(x) \leq 0$
- $\int_{x \in S} f(x) = 1$
- $P(A) = \int_{x \in A} f(x)$



Continuous random variable

Cumulative density function

- $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$
- $P(a < X \leq b) = \int_{a}^{b} f(t)dt = F(b) - F(a)$



Outline

Joint distribution

Joint probability of $P(X_1, X_2, \cdots, X_n)$

- $P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) \in [0, 1]$
- $\sum P(X_1, X_2, \cdots, X_n) = 1$

Marginal distribution

Relative to <u>Joint distribution</u>

Example:

$P(X)$ and $P(Y)$ are marginal distributions to the joint distribution $P(X, Y)$

discrete variable

$P(X) = \sum_Y P(X, y)$

continuous variable

$P(X) = \int_Y P(X, y)$

mixing discrete variable and continuous variable?

Marginal distribution

Example

P(Cavity, Toothache)	
Cavity = false ∧ *Toothache = false*	0.8
Cavity = false ∧ *Toothache = true*	0.1
Cavity = true ∧ *Toothache = false*	0.05
Cavity = true ∧ *Toothache = true*	0.05

P(Cavity)	
Cavity = false	?
Cavity = true	?

P(Toothache)	
Toothache = false	?
Toochache = true	?

Conditional distribution

- $P(A \mid B)$ is the probability of A given B
  - $$P(A \mid B) = \frac{P(A \wedge B)}{P(B)} = \frac{P(A, B)}{P(B)}$$
- B provides information $\longrightarrow$ uncertainty reduce
- $P(A) \leq P(A \mid B)$

Conditional distribution

Example

P(Cavity, Toothache)	
Cavity = false ∧ *Toothache = false*	0.8
Cavity = false ∧ *Toothache = true*	0.1
Cavity = true ∧ *Toothache = false*	0.05
Cavity = true ∧ *Toothache = true*	0.05

P(Cavity)	
Cavity = false	0.9
Cavity = true	0.1

P(Toothache)	
Toothache = false	0.85
Toothache = true	0.15

- $P(Cavity = true \mid Toothache = false)$
- $P(Cavity = false \mid Toothache = true)$

Conditional distribution

discrete variable

$\sum_X P(x \mid y) = 1$

continuous variable

$\int_X P(x \mid y) = 1$

P(Cavity, Toothache)	
Cavity = false ∧ *Toothache = false*	0.8
Cavity = false ∧ *Toothache = true*	0.1
Cavity = true ∧ *Toothache = false*	0.05
Cavity = true ∧ *Toothache = true*	0.05

Select

Toothache, Cavity = false	
Toothache= false	0.8
Toothache = true	0.1

Renormalize

P(Toothache | Cavity = false)	
Toothache= false	0.889
Toothache = true	0.111

Chain rule

From conditional probability

$$
\begin{aligned}
&P(A_1, A_2, \cdots, A_n) \\
=&P(A_1)P(A_2 \mid A_1)\cdots P(A_n \mid A_1, \cdots, A_{n-1}) \\
=&\prod_{i=1}^{n} P(A_i \mid A_1, \cdots, A_{i-1})
\end{aligned}
$$

Example : Toothache

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

- $P(toothache, cavity)$
- $P(toothache \mid cavity)$
- $P(catch \mid toothache)$

Outline

Independence

- $P(A \wedge) = P(A)P(B)$
- $\forall a \in A, b \in B, P(a \wedge b) = P(a)P(b)$
- Equivalence
  - $P(A \mid B) = P(A)$
  - $P(B \mid A) = P(B)$
- Coin flip

Independence

Example:

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P_2(T, W)$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

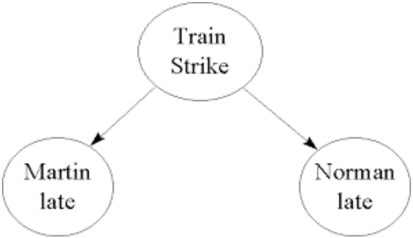
$P(W)$

W	P
sun	0.6
rain	0.4

Conditional independence

- $P(A \wedge B \mid c) = P(A \mid c)P(B \mid c)$
- $\forall a \in A, b \in B, P(a \wedge b \mid c) = P(a \mid c)P(b \mid c)$

Example:



Bayes theorem

Bayes theorem

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}$$

- E - Evidence
- H - Hypothesis
- $P(H \mid E)$ - Posterior
- $P(H)$ - Priori
- $P(E \mid H)$ - Likelihood

ML estimation

Maximum likelihood estimate

ML

$$x^* = \arg\max_X P(e \mid x)$$

Example:

- $P(slim \mid action = diet) = 1$
- $P(slim \mid action = drug) = 0.5$

What should I do to be slim?

MAP estimation

Maximum a posteriori estimate

MAP

$$x^* = \arg\max_X P(x \mid e)$$

Example:

- $P(slim \mid action = diet) = 1$
- $P(slim \mid action = drug) = 0.5$

- $P(action = diet) = 0.1$
- $P(action = drug) = 0.9$

- $P(slim) = 0.5$

What should I do to be slim?

MAP estimation

Maximum a posteriori estimate

MAP

$$x^* = \arg\max_X P(e \mid X)P(X)/P(e)$$
$$= \arg\max_X P(e \mid X)P(X)$$

Example:

- $P(slim \mid action = diet) = 1$
- $P(slim \mid action = drug) = 0.5$

- $P(action = diet) = 0.1$
- $P(action = drug) = 0.9$

What should I do to be slim?