Weekly Notes for EE2012 2014/15 – Week 7

T. J. Lim

March 5, 2015

Book sections covered this week: 4.1 – 4.2.1.

1 Cumulative Distribution Function (CDF) (cont'd)

1.1 Types of Random Variables

The classification of a r.v. into one of the three categories of random variables is based on their CDF, as follows.

- *Discrete* – We studied these r.v.s extensively in the last couple of weeks. A discrete r.v. has a range S_X with a countable number of values, and a CDF that is a staircase function with the height of the k-th step equal to $p_X(x_k)$:

$$F_X(x) = \sum_k p_X(x_k) u(x - x_k), \quad x_k \in S_X. \tag{1}$$

- *Continuous* – A continuous r.v. has a range S_X that is uncountable, and a CDF that is continuous everywhere, i.e.

$$\lim_{\delta x \to 0} F_X(x + \delta x) = \lim_{\delta x \to 0} F_X(x - \delta x) = F_X(x), \quad \forall\, x \in \mathbb{R}. \tag{2}$$

  We will study these types of random variables closely in the coming weeks.

- *Mixed* – A mixed-type random variable has a CDF with jumps at a countable number of points within S_X, and increases smoothly (i.e. with a non-zero gradient) over some range within S_X. Such a CDF can be written as

$$F_X(x) = pF_d(x) + (1 - p)F_c(x) \tag{3}$$

  for some $p \in (0, 1)$, where $F_d(x)$ is the CDF of a discrete r.v., and $F_c(x)$ is the CDF of a continuous r.v.

2 Probability Density Function (PDF)

While the CDF contains a full description of any random variable (continuous, discrete or mixed) in the sense that the probability of any event involving that random variable can be computed using $F_X(x)$, it is not a very convenient visualization tool. This is due to its "cumulative" nature, which requires us to subtract $F_X(a)$ from $F_X(b)$ to find the probability of $\{a < X \leq b\}$. We would prefer to work with a function that more immediately captures the probability of X lying in a certain range.

2.1 Definition

The probability density function or PDF is defined as the derivative of the CDF:

$$f_X(x) = \frac{dF_X(x)}{dx}. \tag{4}$$

By the fundamental theorem of calculus[1], we then have

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt. \tag{5}$$

In other words, $P[X \leq x]$ is the area under the PDF $f_X(t)$ from $t = -\infty$ to $t = x$.

Using the definition of a derivative on (4), we have

$$\begin{aligned} f_X(x) &= \lim_{\delta x \to 0} \frac{F_X(x+\delta x) - F_X(x)}{\delta x} && (6) \\ \Rightarrow \quad \lim_{\delta x \to 0} f_X(x)\delta x &= \lim_{\delta x \to 0} F_X(x+\delta x) - F_X(x) && (7) \\ &= \lim_{\delta x \to 0} P[x < X \leq x + \delta x]. && (8) \end{aligned}$$

In other words, the probability of X lying within the infinitesimal range $(x, x+\delta]$ is the area under $f_X(x)$ within that range.

By splitting any contiguous range of values $(a, b]$ into mutually exclusive strips of width δx, i.e.

$$(a, b] = \bigcup_{k=0}^{N} (x_k, x_{k+1}]$$

with $x_0 = a$ and $x_N = b$, we can use Axiom III to conclude that

$$\begin{aligned} P[a < X \leq b] &= \sum_{k=0}^{N} P[x_k < X \leq x_{k+1}] && (9) \\ &= \sum_{k=0}^{N} f_X(x_k)\delta x && (10) \\ &\overset{\delta x \to 0}{=} \int_a^b f_X(x)dx. && (11) \end{aligned}$$

[1] See http://en.wikipedia.org/wiki/Fundamental_theorem_of_calculus.

From the above discussion, we deduce the following:

1. Unlike the CDF, the PDF is *not* a probability value. It is a density just like linear mass density (measured in kg/m), and is translated into a probability value only after integration over the region of interest.

2. $f_X(x)$ gives us a measure of the probability of X lying in a neighbourhood around x. If $f_X(x_0) > f_X(x_1)$, it does not mean that $P[X = x_0] > P[X = x_1]$ since both of those probabilities may be zero. But it does mean that X is more likely to be in a neighbourhood of width δx around x_0 than it is to be in such a neighbourhood around x_1. Therefore, the shape of $f_X(x)$ gives us an immediate visual sense of what values X is more likely to assume.

Example 1: Consider the CDF

$$F_X(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 < x \le 1 \\ 1, & x > 1 \end{cases} . \tag{12}$$

X is in fact a random variable that is equally likely to be in $(0.1, 0.1 + \Delta)$ as it is to be in $(0.5, 0.5 + \Delta)$, where $\Delta > 0$ is small. But this is not very easy to deduce from the CDF for probability novices. You can imagine that for more complex distributions, the CDF would offer little immediate insight even for experts.

