Course:	Mathematical Statistics
Term:	Fall 2018
Instructor:	Gordan Žitković

Lecture 2
Probability review - continuous random variables

2.1 Probability Density functions (pdfs)

Some random variables naturally take one of a continuum of values, and cannot be associated with a countable set. The simplest example is the uniform random variable Y on $[0,1]$ (also known as a **random number**), which can take any value in the interval $[0,1]$, with the probability of it landing between a and b, where $0 < a < b < 1$, given by

$$\mathbb{P}\Big[Y \in [a,b]\Big] = b - a. \tag{2.1.1}$$

One of the most counterintuitive things about Y is that $\mathbb{P}[Y = y] = 0$ for *any* $y \in [0,1]$, even though we know that Y will take *some value* in $[0,1]$. Therefore, unlike in the discrete case, where the probabilities given by the pmf $p_Y(y) = \mathbb{P}[Y = y]$ contain all the information, in the case of the uniform these are completely uninformative. The right questions to ask is the one of (2.1.1), i.e., one needs to focus on probabilities of values in intervals. The class of random variables where such questions come with an easy-to-represent answer are called *continuous*. More precisely

Definition 2.1.1. A random variable Y is said to have a **continuous distribution** if there exists a function $f_Y : \mathbb{R} \to [0,\infty)$ such that

$$\mathbb{P}\Big[Y \in [a,b]\Big] = \int_a^b f_Y(y)\, dy \text{ for all } a < b.$$

The function f_Y is called the **probability density function (pdf)** of Y.

Not any function can serve as a pdf. The pdf of any random variable will always have the following properties:

1. $f_Y(y) \geq 0$ for all y, and
2. $\int_{-\infty}^{\infty} f_Y(y)\, dy = 1$ since the $\mathbb{P}[Y \in (-\infty,\infty)] = 1$.

It can be shown that any such function is a pdf of some continuous random variables, but we will focus on a small number of important examples in these notes.

Caveat:

1. There are random variables which are neither continuous nor discrete, but we will not encounter them in these notes (even though some important random variables in applications - e.g., insurance - fall into this category.)

2. One should think of the pdf f_Y as an analogue of the pmf in the discrete case, but this analogy should not be stretched too far. For example, we can easily have $f_Y(y) > 1$ at some y, or even on an entire interval. This is the consequence of the fact that $f_Y(y)$ *is not the probability of anything*. It is a probability density, i.e., for small (in the sense of a limit) $\Delta y > 0$ we have

$$\mathbb{P}\Big[Y \in [y, y + \Delta y]\Big] \approx f_Y(y)\Delta y,$$

i.e. $f_Y(y)$ is, approximately, the quotient between the probability of in interval and the size of the same interval.

2.2 The "indicator" notation

Before we list some of the most important examples of continuous random variables, we need to introduce a very useful notation tool.

Definition 2.2.1. For a set $A \subseteq \mathbb{R}$, the function $\mathbf{1}_A : \mathbb{R} \to \mathbb{R}$, given by

$$\mathbf{1}_A(y) = \begin{cases} 1, & y \in A, \\ 0, & \text{otherwise}, \end{cases}$$

is called the **indicator of** A.

As its name already suggests, interval indicators indicate whether their argument y belongs to the set A or not. The graph of a typical indicator - when A is an interval $[a, b]$ - is given in Figure 1.

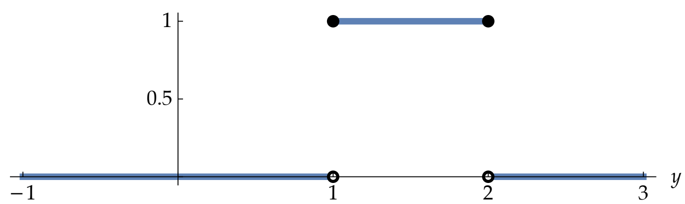

Figure 1. The indicator function $\mathbf{1}_{[1,2]}$ of the interval $[1,2]$.

