# 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)

# 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 **Law of the Unconscious Statistician** (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)


UC SANTA BARBARA
Department of Statistics and Applied Probability

# 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

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

# 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".


UC SANTA BARBARA
Department of Statistics and Applied Probability

# 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...

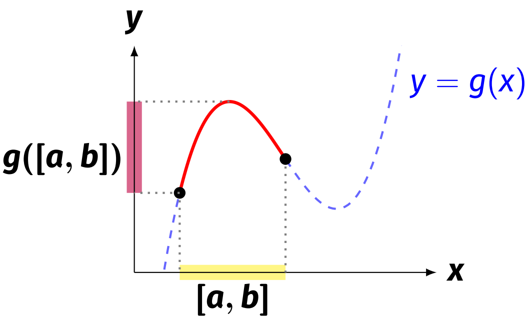

UC SANTA BARBARA
Department of Statistics and Applied Probability

# 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)


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# 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:

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

- 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 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 $2.54\theta$.


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Example

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

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

  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!


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Example

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

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

- 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

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

# Example

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

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


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# 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.

# 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.

# 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:

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

# 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.

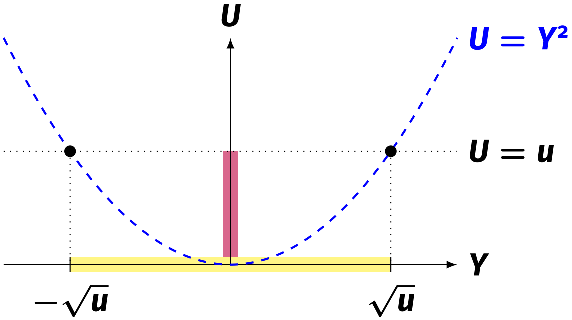

UC SANTA BARBARA
Department of Statistics and Applied Probability

# 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 $U$:


UC SANTA BARBARA
Department of Statistics and Applied Probability

# 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:

# 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, $U \sim \mathrm{Gamma}(1/2\ ,\ 2) \overset{\mathrm{d}}{=} \chi^2_1$!


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.


UC SANTA BARBARA
Department of Statistics and Applied Probability

# 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.

# Method of Transformations (Change of Variable Formula)


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Leadup

- Let's, for a moment, return the example where we squared the standard normal distribution.
- Specifically, we had $Y \sim \mathcal{N}(0, 1)$ and $U := Y^2$.
- After some work, we found that $F_U(u) = 2\Phi(\sqrt{u}) - 1$.
- Then, we differentiated wrt. $u$ to obtain a formula for $f_U(u)$.
- This begs the question - can we perhaps "extend" the CDF method to give us a formula for the *PDF* of $U$ directly?
- The answer turns out to be "yes, under some conditions."

# Leadup

**Goal**

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

- Let's see what happens if we try to apply the CDF method:

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

- Isn't it tempting to apply $g^{-1}(\cdot)$ to both sides of the inequality?
- It is! But we need to be careful. First, remember that we don't have any guarantee that $g^{-1}(\cdot)$ even exists!

## Leadup

- Alright, then - let's add some assumption about our function $g(\cdot)$.

**Goal**

> Given a random variable $Y$ and a strictly increasing function $g()$, we seek to find $f_U(u)$, the PDF of $U$.

- *Now* we are guaranteed the existence of $g^{-1}(\cdot)$.
- Furthermore, since we assumed $g(\cdot)$ itself to be strictly *increasing*, $g^{-1}(\cdot)$ will also be strictly increasing.
- Hence, we "preserve the direction of inequality" when applying $g^{-1}(\cdot)$ to both sides of an inequality.

