Functions of Random Variables

Olanrewaju Akande

Department of Statistical Science, Duke University

February 22, 2016

Lesson Plan

- Questions
- Functions of Continuous Random Variables
- Cumulative Distribution Function (CDF) Technique
- Change of Variable Technique

Objectives

By the end of class, you should be able to derive the marginal distribution of functions of a continuous random variable using the:

- Cumulative Distribution Function (CDF) Technique
- Change of Variable Technique

Motivation

- Let X be a continuous random variable, and Y = h(X), a function of X.
- For simplicity, we will assume that $h(X)$ is a well-behaved function of X. That is, h is continuous, monotone and its inverse h^{-1} exists.
- Notice that Y must also be a random variable that has its own distribution since it is a transformation of a random variable. Can we derive that?

  Example 1:

  Let X be a continuous random variable with pdf $f_X(x) = 4x^3$, $0 < x < 1$ and let $Y = X^2$. What is the pdf of Y?

Illustration

- Sometimes it makes more sense (and can be relatively straight forward) to first find the cdf of Y and then derive the pdf using calculus and what we have learned so far.
- Back to our example, what is the cdf of X?
- Recall: $f_X(x) = 4x^3$, $0 < x < 1$.

Illustration

- $F_Y(y) = P(Y \leq y) = P(X \leq \sqrt[+]{y}) = F_X(\sqrt[+]{y}) = (\sqrt[+]{y})^4 = y^2$

  Do you know why $X \nleq \sqrt[-]{y}$?

- Now that we know what the cdf of Y is, can you figure out what the cdf of X is?

- What is the support for y?

Another Illustration

Example 2:

Let $X_1, X_2, \cdots, X_n$ be a random iid sample from the distribution with pdf $f(x) = 3x^2$ on $0 \leq x \leq 1$

Now, Let $Y = max\{X_1, X_2, \cdots, X_n\}$. Can we find E($Y$)?

First, can we figure out what the pdf of Y is?

(Think about this: what does being the maximum really mean? If the maximum of X random variables is less than y, what is the relationship between each of the X's and y?)

What is the cdf of X?

Another Illustration (Cont'd)

As in the previous example, first find the cdf $F_X(x)$ of X.

Thus, $F(y) = P(Y \leq y) = P(X_1 \leq y, X_2 \leq y, \cdots, X_n \leq y)$

$\Rightarrow P(Y \leq y) = P(X_1 \leq y \text{ and } X_2 \leq y \text{ and } \cdots, \text{and } X_n \leq y)$

$\Rightarrow P(Y \leq y) = P(X_1 \leq y) \times P(X_2 \leq y) \times \cdots \times P(X_n \leq y)$

Can you see why this is the case?

$\Rightarrow F_Y(y) = P(Y \leq y) = F_{X1}(y) \times F_{X2}(y) \times \cdots \times F_{Xn}(y)$

$\Rightarrow F_Y(y) = y^3 \times y^3 \times \cdots \times y^3 = y^{3n}$

Another Illustration (Cont'd)

$\Rightarrow f_Y(y) = (3n)y^{3n-1}$ on $0 \leq y \leq 1$

Do you know why that is the support of Y?

Now,

$$E(Y) = \int_0^1 y(3n)y^{3n-1}dy = 3n\int_0^1 y^{3n}dy$$

Therefore, $E(Y) = \dfrac{3n}{3n+1}$

How about $Z = min\{X_1, X_2, \cdots, X_n\}$, can you find $E(Z)$?

Shortcut to the cdf technique

If the function h(X) is a continuous function with an inverse (which we already assumed), there is a neater and faster way to find its pdf.

Let Y = h(X), then if the inverse exists, $X = h^{-1}(Y)$

Also, the derivative of X with respect to Y is $\dfrac{dX}{dY}$

Thus, the pdf of Y, $f_Y(y) = f_X[h^{-1}(Y)] \times \left|\dfrac{dX}{dY}\right|$

Example 1 again

$f_X(x) = 4x^3$ on $0 \leq x \leq 1$ and $Y = X^2$.

Inverse of $Y = X^2$ is $X = h^{-1}(Y) = \sqrt[+]{y}$ and $\dfrac{dX}{dY} = \dfrac{1}{2\sqrt[+]{y}}$

Then, $f_X[h^{-1}(Y)] = f_X[\sqrt[+]{y}] = 4(\sqrt[+]{y})^3 = 4y(\sqrt[+]{y})$

Therefore, $f_Y(y) = \dfrac{1}{2\sqrt[+]{y}} \times 4y(\sqrt[+]{y}) = 2y$

Do we have the same answer as before?

Recap

You should now be able to:

- Derive the marginal distribution of functions of a continuous random variable using the CDF method
- Use the change of variable technique as a shortcut to the cdf method