Change of Continuous Random Variable

All you are responsible for from this lecture is how to implement the "Engineer's Way" (see page 4) to **compute** how the probability density function changes when we make a change of random variable from a continuous random variable X to Y by a strictly increasing change of variable $y = h(x)$. So for the purpose of surviving Stat 400 you can start reading in §2. I give two examples following the statement of the theorem and its proof(s) where the method gives the correct result, then I give an example where it doesn't work when $h(x)$ is not one-to-one.

Let X be a continuous random variable. I will assume for convenience that the set of values taken by X is the entire real line $\mathbb{R}$. If the set of values taken by X is an interval, for example $[0, 1]$, the formula for the change of density if the same but we don't know the interval where the new density will be nonzero (the support). We will treat this point later. Let $y = h(x)$ be a real-valued *strictly-increasing* function (so h is one-to-one). Since h is one-to-one it has an inverse function $x = g(y)$. We want to define a new random variable $Y = h(X)$. There is **only one possible definition**, to find it we pretend Y exists and compute for each pair of numbers c and d with $c < d$ what the Y-probability $P(c \leq Y \leq d)$ **has to be** in terms of an X-probability.

$$P(c \leq Y \leq d) = P(c \leq h(X) \leq d) = P(g(c) \leq g(h(X)) \leq g(d)) = P(a \leq X \leq b).$$

Here we define a and b by $g(c) = a$ and $g(d) = b$ or equivalently $h(a) = c$ and $h(b) = d$. Note that the last equation holds because $g(h(X)) = X$ since g is the inverse of h. Note $c \leq h(X) \leq d \Rightarrow g(c) \leq g(h(X)) \leq g(d)$ because h is strictly increasing and the inverse of a strictly increasing function is strictly increasing so g preserves inequalities; that is, $a \leq b \leq c \Rightarrow g(a) \leq g(b) \leq g(c)$. So the above calculation forces us to **define** the new random variable Y by

$$P(c \leq Y \leq d) := P(a \leq X \leq b) \text{ where } a = g(c) \text{ and } b = g(d). \tag{1}$$

In other words, the probability that the new random variable Y will be in an interval $[c, d]$ is defined to be the probability that the old random variable X will be in the transformed interval $[g(c), g(d)] = [a, b]$. To begin with, Equation (1) is just a rule for assiging a real number to each interval $[c, d]$. It turns out (it follows from the second proof of the next theorem) that this formula (1) defines a probability measure P on the line. In other words if we define P as above then P satisfies the axioms for a probability measure. Also it follows from the second proof that the new random variable Y (with probabilities defined using Equation (1)) is *continuous* with a new density function related to the density function of the original random variable X in a simple way via the inverse $g(y)$ of the function $h(x)$. The fact that Y as defined by (1) is continuous also follows from the either proof of the next theorem. The point of this lecture is to see how this works and to show you how to make explicit computations in examples.

1 The Theoretical Justification of the Engineer's Way

I will give two proofs of the formula for $f_Y(y)$ The first proof assumes that Equation (1) does in fact define a continuous random variable. It procedes in two stages. First, we compute the cdf F_Y of the new random variable Y in terms of F_X. We then find the density function $f_Y(y)$ of the new random variable Y we differentiate the cdf

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

The second proof uses the "change of variable theorem" from calculus. Don't let the next proof(s) scare you - you won't be tested on them. But they justify the "Engineer's Way", a simple rule to compute the probability density function of the new random variable Y in terms of the probability density function of the original random variable X.

Theorem 1.1 *Suppose X is continuous with probability density function $f_X(x)$. Let $y = h(x)$ with h a strictly increasing continuously differentiable function with inverse $x = g(y)$. Then $Y = h(X)$ defined by (1) is continuous with probability density function $f_Y(y)$ given by*

$$f_Y(y) = f_X(g(y))g'(y) \tag{2}$$

Proof. We will give two proofs of Equation (2). The first proof has the advantage that it is easier to understand and gives a formula for the new cdf as well but involves a tricky point-the appearance of the constant C. To state it loosely, the problem is that we might not have $g(-\infty) = -\infty$ (we state this problem carefully in terms of limits below). The second proof Equation (2) uses the change of variable theorem. It has the advantage of giving a direct computation of $P(c \leq Y \leq d)$. From this formula we see that Equation (1) does in fact define a probability measure and moreover the associated random variable Y is continuous. Indeed, in the second proof we show directly by applying the change of variable formula to $P(a \leq X \leq b)$ that we have

$$P(c \leq Y \leq d) = \int_c^d f_X(g(y))g'(y)dy \tag{3}$$

Equation (3) means that the equation (1) does in fact define a probability measure and the corresponding random variable Y is continuous with probability density function $f_X(g(y))g'(y)$.

