MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J Fall 2008
Lecture 8 10/1/2008

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 PDF of X. We then have, for any Borel set B,

$$\mathbb{P}(X \in B) = \int_B f_X(x)\, dx.$$

Technical remark: Since we have not yet defined the notion of an integral of a measurable function, the discussion in these notes will be rigorous only when we deal with integrals that can be interpreted in the usual sense of calculus,

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

namely, Riemann integrals. For now, let us just concentrate on functions that are piecewise continuous, with a finite number of discontinuities.

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 $f_X(x)$ is allowed to be greater than one for some x. Instead, the proper intuitive interpretation is the fact that if f_X is continuous over a small interval $[x, x+\delta]$, then

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

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

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

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 $x \ge 0$, 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 $x \ge 0$, and $f_X(x) = 0$, for $x < 0$. We denote this distribution by $\mathrm{Exp}(\lambda)$. The exponential distribution can be viewed as a "limit" of a geometric distribution. Indeed, if we fix some δ and consider the values of $F_X(k\delta)$, for $k = 1, 2, \ldots$, these values agree with the values of a geometric CDF. 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.

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 $x, t \ge 0$, we have $\mathbb{P}(X > x + t \mid X > x) = \mathbb{P}(X > t)$.*

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

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

$\square$

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}.

2.3 Normal distribution

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

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

It can be checked that this is a legitimate PDF, i.e., that it integrates to one. Note also that this PDF is symmetric around $x = \mu$. 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 (to be verified later) that if $X \sim N(\mu, \sigma^2)$, then $(X - \mu)/\sigma \sim N(0, 1)$. Thus,

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

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

2.4 Cauchy distribution

Here, $f_X(x) = 1/(\pi(1 + x^2))$, $x \in \mathbb{R}$. 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.

2.5 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 $\beta = c^\alpha$. 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 define

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

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

$$\begin{aligned}
\int_0^\infty (1 - F_X(t))\, dt &= \int_0^\infty \mathbb{P}(X > t)\, dt = \int_0^\infty \int_t^\infty f_X(x)\, dx\, dt \\
&= \int_0^\infty f_X(x) \int_0^x dt\, dx = \int_0^\infty x f_X(x)\, dx = \mathbb{E}[X].
\end{aligned}$$

(The interchange of the order of integration turns out to be justified becuase the integrand is nonnegative.) $\square$

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 such that $g(X)$ is integrable. Then,*

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

Proof: Let us express the function g as the difference of two nonnegative functions,

$$g(x) = g^+(x) - g^-(x),$$

where $g^+(x) = \max\{g(x), 0\}$, and $g^-(x) = \max\{-g(x), 0\}$. In particular, for any $t \geq 0$, we have $g(x) > t$ if and only if $g^+(x) > t$.

We will use the result

$$\mathbb{E}\big[g(X)\big] = \int_0^{\infty} \mathbb{P}\big(g(X) > t\big)\, dt - \int_0^{\infty} \mathbb{P}\big(g(X) < -t\big)\, dt$$

from Proposition 1. The first term in the right-hand side is equal to

$$\int_0^{\infty} \int_{\{x \mid g(x) > t\}} f_X(x)\, dx\, dt = \int_{-\infty}^{\infty} \int_{\{t \mid 0 \leq t < g(x)\}} f_X(x)\, dt\, dx = \int_{-\infty}^{\infty} g^+(x) f_X(x)\, dx.$$

By a symmetrical argument, the second term in the right-hand side is given by

$$\int_0^{\infty} \mathbb{P}\big(g(X) < -t\big)\, dt = \int_{-\infty}^{\infty} g^-(x) f_X(x)\, dx.$$

Combining the above equalities, we obtain

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

$\square$

Note that for this result to hold, the random variable $g(X)$ need not be continuous. The proof is similar to the one for Proposition 1, and involves an interchange of the order of integration; see [GS] for a proof for the special case where $g \geq 0$.

4 JOINT DISTRIBUTIONS

Given a pair of random variables X and Y, defined on the same probability space, we say that they are **jointly continuous** if there exists a measurable[2]

[2] Measurability here means that for every Borel subset of $\mathbb{R}$, the set $f^{-1}(B)$ is a Borel subset of $\mathbb{R}^2$. The Borel σ-field in $\mathbb{R}^2$ is the one generated by sets of the form $[a, b] \times [c, d]$.

$f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$ 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 the joint CDF of X and Y.

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

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

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

$$\mathbb{P}\big((X, Y) \in B\big) = \int_B f_{X,Y}(x, y)\, dx\, dy = \int_{\mathbb{R}^2} I_B(x, y) f_{X,Y}(x, y)\, dx\, dy.$$

Furthermore, if B has Lebesgue measure zero, then $\mathbb{P}(B) = 0$.[3]

We observe that

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

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.

[3]The Lebesgue measure on $\mathbb{R}^2$ is the unique measure μ defined on the Borel subsets of $\mathbb{R}^2$ that satisfies $\mu([a, b] \times [c, d]) = (b - a)(d - c)$, i.e., agrees with the elementary notion of "area" on rectangular sets. Existence and uniqueness of such a measure is obtained from the Extension Theorem, in a manner similar to the one used in our construction of the Lebesgue measure on $\mathbb{R}$.

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,*

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

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 $\mathbb{P}(X \in B_1, Y \in B_2) = \mathbb{P}(X \in B_1)\mathbb{P}(Y \in B_2)$.

Similar to the discrete case (cf. Proposition 1 and Theorem 1 in Section 3 of Lecture 6), 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.

6.436J / 15.085J Fundamentals of Probability
Fall 2008