6: Transformations of Random Variables

PSTAT 120A: Summer 2022

Ethan P. Marzban

July 11, 2022

University of California, Santa Barbara

Where We've Been

- Axioms of Probability, Probability Spaces, Counting
- Conditional Probabilities, independence, etc.
- Basics of Random Variables (classification, p.m.f., c.m.f., moments)
- Discrete Distributions
- Continuous Distributions

Notational Convenience

- First, it will be helpful to introduce some new notation:

Definition: Indicators

The **indicator function** associated with a particular event A is defined as

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \in A^{\complement} \end{cases}$$

Often, the (ω) will be dropped and we will simply write

$$\mathbb{1}_A = \begin{cases} 1 & \text{if } A \\ 0 & \text{otherwise} \end{cases}$$

- For instance, we can write the p.d.f. of the exponential distribution quite succinctly as

$$f_X(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}_{\{x \geq 0\}}$$

Transformations of Random Variables

Introduction

- Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$.
- Consider now a function $g : \mathbb{R} \to \mathbb{R}$.
- Recall from precauclus that $(g \circ X)$ is itself a function.
- Specifically, $(g \circ X) : \Omega \to \mathbb{R}$; that is, $(g \circ X)$ is a random variable!
- Often times we will denote this random variable by $g(X)$; for instance, we will start with a random variable X and define a new random variable $Y := g(X)$. In other words, Y is a **transformation** of X.

Introduction

- Functions of random variables? That sounds awfully abstract...
- Well, suppose T_C measures the temperature as measured in Centigrade of some town, and that $T_C \sim \mathcal{N}(0, 3)$. We may ask ourselves: "what is the distribution of temperatures as measured in Fahrenheit?" That is, if T_F measures the temperature in Fahrenheit then

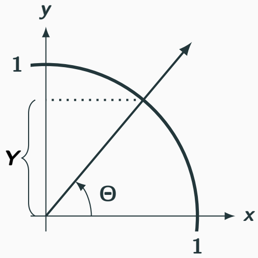
$$T_F = g(T_C) \quad \text{where} \quad g(t) = \frac{9}{5}t + 32$$

  and we seek the distribution of T_F.

Reference Example 2

Reference Example:

A particle is fired from the origin into the first quadrant such that the angle of trajectory (as measured w.r.t. the $x-$axis) is uniformly distributed. The particle travels unobstructed until it collides with circlar ring, placed at a constant radius of 1 away from the origin. We let Y denote the $y-$coordinate of the point of collision (see figure below).



What is the distribution of Y?

The Setup:

- Let Θ denote the angle of trajectory; then $\Theta \sim \text{Unif}[0, \pi/2]$.
- We can see that

$$\sin(\Theta) = \frac{Y}{1} \implies Y = \sin(\Theta)$$

- So, we have $Y = g(\Theta)$ where $g(t) = \sin(t)$, and we want the distribution of Y.

Expectation?

- Finding $\mathbb{E}[Y]$ is relatively easy:

$$\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}[\sin(\Theta)] \\
&= \int_{-\infty}^{\infty} \sin(\theta) \cdot f_\Theta(\theta)\ \mathrm{d}\theta \\
&= \int_0^{\pi/2} \sin(\theta) \cdot \frac{2}{\pi}\ \mathrm{d}\theta \\
&= \frac{2}{\pi} \int_0^{\pi/2} \sin(\theta)\ \mathrm{d}\theta \\
&= \frac{2}{\pi} \left[\cos(0) - \cos(\pi/2)\right] = \frac{2}{\pi}
\end{aligned}$$

- Okay.... now what?
  - $Y \sim \mathcal{N}(2/\pi, 1)$?
  - $Y \sim \mathrm{Exp}(\pi/2)$?
- The point is: **expectation isn't enough!**

A Question

- This leads us to a broader question: what information *is* enough for us to determine the distribution of Y?
- There are at least two answers to this question. The answer we will discuss today is: "pmf's and pdf's."
    - The other answer will be discussed next week, and will lead us to something called Moment-Generating Functions. Again, more on that later.
- What I mean is this: if we can somehow find that the probability density function (p.d.f.) of Y is given by

  $$f_Y(y) = \left(\frac{\pi}{2} \right) e^{-\frac{y\pi}{2}} \cdot \mathbb{1}_{\{y \geq 0\}}$$

  then I can immediately say $Y \sim \text{Exp}(\pi/2)$.