Indicators are useful when dealing with functions that are defined with different formulas on different parts of their domain.

Example 2.2.2. The uniform distribution $U(l,r)$ is a slight generalization of the uniform $U(0,1)$ distribution mentioned above. It models a number randomly chosen in the interval $[l,r]$ such that the probability of getting a point in the subinterval $[a,b] \subseteq [l,r]$ is proportional to its length $b-a$. Since the probability of choosing some point in $[l,r]$ is 1, by definition, we have to have

$$\mathbb{P}[Y \in [a,b]] = \tfrac{1}{r-l}(b-a) \text{ for all } a < b \in [l,r].$$

To show that this is a continuous distribution, we need to show that it admits a pdf, i.e., a function f such that

$$\int_a^b f_Y(y)\,dy = \tfrac{1}{r-l}(b-a) \text{ for all } a < b.$$

For $a,b < l$ or $a,b > r$ we must have $\mathbb{P}[Y \in [a,b]] = 0$, so

$$\int_a^b f_Y(y)\,dy = 0 \text{ for } a,b \in \mathbb{R} \setminus [l,r].$$

These two requirements force that

$$f_Y(y) = \tfrac{1}{r-l} \text{ for } y \in [l,r] \text{ and } f_Y(y) = 0 \text{ for } y \notin [l,r], \tag{2.2.1}$$

and we can easily check that $f_Y(y)$ is, indeed, the pdf of Y.

The indicator notation can be used to write (2.2.1) in a more compact way:

$$f_Y(y) = \tfrac{1}{r-l}\mathbf{1}_{[l,r]}(y).$$

Not only does this give a single formula valid for all y, it also reveals that $[l,r]$ is the "effective" part of the domain of f_Y. We can think of f_Y as "the constant $\frac{1}{r-l}$, but only on the interval $[l,r]$; it is zero everywhere else".

The interval-indicator notation will come into its own a bit later when we discuss densities of several random variables (random vectors), but for now let us comment on how it allows us to write any integral as an integral over $(-\infty, \infty)$. The idea behind is that any function f multiplied by the indicator $\mathbf{1}_{[a,b]}$ stays the same on $[a, b]$, but takes value 0 everywhere else. Therefore,

$$\int_a^b f(y)\, dy = \int_{-\infty}^{\infty} f(y) \mathbf{1}_{[a,b]}(y)\, dy,$$

because the integral of the function 0 is 0, even when taken over infinite intervals.

Finally, let us introduce another notation for the indicator functions. It turns out to be more intuitive, at least for intervals, and will do wonders for the evaluation of iterated integrals. Since the condition $y \in [a, b]$ can be written as $a \leq y \leq b$, we sometimes write

$$\mathbf{1}_{\{a \leq y \leq b\}} \text{ instead of } \mathbf{1}_{[a,b]}(y).$$

2.3 First examples of continuous random variables

Example 2.3.1.

1. **Uniform distribution.** We have already encountered the uniform distribution $U(l, e)$ on the interval $[l, r]$ and we have shown that it is a continuous distribution with the pdf

$$f_Y(y) = \tfrac{1}{r-l} \mathbf{1}_{[l,r]}(y).$$

As always, this is really a whole family of distributions, parameterized by two real parameters a and b.

Figure 2. The density function (pdf) of the uniform $U(a, b)$ distribution.

2. **Normal distribution** The family of normal distributions - denoted by $Y \sim N(\mu, \sigma)$ - is also parameterized by two parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ and its pdf is given by the (at first sight complicated) formula: The normal distribution is symmetric around μ and its standard deviation (as we shall see shortly) is σ; its graph is shown in Figure 3.

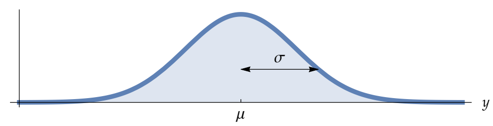

Figure 3. The density function (pdf) of the normal distribution $N(\mu, \sigma)$.

