Probability and Statistics

Raluca Muresan

West University of Timisoara

raluca.muresan@e-uvt.ro

Overview

Continuous random variables I

A continuous random variable differs from a discrete random variable in that it takes on an uncountably infinite number of possible outcomes. For example, if we let X denote the height (in meters) of a randomly selected maple tree, then X is a continuous random variable

Continuous variables can take any value of an interval, (a, b), (a, ∞), $(-\infty, \infty)$, etc. Various times like service time, installation time, download time, failure time, and also physical measurements like weight, height, distance, velocity, temperature, and connection speed are examples of continuous random variables.

The probability mass function (PMF) of a continuous random variable is always equal to zero

$$P(X = x) = 0, \text{ for all } x.$$

Continuous random variables II

As a result, the PMF does not carry any information about a random variable. Rather, we can use the cumulative distribution function (CDF) $F(x)$

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

Proposition

The CDF of a random variable X has the following properties:

- F is increasing: $F(a) \leq F(b)$, for all $a \leq b$
- F is continuous
- $\lim_{t \to -\infty} F(t) = 0$, $\lim_{t \to \infty} F(t) = 1$

Moreover, we will assume that F is differentiable.

Continuous random variables III

Definition

Probability density function (PDF, density) is the derivative of the CDF,

$$f(x) = F'(x).$$

By the Fundamental Theorem of Calculus, the integral of a density from a to b equals to the difference of antiderivatives, i.e.,

$$P(a < X < b) = \int_a^b f(x)dx = F(b) - F(a),$$

where we notice again that the probability in the right-hand side also equals $P(a \leq X < b)$, $P(a < X \leq b)$, $P(a \leq X \leq b)$.

Continuous random variables IV

The integral $\int_a^b f(x)dx$ equals the area under the density curve between the points a and b. Therefore, geometrically, probabilities are represented by areas.

Proposition

We have that

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

and

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

for all $x \in \mathbb{R}$.

Prove it!

Continuous random variables V

Using the above proposition show that $P(X = x) = 0$ (the PMF is constantly equal to 0).

Example 1

The lifetime, in years, of some electronic component is a continuous random variable with the density

$$f(x) = \begin{cases} \frac{k}{x^3}, & x \geq 1 \\ 0, & x < 1 \end{cases}.$$

Find k, draw a graph of the CDF $F(x)$, and compute the probability for the lifetime to exceed 5 years.

Continuous random variables VI

Distribution	Discrete	Continuous
Definition	$P(x) = \boldsymbol{P}\{X = x\}$ (pmf)	$f(x) = F'(x)$ (pdf)
Computing probabilities	$\boldsymbol{P}\{X \in A\} = \sum_{x \in A} P(x)$	$\boldsymbol{P}\{X \in A\} = \int_A f(x)dx$
Cumulative distribution function	$F(x) = \boldsymbol{P}\{X \leq x\} = \sum_{y \leq x} P(y)$	$F(x) = \boldsymbol{P}\{X \leq x\} = \int_{-\infty}^{x} f(y)dy$
Total probability	$\sum_{x} P(x) = 1$	$\int_{-\infty}^{\infty} f(x)dx = 1$

TABLE 4.1: *Pmf* $P(x)$ *versus pdf* $f(x)$.

Moments of a continuous random variable I

Discrete	Continuous
$\mathbf{E}(X) = \sum_x xP(x)$	$\mathbf{E}(X) = \int xf(x)dx$
$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2$ $= \sum_x (x-\mu)^2 P(x)$ $= \sum_x x^2 P(x) - \mu^2$	$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2$ $= \int (x-\mu)^2 f(x)dx$ $= \int x^2 f(x)dx - \mu^2$
$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \sum_x \sum_y (x-\mu_X)(y-\mu_Y)P(x,y)$ $= \sum_x \sum_y (xy)P(x,y) - \mu_x\mu_y$	$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \iint (x-\mu_X)(y-\mu_Y)f(x,y)\,dx\,dy$ $= \iint (xy)f(x,y)\,dx\,dy - \mu_x\mu_y$

TABLE 4.3: *Moments for discrete and continuous distributions.*

Moments of a continuous random variable II

Example 2

Compute the expected value and variance of the random variable defined at Example 1. Determine the expected value and variance of X^2, $\sqrt{X}$.

Uniform distribution I

Uniform distribution plays a unique role in probability theory. As we have seen at the lab, a random variable with any thinkable distribution can be generated from a Uniform random variable. Many computer languages and software are equipped with a random number generator that produces Uniform random variables. Users can convert them into variables with desired distributions and use for computer simulation of various events and processes.

