People
Innovation
Excellence

Course : COMP6577 – Machine Learning
Effective Period : February 2020

Probability and Stochastic Processes 2

Session 07 & 08

BINUS UNIVERSITY

Learning Outcome

- LO2: Student be able to interpret the distribution of dataset using regression method

People
Innovation
Excellence

Outline

- Distribution example of discrete variables
  - The Bernoulli distribution
  - The Binomial distribution
  - The Multinomial distribution
- Distribution example of continuous variables
  - The uniform distribution
  - The Gaussian distribution
  - The exponential distribution
  - The beta distribution
  - The gamma distribution
- Case Study

The Bernoulli distribution

- A random variable is said to be distributed according to a Bernoulli distribution if it is binary, $X = \{0, 1\}$, with

  $$P(x = 1) = p,\ P(x = 0) = 1 - p$$

- In a more compact way, we write $x \sim \text{Bern}(x|p)$ where

  $$\boxed{P(x) = \text{Bern}(x|p) := p^x(1-p)^{1-x}.}$$

- Its mean value is equal to $\mathbb{E}[x] = 1p + 0(1-p) = p$

- and its variance is equal to

  $$\sigma_x^2 = (1-p)^2 p + p^2(1-p) = p(1-p).$$

People
Innovation
Excellence

The Binomial distribution

- A random variable, x, is said to follow a binomial distribution with parameters *n*, *p*, and we write x ~ Bin($x|n$, p) if X = {0, 1, … , *n*} and

$$\boxed{P(\mathrm{x}=k) := \mathrm{Bin}(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n,}$$

- where by definition

$$\binom{n}{k} := \frac{n!}{(n-k)!k!}.$$

People
Innovation
Excellence

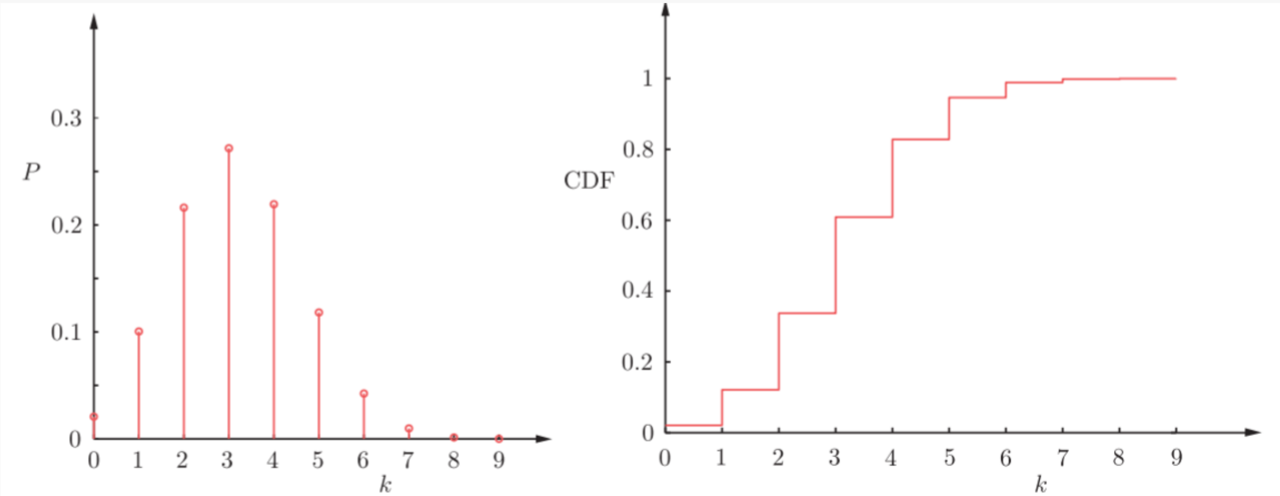
The Binomial distribution (2)

- For example, this distribution models the times that heads occurs in n successive trials, where *P(Heads) = p*. The binomial is a generalization of the Bernoulli distribution, which results if in previous Eq. *P (x=k)* we set n = 1. The mean and variance of the binomial distribution are

$\mathbb{E}[\mathrm{x}] = np$ and $\sigma_x^2 = np(1 - p).$

The Binomial distribution (3)

- The figure (a) shows The probability mass function (pmf) for the binomial distribution for p = 0.4 and n = 9, while in figure (b) shows the respective cumulative probability distribution (cdf).



(a) (b)

- Observe that the latter has a staircase form, as is always the case for discrete variables.