The function f_Y is defined by the above formula for each $y \in \mathbb{R}$ and it is a notrivial task to show that it is, indeed, a pdf of anything. The difficulty lies in evaluating the integral

$$\int_{-\infty}^{\infty} f_Y(y)\, dy$$

and showing that it equals to 1. This is, indeed, true, but needs a bit more mathematics than we care to get into right now. The probabilities $\mathbb{P}[Y \in [a, b]]$ are not any easier to compute for concrete a, b and, in general, do not admit a closed form (formula). That is why we used to use tables of precomputed approximate values (we use software today).

Nevertheless, the normal distribution is, arguably, the most important distribution in probability and statistics. The main reason for that is that it appears in the central limit theorem (which we will talk more about later), and, therefore, shows up whenever a large number of independent random influences act at the same time.

3. **Exponential distribution.** The exponential distribution is a continuous analogue of the geometric distribution and is used in modeling lifetimes of light bulbs or waiting times in the supermarket checkout lines. It comes in a parametric family $E(\tau)$, parameterized by the positive parameter $\tau > 0$. Its pdf is given by

$$f_Y(y) = \tfrac{1}{\tau} e^{-y/\tau} \mathbf{1}_{[0,\infty)}(y).$$

The graph of f_Y is given on the right

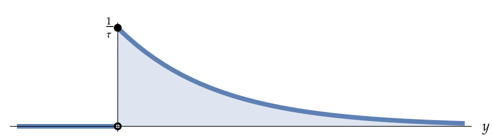

Figure 4. The density function (pdf) of the exponential $E(\tau)$ distribution.

The use of an interval indicator in the expression above signals that f_Y is positive only for $y > 0$, and that, in turn, means that an exponential random variable cannot take negative values.

Caveat: Many books use a different parametrization of the exponential family, namely $Y \sim E(\lambda)$ if

$$f_Y(y) = \lambda e^{-\lambda y} \mathbf{1}_{\{[0,\infty)\}}(y),$$

so that, effectively $\lambda = 1/\tau$. Both parameters have meaningful interpretations, and, depending on the context, one can be more natural than the other. Keep this in mind to avoid unnecessary confusion.

2.4 Expectations and standard deviations

The definitions of the expectation will look similar to that in the discrete case, but sums will be replaced by integrals. Once the expectation is defined, everything else can be repeated verbatim from the previous lecture.

Definition 2.4.1. For a continuous random variable Y with pdf f_Y we define the **expectation** $\mathbb{E}[Y]$ of Y by

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)\, dy, \tag{2.4.1}$$

as long as $\int_{-\infty}^{\infty} |y f_Y(y)|\, dy < \infty$. When this value is $+\infty$, we say that the expectation of Y **is not defined**.

The definition of the variance and the standard deviation are analogous to their discrete versions:

$$\mathrm{Var}[Y] = \int_{-\infty}^{\infty} (y - \mu_Y)^2 f_Y(y)\, dy \text{ where } \mu_Y = \mathbb{E}[Y],$$

and $\mathrm{sd}[Y] = \sqrt{\mathrm{Var}[Y]}$. Theorem **??** and Proposition **??** are valid exactly as written in the continuous case, too.

Let us compute expectations and variances/standard deviations of the distributions from Example 2.3.1.

Example 2.4.2.

1. **Uniform distribution** The computations needed for the expecta-

tion and the variance of the uniform $U(l,r)$ distribution are quite simple:

$$\begin{aligned}\mathbb{E}[Y] &= \int_{-\infty}^{\infty} y f_Y(y)\,dy = \tfrac{1}{r-l}\int_{-\infty}^{\infty} y\mathbf{1}_{[l,r]}(y)\,dy\\ &= \tfrac{1}{r-l}\int_l^r y\,dy = \tfrac{1}{r-l}\tfrac{r^2-l^2}{2} = \tfrac{l+r}{2}.\end{aligned}$$

