Cumulative distribution functions (CDFs) and probability density functions (PDFs) for continuous distributions

A valid pdf $f(x)$ is a nonnegative function such that $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$. The cdf of a continuous random variable X with pdf $f(x)$ is $F(x) = \Pr(X \le x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$.

To find probabilities, you therefore have to integrate the density function. In the general case, if $f(x)$ is some function, then

$$\int_a^b f(x)\mathrm{d}x = F(x)\Big|_a^b = F(b) - F(a), \tag{1}$$

where $F(x)$ is a primitive function or antiderivative of $f(x)$. The list of antiderivatives of some important functions:

- If $f(x) = x^n$, $n \neq -1$, then $F(x) = \frac{x^{n+1}}{n+1} + C$
- If $f(x) = 1/x$, then $F(x) = \ln x + C$
- If $f(x) = e^{\alpha x}$, then $F(x) = \frac{1}{\alpha} e^{\alpha x} + C$
- If $f(x) = \alpha^x$ $(\alpha > 0)$, then $F(x) = \frac{\alpha^x}{\ln \alpha} + C$

Note that C is some constant here as the antiderivative is defined up to an additive constant. $C = 0$ can be used in (1). Some properties of the integrals:

- $\int_a^b \{c_1 f(x) + c_2 g(x)\}\mathrm{d}x = c_1 \int_a^b f(x)\mathrm{d}x + c_2 \int_a^b g(x)\mathrm{d}x$ for any constants c_1 and c_2
- $\int_a^b f(x)\mathrm{d}x = \int_a^c f(x)\mathrm{d}x + \int_c^b f(x)\mathrm{d}x$ where $a < c < b$
- Integration by substitution: $\int_{g(a)}^{g(b)} f(x)\mathrm{d}x = \int_a^b f(g(y))g'(y)\mathrm{d}y$
- Integration by parts formula: $\int_a^b f(x)g'(x)\mathrm{d}x = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\mathrm{d}x$

If $f(x)$ is a pdf of random variable X, then its cdf $F(x)$ is an antiderivative of $f(x)$. In particular, from (1) we get:

$$\Pr(a \le X \le b) = \int_a^b f(x)\mathrm{d}x = F(b) - F(a).$$

Example 1: Random variable X has the pdf $f(x) = \begin{cases} kx + 0.5, & \text{if } 0 < x < 1, \\ 0, & \text{otherwise} \end{cases}$

(a) Find the constant k,

(b) Find the 12% quantile of X.

Solution:

(a) We know that $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$. In this example, $f(x) = 0$ for $x < 0$ or $x > 1$ so we have

$$1 = \int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_0^1 f(x)\mathrm{d}x = \int_0^1 \{kx + 0.5\}\mathrm{d}x = k\int_0^1 x\mathrm{d}x + 0.5\int_0^1 \mathrm{d}x$$

$$= k\frac{x^2}{2}\Big|_0^1 + 0.5x\Big|_0^1 = \frac{k}{2} + 0.5 \quad \Rightarrow \quad \mathbf{k = 1}.$$

(b) We need to find q such that $F(q) = 0.12$ where F is the cdf of X. We therefore have to find $F(x)$. Note that $f(x)$ is positive for $x \in (0,1)$ and therefore X can only take values from $(0,1)$ interval. It implies that $F(x) = \Pr(X \le x) = 0$ for $x < 0$ and $F(x) = 1$ for $x > 1$. If $0 < x < 1$, then

$$F(x) = \int_0^x f(t)\mathrm{d}t = \int_0^x \{t + 0.5\}\mathrm{d}t = \int_0^x t\mathrm{d}t + 0.5\int_0^x \mathrm{d}t = \frac{t^2}{2}\Big|_0^x + 0.5t\Big|_0^x = 0.5(x^2 + x).$$

We now solve $F(q) = 0.5(q^2+q) = 0.12$. This equation has two roots: $q_1 = 0.2$ and $q_2 = 1.2$. We select the first one because $F(q) = 1$ for $q > 1$. Hence, the 12% quantile of X is **0.2**.

Example 2: Random variable X has the pdf $f(x) = \begin{cases} k/x^2, & \text{if } x > 1, \\ 0, & \text{otherwise} \end{cases}$

(a) Find the constant k,

(b) Find $\Pr(2 < X < 4 \,|\, 3 < X < 5)$,

(c) Let $Y = \ln X$. Find the cdf and pdf of Y.

