W203 Statistics for Data Science

Aditya Mengani

aditya.mengani@berkeley.edu

Monday, 4pm

Summer 2020

Unit 3 Homework: Continuous Random Variables

May 17, 2020

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. (3 points)

(b) Using the definition of expectation for a continuous random variable, compute the expected length of the pasta, $E(L)$. (3 points)

(a). Given L is a continuous random variable, which randomly measures the length of a particular piece

& ~~The~~ To find complete expression of cumulative probability function of L, we need to use the following definition:

The cumulative distribution function $F(x)$ for a continuous rv X is defined for every number x by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\,dy$$

for each x, $F(x)$ is the area under the density curve to the left of x.

Thus this rule can be applied to the given probability density function $f(l)$ to obtain the c.d.f of $f(l) = F(l)$

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

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

so, for any number l between 0 and 2,

$$F(x) = \int_{-\infty}^{x} f(y)\,dy = \int_{0}^{x} (y/2)\cdot dy$$

$$= \frac{1}{2}\int_{0}^{x} y\cdot dy = \frac{1}{2}\cdot\left(\frac{y^2}{2}\right)\Bigg|_{y=0}^{y=x}$$

$$= \frac{1}{2}\cdot\left(\frac{x^2}{2}\right) = \frac{x^2}{4}$$

$$F(x) = \frac{x^2}{4} \quad \text{for} \quad 0 \le x \le 2 \qquad \text{——— (a)}$$

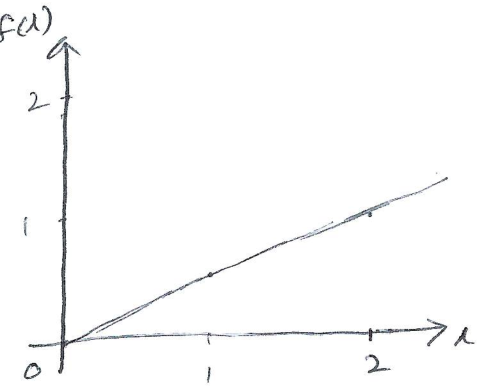
The cumulative probability function will be 0, for all $x < 0$
(Since probability cannot be -ve) ——— (b)

The cumulative probability function will be 1, for all $x > 2$
(Since this is the sum of probabilities upto $x = 2$) ——— (c)

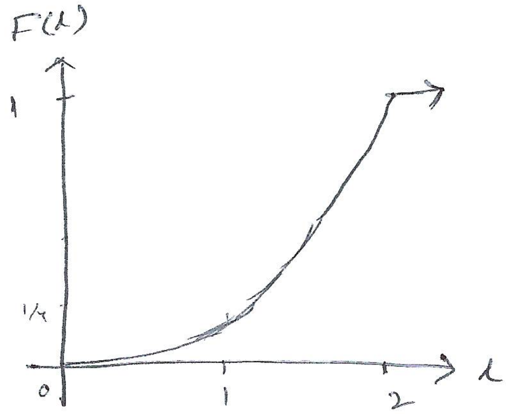
combining (a), (b) and (c), the expression for cumulative probability function $F(x)$ can be written as

$$\boxed{F(x) = \begin{cases} 0 & x < 0 \\ x^2/4 & 0 \le x \le 2 \\ 1 & 2 < x \end{cases}}$$

The possible graphs of pdf and cdf of $f(x)$ and $F(x)$ are :-



PDF



CDF

(b). To compute the expectation for L, which is the expected length of pasta, we will use the following definition: (3)

The expected or mean value of a continuous random variable X with a pdf $f(x)$ is given by μ_x; where

$$\mu_x = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx$$

Thus we can apply above definition to find $E(L)$ as,

$$E(L) = \int_{-\infty}^{\infty} l \cdot f(l)\, dl = \int_{0}^{2} l \cdot (l/2) \cdot dl$$

$$= \frac{1}{2} \int_{0}^{2} (l^2) \cdot dl = \frac{1}{2} \left(\frac{l^3}{3} \right) \Bigg|_{l=0}^{l=2}$$

$$= \frac{1}{2} \cdot \left[\frac{8}{3} \right] = \frac{4}{3} = 1.33$$

Thus $E(L) = 1.33$ which is the expected length or mean value μ_l of the pasta

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, $E(X) = E(g(T))$. (3 points)

Given T is a continuos rv that measures the life span of a server and has a uniform distribution [0, 1]

From the definition of a continuos random variable X, having a uniform distribution on interval [A, B], the pdf of X is

$$f(x:A,B) = \begin{cases} 1/B-A & A \le x \le B \\ 0 & \text{otherwise} \end{cases}$$