Similarly,

$$\begin{aligned}\mathrm{Var}[Y] &= \int_{-\infty}^{\infty}(y-\tfrac{l+r}{2})^2 f_Y(y)\,dy = \tfrac{1}{r-l}\int_l^r (y-\tfrac{l+r}{2})^2\,dy\\ &= \tfrac{1}{r-l}[\tfrac{1}{3}(y-\tfrac{l+r}{2})^3]_a^b = \tfrac{1}{12}(r-l)^2\end{aligned}$$

2. **Normal distribution.** To compute the expectation of the normal distribution $N(\mu,\sigma)$, we need to evaluate the following integral

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-\mu)^2}{2\sigma^2}}\,dy.$$

We change the variable $z=(y-\mu)/\sigma$ to obtain

$$\begin{aligned}\mathbb{E}[Y] &= \int_{-\infty}^{\infty}(\sigma z+\mu)\tfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\,dz\\ &= \tfrac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} ze^{-\frac{1}{2}z^2}\,dz + \mu\int_{-\infty}^{\infty}\tfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\,dz = \mu.\end{aligned}$$

The integral next to μ evaluates to 1, because it is simply the integral of the density function f of the standard normal $N(0,1)$. The integral next to $\frac{\sigma}{\sqrt{2\pi}}$ is 0 because it is an integral of an odd function over the entire $\mathbb{R}$.

To compute the variance, we need to evaluate the integral

$$\mathrm{Var}[Y] = \int_{-\infty}^{\infty}(y-\mu)^2\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-\mu)^2}{2\sigma^2}}\,dy,$$

because we now know that $\mu_Y=\mu$. The same change of variables as above yields:

$$\mathrm{Var}[Y] = \sigma^2\tfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2e^{-\frac{1}{2}z^2}\,dz = \sigma^2,$$

where we used the fact (which can be obtained using integration by parts, but we skip the details here) that $\int_{-\infty}^{\infty} z^2e^{-\frac{1}{2}z^2}\,dz = \sqrt{2\pi}$.

3. **Exponential distribution.** The integrals involved in the evaluation of the expectation and the variance of the exponential distribution are simpler and only involve a bit of integration by parts, so we skip the details. It should be noted that the interval indicator notation we used to define the pdf of the exponential tells use immediately what bounds to use for integration. For $Y \sim E(\tau)$, we have

$$\begin{aligned}\mathbb{E}[Y] &= \int_{-\infty}^{\infty} y f_Y(y)\, dy = \int_{-\infty}^{\infty} y/\tau e^{-y/\tau} \mathbf{1}_{[0,\infty)}(y)\, dy \\ &= \int_0^{\infty} y/\tau e^{-y/\tau}\, dy = \tau.\end{aligned}$$

Therefore $\mu_Y = \tau$ and, so,

$$\begin{aligned}\mathrm{Var}[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 &= \int_{-\infty}^{\infty} y^2 f_Y(y)\, dy - \tau^2 \\ &= \int_0^{\infty} y^2/\tau e^{-y/\tau}\, dy - \tau^2.\end{aligned}$$

To evaluate the first integral on the right, we change variables $z = y/\tau$, so that

$$\mathrm{Var}[Y] = \tau^2 \int_0^{\infty} z^2 e^{-z}\, dz - \tau^2 = 2\tau^2 - \tau^2 = \tau^2,$$

where we used the fact (which can be derived by integration by parts) that $\int_0^{\infty} z^2 e^{-z}\, dz = 2$.

