CONTINUOUS RANDOM VARIABLES

Contents

Readings: For a less technical version of this material, but with more discussion and examples, see Sections 3.1-3.5 of [BT] and Sections 4.1-4.5 of [GS].

1 CONTINUOUS RANDOM VARIABLES

Recall[1] that a random variable $X : \Omega \to \mathbb{R}$ is said to be continuous if its CDF can be written in the form

$$\mathbb{P}(X \leq x) = F_X(x) = \int_{(-\infty,x)} f_X(t)dt,$$

for some nonnegative measurable function $f : \mathbb{R} \to [0, \infty)$, which is called the Probability Density Function (PDF) of X. We then have, for any Borel set B,

$$\mathbb{P}(X \in B) = \int_B f_X(x)\, dx = \int_{\mathbb{R}} I_B(x) f_X(x)\, dx \tag{1}$$

Technical remark: All integrals from now on are understood as Lebesgue integrals, unless stated otherwise. In particular $\int_a^b f(t)dt$ is a shorthand notation for $\int_{\mathbb{R}} 1_{(a,b)}(t) f(t) d\lambda(t)$, where λ is Lebesgue measure on $(\mathbb{R}, \mathcal{B})$.

[1] The reader should revisit Section 4 of the notes for Lecture 5.

We note that f_X should be more appropriately called "a" (as opposed to "the") PDF of X, because it is not unique. For example, if we modify f_X at a finite number of points, its integral is unaffected, so multiple densities can correspond to the same CDF. It turns out, however, that any two densities associated with the same CDF are equal except on a set of Lebesgue measure zero.

A PDF is in some ways similar to a PMF, except that the value $f_X(x)$ cannot be interpreted as a probability. In particular, the value of $f_X(x)$ can to be greater than one for some x. Recall Example 8 from lecture 5. There the density was $1/(2\sqrt{t})$ over $t \in (0,1]$ which is larger than one for small values of t. Instead, the proper intuitive interpretation is the fact that if f_X is continuous over a small interval $[x, x+\delta]$, then

$$\mathbb{P}(x \leq X \leq x+\delta) \approx f_X(x)\delta.$$

Also it is instructive to recall the *fundamental theorem of calculus*: If $F_X(x)$ is continuous and differentiable everywhere except countably many points of $\mathbb{R}$, then

$$F_X(x) = \int_{-\infty}^{x} F'_X(t)dt$$

This provides a simple rule to find PDF from CDF in most cases of practical interest.

Remark: The fact that a random variable X is continuous has no bearing on the continuity of X as a function from Ω into $\mathbb{R}$. In fact, we have not even defined what it means for a function on Ω to be continuous. But even in the special case where $\Omega = \mathbb{R}$, we can have a discontinuous function $X : \mathbb{R} \to \mathbb{R}$ which is a continuous random variable. Here is an example. Let the underlying probability measure on Ω be the Lebesgue measure on the unit interval. Let

$$X(\omega) = \begin{cases} \omega, & 0 \leq \omega \leq 1/2, \\ 1+\omega, & 1/2 < \omega \leq 1. \end{cases}$$

The function X is discontinuous. The random variable X takes values in the set $[0, 1/2] \cup (3/2, 2]$. Furthermore, it is not hard to check that X is a continuous random variable with PDF given by

$$f_X(x) = \begin{cases} 1, & x \in [0, 1/2] \cup (3/2, 2] \\ 0 & \text{otherwise.} \end{cases}$$

2 EXAMPLES

We present here a number of important examples of continuous random variables.

2.1 Uniform

This is perhaps the simplest continuous random variable. Consider an interval $[a, b]$, and let

$$F_X(x) = \begin{cases} 0, & x \leq a, \\ (x-a)/(b-a), & a < x \leq b, \\ 1, & x > b. \end{cases}$$

It is easy to check that F_X satisfies the required properties of CDFs. We denote this distribution by $U(a, b)$. We find that a corresponding PDF is given by $f_X(x) = (dF_X/dx)(x) = \frac{1}{b-a}$ for $x \in [a, b]$, and $f_X(x) = 0$, otherwise. When $[a, b] = [0, 1]$, the probability law of a uniform random variable is just the Lebesgue measure on $[0, 1]$.

