Mathematical Foundations: Probability

Prof. Ziping Zhao

School of Information Science and Technology
ShanghaiTech University, Shanghai, China

CS182: Introduction to Machine Learning (Fall 2021)
http://cs182.sist.shanghaitech.edu.cn

Motivation

Question

Given: We have 25 Male and 15 Female students. If a student is randomly picked from these 2 groups, which group will you guess the student is from?

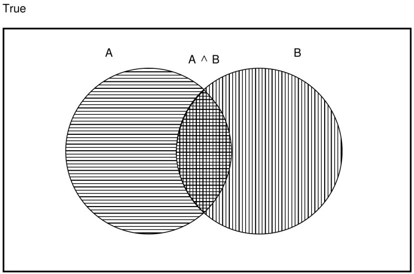
2 classes: A_1 = Male, A_2 = Female

- the state of nature is unpredictable $\rightarrow$ use probability

Axioms for Probability

- All probabilities are between 0 and 1: $0 \leq P(A) \leq 1$
- The certain event has probability 1
- The impossible event has probability 0
- If A and B are any two events,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



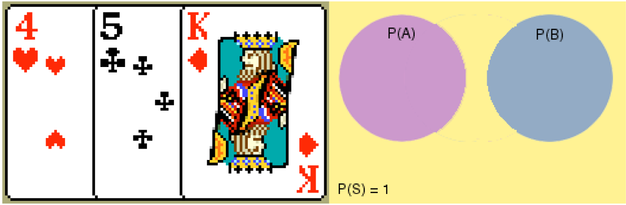
Mutually Exclusive Events

Two events are mutually exclusive if they cannot occur at the same time

Example

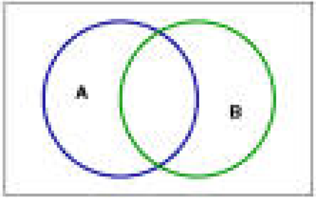
A single card is chosen at random from a standard deck of 52 playing cards

- E1: the card chosen is a five, E2: the card chosen is a king
- mutually exclusive?



Conditional Probability

- Let A and B be two events such that $P(A) > 0$
- $P(B \mid A)$: probability of B given that A has occurred



$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}, \quad P(A \cap B) = P(A)P(B \mid A)$$

- probability that both A and B occur is equal to the probability that A occurs times the probability that B occurs given that A has occurred

Conditional Probability...

For any n events $A_1, A_2, \ldots, A_n$:

$$P(A_1 \cap A_2 \cap \cdots \cap A_{n-1} \cap A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \cdots P(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

(Formula of total probability) If events $A_1, \ldots, A_n$ are mutually exclusive with $\sum_{i=1}^{n} P(A_i) = 1$

$$\begin{aligned} P(B) =& P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n) \\ =& P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \cdots + P(A_n)P(B \mid A_n) \end{aligned}$$

Independence

Two random variables A and B are independent if

$$P(B \mid A) = P(B), \text{ or } P(A \mid B) = P(A)$$

Example

A and B are two coin tosses

- the probability of B occurring is not affected by the occurrence or non-occurrence of A
- knowledge about X contains no information about Y
- this is also equivalent to $P(A \cap B) = P(A)P(B)$

If n Boolean variables $(A_1, \ldots, A_n)$ are independent

$$P(A_1 \cap \cdots \cap A_n) = \prod_{i=1}^{n} P(A_i)$$

Bayes Theorem or Rule

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B \mid A_i)}{P(B)}$$

$$P(\omega_i \mid x) = \frac{P(\omega_i)P(x \mid \omega_i)}{P(x)}$$

- $P(\omega_i)$: prior probability of ω_i
  - initial probability for ω_i, before observing the training data
- $P(\omega_i \mid x)$: posterior probability for ω_i after observing the data x
- $P(x \mid \omega_i)$: likelihood of observing the data x given class ω_i
- $P(x)$: probability that training data x will be observed

Example: Medical Diagnosis

Given:

- $P(\text{Cough} \mid \text{SARS}) = 0.8$
- $P(\text{SARS}) = 0.005$
- $P(\text{Cough}) = 0.05$

Question

Find: $P(\text{SARS} \mid \text{Cough})$

$$
\begin{aligned}
&P(\text{SARS} \mid \text{Cough}) \\
=&\frac{P(\text{Cough} \mid \text{SARS})P(\text{SARS})}{P(\text{Cough})} \\
=&\frac{0.8 \times 0.005}{0.05} = 0.08
\end{aligned}
$$

Discrete Probability Distributions

X: discrete random variable
Probability function or probability distribution

$$P(X = x)$$

Cumulative distribution function (or distribution function):

$$F(x) = P(X \leq x)$$

- if X takes on only a finite number of values $x_1, x_2, \ldots x_n$

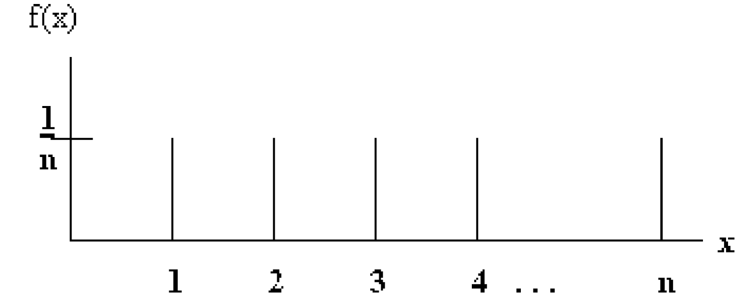
$$F(x) = \begin{cases} 0 & -\infty < x < x_1 \\ P(X = x_1) & x_1 \leq x < x_2 \\ P(X = x_1) + P(X = x_2) & x_2 \leq x < x_3 \\ \vdots & \vdots \\ P(X = x_1) + \cdots + P(X = x_n) & x_n \leq x < \infty \end{cases}$$

