Topic 02: Transformations

Ethan P. Marzban University of California, Santa Barbara PSTAT 120B

Outline

1. Univariate Transformations

2. Method of Distribution Functions (CDF Method)

3. Method of Transformations (Change of Variable Formula)

4. Method of Moment-Generating Functions (MGF Method)

5. Multivariate Transformations

Leadup

- Recall, from PSTAT 120A, that given an appropriate probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we can think of a **random variable** X as a mapping:

$$X : \Omega \to \mathbb{R}$$

- Additionally, recall the following fact from precalculus: given a mapping $f_1 : A \to B$ and another mapping $f_2 : B \to C$, then $(f_2 \circ f_1) : A \to C$.
- This means, given a function $g : \mathbb{R} \to \mathbb{R}$ and a random variable $X : \Omega \to \mathbb{R}$, we have $(g \circ X) : \Omega \to \mathbb{R}$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Leadup

- In this way, we can think of $(g \circ X)$ as a random variable itself!
    - For example, given a random variable X, then the quantity X^2 will also be a random variable.
- Another way of saying this: functions of random variables are themselves random variables.
- "Functions of random variables?" That sounds awfully abstract...
- But, if we think about it a bit more, this isn't as abstract as it may seem!

Leadup

- For example, let H_I denote the height of a randomly-selected individual as measured in inches, and suppose $H_I \sim \mathcal{N}(70, 2)$.
- Let H_C denote the height of a randomly-selected individual as measured in <u>centimeters</u>.
- Clearly, the random variables H_I and H_C are related: specifically, $H_C = g(H_I)$ where $g(t) = 2.54 * t$ [since this is the conversion formula between inches and centimeters].
  - So, **<u>unit conversion</u>** is a fairly simple example of one way transformations (i.e. taking functions of random variables) can be useful.

Leadup

- Transformations can also be used to **summarize** data.
- For example, consider a sequence $\{X_i\}_{i=1}^n := X_1, \cdots, X_n$ of random variables.
  - By the way, I'll be using this notation a lot: $\{X_i\}_{i=1}^n$ is a shorthand for $X_1, \cdots, X_n$.
- The **sample mean** $\overline{X}_n := n^{-1} \sum_{i=1}^n X_i$ [which you hopefully saw in PSTAT 120A!] is actually a *function* of the original sequence of random variables, and is hence an example of a transformation.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Leadup

- Now, these two examples indicate that there are perhaps two sub-cases to consider: transformations of *single* random variables, and transformations of *multiple* random variables.
  - We often refer to a transformation of a single random variable as a **univariate transformation**, and a transformation of multiple random variables as a **multivariate transformation**.

- For simplicity's sake, let's start off with univariate transformations.
  - Specifically, given a random variable Y and a function $g : \mathbb{R} \to \mathbb{R}$, we will seek to explore properties of the random variable $U := g(Y)$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Univariate Transformations

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Goal

Goal

Given a random variable Y and a function $g()$, we seek to describe the random variable $U := g(Y)$.

- What do we mean by "describe" the random variable U?
- Well, there are a couple of things we could seek to do. First, we could try to compute $\mathbb{E}[U]$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

LOTUS

- It turns out... we've already done that!
- Specifically, since $U := g(Y)$, we have that $\mathbb{E}[U] = \mathbb{E}[g(Y)]$.
- The **<u>Law of the Unconscious Statistician</u>** (LOTUS), which we saw in PSTAT 120A, tells us

$$\mathbb{E}[g(Y)] = \int_{\mathbb{R}} g(y) f_Y(y) \ \mathrm{d}y$$

- Similar considerations will allow us to compute $\mathrm{Var}(U)$.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Distributions

- Okay, that's useful! But it's not the whole picture.
- Why don't we get a little more ambitious, and seek to find the *distribution* of U?
- First, let me be a little more clear about what I mean by "distribution".
- Sometimes, we can identify a distribution by name (e.g. "Exponential distribution with parameter $\theta = 0.5$", or "Standard normal distribution").
- But, a distribution could just as easily have been described by any of the following:
  - Its **distribution function** (i.e. CDF)
  - Its **density function** (PDF)
  - Its **MGF** (moment-generating function)

Distributions

- For example, suppose I tell you the random variable W has density function given by
$$f_W(w) = 2e^{-2w} \cdot \mathbb{1}_{\{w \geq 0\}}$$
- You would immediately be able to tell me "oh, W follows the Exponential distribution with parameter $\theta = 1/2$."
- This would, in turn, automatically tell you that W has distribution function
$$F_W(w) = \begin{cases} 1 - e^{-2w} & \text{if } w \geq 0 \\ 0 & \text{otherwise} \end{cases}$$
and MGF
$$M_W(t) = \begin{cases} (1 - t/2)^{-1} & \text{if } t < 1/2 \\ \infty & \text{otherwise} \end{cases}$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Distributions

