EE5110: Probability Foundations for Electrical Engineers **July-November 2015**

Lecture 11: Random Variables: Types and CDF

Lecturer: Dr. Krishna Jagannathan *Scribe: Sudharsan, Gopal, Arjun B, Debayani*

In this lecture, we will focus on the types of random variables. Random variables are categorized into various types, depending on the nature of the measure $\mathbb{P}_X$ induced on the real line (or to be more precise, on the Borel σ-algebra). Indeed, there are three fundamentally different types of measures possible on the real line. According to an important theorem in measure theory, called the Lebesgue decomposition theorem (see Theorem 12.1.1 of [2]), any probability measure on $\mathbb{R}$ can be uniquely decomposed into a sum of these three types of measures. The three fundamental types of measure are

- Discrete,
- Continuous, and
- Singular.

In other words, there are three 'pure type' random variables, namely discrete random variables, continuous random variables, and singular random variables. It is also possible to 'mix and match' these three types to get four kinds of mixed random variables, altogether resulting in *seven* types of random variables.

Of the three fundamental types of random variables, only the discrete and continuous random variables are important for practical applications in the field of engineering and statistics. Singular random variables are largely of academic interest. Therefore, we will spend most of our effort in studying discrete and continuous random variables, although we will define and give an example of a singular random variable.

11.1 Discrete Random Variables

Definition 11.1 *Discrete Random Variable:*
A random variable X is said to be discrete if it takes values in a countable subset of $\mathbb{R}$ with probability 1.

Thus, there is a countable set $E = \{x_1, x_2, \dots\}$, such that $\mathbb{P}_X(E) = 1$. Note that the definition does not necessarily demand that the range of the random variable is countable. In particular, for a discrete random variable, there might exist some zero probability subset of the sample space, which can potentially map to an uncountable subset of $\mathbb{R}$. (Can you think of such an example?)

Definition 11.2 *Probability Mass Function (PMF):*
If X is a discrete random variable, the function $p_X : \mathbb{R} \to [0,1]$ defined by $p_X(x)$=$\mathbb{P}(X = x)$ for every x is called the probability mass function of X.

Although the PMF is defined for all $x \in \mathbb{R}$, it is clear from the definition that the PMF is non-zero only on the set E. Also, since $\mathbb{P}_X(E) = 1$, we must have (by countable additivity)

$$\sum_{i=1}^{\infty} \mathbb{P}(X = x_i) = 1.$$

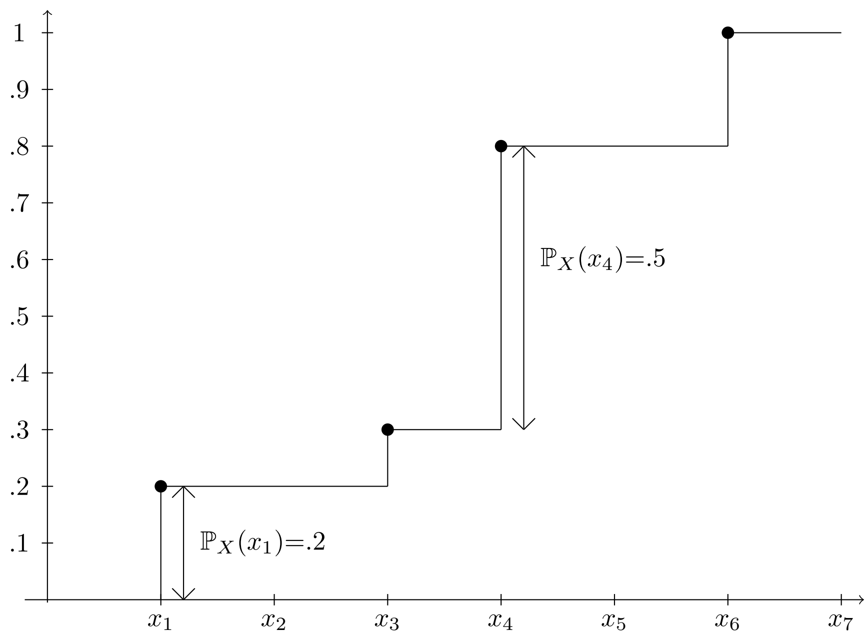

Figure 11.1: CDF of a discrete random variable

Interestingly, for a discrete random variable X, the PMF is enough to get a complete characterization of the probability law $\mathbb{P}_X$. Indeed, for any Borel set B, we can write