2.2 Exponential

Fix some $\lambda > 0$. Let $F_X(x) = 1 - e^{-\lambda x}$, for $x \geq 0$, and $F_X(x) = 0$, for $x < 0$. It is easy to check that F_X satisfies the required properties of CDFs. A corresponding PDF is $f_X(x) = \lambda e^{-\lambda x}$, for $x \geq 0$, and $f_X(x) = 0$, for $x < 0$. We denote this distribution by $\mathrm{Exp}(\lambda)$ and write

$$X \sim \mathrm{Exp}(\lambda).$$

(Recall notation $\stackrel{d}{=}$ and $\sim$ which stand for ”distributed as ...”)

The exponential distribution can be viewed as a “limit” of a geometric distribution. Indeed, if we fix some δ and consider the values of $F_X(k\delta) = 1 - e^{-\lambda\delta k}$, for $k = 1, 2, \ldots$. Check that this is $\mathbb{P}(Y \leq k)$, where Y is geometrically distributed with parameter $\rho = 1 - e^{-\lambda\delta}$. Intuitively, the exponential distribution corresponds to a limit of a situation where every δ time units, we toss a coin whose success probability is $\lambda\delta$, and let X be the time elapsed until the first success. We will revisit this intuition later on in the course.

The distribution $\mathrm{Exp}(\lambda)$ has the following very important **memorylessness** property.

Theorem 1. *Let X be an exponentially distributed random variable. Then, for every $x, t \geq 0$, we have $\mathbb{P}(X > x + t \mid X > x) = \mathbb{P}(X > t)$.*

Proof: Let X be exponential with parameter λ. We have

$$\begin{aligned}\mathbb{P}(X > x+t \mid X > x) &= \frac{\mathbb{P}(X > x+t, X > x)}{\mathbb{P}(X > x)} = \frac{\mathbb{P}(X > x+t)}{\mathbb{P}(X > x)} \\ &= \frac{e^{-\lambda(x+t)}}{e^{-\lambda x}} = e^{-\lambda t} = \mathbb{P}(X > t).\end{aligned}$$

□

Exponential random variables are often used to model memoryless arrival processes, in which the elapsed waiting time does not affect our probabilistic model of the remaining time until an arrival. For example, suppose that the time until the next bus arrival is an exponential random variable with parameter $\lambda = 1/5$ (in minutes). Thus, there is probability e^{-1} that you will have to wait for at least 5 minutes. Suppose that you have already waited for 10 minutes. The probability that you will have to wait for at least another five minutes is still the same, e^{-1}.

Semigroup property of exponential: let $X_1 \sim \mathrm{Exp}(\lambda_1)$, $X_2 \sim \mathrm{Exp}(\lambda_2)$, and $X_1 \perp\!\!\!\perp X_2$. then

$$\min(X_1, X_2) \sim \mathrm{Exp}(\lambda_1 + \lambda_2)\,.$$

note that $\min$ defines a commutative operation on $\mathbb{R}_+ \cup \{+\infty\}$ with $+\infty$ serving the role of identity. We can see that exponential distribution relates $+$ and $\min$ operations on $\mathbb{R}_+ \cup \{+\infty\}$: the addition of λ's is equivalent to $\min$ of X's. For this statement to hold in full, one naturally understands $X \sim \mathrm{Exp}(+\infty)$ as $X = 0$ a.s., and $X \sim \mathrm{Exp}(0)$ as $X = +\infty$ a.s..

