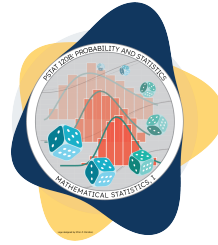


Topic 2.5: Multivariate Transformations

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

Outline

1. Multivariate Transformations

2. Transformations Involving the Normal Distribution

Multivariate Transformations

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Univariate vs. Multivariate

Goal

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

- This was the goal we dealt with over the past several lectures.
- Recall that this is called the **univariate** setting, as we were considering functions of only one random variable.
- Right before the midterm, we started exploring some **multivariate** transformation problems (specifically, sums and linear combinations).

Multivariate

Goal

Given a collection of random variable $\{Y_i\}_{i=1}^n$ and a function $g(\cdot)$, we seek to describe the random variable $U := g(Y_1, \cdots, Y_n)$.

- For example, the **sample mean** of a collection of random variables

$$\overline{Y}_n := \frac{1}{n}\sum_{i=1}^{n} Y_i$$

is an example of a multivariate transformation of the collection $\{Y_i\}_{i=1}^n$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Multivariate

- Perhaps unsurprisingly, dealing with multivariate transformations in pure generality is quite complicated.
- As such, for the purposes of this class, we will only focus on a handful of specific multivariate transformations.
- First, however, it will be enlightening to consider the case of **bivariate** transformations (i.e. transformations of only two random variables).

Bivariate

Goal

Given a pair of random variables $(Y_1, Y_2) \sim f_{Y_1,Y_2}$ and a bivariate function $g(\cdot)$ [e.g. a function that takes two inputs], we seek to describe the distribution of $U := g(Y_1, Y_2)$.

- For example, let Y_1 denote the concentration of a particular compound and Y_2 denote the concentration of a different compound. Then the ratio $U := Y_1/Y_2$ denotes the relative concentration of the first compound as compared to the second.

Bivariate

- Now, as we've seen before, if the function g is a linear function [e.g. $U := aY_1 + bY_2 + c$] and if $Y_1 \perp Y_2$, then the method of MGFs can be used.
- However, if our transformation $g(\cdot)$ is *not* linear, or if our random variables are *not* independent, then the method of MGFs won't be particularly useful.
- Instead, there's one method that will be particularly useful in pretty much any bivariate setting: the CDF method!

Example

- To see what I mean, let's re-derive a familiar result.

Example

Let $Y_1, Y_2 \overset{\text{i.i.d.}}{\sim} \text{Exp}(\theta)$ and $U := (Y_1 + Y_2)$. Find the distribution of U.

- Previously we used the MGF method to show that $U \sim \text{Gamma}(2, \theta)$.
- Let's try and show this again, now using the CDF method.

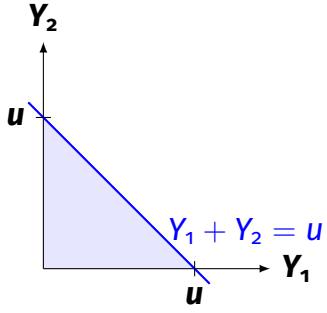
Example

- First note that the support of U is $S_U = [0, \infty)$. Hence, for $u < 0$ we have $F_U(u) = 0$.
- For a fixed $u \geq 0$, we have

$$F_U(u) := \mathbb{P}(U \leq u) = \mathbb{P}(Y_1 + Y_2 \leq u)$$

- And what do you know - this is just a familiar PSTAT 120A-style double integral problem!

Example



- Neither order of integration is particularly harder than the other
- Hence, let's (somewhat arbitrarily) pick the order $dy_1 dy_2$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

$$
\begin{aligned}
F_U(u) &= \mathbb{P}(Y_1 + Y_2 \leq u) \\
&= \int_0^u \int_0^{u-y_2} f_{Y_1,Y_2}(y_1, y_2)\, dy_1\, dy_2 = \int_0^u \int_0^{u-y_2} \frac{1}{\theta^2} e^{-(y_1+y_2)/\theta}\, dy_1\, dy_2 \\
&= \frac{1}{\theta} \int_0^u e^{-y_2/\theta} \left(\int_0^{u-y_2} \frac{1}{\theta} e^{-y_1/\theta}\, dy_1 \right) dy_2 \\
&= \frac{1}{\theta} \int_0^u e^{-y_2/\theta} \left(1 - e^{-(u-y_2)/\theta}\right) dy_2
\end{aligned}
$$

- Can anyone tell me how I was able to compute the blue integral so quickly?