First proof

We first compute $F_Y(y)$ in terms of $F_X(x)$. There is a tricky point here. There is no reason why $\lim_{y \to -\infty} g(y) = -\infty$. But the limit does exist (I leave that to you). Suppose $\lim_{y \to -\infty} g(y) = L$.

$$\begin{aligned} F_Y(y) =&P(Y \leq y) = P(-\infty < Y \leq y) = P(-\infty < h(X) \leq y) \\ =&P(L \leq X \leq g(y)) = F_X(g(y)) - F_X(L) = F_X(g(y)) - C. \end{aligned}$$

So we get

$$F_Y(y) = F_X(g(y)) - C \tag{4}$$

where (roughly) $C = F_X(g(-\infty))$. Note that the third equality holds because $g(y)$ is also strictly increasing (because the inverse of a strictly increasing function is strictly increasing) so g preserves inequalities; that is, $a \le b \Rightarrow g(a) \le g(b)$. So apply g to each side of the inequality $h(X) \le y$ to get $g(h(X)) \le h(y)$. But $g(h(X)) = X$ since $g \circ h = Id$ because g is the inverse of h.

Next we differentiate the function on the right of Equation (4) with respect to y using the Chain Rule to get $f_Y(y)$ (since the derivative of the cdf $F_Y(y)$ with respect to y is the pdf $f_Y(y)$).

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}(F_X(g(y)) - C) = F'_X(g(y))g'(y). \tag{5}$$

But since $F'_X(x) = f_X(x)$ for any number x we get $F'_X(g(y)) = f_X(g(y))$ and substituting into the last term of Equation (5) we get

$$f_Y(y) = f_X(g(y))g'(y).$$

This completes the first proof of the "Engineer's Way".

Second proof

Let a, b be real numbers with $a < b$. By definition

$$P(c \le Y \le d) = P(a \le X \le b) \text{ with } a = g(c) \text{ and } b = g(d).$$

But since X is continuous with density function f_X we have

$$P(a \le X \le b) = \int_a^b f_X(x)dx = \int_{g(c)}^{g(d)} f_X(x)dx = \int_c^d f_X(g(y))g'(y)dy.$$

The last inequality is the "change of variable theorem" for definite integrals. So we get: for every $c, d \in \mathbb{R}$ with $c < d$ we have

$$P(c \le Y \le d) = \int_c^d f_X(g(y))g'(y)dy.$$

But this says that $f_X(g(y))g'(y)$ is the probability density function of Y. $\square$

I didn't prove this theorem in class. The previous proofs are probably a little hard for many of you right now but they justify what I called "The Engineer's Way" in class.

The Engineer's Way from Class

Here is the way I stated the "Engineer's Way" in class. Start with $f_X(x)dx$. Substitute $x = g(y)$ for the x in $f_X(x)$ *and* the x in dx to get $f(g(y))dg(y)$. Now use $dg(y) = g'(y)dy$ to get $f(g(y))g'(y)dy$. Then I told you that the function $f(g(y))g'(y)$ multiplying dy is the probability density function of the new (transformed) random variable Y. This is the "Engineer's Way" from class. But the function $f(g(y))g'(y)$ **really is** the density function of the new random variable Y according to the theorem above. So the simple rule works. There is only one problem, **to implement the Engineer's Way given** X, Y **and** h **you have to compute the inverse function** $x = g(y)$ **to** $y = h(x)$. This amounts to solving the equation

$$h(x) = y \tag{6}$$

for x in terms of y. This can be impossible to do. However the functions I give you on tests will be easy to invert.

2 How to Implement the "Engineer's Way"

Here is where you should start reading for the purpose of preparing for tests. I will work out two examples.

2.1 An easy example

Suppose X has the "linear" density so

$$f_X(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

We will make the change of variable $y = \sqrt{x}$. So $h(x) = \sqrt{x}$ so we want to compute the density function of the random variable $Y = \sqrt{X}$. The inverse function to $h(x)$ is given by $x = y^2 = g(y)$. Now we implement the "Engineer's Way".

- Step One: Multiply the density by dx to get $f_X(x)dx = 2xdx$.