2.3 Normal distribution

Perhaps the most widely used distribution is the normal distribution which is also called Gaussian distribution. It involves parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, and the density

$$X \sim N(\mu, \sigma^2): \quad f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Note also that this PDF is symmetric around $x = \mu$. Namely $f_X(\mu + x) = f_X(\mu - x)$ for every $x \in \mathbb{R}$. We need to show that this is a legitimate PDF, i.e., that it integrates to one. The special case $\mu = 0, \sigma = 1$ corresponds to the claim

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = 1 \tag{2}$$

and will be established later, when we deal with transformation of random variables. For now let us assume this and show that the same applies to the case of general μ, σ. We introduce a change of variables $z = (t-\mu)/\sigma$ implying $dz = dt/\sigma$. The range $t \in (-\infty, +\infty)$ implies the range $z \in (-\infty, +\infty)$. Then

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

We use the notation $N(\mu, \sigma^2)$ to denote the normal distribution with parameters μ, σ. The distribution $N(0,1)$ is referred to as the **standard normal** distribution; a corresponding random variable is also said to be standard normal.

There is no closed form formula for the corresponding CDF, but numerical tables are available. These tables can also be used to find probabilities associated with general normal variables. This is because of the fact that if $X \sim N(\mu, \sigma^2)$, then $(X-\mu)/\sigma \sim N(0,1)$. . Thus,

$$\mathbb{P}(X \le c) = \mathbb{P}\Big(\frac{X-\mu}{\sigma} \le \frac{c-\mu}{\sigma}\Big) = \Phi((c-\mu)/\sigma),$$

where Φ is the CDF of the standard normal, available from the normal tables.

Semigroup property of normal: Let $X_1 \sim N(0, \sigma_1^2)$, $X_2 \sim N(0, \sigma_2^2)$, and $X_1 \perp\!\!\!\perp X_2$. Then

$$X_1 + X_2 \sim N(0, \sigma_1^2 + \sigma_2^2)\,.$$

2.4 Cauchy distribution

Here, there is only one parameter t and

$$X \sim \mathrm{Ca}(t): \quad f_X(x) = \frac{1}{\pi} \frac{t}{t^2 + x^2}, x \in \mathbb{R}$$

It is an exercise in calculus to show that $\int_{-\infty}^{\infty} f(t)dt = 1$, so that f_X is indeed a PDF. The corresponding distribution is called a Cauchy distribution.

Semigroup property of Cauchy: Let $X_1 \sim \mathrm{Ca}(t_1)$, $X_2 \sim \mathrm{Ca}(t_2)$ and $X_1 \perp\!\!\!\perp X_2$. Then

$$X_1 + X_2 \sim \mathrm{Ca}(t_1 + t_2)$$

2.5 Gamma distribution

Gamma distribution is parameterized by two positive reals: shape parameter $a > 0$ and (inverse) scale parameter $c > 0$.

$$X \sim \Gamma(a, c): \quad f_X(x) = \frac{c^a x^{a-1} e^{-cx}}{\Gamma(a)}, x > 0$$

Semigroup property of Gamma: Let $X_1 \sim \Gamma(a_1, c)$, $X_2 \sim \Gamma(a_2, c)$ and $X_1 \perp\!\!\!\perp X_2$. Then

$$X_1 + X_2 \sim \Gamma(a_1 + a_2, c)$$

2.6 Power law

We have already defined discrete power law distributions. We present here a continuous analog. Our starting point is to introduce tail probabilities that decay according to power law: $\mathbb{P}(X > x) = \beta/x^\alpha$, for $x \geq c > 0$, for some parameters $\alpha, c > 0$. In this case, the CDF is given by $F_X(x) = 1 - \beta/x^\alpha$, $x \geq c$, and $F_X(x) = 0$, otherwise. In order for X to be a continuous random variable, F_X cannot have a jump at $x = c$, and we therefore need $\beta = c^\alpha$ and $F_X(x) = 1 - c^\alpha/x^\alpha$. The corresponding density is of the form

$$f_X(t) = \frac{dF_X}{dx}(t) = \frac{\alpha c^\alpha}{t^{\alpha+1}}.$$

3 EXPECTED VALUES

Similar to the discrete case, given a continuous random variable X with PDF f_X, we have a simple rule to compute the expectation:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx.$$