# Leadup

- Then:

$$F_U(u) := \mathbb{P}(U \leq u) = \mathbb{P}(g(Y) \leq u) = \mathbb{P}(Y \leq g^{-1}(u)) = F_Y(g^{-1}(u))$$

- We can now differentiate wrt. $U$ and apply the chain rule (from calculus; we can discuss this further on the chalkboard):

$$\begin{aligned} f_U(u) &:= \frac{\mathrm{d}}{\mathrm{d}u} F_U(u) \\ &= \frac{\mathrm{d}}{\mathrm{d}u} F_Y(g^{-1}(u)) \\ &= f_Y(g^{-1}(u)) \cdot \frac{\mathrm{d}}{\mathrm{d}u} g^{-1}(u) \end{aligned}$$

# Leadup

- If we instead assume that $g(\cdot)$ is strictly decreasing, a similar computation (which I'll be asking you to complete on your homework) yields

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

- So, if we instead simply assume that $g(\cdot)$ is strictly monotonic, we can summarize our findings as:

$$f_U(u) = \begin{cases} f_Y(g^{-1}(u)) \cdot \left[ \frac{\mathrm{d}}{\mathrm{d}u} g^{-1}(u) \right] & \text{if } g(\cdot) \text{ is increasing} \\ f_Y(g^{-1}(u)) \cdot \left[ -\frac{\mathrm{d}}{\mathrm{d}u} g^{-1}(u) \right] & \text{if } g(\cdot) \text{ is decreasing} \end{cases}$$


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Change of Variable Formula

- A bit of simplification (and recollections of how derivatives of increasing/decreasing functions behaves) allows us to rewrite our result above as:

**Theorem (Change of Variable Formula)**

Given a random variable $Y \sim f_Y$ and a function $g(\cdot)$ that is strictly monotonic over the support of $Y$, then the random variable $U := g(Y)$ has density given by

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


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Example

- As an example, let's re-derive the closure under multiplication property of the Exponential distribution, this time using the Change of Variable formula.
- That is: let $Y \sim \text{Exp}(\theta)$, and set $U := cY$ for some positive constant $c > 0$.
- Since the transformation $g(y) = cy$ is strictly monotonic (specifically, it's strictly increasing) it's inverse exists and is calculable as $g^{-1}(u) = u/c$. Hence:

$$\left| \frac{\mathrm{d}}{\mathrm{d}u} g^{-1}(u) \right| = \left| \frac{\mathrm{d}}{\mathrm{d}u} \left( \frac{u}{c} \right) \right| = \left| \frac{1}{c} \right| = \frac{1}{c}$$

where we have dropped the absolute values in the last step since we are assuming $c > 0$.


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Example

- Additionally, since $Y \sim \text{Exp}(\theta)$ we know that

$$f_Y(y) = \frac{1}{\theta} \exp\left\{-\frac{y}{\theta}\right\} \cdot \mathbb{1}_{\{y \geq 0\}}$$

- Therefore, plugging into the change of variable formula, we have

$$\begin{aligned} f_U(u) &= f_Y[g^{-1}(u)] \cdot \left| \frac{\mathrm{d}}{\mathrm{d}u} g^{-1}(u) \right| \\ &= \frac{1}{\theta} \exp\left\{ -\frac{\left(\frac{u}{c}\right)}{\theta} \right\} \cdot \mathbb{1}_{\{\frac{u}{c} \geq 0\}} \cdot \frac{1}{c} \\ &= \frac{1}{c\theta} \exp\left\{-\frac{u}{c\theta}\right\} \cdot \mathbb{1}_{\{u \geq 0\}} \end{aligned}$$


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Clicker Question!

## Clicker Question 1

Given $Y \sim \text{Unif}[1, 2]$ and $U := 2X + 3$, does $U$ also follow a Uniform Distribution?

(A) Yes;  (B) No


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Change of Variable Formula

- Now, note that the only assumption we need to make about $g(\cdot)$ in order for the Change of Variable formula to hold is that it is strictly monotone *over the support of* $Y$.
- For example, suppose $Y \sim \text{Unif}[-1, 0]$ and take $U := Y^2$.
- Though the function $g(y) = y^2$ is not strictly monotone over $\mathbb{R}$, it *is* strictly monotone over $S_Y := [-1, 0]$ (i.e. the support of $Y$), and hence its inverse exists and is given by $g^{-1}(u) = -\sqrt{u}$.
- The Change of Variable formula can therefore safely be applied.