- Step Two: find the inverse function $g(y)$ to $h(x) = \sqrt{x}$, so we have to solve Equation (6), that is we have to solve for x in terms of y in the equation

$$\sqrt{x} = y$$

  The solution is clearly $x = y^2$ so $g(y) = y^2$.

- Step Three: Using the formula $x = y^2$ rewrite $f_X(x)dx = 2xdx$ in terms of y. So substituting y^2 for x *in both places* we get

$$f_X(g(y))dg(y) = 2y^2 d(y^2) = 2y^2 2y dy = 4y^3 dy.$$

- Step Four: The "Engineer's Way" tells us the result must be the new density function $f_Y(y)$ of y multiplied by dy and hence

$$f_Y(y)dy = 4y^3 dy \text{ and so } f_Y(y) = 4y^3$$

- Step Five: Find the support of Y, roughly, the domain where Y is nonzero (see Section 4). From Section 4, we know that the support is $[h(0), h(1)] = [\sqrt{0}, \sqrt{1}] = [0, 1]$. On the complement of $[0, 1]$ the $f_Y(y)$ is zero.

So Y has the "cubic density"

$$f_Y(y) = \begin{cases} 4y^3, & 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

2.2   A much more important example

We will use the Engineer's way to prove that "standardizing a general normal random variable" produces a standard normal variable. This is a result we use over and over in the course so it is nice to understand why it is true. Note that we will be using z instead of y in what follows. We will use the change of variable $z = h(x) = \frac{x-\mu}{\sigma}$ hence $x = g(z) = \sigma z + \mu$.

Theorem 2.1 *Suppose $X \sim N(\mu, \sigma^2)$. Then $Z = \frac{X-\mu}{\sigma} \in N(0,1)$.*

Proof.   We have
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(\frac{x-\mu}{\sigma})^2}.$$
Now we apply the "Engineer's Way" step by step.
This is what you need to learn to do.

- Step One: Multiply the density by dx to get $f_X(x)dx = \frac{1}{\sqrt{2\pi}\sigma} e^{-(\frac{x-\mu}{\sigma})^2} dx$.

- Step Two: find the inverse function $g(z)$ to $h(x) = \frac{x-\mu}{\sigma}$, so we have to solve Equation (6), that is we have to solve for x in terms of y in the equation
$$\frac{x - \mu}{\sigma} = z$$
  The solution is clearly $x = \sigma z + \mu$ so $g(z) = \sigma z + \mu$.

- Step Three: Using the formulas $x = \sigma z + \mu$ and $z = \frac{x-\mu}{\sigma}$ rewrite $\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$ in terms of z. So we get (noting since $z = \frac{x-\mu}{\sigma}$), the argument $(\frac{x-\mu}{\sigma})^2$ of the exponential function is in fact just $(z)^2$)
$$f_X(g(z))dz = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(z)^2} d(\sigma z + \mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(z)^2} \sigma dz.$$

- Step Four: Cancel the σ's to get
$$f_X(g(z))dz = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.$$

- Step Five: We have $h(-\infty) = -\infty$ and $h(\infty) = \infty$ so the support of Z is still $\mathbb{R} = (-\infty, \infty)$.

But the whole point of the "Engineer's Way" is that $f_X(g(z))dz$ is the density function $f_Z(z)$ of the new random variable Z multiplied by dz. So
$$f_Z(z)dz = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz.$$

Cancelling the dz's we get
$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, -\infty < z < \infty.$$

But the right-hand side is the density function of a standard normal random variable, so Z has standard normal distribution. $\square$

Remark 2.2 *Don't forget to substitute $g(z)$ for the x in the dx.*

I will now show that the "Engineer's Way" does not always give the right answer if $h(x)$ is not one-to-one.

3 A Quadratic Change of Variable

We will now prove the following theorem that is very important in statistics.

