MTH 224, Spring 2024

Instructor: Bahman Angoshtari

Lecture 15

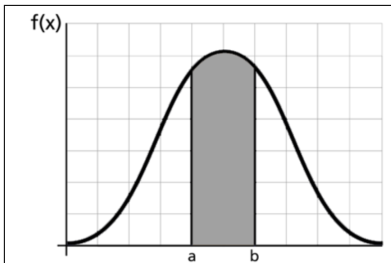
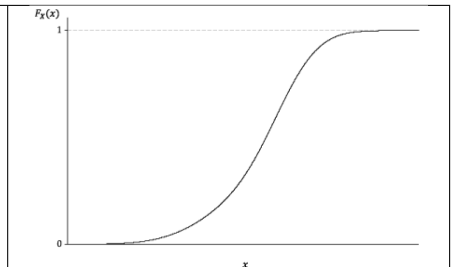
Section 3.1: cdf and expected value of continuous r.v., uniform distribution.

15.1. The cdf of a continuous distribution

- Similar to the discrete case, we define the cdf (cumulative distribution function) of a continuous random variable with pdf $f_X(x)$ as follows: $F_X(a) = \mathbb{P}(X \le a) = \int_{-\infty}^{a} f_X(x)\,dx$.
  Furthermore, we have that $\mathbb{P}(a < X \le b) = \mathbb{P}(X \le b) - \mathbb{P}(X \le a) = F_X(b) - F_X(a)$.

Connection to the pdf of X. By the fundamental theorem of calculus, $F_X'(a) = f_X(a)$ for all $a \in \mathbb{R}$.

Summary:

pdf	cdf
	
$f_X(x) \ge 0, \int_{-\infty}^{\infty} f_X(x)\,dx = 1$	$F_X(-\infty) = 0,\ F_X(\infty) = 1, F_X \nearrow$
$\mathbb{P}(a \le X \le b) = \int_a^b f_X(x)\,dx$	$\mathbb{P}(a \le X \le b) = F_X(b) - F_X(a)$
$f_X(x) = F_X'(x)$	$F_X(a) = \int_{-\infty}^{a} f_X(x)\,dx$

Example 15.1. Assume that X has the pdf $f_X(x) = \frac{3}{8}(4x - 2x^2)$ for $x \in (0, 2)$. Find:
(a) The cdf $F_X(x)$; and (b) $\mathbb{P}(X > 1)$.

Solution. (a) We have $F_X(x) = 0$ for $x \in (-\infty, 0]$ and $F_X(x) = 1$ for $x \in [2, +\infty)$ (why?). For $x \in (0, 2)$, we calculate:

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,dy = \int_0^x \frac{3}{8}\left(4y - 2y^2\right) dy = \frac{3}{8}\left(4 \cdot \frac{y^2}{2}\bigg|_0^x - 2 \cdot \frac{y^3}{3}\bigg|_0^x\right) = \frac{x^2}{4}(3 - x).$$

(b) We have $\mathbb{P}(X > 1) = \mathbb{P}(1 < X < +\infty) = F_X(+\infty) - F_X(1) = 1 - \frac{1^2}{4}(3 - 1) = \frac{1}{2}$.

Example 15.2. The concentration of alcohol in an individual's blood t hours after drinking is e^{-t}. The concentration is measured at a random time X, whose pdf is given by $f_X(x) = 0.5$ for $x \in [3, 5]$ (it is implied that $f_X(x) = 0$ for $x < 3$ and $x > 5$). Find the pdf of $Y =$ the measured concentration of alcohol.