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

$$
\begin{aligned}
F_U(u) &= \frac{1}{\theta} \int_0^u e^{-y_2/\theta} \left(1 - e^{-(u-y_2)/\theta}\right) \, dy_2 \\
&= \frac{1}{\theta} \int_0^u \left(e^{-y_2/\theta} - e^{-u/\theta}\right) \, dy_2 \\
&= 1 - e^{-u/\theta} - \frac{u}{\theta} e^{-u/\theta} = 1 - \left(\frac{u}{\theta} + 1\right) e^{-u/\theta}
\end{aligned}
$$

- So, in all, we have

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

Example

- Differentiating wrt. u and incorporating the support, we find

$$\begin{aligned} f_U(u) &= \left[-\frac{1}{\theta} e^{-u/\theta} + \frac{1}{\theta} \left(\frac{u}{\theta} + 1 \right) e^{-u/\theta} \right] \cdot \mathbb{1}_{\{u \geq 0\}} \\ &= \frac{u}{\theta^2} e^{-u/\theta} \cdot \mathbb{1}_{\{u \geq 0\}} \end{aligned}$$

which, indeed, is the density of the $\text{Gamma}(2, \theta)$ distribution.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Chalkboard Example

Given $Y_1, Y_2 \overset{\text{i.i.d.}}{\sim} \text{Unif}[0, 1]$, find the density of $U := Y_2/Y_1$.

Caution!

- As the previous example illustrates, we need to be very careful about considering all possible cases when drawing our pictures!
- This is *especially* true when one (or both) of the random variables being transformed has a "truncated" support, like $[0, 1]$ or $[2, 4]$.
- As yet another example, try deriving the distribution of $U := (Y_1 + Y_2)$ when Y_1 and Y_2 are i.i.d. Unif$[0, 1]$ random variables.
  - The resulting distribution is a special case of what is known as the **triangular distribution**- can you see why?

Jacobians

- So, we've seen how the method of MGFs extends nicely into the bivariate (indeed, even the multivariate) setting, along with how the CDF method alos extends nicely into the bivariate setting.
- A natural question arises- does there exist an analog of the Change of Variable formula?
- Well, first things first, note that we cannot use the Change of Variable formula as it stands in a bivariate setting - what is the inverse of a bivariate function?
- Now, there *does* exist something called the **method of Jacobians** which can be viewed as a sort of bivariate analog of the Change of Variable formula. It is discussed in Section 6.6 of the textbook.

Order Statistics

- Finally, let's consider one very interesting set of *multivariate* transformations.
- Let's start simple: given an i.i.d. collection of random variables $\{Y_i\}_{i=1}^n$, what can we say about the maximum of these random variables?
- Again, note- *this is a multivariate transformation!* Specifically, with

$$g(y_1, \cdots, y_n) = \max\{y_1, \cdots, y_n\}$$

- The notation we use for the maximum of a collection $\{Y_i\}_{i=1}^n$ of random variables is

$$Y_{(n)} := \max_{1 \leq i \leq n} \{Y_i\}$$

Maximum

Goal

Given a collection $\{Y_i\}_{i=1}^n$ of i.i.d. random variables with density $f_Y(y)$ and CDF $F_Y(y)$, we seek to derive the density of $Y_{(n)} := \max_{1 \leq i \leq n} \{Y_i\}$

- Well, this transformation is very much nonlinear, so the MGF method won't be of much help here.
- This means there's really only one method we can try to use - the CDF method!

Maximum

$$F_{Y_{(n)}}(y) := \mathbb{P}(Y_{(n)} \leq y) = \mathbb{P}\left(\max_{1 \leq i \leq n} \{Y_i\} \leq y\right)$$

- At this point, we need to stop and think.
- What the event on the RHS is saying is: "the largest of all the Y_i's is less than y."
- The key piece of logic here is that this implies *all* the Y_i's must also be less than y!
- Think of it this way - we're asserting that the largest of the Y_i's is less than y. By definition of being the *largest*, all of the other Y_i's must be less than the maximum, which, again, is less than y. Hence, all the Y_i's must be less than y.

Maximum

$$
\begin{aligned}
F_{Y_{(n)}}(y) := \mathbb{P}(Y_{(n)} \leq y) &= \mathbb{P}\left(\max_{1 \leq i \leq n} \{Y_i\} \leq y\right) \\
&= \mathbb{P}\left(\bigcap_{i=1}^{n} \{Y_i \leq y\}\right) \\
&= \prod_{i=1}^{n} \mathbb{P}(Y_i \leq y) = \prod_{i=1}^{n} F_Y(y) = [F_Y(y)]^n
\end{aligned}
$$

- I will, of course, ask us to go through and justify each step carefully together.