(This was shown in Lecture 8.) This integral is well defined and finite if $\int_{-\infty}^{\infty} |x| f_X(x)\, dx < \infty$, in which case we say that the random variable X is **integrable**. The integral is also well defined, but infinite, if one, but not both, of the integrals $\int_{-\infty}^{0} x f_X(x)\, dx$ and $\int_{0}^{\infty} x f_X(x)\, dx$ is infinite. If both of these integrals are infinite, the expected value is not defined.

Practically all of the results developed for discrete random variables carry over to the continuous case. Many of them, e.g., $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, have the exact same form. We list below two results in which sums need to be replaced by integrals.

Proposition 1. *Let X be a nonnegative random variable, i.e., $\mathbb{P}(X < 0) = 0$. Then*

$$\mathbb{E}[X] = \int_{0}^{\infty} (1 - F_X(t))\, dt.$$

Proof: We have

$$\int_0^\infty (1 - F_X(t))\, dt = \int_0^\infty \mathbb{P}(X > t)\, dt = \int_{\mathbb{R}_+} dt \int_\Omega 1\{X(\omega) > t\} d\mathbb{P}(\omega)$$
$$= \int_\Omega d\mathbb{P}(\omega) \left[\int_{\mathbb{R}_+} 1\{t < X(\omega)\} dt\right] = \int_\Omega X(\omega) d\mathbb{P} \triangleq \mathbb{E}[X]$$

(The interchange of the order of integration is by Fubini's theorem for non-negative functions.) □

Proposition 2. *Let X be a continuous random variable with density f_X, and suppose that $g : \mathbb{R} \to \mathbb{R}$ is a (Borel) measurable function. Then*

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(t) f_X(t)\, dt$$

(i.e. the integral and the expectation exist or do not exist simultaneosly, and are equal in the latter case).

Proof: This was shown in Lecture 8. □

Note that for this result to hold, the random variable $g(X)$ need not be continuous.

4 JOINT DISTRIBUTIONS

Definition 1. *Given a pair of random variables X and Y, defined on the same probability space, their* ***joint distribution*** *$\mathbb{P}_{X,Y}$ is a probability measure on $(\mathbb{R} \times \mathbb{R}, \mathcal{B} \times \mathcal{B})$ defined as*

$$\mathbb{P}_{X,Y}[B] \triangleq \mathbb{P}[(X, Y) \in B]$$

for every $B \in \mathcal{B} \times \mathcal{B}$.

Exercise: Show that set $\{\omega : (X(\omega), Y(\omega)) \in B\}$ is measurable for any $B \in \mathcal{B} \times \mathcal{B}$. (This provides another justification for the definition of product σ-algebra.)

The **joint CDF** of X, Y is defined as

$$F_{X,Y}(x, y) = \mathbb{P}[X \leq x, Y \leq y]$$

and we say that X, Y are **jointly continuous** if there exists a measurable $f_{X,Y}$:

$\mathbb{R}^2 \to [0, \infty)$ such that their joint CDF satisfies

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v)\, du\, dv.$$

The function $f_{X,Y}$ is called *a* joint PDF of X and Y.

At those points where a joint PDF is continuous, we have

$$\frac{\partial^2 F}{\partial x \partial y}(x, y) = \frac{\partial^2 F}{\partial y \partial x}(x, y) = f_{X,Y}(x, y).$$

Similar to what was mentioned for the case of a single random variable, for any Borel subset B of $\mathbb{R}^2$, we have

$$\mathbb{P}_{X,Y}[B] = \int_B f_{X,Y}(x, y)\, dx\, dy = \int_{\mathbb{R}^2} 1_B(x, y) f_{X,Y}(x, y)\, dx\, dy. \qquad (3)$$

Furthermore, if B has Lebesgue measure zero, then $\mathbb{P}_{X,Y}(B) = 0$.

