CDF and PDF

1. A random variable X has the PDF
$$f_X(t) = \frac{1}{4}e^{-t/4}, \quad t > 0.$$
   (a) Find the CDF of X.
   Ans: By definition, we have
$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(t)dt && (1)\\ &= \int_0^x \frac{1}{4}e^{-t/4}dt, \quad x \geq 0 && (2)\\ &= e^{-t/4}\Big|_x^0 && (3)\\ &= 1 - e^{-x/4}, \quad x \geq 0. && (4)\end{aligned}$$
   (b) Using the CDF, find $P[1 < X \leq 2]$ and $P[X > 4]$.
   Ans: $P[1 < X \leq 2] = F_X(2) - F_X(1) = e^{-1/4} - e^{-1/2}$. $P[X > 4] = 1 - F_X(4) = e^{-1}$.
   (c) X is a continuous random variable. Can you think of a function $g(x)$ that will make $Y = g(X)$ a discrete random variable?
   Ans: Any type of quantization operation will do. For instance let $g(x) = 1$ if $x > 2$, and $g(x) = -1$ if $x \leq 2$. Then $Y = g(X)$ can take only two values, ± 1, with $P[Y = -1] = F_X(2)$ and $P[Y = 1] = 1 - F_X(2)$.

2. In each of the following functions, find the unknown constant that will make it a valid PDF.
   (a) $f_1(x) = 2ce^{-(x-1)/4}$, $x > 0$.
   Ans: Solving for c in $\int_0^\infty 2ce^{-(x-1)/4}dx = 1$ yields $c = \frac{1}{8}\exp(-1/4)$.
   (b) $f_2(x) = \alpha(1-x)(x-3)$, $1 \leq x \leq 3$.
   Ans: Similarly, $\int_1^3 \alpha(1-x)(x-3)dx = 1 \Rightarrow \alpha = 3/4$.
   (c) $f_3(x) = \beta\cos(x)[u(x + \frac{\pi}{2}) - u(x - \frac{\pi}{2})]$, where $u(x)$ is the unit step function.
   Ans: Finally, $\int_{-\pi/2}^{\pi/2} \beta \cos x dx = 1$ results in $\beta = 1/2$.

3. Explain why the following functions are not valid CDFs:
   (a)
$$F_1(x) = \begin{cases} 0 & x \leq 0 \\ 0.4 & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

Ans: The function is not continuous from the right at $x = 0$ because for any $\epsilon \in (0, 1)$, $F_1(\epsilon) = 0.4 \neq 0 = F_1(0)$.

(b) $F_2(x) = 1 - e^{-x}$.

Ans: The function has a negative value in the range $x < 0$.

(c)

$$F_3(x) = \begin{cases} x & 0 < x \leq 1 \\ 1 - x & 1 < x \leq 2 \end{cases}$$

Ans: The function is decreasing in the range $1 < x \leq 2$.

4. The PDF of a random variable X is given by

$$f_X(x) = 0.2p \sum_{k=1}^{\infty} 0.8^{k-1} \delta(x - k) + (1 - p)e^{-x},$$

in the range $x > 0$.

(a) If $p = 0.5$, sketch the PDF. What type of random variable is this?

Ans: X is a mixed random variable. Its PDF consists of both impulse functions and continuous functions.

(b) For an arbitrary $p \in (0, 1)$, find $P[X > 4.5]$.

Ans: We need $\int_{4.5}^{\infty} f_X(x)$, which is evaluated as follows:

$$\begin{aligned} P[X > 4.5] &= 0.2p \sum_{k=5}^{\infty} 0.8^{k-1} + (1 - p) \int_{4.5}^{\infty} e^{-x} dx && (5) \\ &= \frac{0.2p(0.8^4)}{1 - 0.8} + (1 - p)e^{-4.5} && (6) \\ &= 0.4096p + (1 - p)e^{-4.5}. && (7) \end{aligned}$$

(c) Hence find $f_X(x|X > 4.5)$.

Ans: One way to solve this problem is through the conditional CDF –

$$\begin{aligned} F_X(x|X > 4.5) &= \frac{P[X \leq x, X > 4.5]}{P[X > 4.5]} && (8) \\ &= \alpha \int_{4.5}^{x} f_X(t)dt, \quad x > 4.5 && (9) \\ &= \alpha \left[0.2p \sum_{k=5}^{\lfloor x \rfloor} 0.8^{k-1} + (1 - p) \int_{4.5}^{x} e^{-t} dt \right] && (10) \end{aligned}$$

where $\alpha = 1/P[X > 4.5]$. The first term on the RHS is a staircase function, with the first step at $x = 5$ having magnitude $0.2p(0.8^4) = 0.082p$, the second at $x = 6$ having magnitude $0.2p(0.8^5) = 0.066p$, etc. The second term on the RHS evaluates to

$$(1-p)\int_{4.5}^{x} e^{-t}dt \quad = \quad (1-p)[e^{-4.5} - e^{-x}]. \tag{11}$$

By differentiating both sides and remembering that $du(t)/dt = \delta(t)$ we have

$$f_X(x|X > 4.5) = \alpha \left[0.2p\sum_{k=5}^{\infty} 0.8^{k-1}\delta(x-k) + (1-p)e^{-x}\right], \quad x > 4.5.$$

(d) How would you generate samples of X on a computer, e.g. using Matlab or Scilab?