$$\mathbb{P}_X(B) = \sum_{i:\ x_i \in B} \mathbb{P}(X = x_i).$$

The CDF of a discrete random variable is given by

$$F_X(x) = \sum_{i: x_i \leq x} \mathbb{P}(X = x_i).$$

Figure 15.3 represents the Cumulative Distribution Function of a discrete random variable. One can observe that the CDF plotted in Figure 15.3 satisfies all the properties discussed earlier.

Next, we give some examples of some frequently encountered discrete random variables.

1. Indicator random variable: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $A \in \mathcal{F}$ be any event. Define

$$I_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases}$$

   It can be verified that I_A is indeed a random variable (since A and A^c are $\mathcal{F}$-measurable), and it is clearly discrete, since it takes only two values.

2. Bernoulli random variable: Let $p \in [0, 1]$, and define $p_X(0) = p$, and $p_X(1) = 1 - p$. This random variable can be used to model a single coin toss, where 0 denotes a head and 1 denotes a tail, and the probability of heads is p. The case $p = 1/2$ corresponds to a fair coin toss.

3. Discrete uniform random variable: Parameters are a and b where $a < b$.
   $p_X(m) = 1/(b - a + 1)$, $m = a$, $a + 1$,..b, and $p_X(m)$=0 otherwise.

4. Binomial random variable: $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$, where $n \in \mathbb{N}$ and $p \in [0,1]$. In the coin toss example, a binomial random variable can be used to model the number of heads observed in n independent tosses, where p is the probability of head appearing during each trial.

5. Geometric random variable: $p_X(k) = p(1-p)^{k-1}$, $k = 1, 2, \ldots$ and $0 < p \leq 1$. A geometric random variable with parameter p represents the number of (independent) tosses of a coin until heads is observed for the first time, where p represents the probability of heads during each toss.

6. Poisson: Fix the parameter $\lambda > 0$, and define $p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}$, where $k = 0, 1, \ldots$

Note that except for the indicator random variable, we have described only the PMFs of the random variables, rather than the explicit mapping from Ω.

11.2 Continuous Random Variables

11.2.1 Definitions

Let us begin with the definition of absolute continuity which will allow us to define continuous random variables formally. Let μ and ν be measures on $(\Omega, \mathcal{F})$.

Definition 11.3 *We say ν is* absolutely continuous *with respect to μ if for every $N \in \mathcal{F}$ such that $\mu(N) = 0$, we have $\nu(N) = 0$.*

Now, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ a random variable.

Definition 11.4 *X is said to be a continuous random variable if the law $\mathbb{P}_X$ is absolutely continuous with respect to the Lebesgue measure λ.*

Here, both $\mathbb{P}_X$ and λ are measures on $(\mathbb{R}, \mathcal{B})$. The above definition says that X is a continuous random variable if for any Borel set N set of Lebesgue measure zero, we have $\mathbb{P}_X(N) = \mathbb{P}(\omega | X(\omega) \in N) = 0$.

In particular, it is *not* the case that a random variable is continuous if it takes values in an uncountable set.

Next, we invoke without proof a special case of the *Radon-Nikodym Theorem* [3], which deals with absolutely continuous measures.