2.5 Moments

The expectation is the integral of the first power $y = y^1$ multiplied by the pdf, and the variance involves a similar integral with y replaced by y^2. Integrals of higher powers of y are important in statistics; not as important as the expectation and variance, but still important enough to have names:

Definition 2.5.1. For a random variable Y with pdf f_Y and $k = 1, 2, \dots,$ we define **k-th (raw) moment** μ_k by

$$\mu_k = \mathbb{E}[Y^k] = \int_{-\infty}^{\infty} y^k f_Y(y)\, dy,$$

as well as the **k-th central moment** μ_k^c by

$$\mu_k^c = \mathbb{E}[(Y - \mathbb{E}[Y])^k] = \int_{-\infty}^{\infty} (y - \mu_1)^k f_Y(y)\, dy.$$

We see immediately from the definition that the expectation (mean) is the first (raw) moment and that the variance is the second central moment, i.e.,

$$\mu_1 = \mathbb{E}[Y], \ \mu_2^c = \mathrm{Var}[Y].$$

The third and fourth moment of the standardized random variable, namely,

$$\mathbb{E}\left[\left(\tfrac{Y-\mathbb{E}[Y]}{\mathrm{sd}[Y]}\right)^3\right] \text{ and } \mathbb{E}\left[\left(\tfrac{Y-\mathbb{E}[Y]}{\mathrm{sd}[Y]}\right)^4\right].$$

are called **skeweness** and **kurtosis**, respectively. It is easy to see that, in terms of moments, we can express skeweness as $\mu_3^c/(\mu_2^c)^{3/2}$ and kurtosis as $\mu_4^c/(\mu_2^c)^2$.

Example 2.5.2.

1. **Uniform distribution** We leave this to the reader as an exercise in the Problems section below.

2. **Normal Distribution** Let us compute the central moments, too. Since $Y - \mu \sim N(0, \sigma)$, whenever $Y \sim N(\mu, \sigma)$, central moments of $N(\mu, \sigma)$ are nothing by raw moments of $N(0, \sigma)$. For that, we need to compute the integrals

$$\int_{-\infty}^{\infty} y^k f_Y(y)\, dy = \int_{-\infty}^{\infty} \tfrac{1}{2\pi\sigma} y^k e^{-\frac{1}{2}y^2}\, dy. \tag{2.5.1}$$

For odd k, these are integrals of odd functions over the entire $\mathbb{R}$, and therefore, their value is 0, i.e.,

$$\mu_k = 0 \text{ for } k \text{ odd}.$$

For even k, there is no such a shortcut, and the integral in (2.5.1) can be computed by parts:

$$\int_{-\infty}^{\infty} y^k e^{-\frac{1}{2\sigma^2}y^2}\, dy = \tfrac{1}{k+1} y^{k+1} e^{-\frac{1}{2\sigma^2}y^2}\Big|_{-\infty}^{\infty} - \\ - \int_{-\infty}^{\infty} \tfrac{1}{k+1} y^{k+1}\,(-\tfrac{1}{\sigma^2}y) e^{-\frac{1}{2\sigma^2}y^2}\, dy.$$

Since $\lim_{y\to\pm\infty} y^{k+1} e^{-\frac{1}{2\sigma^2}y^2} = 0$, we obtain

$$\tfrac{1}{2\pi\sigma} \int_{-\infty}^{\infty} y^k e^{-\frac{1}{2\sigma^2}y^2}\, dy = \tfrac{1}{2\pi\sigma} \tfrac{1}{\sigma^2(k+1)} \int_{-\infty}^{\infty} y^{k+2} e^{\frac{1}{2\sigma^2}y^2}\, dy.$$

Written more compactly,

$$\mu_{k+2} = \sigma^2(k+1)\mu_k.$$

Starting from $\mu_2 = \frac{1}{2\pi\sigma}\int_{-\infty}^{\infty} y^2 e^{-\frac{1}{2\sigma^2}y^2} = \sigma^2$, we get

$$\mu_k = \sigma^k \, (k-1)\times(k-3)\times\cdots\times 5\times 3\times 1, \text{ for } k \text{ even.}$$

3. **Exponential distribution.** A similar, integration-by-parts procedure as above, allows us to compute the (raw) moments of the exponential (we skip the details):

$$\mu_k = \int_0^\infty y^k \tfrac{1}{\tau} e^{-y/\tau}\, dy = \tau^k k\times(k-1)\times\cdots\times 2\times 1 = \tau^k\, k!,$$

for $k = 1, 2, 3, \dots$. The central moments are not so important, and do not admit such a nice closed formula.