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Change of Variable Formula

- In general, however, the Change of Variable formula does not work when we are dealing with transformations that are not strictly monotone.
- For example, given $Y \sim \mathcal{N}(0, 1)$ and $U := Y^2$, we *cannot* directly apply the Change of Variable formula.
    - Admittedly, there does exist a way to generalize the Change of Variable formula to work in a situation like this, but we won't cover that in PSTAT 120B. If you're curious, I'm happy to walk you through the general outline during Office Hours.

# Method of Moment-Generating Functions (MGF Method)


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Leadup

**Goal**

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

- So far, we've talked about "describing" the distribution of $U$ by both its CDF (using the CDF method) and its PDF (using the Change of Variable formula).
- We know that there is a third way of classifying distributions - **moment-generating functions** (MGFs).


UC SANTA BARBARA
Department of Statistics and Applied Probability

# MGFs

**Definition (MGF)**

The MGF of a random variable $X$, notated $M_X(t)$, is defined as

$$M_X(t) := \mathbb{E}[e^{tX}]$$

- Recall that this expectation is computed as a sum if $X$ is discrete and as an integral if $X$ is continuous.


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Useful Result

**Theorem**

Given two random variables $X$ and $Y$ with MGFs $M_X(t)$ and $M_Y(t)$, respectively, that are both continuous in a small neighborhood of the origin, then $M_X(t) = M_Y(t)$ implies that $X$ and $Y$ have the same distribution.

- This theorem is essentially just a more formal way of saying "MGFs uniquely determine random variables." For example,

$$M_X(t) = \exp\left\{2t + \frac{1}{2}t^2\right\} \iff X \sim \mathcal{N}(2, 1)$$


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Useful Result

**Theorem**

Given a random variable $Y$ with MGF $M_Y(t)$, and $U := aY + b$ for constants $a, b \in \mathbb{R}$,

$$M_U(t) = e^{bt} M_Y(at)$$


UC SANTA BARBARA
Department of Statistics and Applied Probability

## Proof.

$$
\begin{aligned}
M_U(t) &:= \mathbb{E}[e^{tU}] && \text{[Definition of MGF]} \\
&:= \mathbb{E}[e^{t(aY+b)}] && \text{[Definition of } U\text{]} \\
&:= \mathbb{E}[e^{(at)Y+bt}] && \text{[Algebra]} \\
&:= \mathbb{E}[e^{(at)Y} \cdot e^{bt}] && \text{[Algebra]} \\
&:= e^{bt}\mathbb{E}[e^{(at)Y}] && \text{[Linearity of } \mathbb{E}\text{]} \\
&:= e^{bt}M_Y(at) && \text{[Definition of MGF]}
\end{aligned}
$$

□

- It turns out, we can use this theorem to (again) prove the closure of the exponential distribution under multiplication!


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Example

- Once again, let $Y \sim \text{Exp}(\theta)$, and let $U = cY$ for a positive constant c.
- First recall that the MGF of the exponential distribution is

$$M_Y(t) = \begin{cases} (1-\theta t)^{-1} & \text{if } t < 1/\theta \\ \infty & \text{otherwise} \end{cases}$$

- Hence, by the previous theorem:

$$M_U(t) = e^{0 \cdot t} \cdot M_Y(ct) = 1 \cdot \begin{cases} (1-\theta(ct))^{-1} & \text{if } (ct) < 1/\theta \\ \infty & \text{otherwise} \end{cases}$$

$$= \begin{cases} (1-(c\theta)t))^{-1} & \text{if } t < 1/(c\theta) \\ \infty & \text{otherwise} \end{cases}$$

which is, as expected, the MGF of the $\text{Exp}(c\theta)$ distribution.


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Clicker Question!

## Clicker Question 2

If $Y \sim \text{Pois}(\lambda)$ and $U := cY$ for some positive constant $c$, what is the distribution of $U$?