Thus pdf of T can be written as,

$$f(t;0,1) = \begin{cases} \frac{1}{1-0} & 0 \le t \le 1 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 1 & 0 \le t \le 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{— (a)}$$

where A = 0 and B = 1, which is the range of the server

Given g(t) be a function that represents the amount paid for Server lasting t years where,

$$g(t) = \begin{cases} 100(1-t)^{1/2} & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Let X be a new continuos random variable, where X represents the payout from the contract

$$\Rightarrow X = g(T)$$

we know, if X is a continuos rv with pdf $f(x)$ and $h(X)$ is any function of X, then

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

So, applying the above definition to the random variable X,

$$E(X) = E(g(T)) = \mu_{g(T)}$$

$$= \int_{-\infty}^{\infty} g(t).f(t).dt \quad \text{———} \quad (b)$$

Earlier we have derived from uniform distribution rule in (a),

$$\int_{-\infty}^{\infty} f(t) = \int_{0}^{1} f(t) = 1 \quad \text{———} \quad (c)$$

Substituting equation (c) in (b) above,

$$E(x) = E(g(T)) = \int_{-\infty}^{\infty} g(t)f(t)\,dt = \int_{0}^{1} g(t).1.dt$$

$$= \int_{0}^{1} 100.(1-t)^{1/2}.1.dt = 100.\int_{0}^{1} (1-t)^{1/2}dt$$

Let $u = 1-t$

then $du = -dt$

the expression of $E(x)$ can be written as,

$$E(X) = 100 \int_0^1 u^{1/2} \cdot (-du)$$

$$= -100 \cdot \left[\frac{u^{3/2}}{3/2} \right]_0^1$$

substituting $u = 1-t$

$$E(X) = -100 \left[\frac{(1-t)^{3/2}}{3/2} \right]_0^1$$

$$= -100 \left(0 - 1 \times \frac{2}{3} \right) = \frac{200}{3} = 66.66$$

Thus the Expected value $E(X)$ that manufacturer pays is 66.66

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. (3 points)

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

C is the rv, which is continuous and measures the length of lecture in minutes with pmf $f(t) = e^{-t}$ for $t > 0$ — (a)

As per the definition of a cdf $F(x)$ for continuous random variable X is defined as $F(x) = P(X \leq x) = \int_{-\infty}^{\infty} f(y)\, dy$

so, $F(t) = P(C \leq t) = \int_{-\infty}^{t} e^{-y} dy$

we can zero out any pdf or cdf where there is no support,

hence $F(t)$ can be written as

$$F(t) = P(C \leq t) = \int_{0}^{t} e^{-y} \cdot dy = -e^{-y} \Big|_{y=0}^{y=t}$$

$$= -e^{-t} - (-e^{-0})$$

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

Thus $\boxed{F(t) = \begin{cases} 1-e^{-t} & t \geq 0 \\ 0 & t < 0 \end{cases}}$ — (b)

given $h(t) = f(t) / (1 - F(t))$ — (c)

Substituting (a) and (b) in (c),

$$h(t) = \frac{e^{-t}}{1-(1-e^{-t})}$$

$$= \frac{e^{-t}}{e^{-t}} = 1$$

Thus $\boxed{h(t) = 1}$

The hazard rate for c, which is the value of $f(t)/(1-F(t))$

is $h(t) = 1$

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

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

(Bonus + 3 points)

The problem states X is a continuous random variable with pdf $f(x)$

It further states, Y is also a continuous random variable with $Y = h(X)$

the pdf of Y is defined as $g(y)$, cdf of Y is $G_y(y)$

The cdf of X can be written as,

$$F_x(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt \qquad \text{—— (a)}$$

The cdf of Y can be written as

$$G_y(y) = P(X \leq y) = \int_{-\infty}^{y} f(t).dt \qquad \text{—— (b)}$$

$$G_y(y) = \int_{-\infty}^{y} f(t).dy = P(\cancel{X} \leq y)$$

we know $Y = h(X)$

thus $G_y(y) = P(h(X) \leq y)$

applying inverse rule, this can be written as

$$\cancel{G_y(y) = P(h^{-1}(h(X)) \leq h^{-1}(y))}$$

$$= P(h^{-1}(h(X)) \leq h^{-1}(y))$$

$$= P(X \le h^{-1}(y))$$

This can be written as below from (a), (b) derived earlier,

$$G_y(y) = P(X \le h^{-1}(y)) = \int_{-\infty}^{h^{-1}(y)} f(t) \cdot dt$$

applying differentiation on both sides

$$\frac{d}{dy} G_y(y) = \frac{d}{dy} \int_{-\infty}^{h^{-1}(y)} f(t) \cdot dt$$

According to fundamental theorem of calculus

$$\frac{d}{dx} \int_a^x f(t)dt = f(x)$$

and from the rules of cdf and pdf, we know

$F'(x) = f(x)$ for every x, where F' exists

thus the above can be expressed as,

$$G_y'(y) = \frac{d}{dy} G_y(y) = g(y) = \frac{d}{dy} \int_{-\infty}^{h^{-1}(y)} f(t) \cdot dt$$

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

(from the chain rule of calculus, when $h^{-1}(y)$ is a function of y)

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