# Recitation 5: Continuous Random Variables

## 1 Review

- The *variance* of a random variable (discrete or continuous):

$$\mathrm{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

which has the property that

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$$

- A continuous random variable $X$ can be characterized by its *probability density function* (PDF) $f(x)$, which has the property that

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

- The *cumulative distribution function* (CDF) of a continuous random variable

$$F(x) = P(X \le x) = \int_{-\infty}^x f(u)\,du$$

- Given a CDF (i.e. $F(x)$), we can find the PDF (i.e. $f(x)$) using the following fact from calculus:

$$\frac{d}{dx}F(x) = \frac{d}{dx}\int_{-\infty}^x f(u)\,du = f(x)$$

This works for all $x$ for which $F(x)$ is differentiable. The value of $f(x)$ at the non-differentiable points are usually decided according to some convention – don't worry about them in this class.

- Expectation for continuous random variables

$$E[X] = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,dx$$

- The PDF of a normal random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is

$$\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}.$$

(This is one of the PDFs that you may want to memorize.) If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are **independent**, then $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2,\, \sigma_1^2 + \sigma_2^2)$.

## 2 Exercises

1. Suppose $X$ is a continuous random variable with the following PDF:

$$f(x) = \begin{cases} cx^2 & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $c$.

Since PDFs must integrate to 1, we have that

$$1 = \int_{-\infty}^{\infty} f(x)dx = \int_0^2 cx^2 dx = \frac{cx^3}{3}\bigg|_0^2 = \frac{8c}{3} \qquad \Rightarrow \qquad c = \frac{3}{8}$$

So the PDF of X is

$$f(x) = \begin{cases} \frac{3}{8}x^2 & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

(b) Compute $F(x)$, the cumulative distribution function (CDF) of $X$. Remember that $F(x)$ is defined for all real numbers $x$.

$$F(x) = \begin{cases} 0 & x < 0 \\ \int_{-\infty}^{x} f(u)\,du = \int_0^x \frac{3}{8}u^2\,du = \frac{x^3}{8} & 0 \le x \le 2 \\ 1 & x > 2 \end{cases}$$

(c) Find the *median* of $X$: the value $\gamma$ for which $P(X \le \gamma) = 0.50$

Since $F_X(0) = 0$ and $F_X(2) = 1$, any quantile must lie in the interval [0,2]. For $\gamma \in [0, 2]$, we need to satisfy

$$P(X \le \gamma) = F(\gamma) = \frac{\gamma^3}{8} = 0.50.$$

Thus $\gamma = \sqrt[3]{4} = 2^{2/3} \approx 1.587$.

(d) Compute $E[X]$ and $\text{Var}(X)$.

$$E[X] = \int_{-\infty}^{\infty} xf(x)\,dx = \int_0^2 \frac{3}{8}x^3\,dx = \frac{3}{32}x^4\bigg|_0^2 = \frac{3}{2}$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^2 \frac{3}{8}x^4\,dx = \frac{3}{40}x^5\bigg|_0^2 = \frac{12}{5}$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{3}{20}$$

(e) Compute $E[X^{-2}]$.

$$E[X^{-2}] = \int_{-\infty}^{\infty} x^{-2} f(x)\,dx = \int_0^2 x^{-2}\frac{3}{8}x^2\,dx = \frac{3}{8}x\bigg|_0^2 = \frac{3}{4}$$

(f) Find the CDF, call it $G(y)$, of $Y = X^2$.
Recall:

$$G(y) = P(Y \le y) = P(X^2 \le y)$$

A squared variable must be positive, so $G(y) = 0$ for all $y < 0$. Also, the maximum value of $X$ (with non-zero probability density) is 2. In other words, with probability 1, $X$ is at most 2. Therefore, it follows that with probability 1, $Y$ is less than 4, so $G(y) = 1$ for $y \ge 4$. Consider the case where $4 > y \ge 0$:

$$\begin{aligned} G(y) &= P(X^2 \le y) \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} f(u)\,du \\ &= \int_{0}^{\sqrt{y}} f(u)\,du \\ &= \frac{(\sqrt{y})^3}{8} \end{aligned}$$

(g) Using $G(y)$, find the the PDF of $Y$, call it $g(y)$. You may ignore the cases $g(0)$ and $g(4)$ and consider the PDF undefined at those points.
As described in the review, we can find the PDF by taking the derivative of $G(y)$ at $y$ for which $G(y)$ is differentiable (i.e. every $y$ that isn't 0 or 4):

$$g(y) = \frac{d}{dy} G(y).$$

For $y < 0$, we see $g(y) = \frac{d}{dy} 0 = 0$. For $y > 4$, we see $g(y) = \frac{d}{dy} 1 = 0$. For $0 < y < 4$, we see that:

$$\begin{aligned} g(y) &= \frac{d}{dy} G(y) \\ &= \frac{d}{dy} \int_0^{\sqrt{y}} f(u)\,du \\ &= \frac{1}{2\sqrt{y}} f(\sqrt{y}) \\ &= \frac{1}{2\sqrt{y}} \left( \frac{3}{8} y \right) = \frac{3}{16}\sqrt{y} \end{aligned}$$

2. Harry, a convenience store employee, often takes smoke breaks while on the job. The duration $T$ of these breaks is exponentially distributed with some rate $\lambda$, and thus can be described by the PDF

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & t \ge 0 \\ 0 & \text{otherwise} \end{cases}.$$

You are told that 90% of Harry's breaks last less than 10 minutes.