Result

Theorem (Distribution of Sample Maximum)

Given a collection of i.i.d. random variables $\{Y_i\}_{i=1}^n$ with density $f_Y(y)$ and CDF $F_Y(y)$, the density of

$$Y_{(n)} := \max_{1 \leq i \leq n} \{Y_i\}$$

is given by

$$f_{Y_{(n)}}(y) = n[F_Y(y)]^{n-1} \cdot f_Y(y)$$

Example

Example

Given $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \text{Unif}[0, 1]$, derive the density of $Y_{(n)}$.

- We simply need to plug into the theorem on the previous slide:

$$f_Y(y) = \mathbb{1}_{\{0 \le y \le 1\}}; \qquad F_Y(y) = \begin{cases} 0 & \text{if } y < 0 \\ y & \text{if } 0 \le y < 1 \\ 1 & \text{if } y \ge 1 \end{cases}$$

- Thus,

$$f_{Y_{(n)}}(y) = n y^{n-1} \cdot \mathbb{1}_{\{0 \le y \le 1\}}$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Minimum

Goal

Given a collection $\{Y_i\}_{i=1}^n$ of i.i.d. random variables with density $f_Y(y)$ and CDF $F_Y(y)$, we seek to derive the density of $Y_{(1)} := \min_{1 \leq i \leq n} \{Y_i\}$

- I'd like to highlight the notation once again: $Y_{(1)}$ is the notation we use for the minimum of the collection $\{Y_i\}_{i=1}^n$.
- To achieve our goal, let's start as we did for the maximum, with the CDF method.

Maximum

$$F_{Y_{(1)}}(y) := \mathbb{P}(Y_{(1)} \leq y) = \mathbb{P}\left(\min_{1 \leq i \leq n} \{Y_i\} \leq y\right) \stackrel{?}{=} \mathbb{P}\left(\bigcap_{i=1}^{n} \{Y_i \leq y\}\right)$$

- The last equality **does not hold**!
- Here's why: just because the smallest of a set of numbers is less than y, we cannot immediately conclude that all of the numbers are also less than y.
  - For example, consider the set of numbers $\{2, 4\}$; clearly the smallest of the numbers (i.e. 2) is smaller than 3, however it is *not* the case that both numbers are smaller than 3.

Maximum

- What *is* true is that: if the smallest of a set of numbers is *larger* than y, then all numbers must be larger than y.
  - Again, say we assume the smallest of the Y_i's is larger than 2. By definition of being the smallest, all of the other Y_i's will be larger than $Y_{(1)}$, which is assumed to be bigger than y - hence all of the Y_i's will be bigger than y.
- So, we'd like to convert everything from CDFs to **survival functions** (i.e. one-minus CDFs).

Minimum

$$\begin{aligned}\overline{F_{Y_{(1)}}}(y) &:= 1 - F_{Y_{(1)}}(y) = \mathbb{P}(Y_{(1)} > y) \\ &= \mathbb{P}\left(\bigcap_{i=1}^{n}\{Y_i > y\}\right) \\ &= \prod_{i=1}^{n}\mathbb{P}(Y_i > y) = \prod_{i=1}^{n}[1 - F_Y(y)] = [1 - F_Y(y)]^n\end{aligned}$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Result

Theorem (Distribution of Sample Minimum)

Given a collection of i.i.d. random variables $\{Y_i\}_{i=1}^n$ with density $f_Y(y)$ and CDF $F_Y(y)$, the density of

$$Y_{(1)} := \min_{1 \leq i \leq n} \{Y_i\}$$

is given by

$$f_{Y_{(1)}}(y) = n[1 - F_Y(y)]^{n-1} \cdot f_Y(y)$$

Example

Example

Given $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \text{Exp}(\theta)$, derive the density of $Y_{(1)}$.

- We simply need to plug into the theorem on the previous slide:

$$f_Y(y) = \frac{1}{\theta} e^{-y/\theta} \cdot \mathbb{1}_{\{y \geq 0\}}; \qquad 1 - F_Y(y) = \begin{cases} 1 & \text{if } y < 0 \\ e^{-y/\theta} & \text{if } y \geq 0 \end{cases}$$

- Thus,

$$f_{Y_{(1)}}(y) = n \left(e^{-y/\theta}\right)^{n-1} \cdot \frac{1}{\theta} e^{-y/\theta} \cdot \mathbb{1}_{\{y \geq 0\}} = \frac{n}{\theta} e^{-ny/\theta} \cdot \mathbb{1}_{\{y \geq 0\}}$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Result