The Question

- So, here is our task: given a random variable X that has p.d.f. $f_X(x)$, we seek to find the p.d.f. $f_Y(y)$ of the random variable $Y := g(X)$.

The Change of Variable Formula

Suppose X is a random variable with p.d.f. $f_X(x)$ and state space S_X. Further suppose g is a function that is bijective over S_X, and that $Y := g(X)$. Then the p.d.f. of Y is given by

$$f_Y(y) = f_X\left[g^{-1}(y)\right] \cdot \left| \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) \right|$$

- We'll prove this in a bit.

Reference Example (again)

- Let's return to our reference example:

$$\Theta \sim \text{Unif}(0, \pi/2)$$
$$g(t) := \sin(t)$$
$$Y := g(\Theta)$$

- First, we can check that $S_\Theta = [0, \pi/2]$ and $g(t)$ is indeed bijective over S_Θ with inverse given by

$$g^{-1}(y) = \arcsin(y)$$

- Thus, by calculus:

$$\left|\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right| = \left|\frac{\mathrm{d}}{\mathrm{d}y} \arcsin(y)\right| = \left|\frac{1}{\sqrt{1-y^2}}\right| = \frac{1}{\sqrt{1-y^2}}$$

Reference Example (again)

- Hence:

$$
\begin{aligned}
f_Y(y) &= f_\Theta(\arcsin y) \cdot \frac{1}{\sqrt{1-y^2}} \\
&= \begin{cases} \frac{2}{\pi} & \text{if } 0 \leq \arcsin y \leq \pi/2 \\ 0 & \text{otherwise} \end{cases} \cdot \frac{1}{\sqrt{1-y^2}} \\
&= \begin{cases} \frac{2}{\pi\sqrt{1-y^2}} & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

- By the way, $S_Y = g(S_\Theta)$.
- We can also verify our expectation computation from before:

$$
\mathbb{E}[Y] = \int_0^1 \frac{2y}{\pi\sqrt{1-y^2}} \, \mathrm{d}y = \frac{2}{\pi} \checkmark
$$

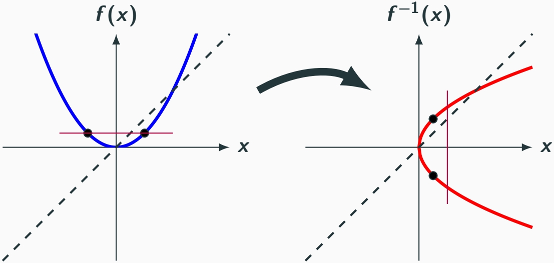
Example (Chalkboard)

Suppose X has p.d.f. $f_X(x)$ and suppose $Y = aX + b$ for $a, b > 0$. Derive an expression [in terms of $f_X(x)$, a, and b] for the p.d.f. $f_Y(y)$ of Y.

Example (cont'd)

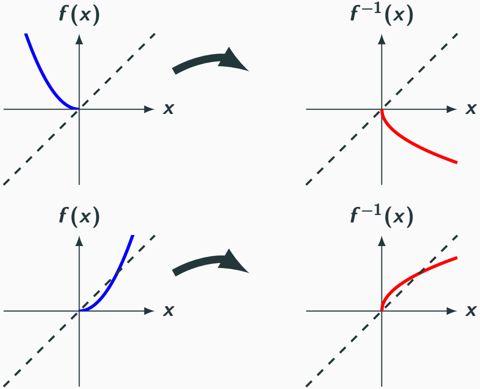
Non-bijective Functions

- Note that the Change of Variable formula requires g to be bijective over the state space of X. What do we do if this is not the case? For instance, suppose $X \sim \mathcal{N}(0,1)$ and $Y = X^2$.
- Firstly, why does the Change of Variable formula not work for non-bijective functions?
- Answer: invertibility. Recall (from Precalculus) that inverting a function is equivalent to performing a reflection about the line $y = x$. Thus, if a function $f(x)$ violates the horizontal line test [i.e. fails to be bijective] its reflection about $y = x$ will fail the *vertical* line test **and not even be a function!**



Non-bijective Functions (cont'd)

- But, a fact that we can leverage is: restrictions of non-bijective functions can be bijective!
- For instance, $f(x) = x^2$ when restricted to only $x \in (-\infty, 0]$ is bijective; similarly, $f(x) = x^2$ when restricted to only $x \in (0, \infty)$ is also bijective:



Non-bijective Functions (cont'd)