(a) Write the CDF of $T$. It will depend on $\lambda$.

$$F(t) = P(T \leq t) = \int_0^t \lambda e^{-\lambda u} du = -e^{-\lambda u}\Big|_0^t = 1 - e^{-\lambda t}$$

(b) Find $\lambda$.

From part (a), we have that

$$P(T < 10) = 1 - e^{-10\lambda} = 0.9$$

Solving for $\lambda$ gives

$$\lambda = 0.23 \text{ per minute}$$

(c) Compute the probability that Harry takes a break lasting more than 20 minutes. Perform this calculation using both the PDF (by setting up an integral) and the CDF (by leveraging your work from parts (a) and (b)).

Using the PDF:

$$\begin{aligned} P(T > 20) &= \int_{20}^{\infty} 0.23 e^{-0.23t} dt \\ &= -e^{-0.23t}\Big|_{20}^{\infty} \\ &= -\lim_{t\to\infty} e^{-0.23t} + e^{-0.23*20} \\ &= e^{-0.23*20} \\ &= 0.01 \end{aligned}$$

Using the CDF:

$$P(T > 20) = 1 - P(T \leq 20) = 1 - F(20) = e^{-0.23*20} = 0.01$$

(d) R's `dexp(x, rate)` and `pexp(x, rate)` can be used to evaluate the PDF and CDF of the exponential distribution, respectively, at a given point $x$. Use (one or both of) these functions to verify your answer from part (c).

`1-pexp(20,0.23)` in R gives $P(T > 20) = 0.01$

3. Intelligence Quotient (IQ) tests are structured so that the score $X$ that a randomly selected individual receives is normally distributed, with a mean of $\mu = 100$ and a variance of $\sigma^2 = 225$.

(a) Write the PDF of $X$.

Given $X \sim N(100, 225)$, the PDF of $X$ is

$$f(x) = \frac{1}{15\sqrt{2\pi}} e^{-(x-100)^2/450}$$

(b) Compute the probability that a randomly selected individual's IQ falls within one standard deviation of the mean (that is, between 85 and 115). Express your answer in terms of the CDF of the standard normal random variable.

$$\begin{aligned} P(85 \le X \le 115) &= P\left(\frac{85-100}{15} \le \frac{X-\mu}{\sigma} \le \frac{115-100}{15}\right) \\ &= P(-1 \le Z \le 1) \\ &= \Phi(1) - \Phi(-1) \\ &= 2\Phi(1) - 1, \end{aligned}$$

where $Z$ is a standard normal random variable.

(c) Use R to evaluate your expression from part (b). The functions `dnorm(x)` and `pnorm(x)` give the PDF and CDF of the standard normal random variable, respectively, evaluated at $x$.

`pnorm(1)-pnorm(-1)` or `2*pnorm(1)-1` in R gives $P(85 \le X \le 115) = 0.6828$

(d) Verify (using R) that your answer from parts (b) and (c) match what you would have obtained without standardizing. The functions `dnorm(x, mu, sigma)` and `pnorm(x, mu, sigma)` give the PDF and CDF of a normal random variable, respectively, evaluated at $x$.

`pnorm(115,100,15)-pnorm(85,100,15)` in R gives $P(85 \le X \le 115) = 0.6828$

(e) Suppose that two individuals are randomly selected to take an IQ test. Assuming their scores are independent, compute the probability that their combined score is at least 200.

Since $X_1 \sim N(100, 225)$ and $X_2 \sim N(100, 225)$ are independent, it follows that $X_1 + X_2 \sim N(\mu', \sigma'^2)$ where $\mu' = 200$ and $\sigma'^2 = 450$. Therefore, the probability for their combined score to be at least 200 is

$$P(X_1 + X_2 \ge 200) = P\left(\frac{X_1 + X_2 - \mu'}{\sigma'} \ge 0\right) = P(Z \ge 0) = 0.5.$$

4. The PDF of a $Beta(\alpha, \beta)$ random variable is

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}$$

where $\alpha > 0$ and $\beta > 0$ are *shape parameters*, and

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$$

is the Gamma function.