2.6 Problems

Problem 2.6.1. Let Y be a continuous random variable whose pdf f_Y is given by

$$f_Y(y) = \begin{cases} cy^2, & y \in [-1,1] \\ 0, & \text{otherwise,} \end{cases}$$

for some constant c.

1. Write down an expression for f_Y using the interval-indicator notation.
2. What is the value of c?
3. Compute $\mathbb{E}[Y]$ and $\mathrm{sd}[Y]$.

Problem 2.6.2. Let Y be a continuous random variable with the pdf

$$f_Y(y) = \tfrac{15}{4} y^2(1-y^2)\mathbf{1}_{\{-1\leq y\leq 1\}}.$$

Compute $\mathbb{P}[Y^2 \leq 1/4]$.

Problem 2.6.3. The random variable Y has the pdf $f_Y(y) = \frac{3}{2}y^2\mathbf{1}_{\{-1\leq y\leq 1\}}$. Compute the probability $\mathbb{P}[2Y^2 \geq 1]$.

Problem 2.6.4. Let Y have the pdf $f(y) = \frac{1}{\pi(1+y^2)}$ for $y \in (-\infty,\infty)$. Compute the probability that Y^{-2} lies in the interval $[1/4, 4]$.

Problem 2.6.5 (The exponential distribution)**.** Suppose that the random variable Y follows an exponential distribution with parameter $\tau > 0$, i.e., $Y \sim E(\tau)$, i.e., Y is a continuous random variable with the density function f_Y given by

$$f_Y(y) = \tfrac{1}{\tau} e^{-y/\tau}\mathbf{1}_{y\geq 0}.$$

Compute the following quantities

1. $\mathbb{P}[Y = 0]$, 2. $\mathbb{P}[Y \leq 0]$, 3. $\mathbb{P}[Y \leq y]$ for $y \in (-\infty, \infty)$,
4. $\mathbb{P}[Y > 1]$, 5. $\mathbb{P}[|Y - 2| > 1]$, 6. $\mathbb{E}[Y]$, 7. $\mathbb{E}[Y^2]$, 8. $\mathrm{Var}[Y]$,
9. The mode of Y (look up *mode* if you don't know what it is)
10. The median of Y (look up *median* if you don't know what it is)
11. (Optional) $\mathbb{P}[\,\lfloor Y \rfloor$ is odd $]''$, where $\lfloor a \rfloor$ denotes the largest integer $\leq a$. Which one is bigger $\mathbb{P}[\,\lfloor Y \rfloor$ is odd $]$ or $\mathbb{P}[\,\lfloor Y \rfloor$ is even $]$? Explain without using any calculations.

Problem 2.6.6 (The triangular distribution)**.** We say that the random variable Y follows the **triangular distribution** with parameters $l < r$ if it is continuous with pdf f_Y given by

$$f_Y(y) = c(y-l)\mathbf{1}_{[l, \frac{1}{2}(l+r)]}(y) + c(r-y)\mathbf{1}_{[\frac{1}{2}(l+r), r]}(y).$$

1. Determine the value of the constant c,
2. Compute the expectation and the standard deviation of Y
3. Assuming that $l = -1$ and $r = 1$, compute $\mathbb{P}\Big[\Big|Y - \mathbb{E}[Y]\Big| \geq \mathrm{sd}[Y]\Big]$.

Problem 2.6.7 (Moments of the uniform distribution)**.** Let Y follow the uniform distribution $U(l, r)$ on the interval $[l, r]$, where $l < r$, i.e., its density is given by

$$f_Y(y) = \tfrac{1}{r-l}\mathbf{1}_{\{l \leq y \leq r\}} = \begin{cases} 1/(r-l), & \text{if } y \in [l, r], \\ 0, & \text{otherwise.} \end{cases}$$

Compute the moments μ_k and central moments μ_k^c, $k = 1, 2, \dots$ of Y,