- So, let us return to our problem of finding the density of $Y := g(X)$ when g is not bijective over S_X.
- Our approach is as follows:
  1. Partition S_X into smaller subintervals $S_X^{(i)}$ such that g is bijective over each $S_X^{(i)}$.
  2. Apply the Change of Variable Formula to each $S_X^{(i)}$, to obtain a series of partial p.d.f.'s $f_Y^{(i)}(y)$ valid over $S_X^{(i)}$
  3. Combine these partial p.d.f.'s into a single piecewise-defined p.d.f. through addition.
  4. CAVEAT: Only add p.d.f's that are valid over the same domain. If you obtain two sub-pdf's $f_Y^{(1)}(y)$ and $f_Y^{(2)}(y)$ that ar valid over two separate domains, say $S_Y^{(1)}$ and $S_Y^{(2)}$, then do not add these but rather simply combine them into the same p.d.f. using piecewise notation:

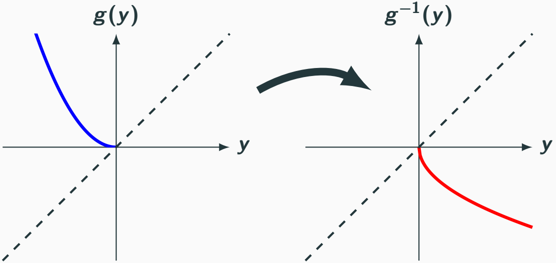
$$f_Y(y) = \begin{cases} f_Y^{(1)}(y) & \text{if } y \in S_Y^{(1)} \\ f_Y^{(2)}(y) & \text{if } y \in S_Y^{(2)} \\ \vdots & \end{cases}$$

     This is important for HW05 question 2.

Worked-Out Example: If $X \sim \mathcal{N}(0,1)$, find the distribution of $Y := X^2$.

- The state space of X is $S_X = \mathbb{R}$; clearly $g(x) = x^2$ is not bijective over S_X. Thus, we partition S_X into
$$S_X^{(1)} := (-\infty, 0]\ ,\ S_X^{(2)} := (0, \infty)$$
because $g(x) = x^2$ *is* bijective over $S_X^{(1)}$ and $S_X^{(2)}$ respectively.

- For $x \in S_X^{(1)}$ we have $g^{-1}(y) = -\sqrt{y}$:



Thus,

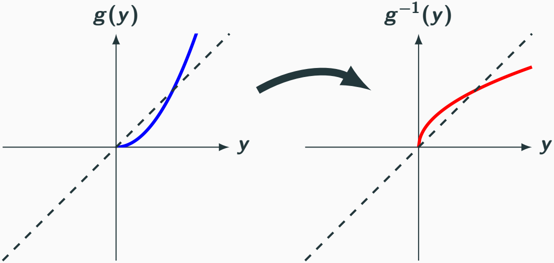
$$\left|\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right| = \left|-\frac{1}{2\sqrt{y}}\right| = \frac{1}{2\sqrt{y}}$$

and so

$$f_Y^{(1)}(y) = \phi(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}$$

Worked-Out Example (cont'd)

- For $x \in S_X^{(2)}$ we have $g^{-1}(y) = \sqrt{y}$:



Thus,

$$\left|\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right| = \left|\frac{1}{2\sqrt{y}}\right| = \frac{1}{2\sqrt{y}}$$

and so

$$f_Y^{(2)}(y) = \phi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}$$

Worked-Out Example (cont'd)

- So, putting everything together,

$$\begin{aligned} f_Y(y) &= f_Y^{(1)}(y) + f_Y^{(2)}(y) \\ &= \phi(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + \phi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} \\ &= \phi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} + \phi(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{y}} \phi(\sqrt{y}) \\ &= \frac{1}{\sqrt{y}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \end{aligned}$$

Worked-Out Example: If $X \sim \mathcal{N}(0,1)$, find the distribution of $Y := X^2$.

- State Space?
- $S_Y(y) = g(S_X)$.
- So, in our example, $S_Y = (\mathbb{R})^2 = [0, \infty)$ meaning

$$f_Y(y) = \frac{1}{\sqrt{y}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \cdot \mathbb{1}_{\{y \geq 0\}}$$

- (This is a special case of the so-called χ^2 **distribution**, a very useful distribution you will return to throughout your Statistics Careers!)