Solution:

(a) $1 = \int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_1^{\infty} f(x)\mathrm{d}x = \int_1^{\infty} \frac{k}{x^2}\,\mathrm{d}x = k\int_1^{\infty} x^{-2}\mathrm{d}x = k\frac{x^{-1}}{-1}\Big|_1^{\infty} = k \;\Rightarrow\; \mathbf{k{=}1}.$

Interestingly, $\mathrm{E}(X)$ is not defined because $\int_1^{\infty} xf(x)\mathrm{d}x = \int_1^{\infty}\frac{\mathrm{d}x}{x} = \ln x\Big|_1^{\infty} = \infty.$

(b) We can use the formula of conditional probability to write:

$$\Pr(2 < X < 4 \,|\, 3 < X < 5) = \frac{\Pr(2 < X < 4, 3 < X < 5)}{\Pr(3 < X < 5)} = \frac{\Pr(3 < X < 4)}{\Pr(3 < X < 5)} = \frac{F(4) - F(3)}{F(5) - F(3)}\,,$$

where $F(x)$ is the cdf of X. X can take values greater than one and therefore $F(x) = 0$ for $x < 1$ and for $x \ge 1$ we find:

$$F(x) = \int_1^x f(t)\mathrm{d}t = \int_1^x \frac{1}{t^2}\,\mathrm{d}t = \int_1^x t^{-2}\,\mathrm{d}t = \frac{t^{-1}}{-1}\Big|_1^x = 1 - \frac{1}{x}\,,$$

and hence

$$\Pr(2 < X < 4 \,|\, 3 < X < 5) = \frac{1 - 1/4 - (1 - 1/3)}{1 - 1/5 - (1 - 1/3)} = \mathbf{0.8}.$$

(c) X can take values greater than one and hence $Y = \ln X$ can take values greater than $\ln 1 = 0$. It implies that $\Pr(Y \leq y) = 0$ for $y < 0$. If $y \geq 0$, the cdf of Y is

$$G(y) = \Pr(Y \leq y) = \Pr(\ln X \leq y) = \Pr(X \leq e^y) = 1 - \frac{1}{e^y} = 1 - e^{-y}.$$

The pdf is $g(y) = G'(y) = e^{-y}$. We can see that Y follows an exponential distribution with $\lambda = 1$.

Example 3: Random variable X has the pdf $f(x) = \begin{cases} C_1, & \text{if } 1 \leq x < 2, \\ C_2, & \text{if } 2 \leq x < 4, \\ 0, & \text{otherwise.} \end{cases}$

(a) Find C_1 and C_2 such that $\mathrm{E}(X) = 2$.

(b) Find the moment generating function (mgf) of X.

Solution:

(a) We have two conditions here:

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_1^4 f(x)\mathrm{d}x = 1, \quad \mathrm{E}(X) = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x = \int_1^4 xf(x)\mathrm{d}x = 1.5.$$

To compute these integrals, we split integration interval, $(1,4)$, into two subintervals, $(1,2)$ and $(2,4)$, and use $f(x) = C_1$ in the first interval and $f(x) = C_2$ in the second interval:

$$\begin{aligned} 1 &= \int_1^4 f(x)\mathrm{d}x = \int_1^2 C_1\,\mathrm{d}x + \int_2^4 C_2\,\mathrm{d}x = C_1 x\Big|_1^2 + C_2 x\Big|_2^4 = C_1 + 2C_2, \\ 2 &= \int_1^4 xf(x)\mathrm{d}x = \int_1^2 xC_1\,\mathrm{d}x + \int_2^4 xC_2\,\mathrm{d}x = C_1\frac{x^2}{2}\Big|_1^2 + C_2\frac{x^2}{2}\Big|_2^4 = 1.5C_1 + 6C_2. \end{aligned}$$

Multiply both sides of the first equation by 1.5 and then subtract it from the second equation to get $0.5 = 3C_2 \;\Rightarrow\; \mathbf{C_2 = 1/6}$ and hence $\mathbf{C_1 = 1 - 2C_2 = 2/3}$.

(b) We again split integration interval into two subintervals, $(1,2)$ and $(2,4)$, to compute the mgf of X:

$$\begin{aligned} M(t) &= \mathrm{E}(e^{Xt}) = \int_1^4 e^{xt} f(x)\mathrm{d}x = \int_1^2 e^{xt}\frac{2}{3}\,\mathrm{d}x + \int_2^4 e^{xt}\frac{1}{6}\,\mathrm{d}x = \frac{2}{3}\frac{e^{tx}}{t}\Big|_1^2 + \frac{1}{6}\frac{e^{tx}}{t}\Big|_2^4 \\ &= \frac{2(e^{2t} - e^t)}{3t} + \frac{e^{4t} - e^{2t}}{6t} = \frac{e^{4t} + 3e^{2t} - 4e^t}{6t}. \end{aligned}$$

Note that we define $M(0)$ as the limit $M(0) = \lim_{t\to 0} M(t)$ in this formula. We use the Taylor expansion for small t: $e^{\alpha t} \approx 1 + \alpha t$ to find $M(t) \approx \frac{1+4t+3(1+2t)-4(1+t)}{6t} = 1$.