Statistics for Data Science
Unit 4 Part 2 Homework: Continuous Random Variables

Kevin Hartman (W203 Wednesday 6:30pm Summer 2019)

6/5/2019

1. **Processing Pasta**

   A certain manufacturing process creates pieces of pasta that vary by length. Suppose that the length of a particular piece, L, is a continuous random variable with the following probability density function.

$$f(l) = \begin{cases} 0, & l \leq 0 \\ l/2, & 0 < l \leq 2 \\ 0, & 2 < l \end{cases}$$

   (a) Write down a complete expression for the cumulative probability function of L.

   <u>**Answer:**</u>

   The CDF of L is defined as

$$F(l) = \int_{y=-\infty}^{l} f(y)dy$$

   When $l \leq 0$,

$$F(l) = \int_{y=-\infty}^{0} 0 \cdot dy = 0$$

   When $0 < l < 2$,

$$F(l) = \int_{y=-\infty}^{0} 0 \cdot dy + \int_{y=0}^{l} \frac{y}{2} dy = 0 + \left.\frac{y^2}{4}\right|_0^l = \frac{l^2}{4}$$

   When $l \geq 2$,

$$F(l) = \int_{y=-\infty}^{0} 0 \cdot dy + \int_{y=0}^{2} \frac{y}{2} dy + \int_{y=2}^{\infty} 0 \cdot dy = 0 + \left.\frac{y^2}{4}\right|_0^2 + 0 = \frac{2^2}{4} = 1$$

   Which gives us

$$F(l) = \begin{cases} 0, & l \leq 0 \\ \frac{l^2}{4}, & 0 < l < 2 \\ 1, & 2 \leq l \end{cases}$$

(b) Using the definition of expectation for a continuous random variable, compute the expected length of the pasta, $E(L)$.
Answer:

$$E(L) = \int_{-\infty}^{\infty} l \cdot f(l) dl$$

Assembling from the different ranges:

$$E(L) = \int_{-\infty}^{0} l \cdot 0 dl + \int_{0}^{2} l \cdot \frac{l}{2} dl + \int_{2}^{\infty} l \cdot 0 dl$$

$$= \int_{0}^{2} \frac{l^2}{2} dl = \frac{l^3}{6}\Big|_0^2 = \frac{8}{6}$$

2. **The Warranty is Worth It**

Suppose the life span of a particular (shoddy) server is a continuous random variable, T, with a uniform probability distribution between 0 and 1 year. The server comes with a contract that guarantees you money if the server lasts less than 1 year. In particular, if the server lasts t years, the manufacturer will pay you $g(t) = \$100(1-t)^{1/2}$. Let $X = g(T)$ be the random variable representing the payout from the contract.

Compute the expected payout from the contract, $E(X) = E(g(T))$.
Answer:

$$E(X) = E(g(T)) = \int g(t) f(t) dt = \int 100(1-t)^{\frac{1}{2}} f(t) dt$$

Since T has a uniform probability distribution between 0 and 1, $f(t) = \frac{1}{(1-0)} = 1$

$$E(X) = \int_{t=0}^{1} 100(1-t)^{\frac{1}{2}} dt = \frac{200}{3}(1-t)^{\frac{3}{2}}\Big|_0^1 = \frac{200}{3} = \$66.67$$

3. **(Lecture)#Fail**

Suppose the length of Paul Laskowski's lecture in minutes is a continuous random variable C, with pmf $f(t) = e^{-t}$ for $t > 0$. This is an example of an exponential random variable, and it has some special properties. For example, suppose you have already sat through t minutes of the lecture, and are interested in whether the lecture is about to end immediately. In statistics, this can be represented by something called the *hazard rate*:

$$h(t) = \frac{f(t)}{1 - F(t)}$$

To understand the hazard rate, think of the numerator as the probability the lecture ends between time t and time $t + dt$. The denominator is just the probability the lecture does not end before time t. So you can think of the fraction as the conditional probability that the lecture ends between t and $t + dt$ given that it did not end before t.

Compute the hazard rate for C.
Givens:
The hazard rate is defined as

$$h(t) = \frac{f(t)}{1 - F(t)}$$

The CDF of C, where t is the length of C in minutes, is defined as

$$F(t) = \int_{y=-\infty}^{t} f(y)dy$$

The PMF of C, where t is the length of C in minutes, is defined as

$$f(t) = \begin{cases} 0, & t \leq 0 \\ e^{-t}, & 0 < t \end{cases}$$

Answer:
When $t \leq 0$,

$$F(t) = \int_{y=-\infty}^{0} 0 \cdot dy = 0$$

When $0 < t$

$$F(t) = \int_{y=0}^{t} e^{-y}dy = (-e^{y} + B)\Big|_{0}^{t} = (-e^{t} + B) - (-e^{0} + B) = -e^{-t} + B + 1 - B = 1 - e^{-t}$$

When combined, $F(t) = 1 - e^{-t}$

Replacing $f(t)$ and $F(t)$ in the hazard rate for C,

$$h_C(t) = \frac{e^{-t}}{1 - (1 - e^{-t})} = \frac{e^{-t}}{e^{-t}} = 1 \text{ for } t > 0$$

4. **Optional Advanced Exercise: Characterizing a Function of a Random Variable**

   Let X be a continuous random variable with probability density function $f(x)$, and let h be an invertible function where h^{-1} is differentiable. Recall that $Y = h(X)$ is itself a continuous random variable. Prove that the probability density function of Y is

$$g(y) = f(h^{-1}(y)) \cdot \left| \frac{d}{dy} h^{-1}(y) \right|$$

Proof:
If h is invertible and $Y = h(X)$ then $X = h^{-1}(Y)$

If h^{-1} is differentiable then so is h, and h will always be increasing or decreasing.

Let h be increasing, as proof for the other side would be similar.

Following the above,

$$\text{if } Y = h(X) \text{ then } P(Y \leq y) = P(h(X) \leq y)$$

$$\text{and if } X = h^{-1}(Y) \text{ then } P(Y \leq y) = P(X \leq h^{-1}(y)$$

Thus, the CDF of Y, G(y), is

$$G(y) = \int_{-\infty}^{h^{-1}(y)} f(x)dx = \int_{-\infty}^{y} f(h^{-1}(u)) \cdot \frac{d}{dy} h^{-1}(u)du \quad \text{by substitution}$$

$$= \int_{-\infty}^{y} f(h^{-1}(u)) \cdot \left| \frac{d}{dy} h^{-1}(u) \right| du \quad \text{since h is increasing}$$

Since $g(y) = G'(y)$,

$$g(y) = f(h^{-1}(y)) \cdot \left| \frac{d}{dy} h^{-1}(y) \right|$$