The C.D.F. Method

- Now, there is another approach to identifying the distributions of transformed random variable. It stems from the following fact:

Fact

CDF's uniquely determine distributions. In other words, if X and Y are random variables with c.d.f.'s F_X and $F_Y(y)$ respectively, then $F_X(t) = F_Y(t)$ for all y implies that X and Y follow the same distribution [sometimes written $X \stackrel{\mathrm{d}}{=} Y$]

The C.D.F. Method

- To see this in action, let's return to our χ^2 example.
- The c.d.f. of Y is given by

$$F_Y(y) := \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(|X| \leq \sqrt{y}) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y})$$

- Since the c.d.f. of X is simply $\Phi(\cdot)$, we have

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

which can be further cleaned up by noting

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

Therefore,

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

- To find the p.d.f. of Y we differentiate and apply the chain rule:

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}\left[2\Phi(\sqrt{y}) - 1\right] = 2 \cdot \frac{1}{2\sqrt{y}} \cdot \phi(\sqrt{y}) = \frac{1}{\sqrt{y}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{y})^2}$$

[recall that ϕ, the lowercase version of Φ, denotes the standard normal density]

Another Example

Let $X \sim \text{Exp}(\lambda)$ and $Y = X + c$ for some $c > 0$. Find the p.d.f. of Y using the c.d.f. method

- First note $S_Y = [c, \infty)$
- So, for $y \in [c, \infty)$ we have that the c.d.f. of Y at y is

$$F_Y(y) = \mathbb{P}(X + c \leq y) = \mathbb{P}(X \leq y - c) = 1 - e^{-\lambda(y-c)}$$

- Differentiating yields

$$f_Y(y) = \frac{1}{\lambda} e^{-\lambda(y-c)}$$

meaning, putting everything together,

$$f_Y(y) = \begin{cases} \frac{1}{\lambda} e^{-\lambda(y-c)} & \text{if } y \geq c \\ 0 & \text{otherwise} \end{cases}$$

which is a special case of what is sometimes known as the **two-parameter exponential distribution**.

Paying our Dues

- Finally, it is time to pay one of our dues and prove something we said we'd prove:

Theorem: Standardization

If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Z := \left(\frac{X - \mu}{\sigma}\right)$, then $Z \sim \mathcal{N}(0, 1)$

Proof.

- Let's use the change of variable formula.

$$g(t) = \frac{t - \mu}{\sigma} \implies g^{-1}(t) = \sigma t + \mu \implies \frac{\mathrm{d}}{\mathrm{d}t} g^{-1}(t) = \sigma$$

- So

$$f_Z(z) = \sigma \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left\{-\frac{1}{2\sigma^2}\left([\sigma z + \mu] - \mu\right)^2\right\}$$

- Once the dust settles, you can see that $f_Z(z) = \phi(z)$ thereby showing $Z \sim \mathcal{N}(0, 1)$.

$\square$

Proving the Change of Variable Formula

- There are two cases we must consider: when g is strictly increasing and when g is strictly decreasing.
- If g is strictly increasing, then g^{-1} will also be strictly increasing and its derivative will be positive.
- Now, let's find the p.d.f. of $Y := g(X)$ using the c.d.f. method:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X\left[g^{-1}(y)\right]$$

  Note that we are guaranteed the existence of g^{-1} by our bijectivity assumption on g.
- Now, we differentiate w.r.t. y:

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_X\left[g^{-1}(y)\right] = f_X[g^{-1}(y)] \cdot \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)$$

  where we have applied the chain rule.

Proving the Change of Variable Formula

- The other case to consider is when g is strictly decreasing.
- If g is strictly decreasing, then g^{-1} will also be strictly decreasing and its derivative will be negative.
- Now, let's find the p.d.f. of $Y := g(X)$ using the c.d.f. method:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \geq g^{-1}(y)) = 1 - F_X\left[g^{-1}(y)\right]$$

Key Point: We had to flip the sign of the inequality because we applied a decreasing function to both sides!

- Now, we differentiate w.r.t. y:

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}\left\{1 - F_X\left[g^{-1}(y)\right]\right\} = f_X[g^{-1}(y)] \cdot \left[-\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right]$$

where we have again applied the chain rule.

Proving the Change of Variable Formula

- So, in piecewise-defined form, we have

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)] \cdot \left[\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right] & \text{if } \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) > 0 \\ f_X[g^{-1}(y)] \cdot \left[-\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right] & \text{if } \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) > 0 \end{cases}$$

$$= f_X[g^{-1}(y)] \cdot \begin{cases} \left[\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right] & \text{if } \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) > 0 \\ \left[-\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right] & \text{if } \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) > 0 \end{cases} \quad = f_X[g^{-1}(y)] \cdot \left|\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right|$$

  which is precisely the Change of Variable formula.