The Multinomial distribution

- This is a generalization of the binomial distribution if the outcome of each experiment is not binary but can take one out of K possible values.
- For example, instead of tossing a coin, a die with K sides is thrown. Each one of the possible K outcomes has probability $P_1, P_2, \ldots, P_K$, respectively, to occur, and we denote P = $[P_1, P_2, \ldots, P_K]^T$
- After n experiments, assume that $x_1, x_2, \ldots, x_K$ times sides x = 1, x = 2, … , x = K occurred, respectively. We say that the random (discrete) vector, $x = [x_1, x_2, \ldots, x_K]^T$

People
Innovation
Excellence

The Multinomial distribution (2)

- follows a multinomial distribution, x ~ Mult(x|n, P), if

$$P(\boldsymbol{x}) = \mathrm{Mult}(\boldsymbol{x}|n, \boldsymbol{P}) := \binom{n}{x_1, x_2, \ldots, x_K} \prod_{k=1}^{K} P_k^{x_k},$$

- Where $$\binom{n}{x_1, x_2, \ldots, x_K} := \frac{n!}{x_1! x_2! \ldots x_K!}$$

- Note that the variables, x_1, ... , x_K , are subject to the constraint $$\sum_{k=1}^{K} \mathrm{x}_k = n$$ and also $$\sum_{k=1}^{K} P_K = 1$$

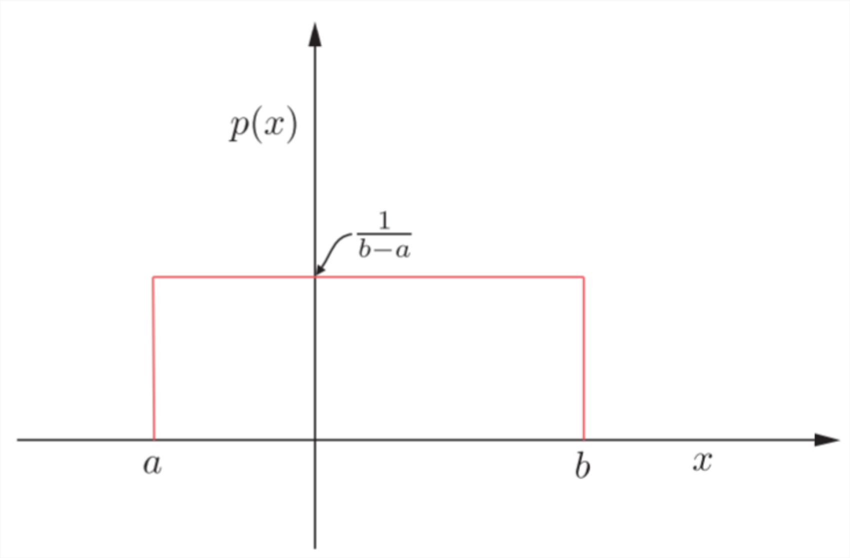
$$\mathbb{E}[\mathbf{x}] = n\boldsymbol{P},\ \sigma_k^2 = nP_k(1 - P_k),\ k = 1, 2, \ldots, K,\ \mathrm{cov}(\mathrm{x}_i, \mathrm{x}_j) = -nP_iP_j,\ i \neq j.$$

The uniform distribution

- A random variable x is said to follow a *uniform* distribution in an interval [*a*, *b*], and we write x ~ U(a, b), with a > −∞ and b < +∞

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

- The pdf of a uniform distribution U(a, b)



People
Innovation
Excellence

The uniform distribution (2)

- The mean value is equal to

$$\mathbb{E}[x] = \frac{a+b}{2}$$

- and the variance is given by

$$\sigma_x^2 = \frac{1}{12}(b-a)^2$$

The Gaussian distribution

- The Gaussian or normal distribution is one among the most widely used distributions in all scientific disciplines.
- A random variable, *x*, is Gaussian or normal with parameters μ and σ^2, and we write $x \sim N(\mu, \sigma^2)$ or $N(x|\mu, \sigma^2)$, if:

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

- It can be shown that the corresponding mean and variance are:

$$\mathbb{E}[x] = \mu \quad \text{and} \quad \sigma_x^2 = \sigma^2$$

People
Innovation
Excellence

The Gaussian distribution (2)

- Indeed, by the definition of the mean value, we have that

$$\mathbb{E}[x] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mathrm{d}x$$

$$= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} (y+\mu) \exp\left(-\frac{y^2}{2\sigma^2}\right) \mathrm{d}y.$$

- Due to the symmetry of the exponential function, performing the integration involving y gives zero and the only surviving term is due to μ. Taking into account that a pdf integrates to one, we obtain the result.