Uniform distribution is used in any situation when a value is picked "at random" from a given interval; that is, without any preference to lower, higher, or medium values.

Uniform distribution II

Definition

The density function of a uniform random variable $U(a, b)$ on the interval (a, b) is constant, i.e,

$$f(x) = \frac{1}{b-a},\ a < x < b.$$

There does not exist a Uniform distribution on the entire real line. In other words, if you are asked to choose a random number from $(-\infty, +\infty)$, you cannot do it uniformly.

The CDF can easily be determined as

$$F(x) = \frac{x-a}{b-a},\ a < x < b$$

Uniform distribution III

By computing $P(t < x < t + h)$, $h > 0$, $t \in [a, b - h]$, we descover the Uniform property: the probability is only determined by the length of the interval, but not by its location.

Example 3.

Compute the expected value and variance of a Uniform random variable on $[a, b]$.

Example 4.

The arrival time of a flight has a uniform distribution on $[4:50, 5:10]$. Compute the probability that the flight does not arrive before 5:05 and the expected time of the arrival.

Standard Uniform distribution I

The Uniform distribution with $a = 0$ and $b = 1$ is called **Standard Uniform distribution**. The Standard Uniform density is $f(x) = 1$ for $0 < x < 1$. Most random number generators return a Standard Uniform random variable (*default value for runif() in R*).

All the Uniform distributions are related in the following way. If X is a Uniform(a, b) random variable, then

$$Y = \frac{X - a}{b - a}$$

is s Standard Uniform.

Standard Uniform distribution II

Likewise, if Y is Standard Uniform, then

$$X = a + (b - a)Y$$

is Uniform(a, b). Check that

$$X \in (a, b) \Leftrightarrow Y \in (0, 1)$$

Example 5.

Compute the expected value and variance of a Standard Uniform random variable. Plot the PDF and CDF of such a random variable.

Exponential distribution I

Exponential distribution is often used to model time: waiting time, inter-arrival time, hardware lifetime, failure time, time between telephone calls, etc. As we shall see below, in a sequence of rare events, when the number of events is Poisson, the time between events is Exponential.

Definition

The Exponential distribution has density

$$f(x) = \lambda e^{-\lambda x}, \text{ for } x > 0.$$

Example 6.

Compute the CDF, expected value and variance of the Exponential distribution with parameter λ.

Exponential distribution II

The parameter λ has the following meaning: if X is time, measured in minutes, between arrivals, then λ is a frequency, measured in min^{-1}.

For example, if arrivals occur every half a minute, on the average, then $E(X) = 0.5$ and $\lambda = 2$, saying that they occur with a frequency (arrival rate) of 2 arrivals per minute. This λ has the same meaning as the parameter of Poisson distribution.

Example 7.

Jobs are sent to a printer at an average rate of 3 jobs per hour.

1. What is the expected time between jobs?
2. What is the probability that the next job is sent within 5 minutes?

Normal distribution I

Normal distribution plays a **vital role** in Probability and Statistics, mostly because of the **Central Limit Theorem**, according to which sums and averages often have approximately Normal distribution. Due to this fact, various fluctuations and measurement errors that consist of accumulated number of small terms appear normally distributed.

Besides sums, averages, and errors, Normal distribution is often found to be a good model for physical variables like **weight, height, temperature, voltage, pollution level**, and for instance, **household incomes or student grades**.

Normal distribution II

Definition

Normal distribution has a density

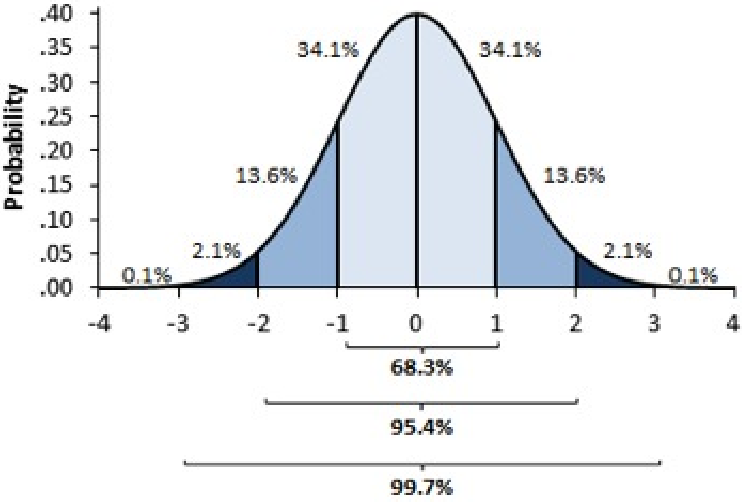
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{-(x-\mu)^2}{2\sigma^2}}, \text{ for all } x \in \mathbb{R},$$

where parameters μ and σ have a simple meaning of the expectation $E(X)$ and the standard deviation $SD(X)$.