Theorem (Minimum of Exponentials)

If $Y_1, \cdots, Y_n \overset{\text{i.i.d.}}{\sim} \text{Exp}(\theta)$, then

$$Y_{(1)} := \min_{1 \leq i \leq n} \{Y_i\} \sim \text{Exp}\left(\frac{\theta}{n}\right)$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

General Order Statistics

- Now, there is nothing restricting us to only considering minima and maxima.
- For example, we might ask: what is the distribution of the *second smallest* of the collection $\{Y_i\}_{i=1}^n$?
- In general, we define the ***k*th order statistic**, notated $Y_{(k)}$, to be the k^{th} smallest of $\{Y_i\}_{i=1}^n$.
- Section 6.7 of your textbook gives a general formula for the density of $Y_{(k)}$; we will not concern ourselves with the general formula this quarter.
  - You should, however, know the formulas for the minimum and the maximum.

Transformations Involving the Normal Distribution

Leadup

- We are almost ready to make our transition from probability to statistics.
- First, there are a few results we should derive.
- Not only will we be utilizing these results going forward, we're also going to be using the techniques used to derive them incredibly often.
  - So, I encourage you not only to pay attention to the results themselves, but also the *methods* used to prove them!

Disclaimer

- I'd like to be a bit upfront - there are a *lot* of results that are going to be coming over the next few slides.
  - We'll actually be using every one of these results again in the coming lectures, so they are all very important!
- However, I'd like to stress (and will continue to stress) that *many of these results are not new!* Rather, they are either simply restatements of PSTAT 120A results, or follow almost directly from previously-derived results in this class.

Standardization

- Let's start off with a result that should be familiar to you, from PSTAT 120A (just restated slightly more formally).

Theorem (Standardization Result)

Given $Y \sim \mathcal{N}(\mu, \sigma^2)$ and $U := (Y - \mu)/\sigma$, we have $U \sim \mathcal{N}(0, 1)$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Direct Proof

- Note that $U := \frac{Y-\mu}{\sigma} = \frac{1}{\sigma}Y - \frac{\mu}{\sigma}$. Hence, by our Useful MGF Result:

$$\begin{aligned} M_U(t) &= e^{-(\mu/\sigma)t} \cdot M_Y\left(\frac{t}{\sigma}\right) \\ &= e^{-(\mu/\sigma)t} \cdot \exp\left\{\mu\left(\frac{t}{\sigma}\right) + \frac{\sigma^2}{2}\left(\frac{t}{\sigma}\right)^2\right\} \\ &= \exp\left\{\cancel{-\frac{\mu t}{\sigma}} + \cancel{\frac{\mu t}{\sigma}} + \frac{\cancel{\sigma^2}}{2} \cdot \frac{t^2}{\cancel{\sigma^2}}\right\} \\ &= e^{t^2/2} \end{aligned}$$

which we recognize as the MGF of the $\mathcal{N}(0, 1)$ distribution.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Result

Theorem (Normal Sum of Squares)

Given $Y_1, Y_2, \cdots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, we have $\sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma} \right)^2 \sim \chi^2_n$

Proof.

By the (Standardization Result), $(Y_i - \mu)/\sigma$ follows a standard normal. We know that the square of a standard normal distribution is χ^2_1. Furthermore, independent χ^2 random variables sum to another χ^2-distributed random variables with degrees of freedom also summing. □

Sample Mean

Theorem (Sample Mean of i.i.d. Normal Sample)

Given $Y_1, Y_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$,

$$\overline{Y}_n := \left(\frac{1}{n} \sum_{i=1}^{n} Y_i \right) \sim \mathcal{N}\left(\mu \, , \, \frac{\sigma^2}{n} \right)$$

Proof.

Appeal to Theorem (Closure of Normal Distribution under Linear Combinations) with $a_i = 1/n$. □

UC SANTA BARBARA
Department of Statistics and Applied Probability

Result

Theorem (Standardization of Sample Mean)

Given $Y_1, Y_2, \cdots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ and $\overline{Y}_n := n^{-1} \sum_{i=1}^{n} Y_i$,

$$U := \sqrt{n} \left(\frac{\overline{Y}_n - \mu}{\sigma} \right) = \frac{\overline{Y}_n - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0, 1)$$

Proof.
Combine results(Sample Mean of i.i.d. Normal Sample) and (Standardization Result). □

Example

Example

The weight of a randomly-selected chocolate bar is normally distributed with standard deviation 1.2 oz, and some unknown mean. Suppose a random sample of 16 chocolate bars is taken - what is the probability that the average weight of these 16 bars lies within 0.1 oz of the true average weight? Assume the weights of different chocolate bars are independent of one another.