Definition 3.1 *A random variable X is said to have chi-squared distribution with ν degrees of freedom abbreviated $X \sim \chi^2(\nu)$, if*

$$f_X(x) = \begin{cases} \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} & x \geq 0 \\ 0 & otherwise \end{cases} \tag{7}$$

We are now going to prove a theorem which is very important in statistics - the square of a standard normal random variable has chi-squared distribution with one degree of freedom. This amounts to solving §4.4, Problem 71 in the text. In terms of equations

Theorem 3.2

$$Z \sim N(0,1) \Rightarrow Y = Z^2 \sim \chi^2(1).$$

We first note that if $\nu = 1$ (and changing x to y and X to Y in the Equation (7)) we have

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2}\ \Gamma(\frac{1}{2})} \frac{1}{\sqrt{y}} e^{-\frac{y}{2}}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

Substituting $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ in Equation (8) we obtain

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-\frac{y}{2}}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

So we have to get the density function on the right-hand side of Equation (9) when we make the change of variable $y = z^2$ starting with the standard normal density

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Note that the change of variable is two-to-one, so there is no guarantee that the "Engineer's Way" will work and in fact it gives $\frac{1}{2}$ times the correct answer (try it). So the answer is off by a factor of $\frac{1}{2}$. It is no coincidence that the map $h(z)$ is **two**-to-one.

So let's prove the theorem.

Proof. The idea of the proof (what I called the "Careful Way" in class) is to first compute the cdf $F_Y(y)$ of the transformed random variable $Y = Z^2$ in terms of the cdf of the original random variable Z. Recall we have denoted the cdf of the standard normal random variable by $\Phi(z)$. Once we have the cdf $F_Y(y)$ of Y we can get the pdf $f_Y(y)$ by differentiating it with respect to y:

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

Away we go. We have

$$F_Y(y) = P(Y \leq y) = P(Z^2 \leq y) = P(-\sqrt{y} \leq Z \leq \sqrt{y}) = 2\Phi(\sqrt{y}) - 1$$

Here the last equation comes from what I called the "handy formula" for the probability that a standard normal random variable is between $\pm a$:

$$P(-a \leq Z \leq a) = 2\Phi(a) - 1.$$

In fact, the key step is the next-to-last equality. The point is that we can solve the *nonlinear* inequality $y^2 \leq c$ for y easily. Indeed we have

$$y^2 \leq c \iff -\sqrt{c} \leq y \leq c \tag{10}$$

Thus we have our desired expression

$$F_Y(y) = 2\Phi(\sqrt{y}) - 1.$$

Now we have to differentiate this equation with respect to y using that the derivative of Φ at z is the standard normal density $\Phi'(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$.

First we get without effort

$$\frac{d}{dy} F_Y(y) = \frac{d}{dy} [2\Phi(\sqrt{y}) - 1] = 2 \frac{d}{dy} [\Phi(\sqrt{y})].$$

Now comes the hard part - the chain rule part. We use the chain rule to get the first equality below. In the third term below the notation $(e^{-\frac{z^2}{2}})|_{z=\sqrt{y}})$ means you take the function $e^{-\frac{z^2}{2}}$ and evaluate it at $z = \sqrt{y}$ to get $e^{-\frac{y}{2}}$ which gives the fourth term. So

$$2 \frac{d}{dy} [\Phi(\sqrt{y})] = 2\Phi'(\sqrt{y}) \frac{d}{dy} [\sqrt{y}] = 2[\frac{1}{\sqrt{2\pi}} (e^{-\frac{z^2}{2}}|_{z=\sqrt{y}})] [\frac{1}{2} \frac{1}{\sqrt{y}}] = 2[\frac{1}{2} \frac{1}{\sqrt{y}}] [\frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}}] = \frac{1}{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}}.$$

But this last expression is the pdf of a chi-squared random variable with one degree of freedom (compare with Equation (9)).

$\square$

I don't expect many people to understand the next remark but I'll put it in for those people who have taken some more advanced math courses.

Remark 3.3 *The fact that the support (see the next definition) of f_Y is $[0, \infty)$ is because the support of f_Z was $(-\infty, \infty)$ and the image of $(-\infty, \infty)$ under the map $h(z) = z^2$ is $[0, \infty)$. The support of the new density is always the image of the support of the old density under the change of variable map.*

4 How the End-Points Change under $h(x)$

We begin this section with a very useful definition (you will learn the definition of closure in Math 410).

Definition 4.1 *The support of a function f on the real line is the closure of set of all points x where $f(x)$ is nonzero.*

In all our examples of density functions the set of points where f is nonzero is either a single closed interval $[a, b]$, a single open interval (a, b) , a single half open interval $(a, b]$ or $[a, b)$ or $[0, \infty), (0, \infty), (-\infty, \infty)$. Taking the closure just adds the missing end points. So, in the first four cases the support is $[a, b]$, in the next two the support is $[0, \infty)$ and the last it is $(-\infty, \infty)$.