Example: Uniform Distribution

Example

outcome of throwing a fair die

- $P(X = 1) = P(X = 2) = \cdots = P(X = 6)$

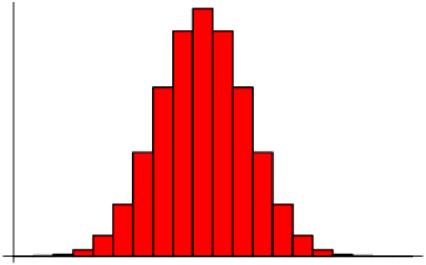


Example: Binomial Distribution

Example

given: probability of getting a head is p, #heads when the biased coin is tossed n times

$$P(X = x) = Bi(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$



Continuous Probability Distributions

X: continuous random variable

- the probability that X takes on any one particular value is generally zero
- the probability that X lies between two different values is more meaningful

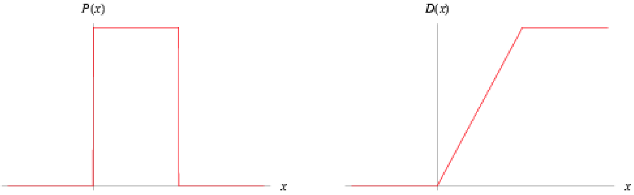
$$P(a < X < b) = \int_a^b p(x)dx$$

  - $p(x)$: probability density function (pdf) (or density function)
- Distribution function:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} p(x)dx \text{ and } \frac{dF(x)}{dx} = p(x)$$

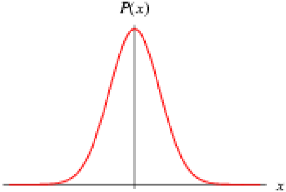
Example: Uniform Distribution

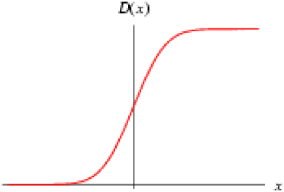
$$p(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x \leq b \\ 0 & \text{otherwise} \end{cases}$$



Example: Normal (Gaussian) Distributions

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$





Joint Distributions: Discrete

- generalization to two or more random variables
- if X and Y are two discrete random variables, we define the joint probability function of X and Y by

  $$P(X = x, Y = y) = p(x, y)$$

  where $p(x, y) \geq 0$ and $\sum_x \sum_y p(x, y) = 1$
- $P(X = x) = \sum_j p(x, y_j)$
  - marginal probability function
- joint distribution function

  $$F(x, y) = P(X \leq x, Y \leq y) = \sum_{u \leq x} \sum_{v \leq y} p(u, v)$$

Joint Distributions: Continuous

- X and Y are continuous random variables

$$P(a < X < b, c < Y < d) = \int_{x=a}^{b} \int_{y=c}^{d} p(x, y) dxdy$$

$$p(x, y) \geq 0 \text{ and } \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) dxdy = 1$$

  - $p(x, y)$: joint density function of X and Y
- marginal density function

$$p(x) = \int_{v=-\infty}^{\infty} p(x, v) dv$$

  - density function of X

Joint Distributions: Continuous...

- joint distribution function

$$F(x, y) = P(X \le x, Y \le y) = \int_{u=-\infty}^{x} \int_{v=-\infty}^{y} p(u, v) du dv$$

$$\frac{\partial^2 F}{\partial x \partial y} = p(x, y)$$

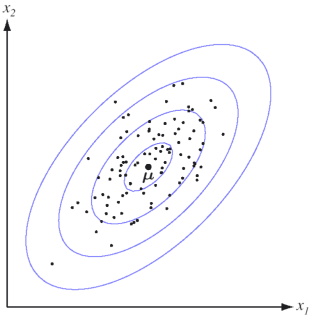
- marginal distribution function

$$P(X \le x) = \int_{u=-\infty}^{x} \int_{v=-\infty}^{\infty} p(u, v) du dv$$

  – distribution function of X

Example

- Random vector: $\mathbf{X} = [X_1, X_2, \dots, X_n]^T$
- multivariate Gaussian: $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$



$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

Mathematical Expectation

- aka expected value or expectation or mean of a random variable X
- X discrete:

$$E(X) = \sum_{j=1}^{n} x_j P(X = x_j)$$

- X continuous :

$$E(X) = \int_{-\infty}^{\infty} x p(x) dx$$

Moments

rth moment: $E(X^r)$

- mean $\mu = E(X)$: 1st moment

rth central moment: $\mu_r = E[(X - \mu)^r]$

- $\mu_0 = 1$, $\mu_1 = 0$, $\mu_2 =$ variance

For multivariate random vector $\mathbf{X}$:

- 2nd central moment: covariance matrix

$$\mathbf{\Sigma} = cov(\mathbf{X}) = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T]$$

Covariance Matrix

For a 2-D vector $\mathbf{X} = [X_1, X_2]^T$:

$$\begin{aligned}
\mathbf{\Sigma} &= E\left(\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix}\begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix}^T\right) \\
&= E\left(\begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 \end{bmatrix}\right) \\
&= \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}
\end{aligned}$$