Continuous Random Variables and their PDF and CDF, Mean, Variance, Covariance and Correlation

From Micheal Barron Book

Recall that any discrete distribution is concentrated on a finite or countable number of isolated values. Conversely, *continuous variables can take any value of an interval*, (a, b), $(a, +\infty)$, $(-\infty, +\infty)$, etc. Various times like service time, installation time, download time, failure time, and also physical measurements like weight, height, distance, velocity, temperature, and connection speed are examples of continuous random variables.

4.1 Probability density

For all continuous variables, the probability mass function (pmf) is always equal to zero,[1]

$$P(x) = 0 \quad \text{for all } x.$$

As a result, the pmf does not carry any information about a random variable. Rather, we can use the *cumulative distribution function* (cdf) $F(x)$. In the continuous case, it equals

$$F(x) = \boldsymbol{P}\{X \le x\} = \boldsymbol{P}\{X < x\}.$$

These two expressions for $F(x)$ differ by $P\{X = x\} = P(x) = 0$.

In both continuous and discrete cases, the cdf $F(x)$ is a non-decreasing function that ranges from 0 to 1. Recall from Chapter 3 that in the discrete case, the graph of $F(x)$ has *jumps* of magnitude $P(x)$. For continuous distributions, $P(x) = 0$, which means no jumps. The cdf in this case is a continuous function (see, for example, Figure 4.2 on p. 77).

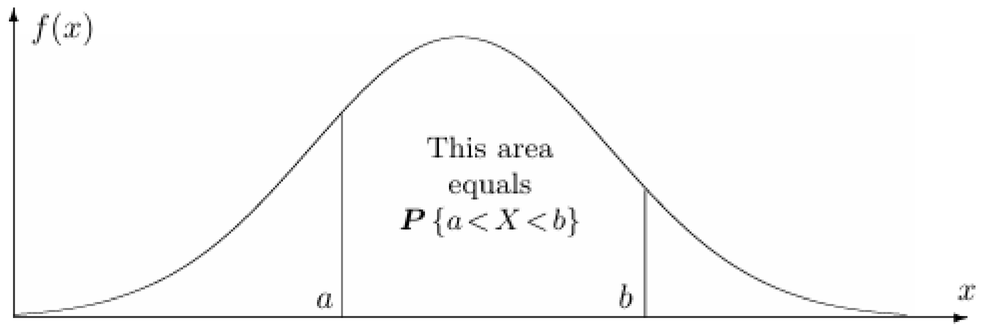

FIGURE 4.1: *Probabilities are areas under the density curve.*

Assume, additionally, that $F(x)$ has a derivative. This is the case for all commonly used continuous distributions, but in general, it is not guaranteed by continuity and monotonicity (the famous Cantor function is a counterexample).

Then, $F(x)$ is an *antiderivative* of a density. By the Fundamental Theorem of Calculus, the integral of a density from a to b equals to the difference of antiderivatives, i.e.,

$$\int_a^b f(x)dx = F(b) - F(a) = \boldsymbol{P}\{a < X < b\},$$

where we notice again that the probability in the right-hand side also equals $\boldsymbol{P}\{a \leq X < b\}$, $\boldsymbol{P}\{a < X \leq b\}$, and $\boldsymbol{P}\{a \leq X \leq b\}$.

Probability density function	$f(x) = F'(x)$ $\boldsymbol{P}\{a < X < b\} = \int_a^b f(x)dx$

Thus, probabilities can be calculated by integrating a density over the given sets. Furthermore, the integral $\int_a^b f(x)dx$ equals the area below the density curve between the points a and b. Therefore, geometrically, probabilities are represented by *areas* (Figure 4.1). Substituting $a = -\infty$ and $b = +\infty$, we obtain

$$\int_{-\infty}^{b} f(x)dx = \boldsymbol{P}\{-\infty < X < b\} = F(b) \text{ and } \int_{-\infty}^{+\infty} f(x)dx = \boldsymbol{P}\{-\infty < X < +\infty\} = 1.$$

That is, the total area below the density curve equals 1.

Example 4.1. The lifetime, in years, of some electronic component is a continuous random variable with the density

$$f(x) = \begin{cases} \dfrac{k}{x^3} & \text{for} \quad x \geq 1 \\ 0 & \text{for} \quad x < 1. \end{cases}$$

Find k, draw a graph of the cdf $F(x)$, and compute the probability for the lifetime to exceed 5 years.