Theorem 11.5 *Suppose $\mathbb{P}_X$ is absolutely continuous with respect to λ, the Lebesgue measure, then there exists a non-negative, measurable function $f_X : \mathbb{R} \to [0, \infty)$, such that for any $B \in \mathcal{B}(\mathbb{R})$, we have*

$$\mathbb{P}_X(B) = \int_B f_X \, d\lambda. \tag{11.1}$$

The integral in the above theorem is not the usual Riemann integral, as B may be any Borel measurable set, such as the Cantor set, for example. We will get a precise understanding of the integral in (15.1) when we study abstract integration later in the course. For the time being, we can just think of the set B as an interval $[a, b]$, so (15.1) essentially says that the probability of X taking values in the interval $[a, b]$ can be written as $\int_a^b f_X dx$ for some non-negative measurable function f_X. Here, when we say f_X is measurable, we mean the pre-images of Borel sets are also Borel sets. In measure theoretic parlance, f_X is called the Radon-Nikodym derivative of $\mathbb{P}_X$ with respect to the Lebesgue measure λ.

In particular, taking $B = (-\infty, x]$, we can write the cumulative distribution function (CDF) as

$$F_X(x) \triangleq \mathbb{P}_X((-\infty, x]) = \int_{-\infty}^{x} f_X(y)\, dy. \tag{11.2}$$

Thus, we can understand f_X as the *probability density function (PDF)* of X, which is nothing but the Radon-Nikodym derivative of $\mathbb{P}_X$ with respect to the Lebesgue measure λ. Also,

$$\mathbb{P}_X(\mathbb{R}) = 1 = \int_{-\infty}^{\infty} f_X(y)\, dy.$$

Unlike the probability mass function in the case of a discrete random variable, the PDF has *no* interpretation as a probability; only integrals of the PDF can be interpreted as a probability.

The function f_X is unique only up to a set of Lebesgue measure zero, as we will understand later. We also remark that many authors (including [4]) define a random variable as being continuous if the CDF satisfies (15.2). This definition can be shown to be equivalent to the one we have given above.

11.2.2 Examples

The following are some common examples of continuous random variables:

1. Uniform: It is a scaled Lebesgue measure on a closed interval $[a, b]$.

   (a) PDF- $f_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \frac{1}{b-a} & \text{for} \quad a \leq x \leq b \\ 0 & \text{for} \quad x > b \end{cases}$

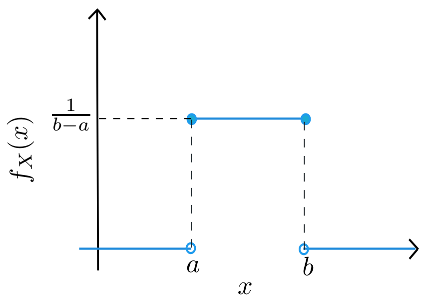

Figure 11.2: The PDF of a uniform random variable

   (b) CDF: $F_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \frac{x-a}{b-a} & \text{for} \quad a \leq x \leq b \\ 1 & \text{for} \quad x > b \end{cases}$

2. Exponential: It is a non-negative random variable, characterized by a single parameter $\lambda > 0$.

   (a) PDF: $f_X(x) = \lambda e^{-\lambda x} \quad \text{for} \quad x \geq 0$

   (b) CDF: $F_X(x) = 1 - e^{-\lambda x} \quad \text{for} \quad x \geq 0$

   (c) The exponential random variable posses an interesting property called the 'memoryless' property. We first give the definition of the memoryless property, and then show that the exponential random variable has this property.

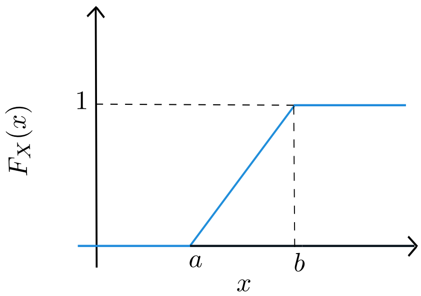

Figure 11.3: The CDF of a uniform random variable

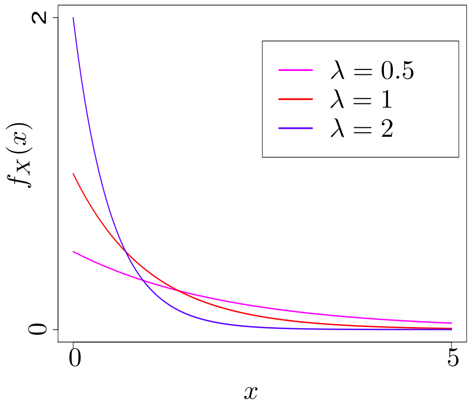

Figure 11.4: The PDF of an exponential random variable, for various values of the parameter λ

Definition 11.6 *A non-negative random variable X is said to be memoryless if* $\mathbb{P}(X > s+t|X > t) = \mathbb{P}(X > s) \quad \forall s, t \geq 0$.

