Pattern Recognition

Lecture 03(a). Review of Probability Theory

Dr. Shanshan ZHAO & Dr. Yuxuan ZHAO

School of AI and Advanced Computing
Xi'an Jiaotong-Liverpool University

Academic Year 2023-2024

Table of Contents

Review of Probability Theory

Probability theory is the study of uncertainty.

Elements of probability

- Sample space Ω :The set of all the outcomes of a random experiment. Here, each outcome $\omega \in \Omega$ can be thought of as a complete description of the state of the real world at the end of the experiment.
- Set of events (or event space) $\mathscr{F}$: A set whose elements $A \in \mathscr{F}$ (called events) are subsets of Ω (i.e., $A \subseteq \Omega$ is a collection of possible outcomes of an experiment)
- Probability measure : A function $P : \mathscr{F} \rightarrow \mathbb{R}$ that satisfies the following properties,
  - $P(A) \geq 0$ for all $A \in \mathscr{F}$
  - $P(\Omega) = 1$
  - If A_1 , A_2 , . . . are disjoint events (i.e., $A_i \cap A_j = \emptyset$ whenever $i \neq j$), then $P(\cup_i A_i) = \sum_i P(A_i)$

Conditional probability

Let B be an event with non-zero probability. The conditional probability of any event A given B is defined as

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

In other words, $P(A|B)$ is the probability measure of the event A after observing the occurrence of event B.

Random Variables

- **discrete random variable:** takes either finite or countably infinite values.
- **continuous random variable:** takes infinite number of possible values.

(a)

(b)

Discrete Random Variable

Experiment:
Consider an experiment in which we flip 10 coins, and we want to know the number of coins that come up heads. Here, the elements of the sample space Ω are 10-length sequences of heads and tails. For example, we might have $\omega_0 =< H, H, T, H, T, H, H, T, T, T >\in \Omega$.
Example: In our experiment above, suppose that X is the number of heads which occur in the sequence of tosses ω. Given that only 10 coins are tossed, X can take only a finite number of values, so it is known as a **discrete random variable**. Here, the probability of the set associated with a random variable X taking on some specific value k is.

$$P(X = k)$$

Continuous Random Variable

Suppose that X is a random variable indicating the amount of time it takes for a radioactive particle to decay. In this case, X takes on an infinite number of possible values, so it is called a **continuous random variable**. The probability that X takes on a value between two real constants a and b (where $a < b$) is denoted as

$$p(a \leq X \leq b)$$

CDFs, PDFs, PMFs

- CDF cumulative distribution function
- PDF Probability Density Function
- PMF probability mass function

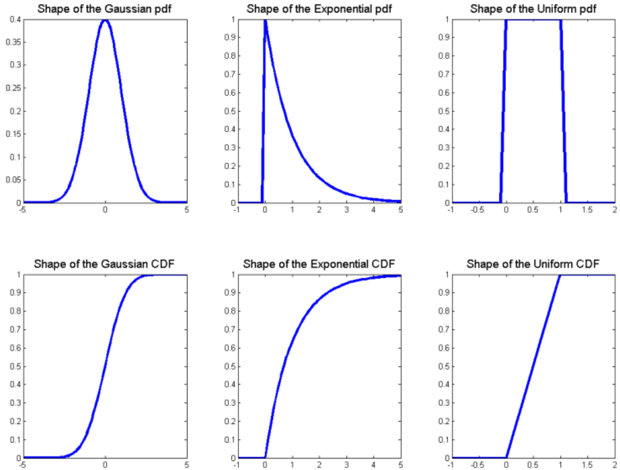

Figure 1: PDF and CDF of a couple of random variables.