We observe that by (3) and Fubini's theorem

$$\mathbb{P}(X \leq x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{X,Y}(u, v)\, du\, dv.$$

Thus, X itself is a continuous random variable, with **marginal PDF**

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy.$$

We have just argued that if X and Y are jointly continuous, then X (and, similarly, Y) is a continuous random variables. The converse is not true. For a trivial counterexample, let X be a continuous random variable, and let and $Y = X$. Then the set $\{(x, y) \in \mathbb{R}^2 \mid x = y\}$ has zero area (zero Lebesgue measure), but unit probability, which is impossible for jointly continuous random variables. In particular, the corresponding probability law on $\mathbb{R}^2$ is neither discrete nor continuous, hence qualifies as "singular."

Proposition 2 has a natural extension to the case of multiple random variables.

Proposition 3. *Let X and Y be jointly continuous with PDF $f_{X,Y}$, and suppose that $g : \mathbb{R}^2 \to \mathbb{R}$ is a (Borel) measurable function such that $g(X)$ is integrable. Then,*

$$\mathbb{E}[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u, v) f_{X,Y}(u, v)\, du\, dv.$$

4.1 Multivariate measureable functions

A non-trivial assumption is for the joint PDF $f_{X,Y}$ to be a measurable function on $\mathbb{R}^2$. How can we ensure that? Of course simple functions are measurable, as are their limits, limsup's and liminf's. However, a criterion frequently used in practice is the following. (It is also an excellent exercise for getting some practice with product σ-algebras!)

Proposition 4. *Let $\phi(x, y)$ be a function on $\mathbb{R}^2$ such that*

1. *$y \mapsto \phi(x, y)$ is measurable for every fixed $x \in \mathbb{R}$*
2. *$x \mapsto \phi(x, y)$ is right-continuous for every fixed $y \in \mathbb{R}$*

Then ϕ is jointly measurable in (x, y).

Proof. First, it is instructive to understand the proof of Borel measurability of any right-continuous function $x \mapsto f(x)$. Let $R_a = \{x \in \mathbb{Q} : f(x) < a\}$. Then for any $a \in \mathbb{R}$ it follows

$$\{f(x) < a\} = \bigcup_{\epsilon_1 > 0} \bigcap_{\epsilon_2 > 0} \bigcup_{r \in R_{a-\epsilon_1}} [r - \epsilon_2, r] \,. \tag{4}$$

Here we write (abusing notation) $\cup_{\epsilon>0}$ to mean union over arbitrary sequence of $\epsilon_n \searrow 0$, so that resulting operations are countable. If (4) holds then $\{f(x) < a\}$ belongs to $\mathcal{B}$ and thus f is Borel.

To understand (4)[2] note that the set

$$L_b \triangleq \bigcap_{\epsilon_2 > 0} \bigcup_{r \in R_b} [r - \epsilon_2, r]$$

corresponds to all points on the real-line $\mathbb{R}$ that are decreasing limits of elements of R_b. For a right continuous function f

$$f(x) < b \Rightarrow x \in L_b \Rightarrow f(x) \le b$$

And thus (4) follows.