Ans: First generate a uniform random variable in $[0, 1]$. If it is smaller than p, let X be drawn from a geometric distribution with $p = 0.2$; if it is not smaller than p, let X be drawn from an exponential distribution with $\lambda = 1$.

5. Find $P[X > 2]$ if X has the following distributions (if there is an unknown constant in the PDF or CDF, you first have to find its value):

(a) $f_X(x) = \alpha \sum_{k=0}^{4} k\delta(x-k)$.

Ans: This is a discrete random variable, with PMF $p_X(k) = \alpha k$, $k = 0, 1, 2, 3, 4$. Therefore, $\alpha(1 + 2 + 3 + 4) = 1$ or $\alpha = 0.1$, and $P[X > 2] = P[X = 3] + P[X = 4] = 0.7$.

(b) $F_X(x) = 1 - e^{-(x-1)}$, $x > 1$.

Ans: $P[X \le 2] = F_X(2) = 1 - e^{-1}$ and thus $P[X > 2] = e^{-1}$.

(c) $f_X(x) = ce^{-|x|/2}$.

Ans: We first find c as follows.

$$\int_{-\infty}^{0} ce^{x/2}dx + \int_{0}^{\infty} ce^{-x/2}dx \quad = \quad 1 \tag{12}$$

$$2c\int_{0}^{\infty} e^{-x/2}dx \quad = \quad 1 \quad \text{(symmetry)} \tag{13}$$

$$c \quad = \quad \frac{1}{4}. \tag{14}$$

Now we can evaluate

$$P[X > 2] \quad = \quad \int_{2}^{\infty} \frac{1}{4}e^{-x/2}dx \tag{15}$$

$$= \quad \frac{e^{-1}}{2} \tag{16}$$

(d) $F_X(x) = \begin{cases} 0 & x < 0 \\ 2x & 0 \le x < 0.5 \\ 1 & x \ge 1 \end{cases}$

Ans: Since $P[X \le 2] = F_X(2) = 1$, we have $P[X > 2] = 0$.

6. Consider the PDF $f_X(x) = cx(1-x)$, $0 < x < 1$, where c is a constant.

   (a) Evaluate c.
   Ans: Since $\int_{-\infty}^{\infty} f_X(x)dx = 1$, we have

$$\begin{aligned} c\int_0^1 x - x^2 dx &= 1 && (17)\\ c\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1 &= 1 && (18)\\ c &= 6. && (19) \end{aligned}$$

   (b) Find the CDF $F_X(x)$.
   Ans: The CDF is

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(t)dt && (20)\\ &= \begin{cases} 0 & x < 0 \\ \int_0^x 6(t-t^2)dt & 0 \le x \le 1 \\ 1 & x > 1 \end{cases} && (21)\\ &= \begin{cases} 0 & x < 0 \\ 6\left(\frac{x^2}{2} - \frac{x^3}{3}\right) & 0 \le x \le 1 \\ 1 & x > 1 \end{cases} && (22) \end{aligned}$$

   (c) Sketch both the PDF and the CDF.

7. Let X be a mixed-type random variable with PDF

$$f_X(x) = 0.25\delta(x+1) + 0.25\delta(x-1) + \beta e^{-2|x|}[u(x+1) - u(x-1)].$$

   (a) Sketch the PDF and find β.
   Ans: Two impulses of strength 0.25 at $x = -1$ and $x = 1$ respectively, with a symmetric exponential function in between. β needs to satisfy

$$\begin{aligned} \beta\int_{-1}^{1} e^{-2|x|}dx &= 0.5 && (23)\\ 2\beta\int_0^1 e^{-2x}dx &= 0.5 && (24)\\ \beta(1-e^{-2}) &= 0.25 && (25)\\ \beta &= \frac{1}{4(1-e^{-2})} && (26) \end{aligned}$$

   (b) Give an example of a physical experiment that may result in such a random variable.

Ans: The signal X is clipped when it falls outside the range $(-1, 1)$ whereas when it is inside that range, it has a Laplacian distribution. This corresponds to a signal with a Laplacian distribution being passed through a clipper with the input-output transfer function

$$y = \begin{cases} -1 & x < -1 \\ x & -1 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

For instance, a voice signal has a Laplacian distribution and when passed through the clipper, it will have the given distribution.