SOLUTION. We have $Y = \mathrm{e}^{-X}$, and we are asked to find $f_Y(y)$, the pdf of Y. The trick for finding the pdf of a function of a random variable is to first find its **cdf** $F_Y(y) = \mathbb{P}(Y \leq y)$. Then, differentiate $F_Y(y)$ to obtain the pdf: $f_Y(y) = F_Y'(y)$. Let us see how this works.
First, we calculate $F_Y(y)$

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(\mathrm{e}^{-X} \leq y) = \mathbb{P}\left(X \geq -\log(y)\right) \\ &= 1 - \mathbb{P}\left(X \leq -\log(y)\right) = 1 - F_X\left(-\log(y)\right). \end{aligned}$$

At this point, it is tempting to explicitly find $F_X\Big(-\log(y)\Big) = \int_{-\infty}^{-\log(y)} f_X(x)\mathrm{d}x$. However, we don't need to do so! In fact, since we are interested in $f_Y(y)$, we can differentiate the above equation to obtain:

$$f_Y(y) = F_Y'(y) = \frac{\mathrm{d}}{\mathrm{d}y}\left(1 - F_X\big(-\log(y)\big)\right) = -F_X'\big(-\log(y)\big)\frac{\mathrm{d}}{\mathrm{d}y}\Big(-\log(y)\Big) = f_X\big(-\log(y)\big) \times \frac{1}{y}.$$

We also need to figure out the support of $Y = \mathrm{e}^{-X}$ (the support of a r.v. is the largest subset of $\mathbb{R}$ where its pdf is non-zero). Note that the support of X is $(3,5)$. So, $f_X\big(-\log(y)\big) > 0$ for

$$3 < -\log(y) < 5 \quad \Longleftrightarrow \mathrm{e}^{-5} < y < \mathrm{e}^{-3}.$$

We then conclude that the pdf of Y is $f_Y(y) = \frac{1}{2y}$ for $y \in \big(\mathrm{e}^{-5}, \mathrm{e}^{-3}\big)$.

15.2. Expected value of continuous r.v.

- The expected value of a discrete r.v. was defined as $\mathbb{E}\left(X\right) = \sum_x x \cdot p_X\left(x\right)$. For a continuous r.v., the sum is replaced by an integral and the pmf is replaced by the pdf.

DEFINITION 15.3. Let X be a r.v. with pdf $f_X(x)$. We define the **mean** or **expected value** of X, denoted by $\mathbb{E}[X]$, as follows

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x)\mathrm{d}x,$$

assuming that the integral on the right side exists. More generally, for "any" function $g(x)$, we define the expected value of $g(X)$ by

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x)\mathrm{d}x,$$

assuming that the integral on the right side exists.

DEFINITION 15.4. Let X be a r.v. with pdf $f_X(x)$. The **variance** of X is given by

$$\mathrm{Var}(X) = \mathbb{E}\Big[\big(X - \mathbb{E}[X]\big)^2\Big] = \int_{-\infty}^{+\infty} \big(x - \mathbb{E}[X]\big)^2 f_X(x)\mathrm{d}x,$$

assuming that the integral on the right side exists.

- The properties that we discussed for $\mathbb{E}[X]$ and $\mathrm{Var}(X)$ of discrete r.v. also hold in the continuous case. Specifically:
  - $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ and $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.
  - $\mathbb{E}(aX + bY) = a\mathbb{E}[X] + b\mathbb{E}[Y]$
  - $\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}(X,Y)$. $\mathrm{Cov}(X,Y)$ for continuous r.v. are defined, as in the discrete case, using continuous joint distributions. We will not present the details, since doing so requires knowledge of multiple integrals, which is outside the scope of our class.
  - $\mathrm{Var}(X) = \mathbb{E}\Big[X^2\Big] - \Big(\mathbb{E}[X]\Big)^2$.

- You can prove all the above properties using the definition of expectation and variance.

EXAMPLE 15.5. In Example 15.2, find the mean and variance of the measured concentration of alcohol.