Let's now look at the PDF of X, which can be easily derived as

$$f_X(x) = 1, \quad 0 < x \le 1, \tag{13}$$

and $f_X(x) = 0$ elsewhere[2].

A sketch of $f_X(x)$ shows that it is flat from 0 to 1, corresponding nicely to the fact that X is equally likely to be in any Δ-wide range within $[0, 1]$. ■

2.2 Properties and Uses of the PDF

Equation (11) gives us a convenient visualization of the main use of the PDF, i.e. that the probability of X lying in the range $(a, b]$ is the *area under the PDF* between $x = a$ and $x = b$. For some cases, this geometrical interpretation allows us to avoid computing integrals.

Example 2: Consider the PDF

$$f_X(x) = \begin{cases} x, & 0 < x \le 1 \\ 2 - x, & 1 < x \le 2 \\ 0, & \text{elsewhere} \end{cases} . \tag{14}$$

[2]When $F_X(x)$ is continuous everywhere, we don't really care whether we include or exclude the boundary values, in this case 0 and 1, because the value of an integral of a function without impulses is unchanged when we include or exclude a finite number of isolated values.

A sketch of the PDF[3] reveals that the PDF is a triangle with vertices at (0,0), (1,1) and (2,0). Therefore, instead of integrating $f_X(x)$ between the desired limits, we can just use geometry to compute the following probabilities:

$$\begin{aligned} P[0 < X \leq 0.5] &= \frac{1}{2} \times 0.5 \times 0.5 = \frac{1}{8}; \\ P[0.5 < X \leq 1.5] &= 2 \times \left(\frac{1}{2}(0.5 + 1)(0.5) \right) = 0.75. \quad \blacksquare \end{aligned}$$

From (7), for any $\delta x > 0$, $f_X(x)\delta x$ is a probability value which must be non-negative, and hence $f_X(x)$ must be *non-negative* for all x. And finally, by setting a to $-\infty$ and b to ∞ in (11) and noting that the event $\{-\infty < X \leq \infty\}$ contains the entire sample space, we have

$$\int_{-\infty}^{\infty} f_X(x) dx = 1. \tag{15}$$

Example 3: Find the normalization constant in the functions below that can be considered valid PDFs. For those that cannot, explain why not.

1. $f_1(x) = ce^{-x}$, $x \in \mathbb{R}$.
2. $f_2(x) = cx(2 - x)$, $0 < x \leq 2$.
3. $f_3(x) = 1 - cx$, $0 < x \leq 1/c$.

The first function $f_1(x)$ cannot be a PDF because over its domain $\mathbb{R}$, its area is infinite for all $c > 0$; it will be uniformly zero everywhere if $c = 0$; and it will be negative everywhere if $c < 0$. Thus no value of c will make $f_1(x)$ satisfy all the properties of a PDF.

A quick sketch reveals that the function $f_2(x) > 0$ for all $x \in (0, 2]$ when $c > 0$, and it has a finite area equal to

$$A = c \int_0^2 2x - x^2 dx = c \left[x^2 - \frac{x^3}{3} \right]_0^2 = \frac{4c}{3}. \tag{16}$$

To make $f_2(x)$ a valid PDF, we choose c so that $A = 1$, i.e. $c = \frac{3}{4}$.

Sketching $f_3(x)$ shows us that $f_3(x) > 0$ over its domain. The region enclosed by $f_3(x)$ and the x-axis in $(0, 1/c]$ is a triangle with area

$$B = \frac{1}{2c}. \tag{17}$$

Setting $B = 1$ to obtain a valid PDF, we have $c = 0.5$. $\blacksquare$

[3] We always start with a sketch of the PDF when given a problem.

2.3 PDF of Discrete Random Variables

Recall that the CDF of a discrete random variable is a staircase function:

$$F_X(x) = \sum_{k=1}^{n} p_X(x_k) u(x - x_k) \tag{18}$$

where $S_X = \{x_1, \ldots, x_n\}$ and $u(x)$ is the unit step function. The PDF of a discrete r.v. would then be

$$f_X(x) = \sum_{k=1}^{n} p_X(x_k) \frac{d}{dx} u(x - x_k). \tag{19}$$

Given that $u(x)$ has a jump at $x = 0$, it would appear that it is a non-differentiable function. However, we can define the *Dirac delta* or impulse function $\delta(x)$ to circumvent the problem:

$$\delta(x) = \frac{d}{dx} u(x), \quad u(x) = \int_{-\infty}^{x} \delta(t) dt. \tag{20}$$