The Gaussian distribution (3)

- To derive the variance, from the definition of the Gaussian pdf, we have that

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mathrm{d}x = \sqrt{2\pi}\sigma$$

- Taking the derivative of both sides with respect to σ, we obtain

$$\int_{-\infty}^{+\infty} \frac{(x-\mu)^2}{\sigma^3} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mathrm{d}x = \sqrt{2\pi}$$

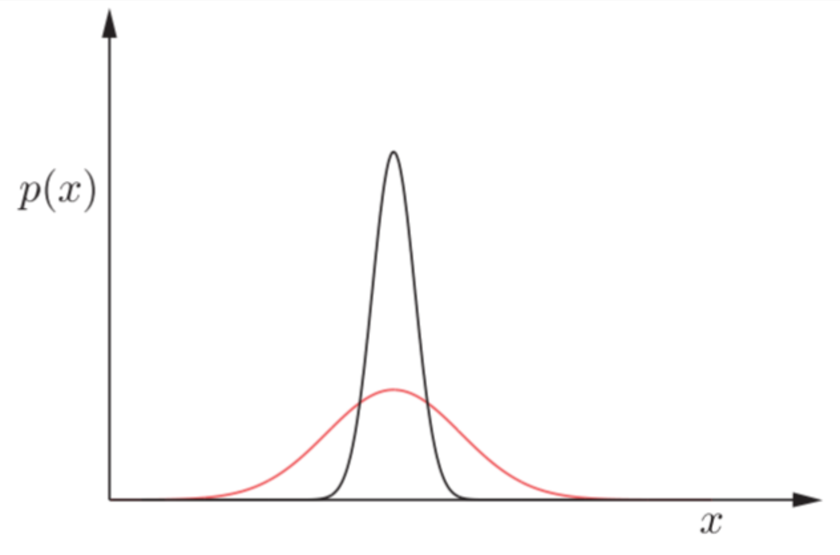
- Or

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (x-\mu)^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \mathrm{d}x = \sigma^2$$

- which proves the claim.

The Gaussian distribution (4)

- The figure below shows the graph for two cases, *N* (x|1, 0.1) and *N* (x|1, 0.01). Both curves are symmetrically placed around the mean value μ = 1. Observe that the smaller the variance is, the sharper around the mean value the pdf becomes.



People
Innovation
Excellence

The exponential distribution

- A random variable follows an exponential distribution with parameter λ > 0, if:

$$p(x) = \begin{cases} \lambda \exp(-\lambda x), & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

- The distribution has been used, for example, to model the time between arrivals of telephone calls or of a bus at a bus stop. The mean and variance can be easily computed by following simple integration rules, and they are

$$\mathbb{E}[x] = \frac{1}{\lambda}, \quad \sigma_x^2 = \frac{1}{\lambda^2}$$

People
Innovation
Excellence

The beta distribution

- A random variable, x ∈ [0, 1], follows a beta distribution with positive parameters, *a*, *b*, and we write, x ~ Beta(x|*a*, *b*,), if:

$$p(x) = \begin{cases} \dfrac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, & \text{if } 0 \le x \le 1, \\ 0, & \text{otherwise,} \end{cases}$$

- where B(*a*, *b*) is the beta function, defined as

$$B(a,b) := \int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x.$$

People
Innovation
Excellence

The beta distribution (2)

- The mean and variance of the beta distribution are given by

$$\mathbb{E}[x] = \frac{a}{a+b}, \quad \sigma_x^2 = \frac{ab}{(a+b)^2(a+b+1)}$$

- Moreover, it can be shown

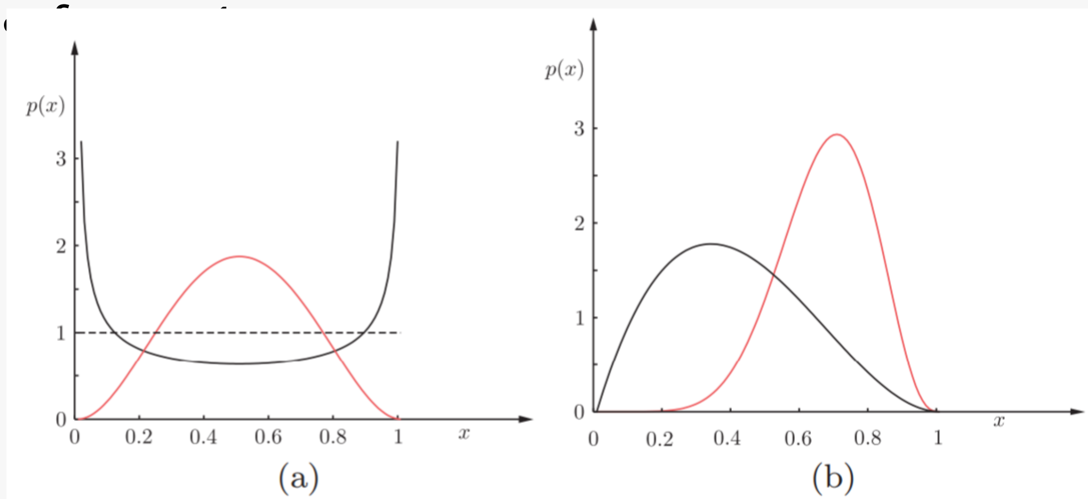
$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