SOLUTION. **Solution 1:** Using the result of part (a), we calculate the expectation of Y as follows:

$$\mathbb{E}[Y] = \int_{-\infty}^{+\infty} y f_Y(y)\mathrm{d}y = \int_{\mathrm{e}^{-5}}^{\mathrm{e}^{-3}} y\frac{1}{2y}\mathrm{d}y = \frac{1}{2}\left(e^{-3} - e^{-5}\right) \approx 0.022.$$

To find the variance, we first calculate

$$\mathbb{E}[Y^2] = \int_{-\infty}^{+\infty} y^2 f_Y(y)\mathrm{d}y = \int_{\mathrm{e}^{-5}}^{\mathrm{e}^{-3}} \frac{y}{2}\mathrm{d}y = \frac{1}{4}\left(e^{-6} - e^{-10}\right).$$

So, we get $\mathrm{Var}(Y) = \mathbb{E}[Y^2] - \left(\mathbb{E}[Y]\right)^2 = \frac{1}{4}\left(e^{-6} - e^{-10}\right) - \frac{1}{4}\left(e^{-3} - e^{-5}\right)^2 \approx 0.000145.$

Solution 2: Alternatively, we can use the pdf of X to find $\mathbb{E}[Y]$ and $\mathrm{Var}(Y)$. By the definition of expected value of a function of a r.v., we have

$$\mathbb{E}[Y] = \mathbb{E}[\mathrm{e}^{-X}] = \int_{-\infty}^{+\infty} \mathrm{e}^{-x} f_X(x)\mathrm{d}x = \int_3^5 e^{-x}\frac{1}{2}dx = \frac{1}{2}\left(e^{-3} - e^{-5}\right) \approx 0.022.$$

Similarly,

$$\mathbb{E}[Y^2] = \mathbb{E}[\mathrm{e}^{-2X}] = \int_{-\infty}^{+\infty} \mathrm{e}^{-2x} f_X(x)\mathrm{d}x = \int_3^5 e^{-2x}\frac{1}{2}dx = \frac{1}{4}\left(e^{-6} - e^{-10}\right).$$

Therefore, $\mathrm{Var}(Y) = \mathbb{E}[Y^2] - \left(\mathbb{E}[Y]\right)^2 = \frac{1}{4}\left(e^{-6} - e^{-10}\right) - \frac{1}{4}\left(e^{-3} - e^{-5}\right)^2 \approx 0.000145.$

15.3. Uniform distribution

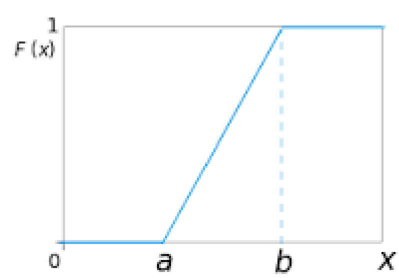
DEFINITION 15.6. We say that a r.v. X has the uniform distribution on the interval $[a, b]$ if its pdf is

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{otherwise} \end{cases}$$

Notation: $X \sim \mathrm{U}(a, b)$.

- If $X \sim U(a, b)$, then X is equally likely to take any value in $[a, b]$.
- Where does the value $\frac{1}{b-a}$ come from? One must have $\int_{-\infty}^{\infty} f_X(x)\, dx = \int_a^b \frac{1}{b-a} dx = 1$.
- Note that if $X \sim U(a, b)$, then the cdf $F_X(x)$ is

$$F_X(x) = \mathbb{P}(X \leq x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a < x < b \\ 1, & x > b \end{cases}$$



EXAMPLE 15.7. Find $\mathbb{E}[Z]$ and $\mathrm{Var}(Z)$ if $Z \sim U(0, 1)$.

SOLUTION. $\mathbb{E}[Z] = \int_0^1 z \cdot 1 dz = \frac{z^2}{2}\Big|_0^1 = \frac{1}{2}$. Similarly, $\mathbb{E}\left[Z^2\right] = \int_0^1 z^2 dx = \frac{z^3}{3}\Big|_0^1 = \frac{1}{3}$. Therefore, $\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}[X]\right)^2 = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$.