This density is known as the bell-shaped curve, symmetric and centered at μ, its spread being controlled by σ.

Changing μ shifts the curve to the left or to the right without affecting its shape, while changing σ makes it more concentrated or more flat. Often μ and σ are called location and scale parameters.

Normal distribution III



Standard Normal Distribution I

Definition

Normal distribution with "standard parameters" $\mu = 0$ and $\sigma = 1$ is called Standard Normal distribution.

A Standard Normal variable, usually denoted by Z, can be obtained from a non-standard $Normal(\mu, \sigma)$ random variable X by standardizing, that is, subtracting the mean and dividing by the standard deviation,

$$Z = \frac{X - \mu}{\sigma}.$$

Un-standardizing Z, we can reconstruct the initial variable X as

$$X = \mu + \sigma * Z.$$

Standard Normal Distribution II

Using these transformations, any Normal random variable can be obtained from a Standard Normal variable Z; therefore, we need a table of Standard Normal Distribution only.

Definition

The PDF of the Standard Normal Distribution is

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

and the CDF is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Standard Normal Distribution III

Example 8.

Suppose that the average household income in some country is 900 coins, and the standard deviation is 200 coins. Assuming the Normal distribution of incomes, compute the proportion of "the middle class", whose income is between 600 and 1200 coins.

Example 9.

The government of the country in Example 8. decides to issue food stamps to the poorest 3% of households. Below what income will families receive food stamps?

Central Limit Theorem I

Theorem

Let $X_1, X_2, \ldots, X_n$ be independent random variables with the same expectation $\mu = E(X_i)$ and the same standard deviation $\sigma = SD(X_i)$, and let

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n} \sim N(\mu, \frac{\sigma}{\sqrt{n}}).$$

As $n \to \infty$ the standardized variable

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

converges in distribution to a Standard Normal random variable, that is

$$Z \sim N(0, 1)$$

Central Limit Theorem II

This theorem is very powerful because it can be applied to random variables $X_1, X_2, \ldots$ having virtually any thinkable distribution with finite expectation and variance. As long as n is large (the rule of thumb is $n > 30$), one can use Normal distribution to compute probabilities about $X_1 + \cdots + X_n$.

Example 10.

A disk has free space of 330 megabytes. Is it likely to be sufficient for 300 independent images, if each image has expected size of 1 megabyte with a standard deviation of 0.5 megabytes?

Central Limit Theorem III

Example 11.

You wait for an elevator, whose capacity is 2000 pounds. The elevator comes with ten adult passengers. Suppose your own weight is 150 lbs, and you heard that human weights are normally distributed with the mean of 165 lbs and the standard deviation of 20 lbs. Would you board this elevator or wait for the next one?

Normal approximation to the Binomial distribution

Binomial variables represent a special case of $S_n = X_1 + \cdots + X_n$, where all X_i have Bernoulli distribution with some parameter p. We know that small p allows to approximate Binomial distribution with Poisson, and large p allows such an approximation for the number of failures. For the moderate values of p (say, $0.05 \leq p \leq 0.95$) and for large n, we can use the Central Limit Theorem:

$$B(n, p) \sim N(np, \sqrt{npq})$$

Continuity correction I

This correction is needed when we approximate a discrete distribution (Binomial in this case) by a continuous distribution (Normal). Recall that the probability $P(X = x)$ may be positive if X is discrete, whereas it is always 0 for continuous X. Thus, a direct use of the normal approximation will always approximate this probability by 0. It is obviously a poor approximation.

This is resolved by introducing a continuity correction. Expand the interval by 0.5 units in each direction, then use the Normal approximation.

Notice that

$$P(X = x) = P(x - 0.5 < X < x + 0.5)$$

is true for a Binomial variable X; therefore, the continuity correction does not change the event and preserves its probability.

Continuity correction II

It makes a difference for the Normal distribution, so every time when we approximate some discrete distribution with some continuous distribution, we should be using a continuity correction.

Example 12.

A new computer virus attacks a folder consisting of 200 files. Each file gets damaged with probability 0.2 independently of other files. What is the probability that fewer than 50 files get damaged?

The End