For an exponential random variable,

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

Therefore, the exponential random variable is memoryless. For example, if the failure time of a light bulb is distributed exponentially, then the further time to failure, given that the bulb has not failed until time t, has the same distribution as the unconditional failure time of a new light bulb! Interestingly, it can also be shown that the exponential random variable is the *only* continuous random variable which possesses the memoryless property.

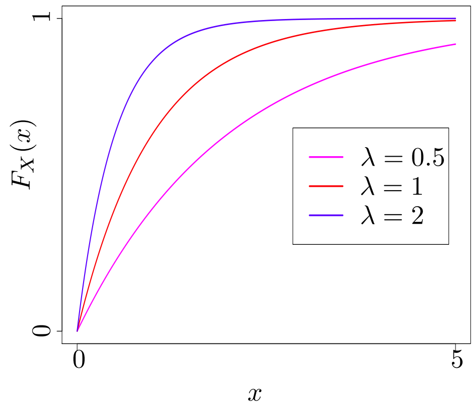

Figure 11.5: The CDF of an exponential random variable, for various values of the parameter λ

3. Gaussian (or Normal): This is a two parameter distribution, and as we shall interpret later, these parameters are the mean $\mu \in \mathbb{R}$ and standard deviation $\sigma > 0$. It has wide applications in engineering and statistics, owing to a 'stable-attractor" property of Gaussian random variables. We will study these properties later.

   (a) PDF: The probability density function of a Gaussian random variable is given by $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \text{for} \quad x \in \mathbb{R}$.
   The above distribution is denoted $N(\mu, \sigma^2)$. In particular, when $\mu = 0$ and $\sigma^2 = 1$, we get the *standard Gaussian* PDF: $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

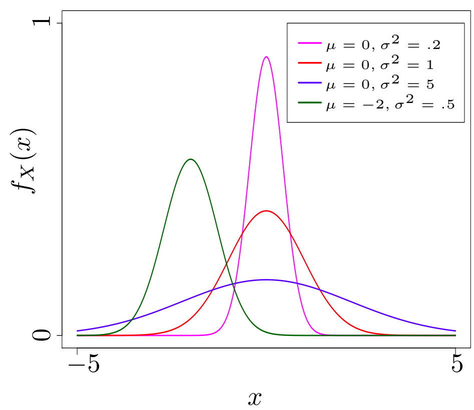

Figure 11.6: The PDF of a normal random variable, for various parameters

(b) CDF: There is no closed-form expression for the CDF of a Gaussian distribution (although the notion of a 'closed-form' is itself rather arbitrary, and over-rated!). For convenience, we call the CDF of the standard Gaussian the "error-function" $\text{Erf}(x) \triangleq \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \, dy$.

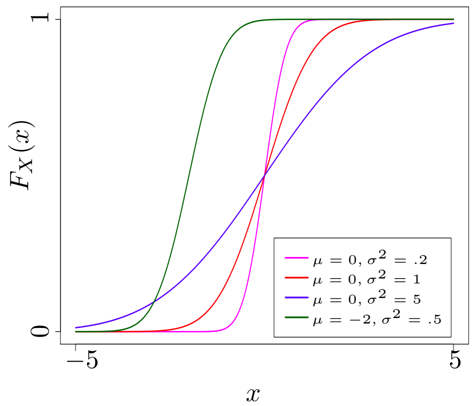

Figure 11.7: The CDF of a normal random variable, for various parameters

4. Cauchy: This is a two-parameter distribution parametrised by $x_0 \in \mathbb{R}$, the centering parameter, and $\gamma > 0$, the scale parameter. It is qualitatively very different from the previous distributions, because it is "heavy-tailed," i.e., its complementary CDF $1 - F_X(x)$ decays slower than any exponential. Heavy-tailed random variables tend to take very large values with non-negligible probability, and are used to model high variability and burstiness in engineering applications.