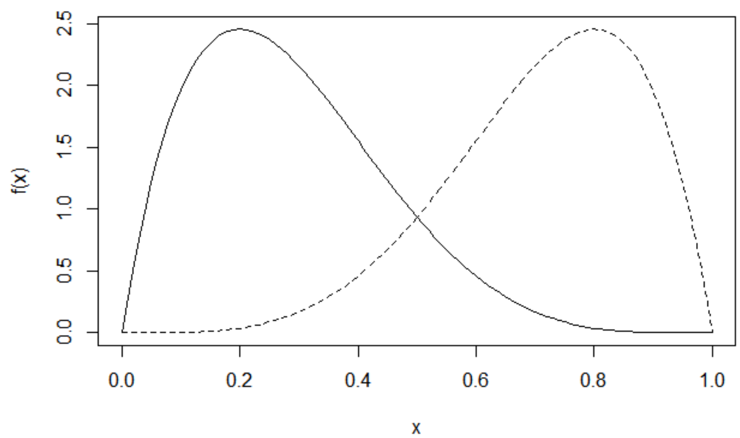
(a) Plot the $Beta(2, 5)$ and the $Beta(5, 2)$ PDF using R. This can be done, for instance, using the commands

```
x = seq(0, 1, 0.01)
y = dbeta(x, alpha, beta)
```

```
y2 = dbeta(x, alpha2, beta2)
plot(x, y, type="l")
lines(x, y2)
```

Note that the two densities are symmetric, in that we can rotate one PDF about the line $x = 0.5$ to obtain the other. This holds anytime we interchange $\alpha$ and $\beta$.

See the file `rec4.R`. We obtain the following plot:




The solid line denotes the $Beta(2, 5)$ PDF, and the dashed line the $Beta(5, 2)$ PDF.

(b) Suppose $\alpha > 1$ and $\beta > 1$. Show that $f(x)$ is maximized at the point

$$x^* = \frac{\alpha - 1}{\alpha + \beta - 2}$$

Ignoring the constant in front of the PDF, we instead show that the function

$$h(x) = x^{\alpha-1}(1-x)^{\beta-1}$$

is maximized at the desired point. Taking a derivative and setting it to zero, we want $x$ to solve

$$\begin{aligned} h'(x) &= (\alpha-1)x^{\alpha-2}(1-x)^{\beta-1} - (\beta-1)x^{\alpha-1}(1-x)^{\beta-2} \\ &= x^{\alpha-2}(1-x)^{\beta-1}\Big[(\alpha-1)(1-x) - (\beta-1)x\Big] = 0 \end{aligned}$$

Solving the equation $(\alpha-1)(1-x) - (\beta-1)x = 0$ for $x$, we indeed obtain

$$x^* = \frac{\alpha - 1}{\alpha + \beta - 2}.$$

To verify this is indeed a global maximum, we observe that

$$(\alpha-1)(1-x) - (\beta-1)x = (\alpha-1) - (\alpha+\beta-2)x,$$

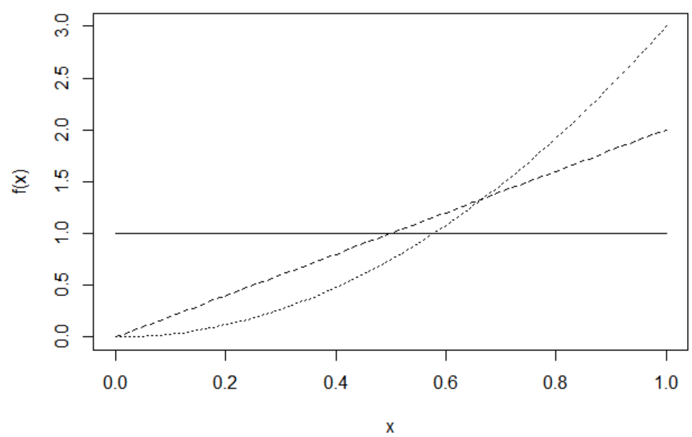
which is positive when $x < x^*$, and negative when $x > x^*$.

(c) If $\alpha = \beta$, where does the PDF $f(x)$ attain its peak value? What can you say about the location of the peak when $\alpha < \beta$? When $\alpha > \beta$?

If $\alpha = \beta$, then $x^* = 1/2$, and so $f(x)$ attains its peak value at the point $x = 1/2$. If $\alpha < \beta$, then the peak will occur at a point $x < 1/2$, and if $\alpha > \beta$, the peak occurs at $x > 1/2$. Small values of $\alpha$ drag the peak closer to 0, whereas small values of $\beta$ drag the peak closer to 1.

(d) To get an idea of what happens when $\alpha \leq 1$ or $\beta \leq 1$, plot the PDFs of the $Beta(1, 1)$, $Beta(2, 1)$, $Beta(3, 1)$, and $Beta(0.5, 0.5)$ distributions in R.

See the file `rec4.R`. For the first three PDFs, we obtain the plots




Note how the $Beta(1, 1)$ PDF is exactly the $Uniform(0, 1)$ PDF, , the $Beta(2, 1)$ PDF increases linearly, and the $Beta(3, 1)$ PDF increases quadratically. For the $Beta(0.5, 0.5)$ PDF, we obtain the plot