<u>Solution</u>. Find k from the condition $\int f(x)dx = 1$:

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{1}^{+\infty} \frac{k}{x^3}dx = -\left.\frac{k}{2x^2}\right|_{x=1}^{+\infty} = \frac{k}{2} = 1.$$

Hence, $k = 2$. Integrating the density, we get the cdf,

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{1}^{x} \frac{2}{y^3}dy = -\left.\frac{1}{y^2}\right|_{y=1}^{x} = 1 - \frac{1}{x^2}$$

for $x > 1$. Its graph is shown in Figure 4.2.

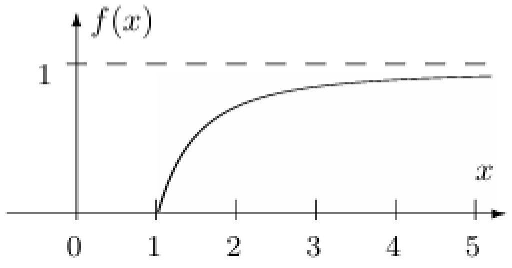

FIGURE 4.2: *Cdf for Example 4.1.*

Next, compute the probability for the lifetime to exceed 5 years,

$$\boldsymbol{P}\{X > 5\} = 1 - F(5) = 1 - \left(1 - \frac{1}{5^2}\right) = 0.04.$$

We can also obtain this probability by integrating the density,

$$\boldsymbol{P}\{X > 5\} = \int_{5}^{+\infty} f(x)dx = \int_{5}^{+\infty} \frac{2}{x^3}dx = -\left.\frac{1}{x^2}\right|_{x=5}^{+\infty} = \frac{1}{25} = 0.04.$$

$\Diamond$

Distribution	**Discrete**	**Continuous**
Definition	$P(x) = P\{X = x\}$ (pmf)	$f(x) = F'(x)$ (pdf)
Computing probabilities	$\boldsymbol{P}\{X \in A\} = \sum_{x \in A} P(x)$	$\boldsymbol{P}\{X \in A\} = \int_A f(x)dx$
Cumulative distribution function	$F(x) = \boldsymbol{P}\{X \leq x\} = \sum_{y \leq x} P(y)$	$F(x) = \boldsymbol{P}\{X \leq x\} = \int_{-\infty}^{x} f(y)dy$
Total probability	$\sum_{x} P(x) = 1$	$\int_{-\infty}^{\infty} f(x)dx = 1$

TABLE 4.1: *Pmf* $P(x)$ *versus pdf* $f(x)$.

Joint and marginal densities

DEFINITION 4.2

For a vector of random variables, the **joint cumulative distribution function** is defined as

$$F_{(X,Y)}(x,y) = \boldsymbol{P}\left\{X \leq x \ \cap Y \leq y\right\}.$$

The **joint density** is the *mixed derivative* of the joint cdf,

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

Expectation and variance

Continuing our analogy with the discrete case, *expectation* of a continuous variable is also defined as a center of gravity,

Distribution	**Discrete**	**Continuous**
Marginal distributions	$P(x) = \sum_y P(x,y)$ $P(y) = \sum_x P(x,y)$	$f(x) = \int f(x,y)dy$ $f(y) = \int f(x,y)dx$
Independence	$P(x,y) = P(x)P(y)$	$f(x,y) = f(x)f(y)$
Computing probabilities	$\boldsymbol{P}\{(X,Y) \in A\}$ $= \sum\sum_{(x,y)\in A} P(x,y)$	$\boldsymbol{P}\{(X,Y) \in A\}$ $= \iint_{(x,y)\in A} f(x,y)\,dx\,dy$