- where $\Gamma(\cdot)$ is the gamma functions defined as

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\,\mathrm{d}x.$$

The beta distribution (3)

- The beta distribution is very flexible and one can achieve various shapes by changing the parameters a, b. For example, if a = b = 1, the uniform distribution results. If a = b, the pdf has a symmetric graph around 1/2. If a > 1, b > 1 then p(x)→0 both at x = 0 and x = 1. If a < 1 and b < 1, it is convex with a unique minimum. If a < 1, it tends to ∞ as x→0, and if b < 1, it tends to



- Figures (a) and (b) show the graph of the beta distribution for different values of the parameters.

The gamma distribution

- A random variable follows the gamma distribution with positive parameters a, b, and we write x ~ Gamma($x|a, b$) if:

$$p(x) = \begin{cases} \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, & x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

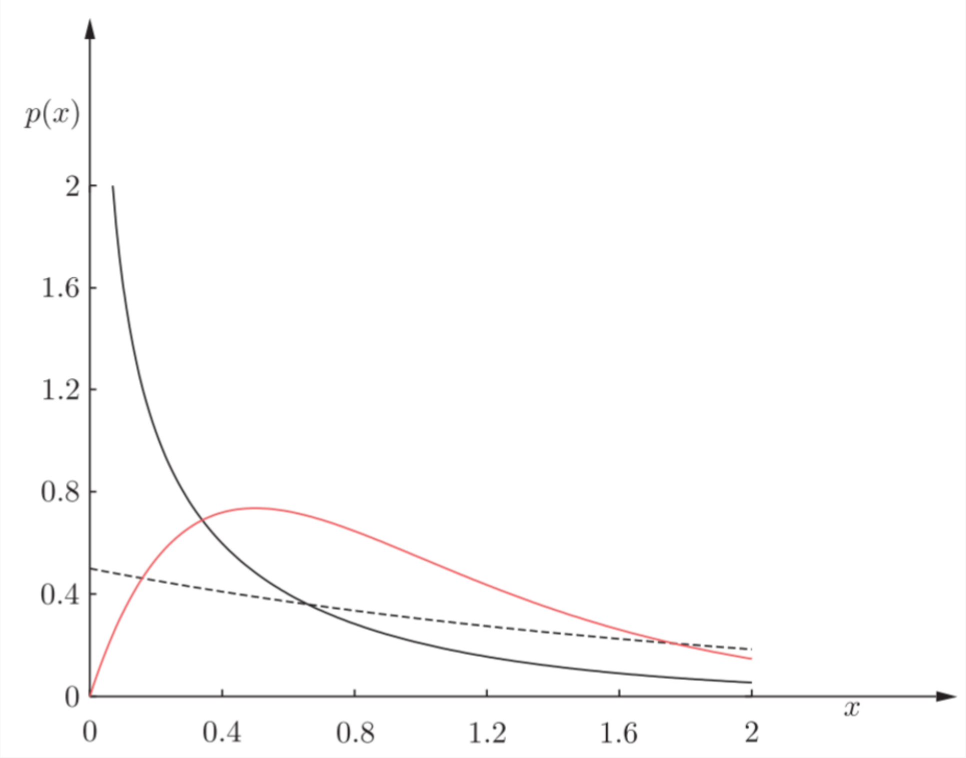
$$\mathbb{E}[\mathrm{x}] = \frac{a}{b}, \quad \sigma_x^2 = \frac{a}{b^2}$$

- The mean and variance are given by

- The gamma distribution also takes various shapes by varying the parameters. For a < 1, it is strictly decreasing and p(x)→∞ as x→0 and p(x)→0 as x→∞.

The gamma distribution (2)

- The figure below shows the resulting graphs for various values of the parameters.



People
Innovation
Excellence

Case Study

Given data of Singapore Airbnb which can be downloaded in this link

https://www.kaggle.com/jojoker/singapore-airbnb

1. You have identified the discrete and continuous random variables in the previous section. Now, you can also identify the distribution for both discrete and continuous variables.

2. Try in Google Collaboratory:
   – Plot the data on a histogram
   – Find a well known distribution
   – Generate and plot the *pdf* on top of your histogram

End of Session 07 & 08

References

- Sergios Theodoridis. (2015). *Machine Learning: a Bayesian and Optimization Perspective*. Jonathan Simpson. ISBN: 978-0-12-801522-3. Chapter 2.
- https://www.kaggle.com/jojoker/singapore-airbnb

People
Innovation
Excellence