Conditional Distributions

1. An exponential random variable has the PDF

$$f_X(x) = \lambda e^{-\lambda x}, \quad x > 0,$$

where $\lambda > 0$ is a parameter of the distribution.

(a) Prove the "memory-less property" of the exponential distribution, i.e. for any $t, h > 0$,

$$P[X > t + h | X > h] = P[X > t]$$

Ans: From first principles,

$$\begin{aligned} P[X > t + h | X > h] &= \frac{P[X > t + h, X > h]}{P[X > h]} &(27) \\ &= \frac{P[X > t + h]}{P[X > h]} &(28) \\ &= \frac{e^{-\lambda(t+h)}}{e^{-\lambda h}} &(29) \\ &= e^{-\lambda t} = P[X > t]. &(30) \end{aligned}$$

(b) Using the memory-less property, show that the conditional CDF of X given $\{X > h\}$ is

$$F_X(x | X > h) = 1 - e^{-\lambda(x-h)}, \quad x > h.$$

Ans: We have $F_X(x|X > h) = 1 - P[X > x|X > h]$. Using the memoryless property,

$$P[X > x | X > h] = e^{-\lambda(x-h)}, \quad x > h,$$

and the result follows.

2. At a certain location, the temperature in summer follows a $\mathcal{U}(25, 32)$ distribution, in autumn it has a $\mathcal{U}(10, 25)$ distribution, in winter it is $\mathcal{U}(-10, 10)$ and in spring it is $\mathcal{U}(10, 25)$. The $\mathcal{U}(a, b)$ distribution has the PDF

$$f_U(u) = \frac{1}{b-a}, \quad a < u < b.$$

Find the PDF of the temperature of a randomly selected day of the year, assuming that autumn and spring each occupy 20 percent of the year, and summer and winter each occupy 30 percent of the year.

Ans: We have the following conditional distributions for the temperature T:

$$\begin{aligned} f_T(t|\text{summer}) &= \frac{1}{7}, \quad 25 < t < 32 && (31) \\ f_T(t|\text{autumn}) &= \frac{1}{15}, \quad 10 < t < 25 && (32) \\ f_T(t|\text{winter}) &= \frac{1}{20}, \quad -10 < t < 10 && (33) \\ f_T(t|\text{spring}) &= \frac{1}{15}, \quad 10 < t < 25. && (34) \end{aligned}$$

The probabilities of picking a day in summer, autumn, winter and spring are respectively 0.3, 0.2, 0.3 and 0.2. Therefore by the theorem on total probability,

$$\begin{aligned} f_T(t) &= 0.3 f_T(t|\text{summer}) + 0.2 f_T(t|\text{autumn}) + 0.3 f_T(t|\text{winter}) + 0.2 f_T(t|\text{spring}) \\ &= \begin{cases} \frac{3}{200} & -10 < t < 10 \\ \frac{2}{75} & 10 < t < 25 \\ \frac{3}{70} & 25 < t < 32 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

3. Imagine that the 3G network is available with probability 0.8; independently, the 4G network is available with probability 0.8. A smartphone first tries to connect to the 4G network, then if that is not available, it connects to the 3G network. If neither network is available then there is no connectivity. Let X be some performance measure of the network, with the following conditional distributions:

$$\begin{aligned} f_X(x|\text{“connected to 4G”}) &= 0.1[u(x-10) - u(x-20)] && (35) \\ f_X(x|\text{“connected to 3G”}) &= 0.2[u(x-7) - u(x-12)] && (36) \\ f_X(x|\text{“neither”}) &= \delta(x). && (37) \end{aligned}$$

Find the PDF of X.

Ans: Let A_3 and A_4 denote the events “3G network is available” and “4G network is available”, respectively. We are told that $P[A_3] = 0.8$, $P[A_4] = 0.8$ and A_3 and A_4 are independent.

A user can be in one of three states: connected to 4G, connected to 3G, or not connected. In terms of A_3 and A_4, these are A_4, $A_4^c \cap A_3$ and $A_4^c \cap A_3^c$ respectively, with probabilities

$$P[A_4] = 0.8, \quad P[A_4^c A_3] = 0.16, \quad P[A_4^c A_3^c] = 0.04.$$

Therefore, the PDF of X is given by the theorem on total probability as follows:

$$\begin{aligned} f_X(x) &= f_X(x|\text{``connected to 4G''})P[A_4] + f_X(x|\text{``connected to 3G''})P[A_4^c A_3] \\ &\quad + f_X(x|\text{``neither''})P[A_4^c A_3^c] \\ &= 0.08[u(x-10) - u(x-20)] + 0.032[u(x-7) - u(x-12)] + 0.04\delta(x). \end{aligned}$$