We now state

Theorem 4.2 *Suppose the density function $f_X(x)$ has support the interval $[a, b]$ and $y = h(x)$ with h strictly increasing. Then the support of the density function $f_Y(y)$ is the (image) interval $[h(a), h(b)]$.*

Proof.   The next proof is not quite correct but it gives the main idea. For convenience we assume $h'(x)$ and hence $g'(y)$,is never zero (this isn't true for the strictly increasing function $h(x) = x^3$ but I want to keep things easy here). Suppose also for convenience that we are in the first case. Then since $f_Y(y) = f_X(g(y)))g'(y))$ and $g'(y)$ is never zero we find that

$$f_Y(y) \neq 0 \iff f_X(g(y)) \neq 0 \iff a \leq g(y) \leq b \iff h(a) \leq h(g(y)) \leq h(b).$$

The last step follows since h(y) is (strictly) increasing iand any increasing function preserves inequalities. But h(g(y)) =y since g is the inverse function to h. Hence we obtain

$$f_Y(y) \neq 0 \iff h(a) \leq y \leq h(b).$$

In other words the set where $f_Y(y)$ is nonzero is exactly the closed interval $[h(a), h(b)]$.
$\square$

Remark 4.3 *The point is that h maps the set (no matter what it is) where f_X is nonzero to the set where $f_X \circ g$ is nonzero.*

5 Linear change of a uniform random variable

We now do an example. Suppose $X \sim U(0, 1)$, that is X has uniform distribution on $[0, 1]$ so

$$f_X(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Let $y = h(x) = ax + b$ with $a > 0$ so $x = g(y) = \frac{y-b}{a}$. So we are making the linear change of random variable $Y = aX + b$. So the support of f_X is the interval $[0, 1]$. Now $h(0) = b$ and $h(1) = a + b$ so the support of the density function of the transformed

random variable $Y = aX + b$ is $[a, a + b]$ by Theorem 4.2. We now compute the density of Y. Assuming $y \in [0, 1]$ we have

$$f_Y(y) \cdot dy = f_X(\frac{y - b}{a}) \cdot d(\frac{y - b}{a}) = 1 \cdot \frac{1}{a} \cdot dy.$$

So

$$f_Y(y) = \begin{cases} \frac{1}{a}, & b \leq y \leq a + b \\ 0 & \text{otherwise} \end{cases}$$

We have proved

Theorem 5.1 *The linear change $y = ax + b$ of a random variable X with uniform distribution on $[0, 1]$ produces a random variable Y with uniform distribution on $[a, a+ b]$.*

6   The Law of the Unconscious Statistican

Theorem 6.1 *Suppose X is a continuous random variable with density $f_X(x)$ with support $[a, b]$. Suppose we change variables to $Y = h(X)$. Then the expected value of the new random variable Y can be computed from the density of the original random variable X according to the formula*

$$E(Y) = \int_a^b h(x) f_X(x) dx. \tag{11}$$

The theorem gets its name because a statistician who didn't know what he was doing would get the right answer by plugging $h(x)$ into the integral on the right-hand side of Equation (11) and would thereby "unconsciously" compute the expected value of the new random variable Y.

One of the main points of the theorem is that you can compute $E(Y)$ *without* computing $f_Y(y)$. I want to emphasize that given $y = h(x)$ it can be impossible to solve for the inverse function $x = g(y)$ so you can't use the "Engineer's Way". Even in this case you can still compute $E(Y)$.

7   Comparison with the Discrete Case

The corresponding result for how the probability mass function changes under a one-to-one change of variable $y = h(x)$ is very easy.

Theorem 7.1 *Suppose X is a discrete random variable with probability mass function $p_X(x)$. Suppose $h(x)$ is a one-to-one function. Put $Y = h(X)$. Then*

$$p_Y(y) = p_X(g(y)). \tag{12}$$

There is no factor of $g'(y)$ multiplying $p_X(g(y))$ in the discrete case.

Proof. By definition

$$p_Y(y) = P(Y = y) = P(h(X) = y) = P(X = g(y)) = p_X(g(y)).$$

$\square$

The reason the continuous case is so difficult is because in the continuous case

$$P(Y = y) = 0 \text{ for all } y.$$

The density function $f_Y(y)$ *does not have a description as a probability of an event involving* Y.

Also *it is not true in the discrete case that*

$$p_Y(y) = \frac{d}{dy} F_Y(y).$$

So there is no "chain rule" way to compute $p_Y(y)$.