   (a) PDF: $f_X(x) = \frac{1}{\pi} \frac{\gamma}{(x-x_0)^2+\gamma^2}$.

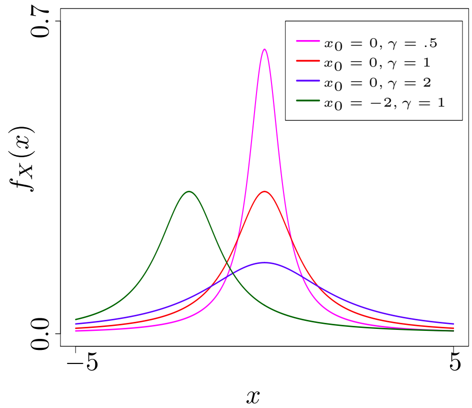

Figure 11.8: The PDF of a Cauchy random variable, for different parameters

11.3 Singular Random Variable

Singular random variables are rather bizzare, and in some sense, they occupy the 'middle-ground' between discrete and continuous random variables. In particular, singular random variables take values with probability one on an uncountable set of Lebesgue measure zero!

Definition 11.7 *A random variable X is said to be singular if, for every $x \in \mathbb{R}$, we have $\mathbb{P}_X(\{x\}) = 0$, and there exists a zero Lebesgue measure set $F \in \mathcal{B}(\mathbb{R})$, such that $\mathbb{P}_X(F) = 1$.*

Although it is not stated explicitly in the definition, it is clear that F must be an *uncountable* set of Lebesgue measure zero. (Why?)

Example A random variable having the Cantor distribution as its CDF is an example of a Singular random variable. The range of this random variable is the Cantor Set, C, which is a Borel set with Lebesgue measure zero. Further, if $x \in C$, then x has a ternary expansion of the following form

$$x = \sum_{i=1}^{\infty} \frac{x_i}{3^i}, \quad \text{where} \quad x_i \in \{0, 2\}. \tag{11.3}$$

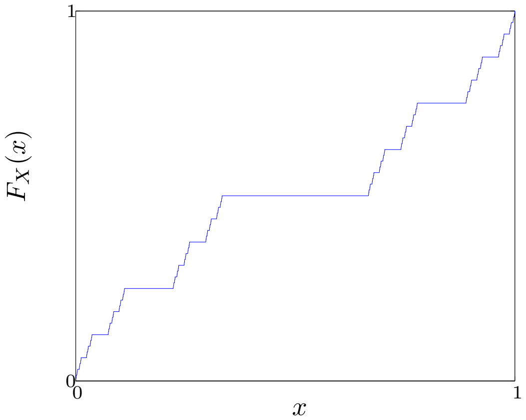

Figure 11.9: The Cantor Function

To look at a concrete example, consider an infinite sequence of independent tosses of a fair coin. When the outcome is a head, we record $x_i = 2$, otherwise, we record $x_i = 0$. Using these values of x_i we form a number x using (15.3). This results in a random variable X. This random variable satisfies the two properties that make it a Singular Random variable, namely $\mathbb{P}_X(C) = 1$, and $\mathbb{P}_X(\{x\}) = 0$, $\forall\ x \in [0, 1]$. The cumulative distribution function of this random variable, shown in Figure 15.9, is the *Cantor function* (which is sometimes referred to as the Devil's staircase). The Cantor function is continuous everywhere, since all singletons have zero probability under this distribution. Also, the derivative is zero wherever it exists, and the derivative does not exist at points in the Cantor set. The CDF only increases at these Cantor points, but does so without a well defined derivative, or any jump discontinuities for that matter!