- Similarly, if I tell you that the random variable T has MGF given by

$$M_X(t) = \exp\left\{2t + \frac{1}{2}t^2\right\}$$

you would immediately be able to say

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x-2)^2\right\}$$

and

$$F_X(x) = \Phi\left(x - 2\right); \qquad \Phi(x) := \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,\mathrm{d}z$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Distributions

- Now, what if we have a random variable X whose density is given by

$$f_X(x) = \cos(x) \cdot \mathbb{1}_{\{0 \leq x \leq \pi/2\}}$$

- What is the distribution of X?
- Well… it's just the density above!
- What I mean is this - the distribution of X doesn't have a name, like "Exponential" or "Gamma". But it certainly *has* a distribution!
- All of this is to say: I encourage you to get into the habit of thinking about "distributions" fairly broadly, and thinking of a distribution as either a density function, distribution function, or MGF (or all three).

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Back to Transformations

Goal

Given a random variable Y and a function $g()$, we seek to describe the random variable $U := g(Y)$.

- Now, our discussion on the previous few slides tells us that there are three approaches to achieving our goal above.
- We could go after the density function of U.
- Or we could go after the distribution function of U.
- Or we could go after the MGF of U.
- Indeed, each of these three approaches are what our textbook calls different "methods".

Support

- Before we dive into these three methods, let's talk a bit about **support**.
- Recall that the support (aka "state space") of a random variable Y is the set of all values that Y maps to: i.e. $S_Y := Y(\Omega)$. Equivalently, it's the set of all values y for which the density $f_Y(y)$ is nonzero.
- Then, given a random variable $U := g(Y)$, we have $S_U = g(S_Y)$.
  - That is, the support of a transformed random variable is the image of the original support under the transformation.

- Though this formula seems inoccuous enough, finding the support of a transformed random variable can be trickier than it first appears...

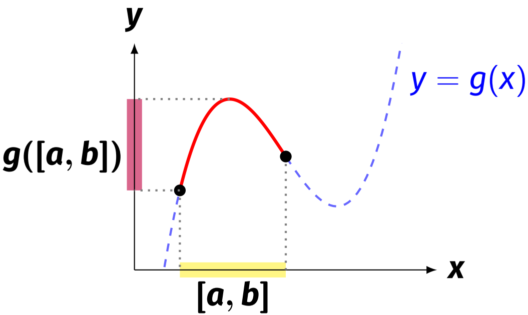
UC SANTA BARBARA
Department of Statistics and Applied Probability

Support

- A simple way I like to think about things is to draw a picture.
- Specifically, let's say we have an interval $[a, b]$ and a transformation $g : \mathbb{R} \to \mathbb{R}$.
- To figure out what $g([a, b])$ looks like, simply graph the function $y = g(x)$, indicate $[a, b]$ on the $x-$axis, and figure out what the corresponding values on the $y-$axis are.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Support



- Note: in general, $g([a, b]) \neq [g(a), g(b)]$!

UC SANTA BARBARA
Department of Statistics and Applied Probability

Clicker Question!

Clicker Question 1

For $A = [0\ , 6]$ and $g(x) = \cos(\pi x)$, what is the correct expression for $g(A)$?

(A) $[0\ ,\ 1]$ (B) $[0\ ,\ 6]$ (C) $[-1\ , 1]$ (D) $\{0\}$
(E) None of the above

Try this On Your Own:

Example

For $A = [-1\ ,\ 1]$ and $g(x) = x^2$, what is the correct expression for $g(A)$?

Method of Distribution Functions (CDF Method)

CDF Method

- Let's consider the following rephrasing of our goal:

Goal

Given a random variable Y and a function $g()$, we seek to derive an expression for $F_U(u) := \mathbb{P}(U \leq u)$, the CDF of U.

- As a concrete example, let $Y \sim \text{Exp}(\theta)$ and let $U := cY$ for a positive constant c.
  - If it helps, you can think of this in terms of our inches-to-centimeter conversion example from the start of this lecture: Y can denote the heights in inches and U can denote the heights in cenimeters.