Now to prove Proposition, we only need to notice that sets $\{y : \phi(r, y) < a\}$ are measurable subsets of $\mathbb{R}$ by assumption. Hence, by setting

$$L_b \triangleq \bigcap_{\epsilon_2 > 0} \bigcup_{r \in \mathbb{Q}} [r - \epsilon_2, r] \times \{y : \phi(r, y) < b\}$$

[2] It may also be helpful to remember that right-continuity of $f : \mathbb{R} \to \mathbb{R}$ is equivalent to usual continuity when topology on the domain is refined declaring sets $[a, b)$ open.

we infer that

$$\phi(x,y) < b \Rightarrow (x,y) \in L_b \Rightarrow \phi(x,y) \leq b\,.$$

Thus, we have

$$\{(x,y) : \phi(x,y) < a\} = \bigcup_{\epsilon>0} L_{a-\epsilon}$$

which is a countable combination of measurable rectangles. □

5 INDEPENDENCE

Recall that two random variables, X and Y, are said to be independent if for any two Borel subsets, B_1 and B_2, of the real line, we have $\mathbb{P}(X \in B_1, Y \in B_2) = \mathbb{P}(X \in B_1)\mathbb{P}(Y \in B_2)$. This is equivalent to saying $\mathbb{P}_{X,Y} = \mathbb{P}_X \times \mathbb{P}_Y$, which explains why product of measures corresponds to independence.

Similar to the discrete case (cf. Proposition 1 and Theorem 1 in Section 3 of Lecture 5), simpler criteria for independence are available.

Theorem 2. *Let X and Y be jointly continuous random variables defined on the same probability space. The following are equivalent.*

(a) The random variables X and Y are independent.

(b) For any $x, y \in \mathbb{R}$, the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent.

(c) For any $x, y \in \mathbb{R}$, we have $F_{X,Y}(x,y) = F_X(x)F_Y(y)$.

(d) If f_X, f_Y, and $f_{X,Y}$ are corresponding marginal and joint densities, then $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, for all (x,y) except possibly on a set that has Lebesgue measure zero.

The proof parallels the proofs in Lecture 6, except for the last condition, for which the argument is simple when the densities are continuous functions (simply differentiate the CDF), but requires more care otherwise.

6 RADON-NIKODYM DERIVATIVE

In this section, we address a natural question: *Given a random variable X (or X, Y) how do we know if it is (jointly) continuous?*

Notice that Lebesgue measure plays a distinguished role in the definition of continuity. Thus a more general approach requires the following definition:

Definition 2. *Let $(\Omega, \mathcal{F}, \lambda)$ be a measure space. Let μ be another measure on $(\Omega, \mathcal{F})$. Then function $f : \Omega \to \mathbb{R}_+$ is called a Radon-Nikodym derivative $\frac{d\mu}{d\lambda}$ if*

$$\mu[E] = \int_E f \, d\lambda$$

for any $E \in \mathcal{F}$.

According to this definition: X is a continuous random variable if and only if there exists a Radon-Nikodym derivative $\frac{d\mathbb{P}_X}{d\mathrm{Leb}}$ on $\mathbb{R}$. Similarly, X and Y are jointly continuous if $\frac{d\mathbb{P}_{X,Y}}{d\mathrm{Leb}}$ exists on $\mathbb{R}^2$, etc. One simple consequence of (1) is that X cannot be a continuous random variable if $\mathbb{P}_X$ has *atoms*, namely if $\mathbb{P}[X = a] \neq 0$ for some $a \in \mathbb{R}$. However, as "singular" example in Section 4 shows the absence of atoms is not sufficient for continuity. The following definition and Theorem describe the necessary and sufficient condition:

Definition 3. *Measure μ is absolutely continuous with respect to λ (notation: $\mu \ll \lambda$), if for every E*

$$\lambda(E) = 0 \Rightarrow \mu(E) = 0 \,.$$

Note that from (1) we see: if X is continuous then $\mathbb{P}_X \ll \mathrm{Leb}$ and similarly for joint continuity. Remarkably, the converse holds as well:

Theorem 3 (Radon-Nikodym)**.** *Let μ and λ be σ-finite measures on $(\Omega, \mathcal{F})$. There exists a Radon-Nikodym derivative $\frac{d\mu}{d\lambda}$ if and only if $\mu \ll \lambda$.*

Proof of this theorem is outside of the scope of this class (mainly it relies on some basic properties of Hilbert spaces).