Determining the new probability density function (PDF) after some transformation of a PDF

Question asked:

There is a number generator (abbreviated to NG) that generates a number "n" based on a specific PDF, which I will call P(x). Whatever the NG spits out you plug into another function I will call f(x). (This can be anything, x^2, 1/x, and so on). What will the new PDF that represents the distribution of the output of f(x) be?

Solving:

First things first, NAMES. The PDF of the NG is called P(x) and the output of the NG is called "n", f(x) is the function that takes in "n" and gives out another number that I will call "u" (basically, u=f(n)), and the new PDF that shows the probability density of "u". I will call this new PDF G(x).

I will start with what I believe is the most important concept needed to solve this question. Say f(x)=x^2. If the NG spits out 3, then "u" is 9. If the NG spits out 4, then "u" is 16. Now, if the NG generates something between 3 and 4, then "u" is between 9 and 16. If we generalize this pattern we find that if "n" is between some two numbers "a" and "b" then "u" is between f(a) and f(b).

Here, we have to be careful. This only applies if f(x) is continuous between "a" and "b", and if the gradient of f(x) does NOT flip signs between "a" and "b". (You can imagine that the output of n^2 between -5 and 5 is NOT 25 to 25, but 0 to 25. This happens because of the gradient flip at 0).

So far we are at:

If a ≤ n ≤ b, the f(a) ≤ u ≤ f(b), given that f(x) is continuous between "a" and "b", and does not flip gradient in that domain. To proceed we will make one more assumption: the probability that "n" is between "a" and "b" = the probability that "u" is between f(a) and f(b). By continuing that line of logic, we can deduce that the probability that "u" is between any two values "c" and "d" = the probability that "n" is between $f^{-1}(c)$ and $f^{-1}(d)$. From this we can say that:

$$\int_c^d G(x)dx = \int_{f^{-1}(c)}^{f^{-1}(d)} P(x)dx$$

With 3 assumptions so far. But this isn't quite right. Say that our f(x)=-x, then $f^{-1}(x)$=-x. In that case when you plug in c and d, they change order, and what I mean by that is that if c is smaller than d, then -d is smaller than -c meaning that the integral on the right side of the equation is negative in that case and we know that probabilities cannot be negative, so we have to add an absolute sign to it.

$$\int_c^d G(x)dx = \left|\int_{f^{-1}(c)}^{f^{-1}(d)} P(x)dx\right|$$

Great so far. Now, how do we turn this into a formula for G(x)? It would seem best to use derivation on both sides, but the situation on the right is a bit more complex. It might not be immediately clear what will happen when you derive the right side, so let's get rid of the integral using a few steps:

We say that d=c+s

$$\int_{c}^{c+s} G(x)dx = \left|\int_{f^{-1}(c)}^{f^{-1}(c+s)} P(x)dx\right|$$

Then we divide both sides by s

$$\frac{1}{s}\int_{c}^{c+s} G(x)dx = \frac{1}{s}\left|\int_{f^{-1}(c)}^{f^{-1}(c+s)} P(x)dx\right|$$

Then we set a limit for s to approach 0

$$\lim_{s\to 0}\frac{1}{s}\int_{c}^{c+s} G(x)dx = \lim_{s\to 0}\frac{1}{s}\left|\int_{f^{-1}(c)}^{f^{-1}(c+s)} P(x)dx\right|$$

The I can substitute in Gi(x) as the indefinite integral of G(x) with respect to x and Pi(x) as the indefinite integral of P(x) with respect to x.

$$\lim_{s\to 0}\frac{Gi(c+s)-Gi(c)}{s} = \lim_{s\to 0}\frac{\left|Pi\left(f^{-1}(c+s)\right)-Pi(f^{-1}(c))\right|}{s}$$

There we go, the definition of a derivative. Also, since c is just a random number, I can replace it with x.

$$\frac{d}{dx}Gi(x) = \lim_{s\to 0}\frac{\left|Pi\left(f^{-1}(x+s)\right)-Pi(f^{-1}(c))\right|}{s}$$

$$G(x) = \lim_{s\to 0}\frac{\left|Pi\left(f^{-1}(x+s)\right)-Pi(f^{-1}(c))\right|}{s}$$

So, the right side also looks like the derivative definition (in this case, the function being derived is "Pi(f^{-1}(x))") but there is that absolute value there. It is important to ask how it will affect the derivative. See without the absolute value there, if Pi(f^{-1}(x)) is ascending at point x then the x+s part will be greater than the x part and the difference will be positive, hence the derivative will be positive. If however Pi(f^{-1}(x)) is descending at point x, then the x part is greater than the x+s part and the difference will be negative hence, the derivative is negative. If you think about it that way, then adding an absolute sign there means that the difference will always be positive and so the derivative will always be positive and so we can say that:

$$G(x) = \left|\frac{d}{dx}Pi(f^{-1}(x))\right|$$

Side note:

from this we can conclude that:

$$\int_{f^{-1}(c)}^{f^{-1}(d)} P(x)dx = \int_{c}^{d} P(f^{-1}(x))dx$$

End side note

From here we can proceed as follows:

Using the chain rule: $\frac{d}{dx} f\big(g(x)\big) = f'\big(g(x)\big) * g'(x)$, we can say that

$$G(x) = \left|\frac{d}{dx} Pi(f^{-1}(x))\right|$$

$$G(x) = \left|Pi'\big(f^{-1}(x)\big) * \frac{d}{dx} f^{-1}(x)\right|$$

Pi' is the derivative of the integral of P, so we can put that in

$$G(x) = \left|P\big(f^{-1}(x)\big) * \frac{d}{dx} f^{-1}(x)\right|$$

Next, Since P(x) is a PDF and probabilities are never negative, we are certain that P(x) can never produce a value smaller than 0 and so we can factor it out of the absolute value.

$$G(x) = P\big(f^{-1}(x)\big) * \left|\frac{d}{dx} f^{-1}(x)\right|$$

And there we go. We can still try to simplify it a bit further, but it is in usable form as is. Do beware how you deal with $P(f^{-1}(x))$ though, if P(x) = 1, 1 ≤ x ≤ 2, and is 0 everywhere else then $P(f^{-1}(x)) = 1$, $1 \le f^{-1}(x) \le 2$, and is 0 everywhere else. Then you will have to use that to find the new domain. Example: say f(x)=1/x, then $f^{-1}(x)=1/x$, so the domain of $P(f^{-1}(x))$ is 1 ≤ 1/x ≤ 2, which becomes 1/2 ≤ x ≤ 1. (if you are unsure about what just happened you can try to look up some help online).

Now, you can follow along for some more simplification but after that we mention our 3 assumptions from the beginning and then discuss some insight.

Some simplifications: --

As for the remaining derivative, inverse functions are mirror images of their corresponding functions along the y=x line on a plane. It means that $f^{-1}(x)$ is the same as f(x) if you exchange the x and y axes. This means the gradient of the inverse function is 1/gradient of the normal function at the "same point" BUT, even the inputs and outputs of the functions are flipped when inverted so the "same point" has flipped x and y coordinates too. Look, the whole thing is a bit hard to keep track of while doing the manipulations so, I'll rename things. In other words:

At input value a, f(x) produces an output value b=f(a), and the gradient there is dy/dx, and

At input value b, $f^{-1}(x)$ produces and output value a=$f^{-1}(b)$, and the gradient there is dx/dy.

I know that this doesn't help much, but it is all I can think of. If you have a better way of keeping track of what is input and what is output then use that for the next bit.

So, $\left|\frac{d}{dx}f^{-1}(x)\right|$ is calculating the gradient of inverse f(x) at b. We can then say that that equals 1/gradient of normal function when you input a which is $f^{-1}(b)$ in the case of the original equation $f^{-1}(x)$. Note that we said the GRADIENT of f(x) when you input $f^{-1}(x)$, so you differentiate f(x) and then plug in the $f^{-1}(x)$, not the other way around. So we can deduce that:

$$\left|\frac{d}{dx}f^{-1}(x)\right| = 1/|f'(f^{-1}(x))|$$

Putting that back into the original equation, we get

$$G(x) = \frac{P\left(f^{-1}(x)\right)}{|f'(f^{-1}(x))|}$$

Now let's discus our three assumptions:

1. f(x) is continuous along the domain you are interested in. Say for example that P(x) = 1, 1 ≤ x ≤ 2, and is zero everywhere else, then we may use f(x)=1/x because it is continuous between 1 and 2, and P(x) is zero at the point where 1/x is discontinuous. In the case that you have P(x) = 1/2 ranging from -1 to 1 and f(x)=1/x, then you can use a trick saying that p(x) = 1/2 ranging from -1 to 0 + ½ ranging from 0 to 1. This way each part will be transformed separately and you will get a G(x) = so and so ranging so and so + so and so ranging so and so.
2. The gradient of f(x) does NOT flip signs in the domain you are interested in. This problem can be solved with the same trick used above. Just find the point where the gradient flips and split P(x) around it.
3. The probability that "n" is between "a" and "b" = the probability that "u" is between f(a) and f(b). I do not have the math skills to prove this. I have tried many examples trying to find a case where this does not hold and the only situations I found were when one of the first two assumptions was broken. Even then you can still fix it with the same trick used above.

In this last bit I wish to discuss some insight. After looking at a few examples I have to appreciate the parts of this equation

$$G(x) = P\left(f^{-1}(x)\right) * \left|\frac{d}{dx}f^{-1}(x)\right|$$

The first part $P\left(f^{-1}(x)\right)$ seems to take the original PDF and stretch it and squeeze it along the x axis based on – strangely – the gradient of f(x). Imagine f(x)=x^2, the value at p(3) is moved to become G(9), and the value at p(4) is moved to be at G(16) and everything between them is stretched. I believe this relation to the gradient is because the greater the gradient of f(x), the farther apart two nearby points will move. The second part $\left|\frac{d}{dx} f^{-1}(x)\right|$ seems to stretch and squeeze P(x) in the vertical direction, and that makes sense. Since the total area cannot be greater than 1, then the areas that are stretched horizontally have to be squeezed vertically by some similar amount, and the relation to the gradient of f(x) is clearer in this case.