CDF Method

- Now, we know everything we could want to know about Y.
- Specifically, we have the CDF of Y:

$$F_Y(y) = \begin{cases} 1 - e^{-y/\theta} & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

- So, if we can relate $F_U(u)$ to $F_Y(y)$, we'd be done.
- Note:

$$F_U(u) := \mathbb{P}(U \leq u) = \mathbb{P}(cY \leq u)$$

- Divide through by c:

$$F_U(u) = \mathbb{P}\left(Y \leq \frac{u}{c}\right) = F_Y\left(\frac{u}{c}\right)$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

CDF Method

- So, plugging into our expression for $F_Y(y)$, we have:

$$
\begin{aligned}
F_U(u) &= F_Y\left(\frac{u}{c}\right) \\
&= \begin{cases} 1 - e^{(u/c)/\theta} & \text{if } (u/c) \geq 0 \\ 0 & \text{otherwise} \end{cases} \\
&= \boxed{\begin{cases} 1 - e^{u/(c\theta)} & \text{if } u \geq 0 \\ 0 & \text{otherwise} \end{cases}}
\end{aligned}
$$

- And we're done! We've accomplished our goal, and found an expression for $F_U(u)$, the CDF of U.

Going Further

- Now, in this particular case, we can take things a step further.
- Specifically, doesn't that CDF look awfully familiar?
- Indeed, it is the CDF of the $\text{Exp}(c\theta)$ distribution!
- So, what we've essentially shown is:

Theorem (Closure of Exponential Distribution under Multiplication)

Given $Y \sim \text{Exp}(\theta)$ and a positive constant c, then $(cY) \sim \text{Exp}(c\theta)$.

- We're going to use this result a **LOT**!

UC SANTA BARBARA
Department of Statistics and Applied Probability

Interpretation

- I know this might seem a little abstract - what does it mean to "multiply the exponential distribution by a constant?"
- Again, if it helps, you can always think in terms of our inches-to-centimeter problem from the start of these slides.
- If $Y \sim \text{Exp}(\theta)$ denotes the height of a randomly selected person in inches, then the distribution of heights in centimeters will *also* be exponential, this time with mean 2.54θ.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- Let's do another example together.
- Suppose Y has density function given by

$$f_Y(y) = 2y^2 \cdot \mathbb{1}_{\{0 \leq y \leq 1\}}$$

  and again define $U := cY$ for a positive constant c.
- Now, before we got lucky because we immediately knew what the CDF of Y was.
- But, even though we can't *immediately* recognize the CDF of Y in this example, we can still derive it!

Example

- By definition, for a $y \in [0, 1]$,

$$\begin{aligned} F_Y(y) &= \int_{-\infty}^{y} f_Y(t)\,\mathrm{d}t \\ &= \int_{-\infty}^{y} 2t \cdot \mathbb{1}_{\{0 \le t \le 1\}}\,\mathrm{d}t = \int_0^y 2t\,\mathrm{d}t = y^2 \end{aligned}$$

- Clearly, for $y < 0$ we have $F_Y(y) = \mathbb{P}(Y \le y) = 0$ and for $y > 1$ we have $\mathbb{P}(Y \le y) = 1$, meaning

$$F_Y(y) = \begin{cases} 0 & \text{if } y < 0 \\ y^2 & \text{if } 0 \le y < 1 \\ 1 & \text{if } y \ge 1 \end{cases}$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Example

- And now we're in the same position as before!

$$
\begin{aligned}
\mathbb{P}(U \leq u) &= \mathbb{P}(cY \leq u) = \mathbb{P}\left(Y \leq \frac{u}{x}\right) \\
&= F_Y\left(\frac{u}{c}\right) \\
&= \begin{cases} 0 & \text{if } (u/c) < 0 \\ (u/c)^2 & \text{if } 0 \leq (u/c) < 1 \\ 1 & \text{if } (u/c) \geq 1 \end{cases} = \boxed{\begin{cases} 0 & \text{if } u < 0 \\ u^2/c^2 & \text{if } 0 \leq u < c \\ 1 & \text{if } u \geq c \end{cases}}
\end{aligned}
$$

Example

- One more example before we summarize.
- Let $Y \sim \mathcal{N}(0, 1)$ and $U := Y^2$.
- A quick sketch (see chalkboard) reveals that $S_U = [0, \infty)$. So, $F_U(u) = 0$ whenever $u < 0$.
- Additionally, we (again) have the CDF of Y: $F_Y(y) = \Phi(y)$, where $\Phi(\cdot)$ denotes the standard normal CDF.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- So, let's try and proceed like we did before! For a fixed $u \geq 0$,

$$F_U(u) := \mathbb{P}(U \leq u) = \mathbb{P}(Y^2 \leq u)$$

- Now, it's tempting to continue this as

$$F_U(u) = \mathbb{P}(Y^2 \leq u) = \mathbb{P}(Y \leq \sqrt{u})$$

This is, however, <u>INCORRECT</u>.