(A) $\text{Pois}(c\lambda)$

(B) $\text{Pois}(c/\lambda)$

(C) $\text{Pois}(\lambda/c)$

(D) None of the above


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Leadup

- Now, it may seem a little pointless to use the method of MGFs to identify distributions of transformed random variables.
- I admit - the method of MGFs is not always the best choice, especially if you're looking for a density or distribution function. (If you're only interested in moments then the MGF method is a good shout, but you can always use the LOTUS for that as well!)
- However, the method of MGFs really shines when we start taking linear combinations of *multiple* random variables.
- We'll talk about multivariate transformations more after the first midterm, but let's get a quick flavor of some of them now.

# Leadup

- Suppose two friends, Jack and Jill, each enter a separate checkout lane at a grocery store.
- It makes sense to model their wait times as two separate random variables: say, $Y_1$ and $Y_2$.
  - That is, $Y_1$ denotes one wait time (in minutes) and $Y_2$ denotes another wait time (in minutes).
- The random variable $U := (Y_1 + Y_2)$ then represents the *combined* wait times of Jack and Jill (in minutes).
- If, for example, $Y_1, Y_2 \overset{\text{i.i.d.}}{\sim} \text{Exp}(\theta)$, then what distribution does $U$ follow?

# Leadup

- As a bit of a spoiler, we *could* try to find the distribution of $U$ using the CDF method. (Doing so would involve computing a double integral - these are the sorts of things we'll be doing after MT01).
- But, instead, note:

$$\begin{aligned} M_U(t) &:= \mathbb{E}[e^{tU}] \\ &= \mathbb{E}[e^{t(Y_1+Y_2)}] \\ &= \mathbb{E}[e^{tY_1} \cdot e^{tY_2}] \\ &= \mathbb{E}[e^{tY_1}] \cdot \mathbb{E}[e^{tY_2}] \\ &= M_{Y_1}(t) \cdot M_{Y_2}(t) \end{aligned}$$

# Leadup

- So, plugging in the MGF of the Exp($\theta$) distribution, we have

$$\begin{aligned} M_U(t) &= \left( \begin{cases} (1-\theta t)^{-1} & \text{if } t < 1/\theta \\ \infty & \text{otherwise} \end{cases} \right) \cdot \left( \begin{cases} (1-\theta t)^{-1} & \text{if } t < 1/\theta \\ \infty & \text{otherwise} \end{cases} \right) \\ &= \begin{cases} (1-\theta t)^{-2} & \text{if } t < 1/\theta \\ \infty & \text{otherwise} \end{cases} \end{aligned}$$

  which is the MGF of the Gamma$(2, \theta)$ distribution!
- So, we've shown that $(Y_1 + Y_2) \sim \text{Gamma}(2, \theta)$.


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Useful Result

**Theorem (Important MGF Formula)**

Given a collection of independent random variables $\{Y_i\}_{i=1}^{n}$, we have

$$M_U(t) = \prod_{i=1}^{n} M_{Y_i}(a_i t) \qquad \text{where } U := \sum_{i=1}^{n} a_i Y_i$$


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Useful Result

**Theorem (Closure of Gamma Distribution under Sums)**

Given $\{Y_i\}_{i=1}^n$ with $Y_i \sim \text{Gamma}(\alpha_i, \beta)$ and constants $\{a_i\}_{i=1}^n$, we have

$$\left( \sum_{i=1}^n Y_i \right) \sim \text{Gamma}\left( \sum_{i=1}^n \alpha_i \ , \ \beta \right)$$


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Proof

We use the formula from the previous slide:

$$M_{\sum_{i=1}^{n} a_i Y_i}(t) = \prod_{i=1}^{n} M_{Y_i}(a_i t)$$

Recall that the MGF of the Gamma$(\alpha_i, \beta)$ distribution is given by

$$M_{Y_i}(t) = \begin{cases} (1 - \beta t)^{-\alpha_i} & \text{if } t < 1/\beta \\ \infty & \text{otherwise} \end{cases}$$

Hence, plugging in, we find:

## Proof

$$
\begin{aligned}
M_{\sum_{i=1}^n a_i Y_i}(t) &= \prod_{i=1}^n M_{Y_i}(a_i t) \\
&= \prod_{i=1}^n \left( \begin{cases} (1-\beta t)^{-\alpha_i} & \text{if } t < 1/\beta \\ \infty & \text{otherwise} \end{cases} \right) \\
&= \begin{cases} (1-\beta t)^{-\sum_{i=1}^n \alpha_i} & \text{if } t < 1/\beta \\ \infty & \text{otherwise} \end{cases}
\end{aligned}
$$

which we recognize as the MGF of the Gamma$(\sum_{i=1}^n \alpha_i, \beta)$ distribution. Hence, we are done.