Since $u(x)$ is flat everywhere except at $x = 0$, it is clear from the definition that $\delta(x) = 0$ for all $x \neq 0$. In addition, since the derivative of $u(x)$ at $x = 0$ is undefined, we will have an undefined value for $\delta(0)$. But due to the integral part of the definition above, we have that

$$\int_{-\infty}^{0-\epsilon} \delta(t) dt = 0, \quad \int_{-\infty}^{0+\epsilon} \delta(t) dt = 1 \tag{21}$$

for any $\epsilon > 0$, including very tiny values of ϵ. Therefore,

$$\int_{-\epsilon}^{+\epsilon} \delta(t) dt = 1, \quad \forall \, \epsilon > 0. \tag{22}$$

So the area under the impulse function over the infinitesimal interval $(-\epsilon, \epsilon)$ is one!

Now what happens when we multiply $\delta(t)$ by a function $g(t)$ which is continuous at $t = 0$? Since $\delta(t) = 0$ for all $t \neq 0$, we must have

$$g(t)\delta(t) = 0, \quad t \neq 0, \tag{23}$$

and hence for any $\epsilon > 0$,

$$\int_{-\infty}^{\infty} g(t)\delta(t) dt = \int_{-\epsilon}^{\epsilon} g(t)\delta(t) dt. \tag{24}$$

By taking the limit $\epsilon \to 0$, and using the continuity of $g(t)$ at $t = 0$, we have

$$\int_{-\infty}^{\infty} g(t)\delta(t) dt = g(0) \int_{-\epsilon}^{\epsilon} \delta(t) dt = g(0). \tag{25}$$

Repeating the above steps for $g(t-t_0)$ rather than $g(t)$, we have the **sifting property** of the impulse function:

$$\int_{-\infty}^{\infty} g(t)\delta(t-t_0)dt = g(t_0) \tag{26}$$

as long as $g(t)$ is continuous at $t = t_0$.

Having defined the impulse function $\delta(t)$, we can now go back to (19) and obtain

$$f_X(x) = \sum_{k=1}^{n} p_X(x_k)\delta(x-x_k) \tag{27}$$

as the PDF of a discrete r.v. with PMF $p_X(x_k)$, $k = 1, \ldots, n$. In other words, the PDF of a discrete r.v. is a train of impulses, each with a strength of $p_X(x_k)$.

Example 4: Let X have the PMF

$$p_X(0) = p_X(2) = 0.2, \;\; p_X(1) = p_X(3) = 0.3. \tag{28}$$

Then the CDF of X is

$$F_X(x) = 0.2u(x) + 0.3u(x-1) + 0.2u(x-2) + 0.3u(x-3), \tag{29}$$

and its PDF is

$$f_X(x) = 0.2\delta(x) + 0.3\delta(x-1) + 0.2\delta(x-2) + 0.3\delta(x-3). \tag{30}$$

To find for instance $P[0.5 < X \leq 2.1]$, we can use the PMF, CDF or PDF:

$$\begin{aligned} P[0.5 < X \leq 2.1] &= P[X=1] + P[X=2] = p_X(1) + p_X(2) = 0.5; && (31)\\ &= F_X(2.1) - F_X(0.5) = 0.7 - 0.2 = 0.5; && (32)\\ &= \int_{0.5}^{2.1} f_X(x)dx = 0.3 + 0.2 = 0.5. && (33) \end{aligned}$$

When dealing with discrete random variables, it is common practice to use the PMF. However in some problems, we have a mixture of discrete and continuous random variables, and to unify their description we may want to use the PDF of a discrete r.v. ■

2.4 Mixed Random Variables

A random variable of mixed type is one with jumps at a countable number of points in its CDF, as well as strictly increasing segments. In this case, the PDF will be a mix of impulse functions and conventional functions.

Example 5: In a small variation of the taxi waiting time example encountered earlier, the waiting time X has the PDF

$$f_X(x) = 0.5\delta(x) + x[u(x) - u(x-1)]. \tag{34}$$

A sketch of the PDF is (as always) helpful for visualization. From the sketch we can obtain

$$P[X \le 0.2] = \int_{-\infty}^{0.2} f_X(x)dx = 0.5 + \frac{1}{2}(0.2^2) = 0.52. \tag{35}$$

The CDF can also be obtained with careful consideration of the critical sections of the real number line. (i) For $x < 0$, it should be clear that $F_X(x) = 0$. (ii) At $x = 0$, due to the impulse with a strength of 0.5, we have $F_X(0) = 0.5$. (iii) When $x \in (0, 1]$, we have