TABLE 4.2: *Joint and marginal distributions in discrete and continuous cases.*

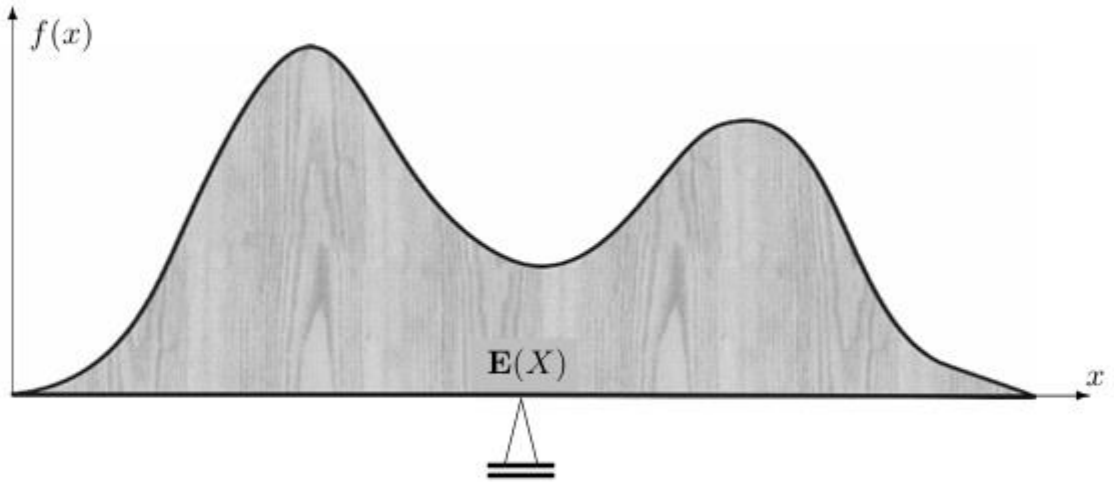

FIGURE 4.3: *Expectation of a continuous variable as a center of gravity.*

$$\mu = \mathbf{E}(X) = \int x f(x) dx$$

(compare with the discrete case on p. 48, Figure 3.4). This time, if the entire region below the density curve is cut from a piece of wood, then it will be balanced at a point with coordinate $\mathbf{E}(X)$, as shown in Figure 4.3.

Variance, standard deviation, covariance, and correlation of continuous variables are defined similarly to the discrete case, see Table 4.3. All the properties in (3.5), (3.7), and (3.8) extend to the continuous distributions. In calculations, don't forget to replace a pmf with a pdf, and a summation with an integral.

Example 4.2. A random variable X in Example 4.1 has density

$$f(x) = 2x^{-3} \text{ for } x \geq 1.$$

Its expectation equals

$$\mu = \mathbf{E}(X) = \int x\, f(x)dx = \int_1^\infty 2x^{-2}dx = -2x^{-1}\Big|_1^\infty = 2.$$

Discrete	**Continuous**
$\mathbf{E}(X) = \sum_x xP(x)$	$\mathbf{E}(X) = \int xf(x)dx$
$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2$ $= \sum_x (x-\mu)^2 P(x)$ $= \sum_x x^2 P(x) - \mu^2$	$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2$ $= \int (x-\mu)^2 f(x)dx$ $= \int x^2 f(x)dx - \mu^2$
$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \sum_x \sum_y (x-\mu_X)(y-\mu_Y)P(x,y)$ $= \sum_x \sum_y (xy)P(x,y) - \mu_x\mu_y$	$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \iint (x-\mu_X)(y-\mu_Y)f(x,y)\,dx\,dy$ $= \iint (xy)f(x,y)\,dx\,dy - \mu_x\mu_y$

TABLE 4.3: *Moments for discrete and continuous distributions.*

Exercises

4.1. The lifetime, in years, of some electronic component is a continuous random variable with the density

$$f(x) = \begin{cases} \dfrac{k}{x^4} & \text{for} \quad x \geq 1 \\ 0 & \text{for} \quad x < 1. \end{cases}$$

Find k, the cumulative distribution function, and the probability for the lifetime to exceed 2 years.

4.2. The time, in minutes, it takes to reboot a certain system is a continuous variable with the density

$$f(x) = \begin{cases} C(10-x)^2, & \text{if } 0 < x < 10 \\ 0, & \text{otherwise} \end{cases}$$

(a) Compute C.

(b) Compute the probability that it takes between 1 and 2 minutes to reboot.

4.3. The installation time, in hours, for a certain software module has a probability density function $f(x) = k(1 - x^3)$ for $0 < x < 1$. Find k and compute the probability that it takes less than 1/2 hour to install this module.

4.4. Lifetime of a certain hardware is a continuous random variable with density

$$f(x) = \begin{cases} K - x/50 & \text{for } 0 < x < 10 \text{ years} \\ 0 & \text{for all other } x \end{cases}$$

(a) Find K.

(b) What is the probability of a failure within the first 5 years?

(c) What is the expectation of the lifetime?

4.5. Two continuous random variables X and Y have the joint density

$$f(x, y) = C(x^2 + y), \quad -1 \leq x \leq 1, \quad 0 \leq y \leq 1.$$

(a) Compute the constant C.

(b) Find the marginal densities of X and Y. Are these two variables independent?

(c) Compute probabilities $P\{Y < 0.6\}$ and $P\{Y < 0.6 \mid X < 0.5\}$.

Exercise questions mentioned in updated outline which was shared on GCR