Manipulating Continuous Random Variables
Class 5, 18.05

Jeremy Orloff and Jonathan Bloom

1 Learning Goals

1. Be able to find the pdf and cdf of a random variable defined in terms of a random variable with known pdf and cdf.

2 Transformations of Random Variables

We frequently transform a known random variable into a new one by applying a formula. For example we might look at $Y = aX + b$ or $Y = X^2$. In this section we will see how to find the probability density and cumulative distribution of Y from those of X.

For discrete random variables it was often possible do this by looking at probability tables. For continuous random variables we will need to use systematic algebraic techniques. We will see that transforming the pdf is just the change of variables ('u-substitution') from calculus. To transform the cdf directly we will rely on its definition as a probability.

Let's remind ourselves of the basics:

- The cdf of X is $F_X(x) = P(X \leq x)$.
- The pdf of X is related to F_X by $f_X(x) = F_X'(x)$.

2.1 Transforming the cdf

Example 1. Suppose X has range $[0, 2]$ and cdf $F_X(x) = x^2/4$. What is the range, pdf and cdf of $Y = X^2$?

Solution: The range is easy: $[0, 4]$.

To find the cdf we work systematically from the definition. For this example we will break it down into tiny steps, so you can see the thought process in detail.

Step 1. Use definition:

$$F_Y(y) = P(Y \leq y).$$

Step 2. Replace Y by its formula in X:

$$P(Y \leq y) = P(X^2 \leq y).$$

Step 3. Algebraically manipulate this to isolate the X:

$$P(X^2 \leq y) = P(X \leq \sqrt{y})$$

Step 4. Notice that this is exactly the definition of F_X:

$$P(X \leq \sqrt{y}) = F_X(\sqrt{y})$$

Step 5. Use the known formula for F_X:

$$F_X(\sqrt{y}) = (\sqrt{y})^2/4 = y/4.$$

Following the chain from step 1 to step 5 we have the cdf:

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(X \le \sqrt{y}) = F_X(\sqrt{y}) = y/4.$$

Finally, to find the pdf we can just differentiate the cdf:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{4}.$$

2.2 Transforming the pdf directly

An alternative way to find the pdf directly is by change of variables. We present this for completeness and for anyone who prefers it as a method. Our observation is that most people find the cdf easier to transform.

In calculus you learned the 'u'-substitution. We'll do a calculus example to remind you how this goes and then apply it to the pdf.

Example 2. Calculus example. Convert the integral $\int (x^2+1)^7\,dx$ into an integral in $u = x^2 + 1$.

Solution: We have to convert each part of the integral from x to u:

$$(x^2+1)^7 = u^7$$
$$du = 2x\,dx, \qquad \text{therefore} \qquad dx = \frac{du}{2x} = \frac{du}{2\sqrt{u-1}}$$

Now we replacing each piece in the integral we get

$$\int (x^2+1)^7\,dx = \int u^7 \frac{du}{2\sqrt{u-1}}.$$

Example 3. Find the pdf of Y in Example 1 directly using the method of 'u'-substitution. (In this case, 'u' will actually be 'y'.)

Solution: The trick is to remember that probability is given by an integral $\int f_X(x)dx$. We are given the change of variable $y = x^2$, so we change the integral from one in x to one in y.

$$y = x^2 \;\Rightarrow\; dy = 2x\,dx, \qquad \text{therefore} \qquad dx = \frac{dy}{2\sqrt{y}}.$$

We are given $F_X(x) = x^2/4$, so we can compute $f_X(x) = F_X'(x) = x/2$. Changing this to y we have

$$f_X(x) = \sqrt{y}/2.$$

Putting the two pieces together we have the transformation

$$f_X(x)\,dx \;=\; \frac{\sqrt{y}}{2}\,\frac{dy}{2\sqrt{y}} \;=\; \frac{1}{4}\,dy$$

Since this is a probability, the factor in front of dy is the probability density. That is, $f_Y(y) = 1/4$, exactly as in Example 1.

Here are a few more examples. We do them a little more quickly than the above examples.

Example 4. Let $X \sim \exp(\lambda)$, so $f_X(x) = \lambda e^{-\lambda x}$ on $[0, \infty]$. What is the probability density of $Y = X^2$?

Solution: We will do this using the change of variables for the pdf.

$$\begin{aligned} &y = x^2 \;\Rightarrow\; dy = 2x\,dx, \qquad \text{therefore} \qquad dx = \tfrac{dy}{2\sqrt{y}} \\ &f_X(x) = \lambda e^{-\lambda x} = \lambda e^{-\lambda\sqrt{y}}. \end{aligned}$$

Combining these we get,

$$f_X(x)\,dx = \lambda e^{-\lambda\sqrt{y}}\,\frac{dy}{2\sqrt{y}} = f_Y(y)\,dy.$$

So we conclude that $\boxed{f_Y(y) = \dfrac{\lambda}{2\sqrt{y}}\,e^{-\lambda\sqrt{y}}}$.

Example 5. Redo the previous example using the cdf.

Solution: The cdf for the exponential random variable X is $F_X(x) = 1 - e^{-\lambda x}$. Therefore, for $Y = X^2$ we have

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(X \le \sqrt{y}) = F_X(\sqrt{y}) = 1 - e^{-\lambda\sqrt{y}}.$$

We have found $F_Y(y)$. If we wanted $f_Y(y)$ we could take the derivative. We would get the same answer as in the previous example.

Example 6. Assume $X \sim N(5, 3^2)$ then $Z = \dfrac{X-5}{3}$ is standard normal, i.e., $Z \sim N(0, 1)$.

Solution: Again using the change of variables and the formula for $f_X(x)$ we have

$$z = \frac{x-5}{3} \;\Rightarrow\; dz = \frac{dx}{3}, \qquad \text{therefore} \qquad dx = 3\,dz$$

For this example we will transform $f_X(x)\,dx$ in one line instead of two.

$$f_X(x)\,dx = \frac{1}{3\sqrt{2\pi}}e^{-(x-5)^2/(2\cdot 3^2)}\,dx = \frac{1}{3\sqrt{2\pi}}e^{-z^2/2}\,3\,dz = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,dz = f_Z(z)\,dz$$

Therefore $f_Z(z) = \dfrac{1}{\sqrt{2\pi}}\,e^{-z^2/2}$. Since this is exactly the density for $N(0, 1)$ we have shown that Z is standard normal.

This example shows an important general property of normal random variables which we state as a theorem.

Theorem. Standardization of normal random variables.
Assume $X \sim N(\mu, \sigma^2)$. Show that $Z = \dfrac{X-\mu}{\sigma}$ is standard normal, i.e., $Z \sim N(0, 1)$.

Proof. This is exactly the same computation as the previous example with μ replacing 5 and σ replacing 3. We show the computation without comment.

$$z = \frac{x-\mu}{\sigma} \;\Rightarrow\; dz = \frac{dx}{\sigma} \;\Rightarrow\; dx = \sigma\, dz$$

$$f_X(x)\, dx = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(x-\mu)^2/(2\cdot\sigma^2)}\, dx = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-z^2/2}\, \sigma\, dz = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}\, dz = f_Z(z)\, dz$$

Therefore $f_Z(z) = \dfrac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-z^2/2}$. This shows Z is standard normal.

We call the change from X to Z in this theorem standardization because it converts X from an arbitrary normal random variable to a standard normal variable.