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{0^+} f_X(t)dt + \int_{0^+}^{x} f_X(t)dt & (36)\\ &= 0.5 + \left.\frac{t^2}{2}\right|_0^x & (37)\\ &= 0.5(1 + x^2). & (38) \end{aligned}$$

(iv) When $x > 1$, we have $F_X(x) = 1$. Therefore, we can write

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 0.5(1 + x^2) & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}. \tag{39}$$

3 Conditional CDF and PDF

Recall that for discrete random variables, we can define a conditional PMF $p_X(k|C)$ where C is an event such as $\{X > 5\}$ involving X, or one that does not involve X e.g. "there are no clouds in the sky". Similarly, for continuous and other types of random variables, the conditional distribution is given by the conditional CDF and conditional PDF, defined as

$$\begin{aligned} F_X(x|C) &= P[X \le x|C] = \frac{P[\{X \le x\} \cap C]}{P[C]} & (40)\\ f_X(x|C) &= \frac{d}{dx} F_X(x|C). & (41) \end{aligned}$$

The **theorem on total probability** applies directly to the CDF since $F_X(x)$ is a probability value, i.e., given a partition $\{B_1, \ldots, B_n\}$ of the sample space S, we have

$$F_X(x) = \sum_{i=1}^{n} F_X(x|B_i)P(B_i) \tag{42}$$

and by differentiating both sides, we get

$$f_X(x) = \sum_{i=1}^{n} f_X(x|B_i)P(B_i). \tag{43}$$

Example 6: Let T be the waiting time for a webpage to be downloaded. When the webpage is hosted locally, the PDF of T is e^{-t}, $t > 0$; when the webpage is hosted in another country, the PDF of T is $0.2e^{-t/5}$, $t > 0$. The probability of visiting a locally-hosted website is p. We can find the PDF of T using (43) as

$$\begin{aligned} f_T(t) &= f_T(t|\text{local})P[\text{local}] + f_T(t|\text{foreign})P[\text{foreign}] && (44) \\ &= pe^{-t} + 0.2(1-p)e^{-t/5}, \quad t > 0. && (45) \end{aligned}$$

The PDF $f_T(t)$ represents the distribution of T values overall. For instance, if T is measured over a large number of page accesses by someone who visits local and foreign websites in the frequency ratio $p : 1-p$, then the probability of $T \in A$ is given by $\int_A f_T(t)dt$. If we only measure T when he visits local websites, then the probability of $T \in A$ will become $\int_A f_T(t|\text{local})dt$. ■

Example 7: Suppose the event being conditioned on involves X, e.g. $C = \{X > t\}$. Then the conditional CDF and PDF take on particularly simple and intuitive forms.

$$\begin{aligned} F_X(x|C) &= \frac{P[\{X \le x\} \cap \{X > t\}]}{P[X > t]} \\ &= \begin{cases} \frac{F_X(x) - F_X(t)}{1 - F_X(t)} & x > t \\ 0 & x \le t. \end{cases} && (46) \end{aligned}$$

By differentiating the above expression (remembering that $F_X(t)$ is a constant with respect to x), we get the conditional PDF as

$$f_X(x|X > t) = \begin{cases} \frac{f_X(x)}{1 - F_X(t)} & x > t \\ 0 & x \le t. \end{cases} \tag{47}$$

In other words, after being told that $X > t$, we set the PDF in the range $x \le t$ to zero since there is no longer any possibility of X being smaller than t. For $x > t$, we keep the same PDF shape but normalize it so that the resulting function has unit area. ■

The result of this last example can be generalized to the case of conditioning on any event involving X, i.e. let $C = \{X \in A\}$, where $A \subset \mathbb{R}$, then

$$f_X(x|C) = \begin{cases} \frac{f_X}{P[C]} & x \in A \\ 0 & x \notin A \end{cases} \tag{48}$$

4 Diagnostic Questions

1. A random variable X has the CDF

$$F_X(x) = \begin{cases} \beta \sin(\pi x) & 0 < x \le 0.5 \\ 0 & \text{elsewhere} \end{cases} \tag{49}$$

(i) Find β. (ii) What is the PDF of X?

2. A random variable Y has the PDF

$$f_Y(y) = 0.4 f_1(y) + 0.6 f_2(y)$$

where $f_1(y) = 0.2 \sum_{k=1}^{\infty} 0.8^{k-1} \delta(y-k)$ and $f_2(y) = 0.1[u(y) - u(y-10)]$. Find $P[Y \leq 2]$.

3. An experiment consists of picking a ball out of a bag of 3 black balls and 2 white balls. If a black ball is picked, X is generated using the PDF $f_1(x) = u(x) - u(x-1)$; if a white ball is picked, X is equally likely to be -1 or $+1$. Find the PDF and CDF of X.