- Let's understand why.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- There are a couple of ways to understand why the above is incorrect.
- One is to recall a fact from algebra/precalculus that you might have forgotten: $\sqrt{\cdot}$ means the *principal* square root, and so, for any real number x, we have $\sqrt{x^2} = |x|$.
  - Remember, both -3 and 3 have squares equal to 9! But, when we write $\sqrt{9}$, we implicitly mean the principal square root which is why we write $\sqrt{9} = 3$.
- So, what we really have is:

$$F_U(u) := \mathbb{P}(U \leq u) = \mathbb{P}(Y^2 \leq u) = \mathbb{P}(|Y| \leq \sqrt{u}) = \mathbb{P}(-\sqrt{u} \leq Y \leq \sqrt{u})$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example (cont'd)

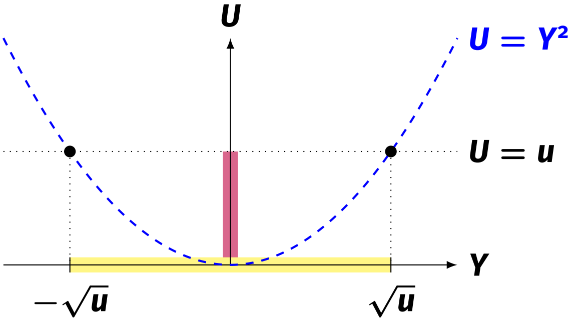
- Now, there's another way to see how to get from $\mathbb{P}(Y^2 \leq u)$ to $\mathbb{P}(-\sqrt{u} \leq Y \leq \sqrt{u})$; one that doens't require us to dig into our memory banks and dredge up something from algebra/precalculus, and instead uses pictures.

Video

https://www.youtube.com/watch?v=HtzqjHfoRbw

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Static Image



UC SANTA BARBARA
Department of Statistics
and Applied Probability

Example (cont'd)

- So, let's finish up our example!

$$
\begin{aligned}
F_U(u) &= \cdots = \mathbb{P}(-\sqrt{u} \leq Y \leq \sqrt{u}) \\
&= F_Y(\sqrt{u}) - F_Y(-\sqrt{u}) = \Phi(\sqrt{u}) - \Phi(-\sqrt{u}) \\
&= \Phi(\sqrt{u}) - [1 - \Phi(\sqrt{u})] = 2\Phi(\sqrt{u}) - 1
\end{aligned}
$$

- That's a bit anticlimactic... Let's differentiate wrt. u and obtain the PDF of U:

Example (cont'd)

$$
\begin{aligned}
f_U(u) &= \frac{\mathrm{d}}{\mathrm{d}u} F_U(u) \\
&= \frac{\mathrm{d}}{\mathrm{d}u} [2\Phi(\sqrt{u}) - 1] \\
&= 2 \cdot \frac{1}{2\sqrt{u}} \cdot \phi(\sqrt{u}) = \frac{1}{\sqrt{u}} \phi(\sqrt{u})
\end{aligned}
$$

- Let's incorporate the support of U, and simplify:

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Example (cont'd)

$$\begin{aligned} f_U(u) &= \frac{1}{\sqrt{u}} \phi(\sqrt{u}) \cdot \mathbb{1}_{\{u \geq 0\}} \\ &= \frac{1}{\sqrt{u}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\sqrt{u})^2} \cdot \mathbb{1}_{\{u \geq 0\}} \\ &= \frac{1}{\sqrt{\pi} \cdot 2^{1/2}} \cdot u^{1/2 - 1} \cdot e^{-u/2} \cdot \mathbb{1}_{\{u \geq 0\}} \end{aligned}$$

- One useful fact: $\Gamma(1/2) = \sqrt{\pi}$. Hence:

$$f_U(u) = \frac{1}{\Gamma(1/2) \cdot 2^{1/2}} \cdot u^{1/2 - 1} \cdot e^{-u/2} \cdot \mathbb{1}_{\{u \geq 0\}}$$

- Indeed, $U \sim \text{Gamma}(1/2\ ,\ 2) \stackrel{\mathrm{d}}{=} \chi_1^2$!

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Theorem

- This is an ***extremely*** important result which we will use repeatedly throughout this course. Let's make it more formal by rephrasing it as a theorem:

Theorem (Square of Standard Normal)

If $Y \sim \mathcal{N}(0, 1)$ and $U := Y^2$, then $U \sim \chi^2_1$.

- The proof of this theorem is exactly the work we did on the previous slides.

Recap

- Whew- that was a lot of work! Let's recap.
- Given a random variable Y, and $U := g(Y)$ for some function $g : \mathbb{R} \to \mathbb{R}$, we can use the **method of distribution functions** (aka the **CDF**) method to find the distribution of U.
- Specifically, this entails:
  (1) Writing $F_U(u)$, the CDF of U, in terms of $F_Y(y)$, the CDF of Y, by basically finding an equivalent formulation for the event $A_U := \{U \leq u\}$ that is in terms of Y
  (2) Plugging into the CDF of Y, and simplifying as necessary.

UC SANTA BARBARA
Department of Statistics
and Applied Probability