measures

- expectation
- variance

Summary of some of the properties of common distributions

Distribution	PDF or PMF	Mean	Variance
$Bernoulli(p)$	$\begin{cases} p, & \text{if } x = 1 \\ 1-p, & \text{if } x = 0. \end{cases}$	p	$p(1-p)$
$Binomial(n,p)$	$\binom{n}{k} p^k (1-p)^{n-k}$ for $0 \leq k \leq n$	np	npq
$Geometric(p)$	$p(1-p)^{k-1}$ for $k = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$Poisson(\lambda)$	$e^{-\lambda}\lambda^x/x!$ for $k = 1, 2, \ldots$	λ	λ
$Uniform(a,b)$	$\frac{1}{b-a}$ $\forall x \in (a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$Gaussian(\mu,\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
$Exponential(\lambda)$	$\lambda e^{-\lambda x}$ $x \geq 0, \lambda > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

The Gaussian distribution

The normal distribution, also called the Gaussian distribution, is a probability distribution commonly used to model phenomena such as physical characteristics (e.g. height, weight, etc.) and test scores.

Why always Gaussian?

- it fits many natural phenomena
- the normal distribution exhibits a number of nice simplifying characteristics
- central limit theorem: the distribution of sample means approximates a normal distribution as the sample size gets larger, regardless of the population's distribution

The Gaussian distribution

Question:

Which of the following is/are true of normal distribution?

A. They are always symmetric

B. They are never fat-tailed

C. They always have a mean of 0

The Gaussian distribution

The *Gaussian(or Normal)* distribution is the most commonly encountered (and easily analysed) continuous distribution.

$$p(x|\mu,\sigma^2) = N(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(\frac{-(x-\mu)^2}{2\sigma^2})$$

The Gaussian is described by two parameters: the mean μ(location), and the variance σ^2(dispersion).

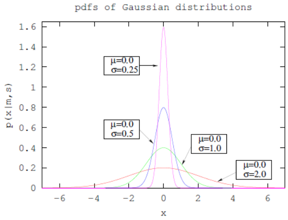

Figure 2: Four Gaussian pdfs with zero mean and different standard deviations.

The multivariate Gaussian distribution and covariance

The univariate(one-dimensioanl) Gaussian may be extended to the multivariate(multi-dimensional) case. The D-dimensional Gaussian is parameterised by a mean vector, $\mu = (\mu_1, ..., \mu_D)^T$, and a *covariance matrix* $\Sigma = (\sigma_{ij})$ (Note that Σ is a D by D dquare matrix, and σ_{ij} or Σ_{ij} denotes its element at i'th row and j'th column.) The probability density is

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

The multivariate Gaussian distribution and covariance

multivariate Gaussian distribution pdf:

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

Exercise:

Consider a 2-dimensional Gaussian distribution with a mean vector $\mu = (0,0)^T$ and a diagonal covariance matrix,i.e.$\Sigma = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix}$, show that its pdf can be simplified to the product of two pdfs, each of which corresponds to a one-dimensional Gaussian distribution.

$$p(x|\mu,\Sigma) = p(x_1|\mu_1,\sigma_{11})p(x_2|\mu_2,\sigma_{22})$$

Solution

hint:

With

$$\Sigma^{-1} = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{\sigma_{11}} & 0 \\ 0 & \frac{1}{\sigma_{22}} \end{bmatrix}$$

We have,

$$(x-\mu)^T\Sigma^{-1}(x-\mu) = \begin{bmatrix} x_1-\mu_1, x_2-\mu_2 \end{bmatrix} \Sigma^{-1} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$$

$$= \begin{bmatrix} x_1, x_2 \end{bmatrix} \Sigma^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{x_1^2}{\sigma_{11}} + \frac{x_2^2}{\sigma_{22}}$$

Then,

$$\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) = \exp\left(-\frac{1}{2}\left(\frac{x_1^2}{\sigma_{11}} + \frac{x_2^2}{\sigma_{22}}\right)\right)$$

$$= \exp\left(-\frac{1}{2}\frac{x_1^2}{\sigma_{11}}\right)\exp\left(-\frac{1}{2}\frac{x_2^2}{\sigma_{22}}\right)$$

...

Thank You !

$\mathscr{Q}$ & $\mathscr{A}$