11.4 Exercises:

1. (a) Prove Theorem 15.4.

(b) Verify that $\pi(\mathbb{R})$, defined in the lecture on Random Variables is indeed a π-system over $\mathbb{R}$.

(c) Prove Lemma 15.6.

(d) Plot the CDF of the indicator random variable.

2. For a random variable X, prove that $\mathbb{P}_X(\{y\}) = F_X(y) - \lim_{x \uparrow y} F_X(x)$. Hence show that F_X is continuous at y if and only if $\mathbb{P}_X(\{y\}) = 0$.

3. Among the functions given below, find the functions that are valid CDFs and find their respective densities. For those that are not valid CDFs, explain what fails.

(a)

$$F(x) = \begin{cases} 1 - e^{-x^2} & x \geq 0 \\ 0 & x < 0. \end{cases} \tag{11.4}$$

(b)

$$F(x) = \begin{cases} e^{-\frac{1}{x}} & x > 0 \\ 0 & x \leq 0. \end{cases} \tag{11.5}$$

(c)

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{3} & 0 < x \leq \frac{1}{2} \\ 1 & x > \frac{1}{2}. \end{cases} \tag{11.6}$$

4. **Negative Binomial Random Variable.** Consider a sequence of independent Bernoulli trials $\{X_i\}_{i \in \mathbb{N}}$ with parameter of success $p \in (0, 1]$. The number of successes in first n trials is given by

$$Y_n = \textstyle\sum_{i=1}^{n} X_i.$$

Y_n is distributed as Binomial with parameters n and p.
Consider the random variable defined by

$$V_k = \min\{n \in \mathbb{N}_+ : Y_n = k\}.$$

Note that V_1 is distributed as Geometric with parameter p.

(a) Give a verbal description of the random variable V_k.

(b) Show that the probability mass function of the random variable V_k is given by

$$\mathbb{P}(V_k = n) = (\tfrac{n-1}{k-1}) p^k (1-p)^{(n-k)}$$

where $n \in \{k, k+1, ...\}$. This is known as Negative Binomial Distribution with parameters k and p.

(c) Argue that Binomial and Negative Binomial Distributions are inverse to each other in the sense that

$$Y_n \geq k \Leftrightarrow V_k \leq n.$$

5. **Radioactive decay.** Assume that a radioactive sample emits a random number of α particles in any given hour, and that the number of α particles emitted in an hour is Poisson distributed with parameter λ. Suppose that a faulty Geiger-Muller counter is used to count these particle emissions. In particular, the faulty counter fails to register an emission with probability p, independently of other emissions.

(a) What is the probability that the faulty counter will register exactly k emissions in an hour?

(b) Given that the faulty counter registered k emissions in an hour, what is the PMF of the actual number of emissions that happened from the source during that hour?

6. Buses arrive at ten minute intervals starting at noon. A man arrives at the bus stop at a random time X minutes after noon, where X has the CDF:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{60} & 0 \leq x \leq 60 \\ 1 & x > 60. \end{cases} \tag{11.7}$$

What is the probability that he waits less than five minutes for a bus?

7. Find the values of a and b such that the following function is a valid CDF:

$$F(x) = \begin{cases} 1 - ae^{-x/b} & x \geq 0 \\ 0 & x < 0. \end{cases} \tag{11.8}$$

Also, find the values of a and b such that the function above corresponds to the CDF of some

(a) Continuous Random Variable

(b) Discrete Random Variable

(c) Mixed type Random Variable

8. Let X be a continuous random variable. Show that X is memoryless iff X is an exponential random variable.

References

[1] David Williams, “Probability with Martingales”, *Cambridge University Press*, Fourteenth Printing, 2011.

[2] Paul Halmos, “Measure Theory”, *Springer-Verlog*, Second Edition, 1978.

[3] David Gamarnick and John Tsitsiklis, “Introduction to Probability”, *MIT OCW*, , 2008.

[4] Geoffrey Grimmett and David Stirzaker, “Probability and Random Processes”, *Oxford University Press*, Third Edition, 2001.