UC SANTA BARBARA
Department of Statistics and Applied Probability

# Note

- Note: I am a bit of a stickler when it comes to ending proofs using the MGF method.
- Specifically, I am adamant that you end with some sort of concluding *statement* - don't just leave the MGF without saying something about the underlying distribution!
    - For example, in the previous proof, notice how I ended with "which we recognize as...". Just make sure you end your MGF-related proofs with something similar!

# Another Useful Result

**Theorem (Closure of Normal Distribution under Linear Combinations)**

Given a collection of independent random variables $\{Y_i\}_{i=1}^n$ with $Y_i \sim \mathcal{N}(\mu_i\ ,\ \sigma_i^2)$ and constants $\{a_i\}_{i=1}^n$, we have

$$U := \left( \sum_{i=1}^n a_i Y_i \right) \sim \mathcal{N}\left( \sum_{i=1}^n a_i \mu_i\ ,\ \sum_{i=1}^n a_i^2 \sigma_i^2 \right)$$


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# Proof

- I leave the proof to you.
- One word of extreme caution: we can get the expectation and variance of $U$ using 120A-related formulas.
- But, the *normality* of $U$ is something that we cannot take for granted - this is why we need to use the MGF method to complete the proof!

# Summary: Univariate Transformations

**Goal**

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

- So far, we've accomplished this goal in three different ways:
  - The CDF Method (Method of Distribution Functions)
  - The Change of Variable formula (Method of Transformations)
  - The MGF Method.


UC SANTA BARBARA
Department of Statistics
and Applied Probability

# CDF Method

(1) Write $F_U(u) := \mathbb{P}(A_U)$, where $A_U := \{U \leq u\}$.

(2) Find an equivalent expression for $A_U$ in terms of $Y$; call this $A_Y$.

(3) Compute $\mathbb{P}(A_Y)$ using the distribution of $Y$ (which is known), to then find the CDF of $U$.

- Remember: when carrying out step 3, drawing a picture can be incredibly helpful!

# Change of Variable Formula

(1) Compute $g^{-1}(u)$ [remember that this can be done by solving the equation $u = g(y)$ for $u$ in terms of $y$].

(2) Plug into the Change of Variable formula:

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

- Remember: this method only works when the transformation $g(\cdot)$ is strictly monotonic over the support of $Y$!
- Also, a side note: so long as you are careful, the change of variable formula will give you the support of $U$. But, in some cases, it might be easier to find the support first (by drawing a picture), and then incorporating that into your answer later.

# MGF Method

(1) Compute the MGF $M_U(t)$ of $U$ by writing it in terms of the MGF $M_Y(t)$ of $Y$, and then recognize the resulting MGF as belonging to a particular distribution.

- This works well for linear transformations and linear combinations of random variables, but not too well for nonlinear transformations.
- Also, the MGF method won't (typically) give you a PDF/CDF, so if you really want the PDF/CDF of $U$ you should use a different method [unless you believe you will be able to recognize the resulting distribution as one that has a name].

# Chalkboard Example

## Example

The **kinetic energy** of a particle with mass $m$ traveling at a velocity $V$ is given by

$$E = \frac{1}{2} m V^2$$

Consider a particle selected at random, whose velocity is a random variable $V$ with density

$$f_V(v) = 2v^3 e^{-v^2} \cdot \mathbb{1}_{\{v > 0\}}$$

Find the distribution of the kinetic energy of this particle once using the CDF method and once using the Change of Variable formula.