Solutions

- Let Y_i denote the weight (in oz) of the i^{th} randomly-selected chocolate bar; then, from the problem statement,

$$Y_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu\ ,\ 1.44)$$

  where μ denotes the true average weight of a chocolate bar.
- Define the sample mean weight of the 16 randomly-selected chocolate bars to be

$$\overline{Y}_{16} := \frac{1}{16} \sum_{i=1}^{16} Y_i$$

Solutions

- By the previous theorem,

$$\overline{Y}_{16} \sim \mathcal{N}\left(\mu\,,\ \frac{1.44}{16}\right) \sim \mathcal{N}(\mu, 0.09)$$

  or, equivalently (by our standardization result)

$$\frac{\overline{Y}_{16} - \mu}{0.09} \sim \mathcal{N}(0, 1)$$

- We seek the probability that $\overline{Y}_{16}$ lies within 0.1 oz of μ; i.e. we seek

$$\mathbb{P}(|\overline{Y}_{16} - \mu| < 0.1) = \mathbb{P}(-0.1 < \overline{Y}_{16} - \mu < 0.1)$$

- We're now in 120A territory! Specifically, we standardize and then write our answer in terms of $\Phi(\cdot)$.

Solutions

$$\begin{aligned}
\mathbb{P}(|\overline{Y}_{16} - \mu| < 0.1) &= \mathbb{P}(-0.1 < \overline{Y}_{16} - \mu < 0.1) \\
&= \mathbb{P}\left(-\frac{0.1}{0.09} < \frac{\overline{Y}_{16} - \mu}{0.09} < \frac{0.1}{0.09}\right) \\
&= \Phi\left(\frac{0.2}{0.09}\right) - \Phi\left(-\frac{0.2}{0.09}\right) \\
&= 2\Phi\left(\frac{10}{9}\right) - 1 \approx 0.7335 = 73.35\%
\end{aligned}$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Result

- Now, the result (Standardization of Sample Mean) is an *incredibly* useful result which we will leverage several times going forward.
- But, it does depend on knowing the population variance σ^2.
- In some real-world cases (as we will see soon), σ^2 is not known - instead, all that is computable is the **sample variance**

$$S_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2$$

- Note that S_n^2 is a random variable. Intuitively, this makes sense- different random samples will have (potentially) different observed sample variances!

Result

Theorem (Distribution of Sample Variance)

Given $Y_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, define the sample variance as

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2$$

We have that

$$\frac{n-1}{\sigma^2} S_n^2 \sim \chi^2_{n-1}$$

Proof

- The proof of this fact is fairly involved, so we will skip it for this class.
- Please note the degrees of freedom: $(n-1)$, not n.
- As an aside, there are two main ways of thinking about degrees of freedom: one is simply as the parameter of a given distribution (e.g. χ^2, t, etc.).
- However, another way to view degrees of freedom is as the number of terms in a given sum that are allowed to vary freely.

Degrees of Freedom

- Consider, for example, the formula for S_n^2 again:

$$S_n^2 := \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2$$

- Once we know the values of $Y_1, \cdots, Y_{n-1}$ and the value of $\overline{Y}_n$ (which must be computed before computing S_n^2), the value of Y_n is uniquely determined. This is because $n\overline{Y}_n = Y_1 + \cdots + Y_{n-1} + Y_n$!
- So, in the summation definition of S_n^2, there are really only $(n-1)$ terms that are free to vary - the n^{th} term is uniquely determined by the other $(n-1)$.

Degrees of Freedom

- As yet another example, consider the following quantity:

$$S_{n,2}^2 := \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu)^2$$

  where $\mu := \mathbb{E}[Y_i]$.
- This quantity has n degrees of freedom, as all n terms in the sum are free to vary.

Example

Example

Given $Y_1, Y_2, \cdots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$, compute $\mathbb{P}(0.9 \leq S_5^2 \leq 1.2)$ where

$$S_5 := \frac{1}{4}\sum_{i=1}^{5}(Y_i - \overline{Y}_5)^2 \quad \text{and} \quad \overline{Y}_5 := \frac{1}{5}\sum_{i=1}^{5} Y_i$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Solutions

- By the result (Distribution of Sample Variance), we have that

$$\frac{5-1}{1}S_5^2 = 4S_5^2 \sim \chi^2_{5-1} \sim \chi^2_4$$

- Hence,

$$\begin{aligned}\mathbb{P}(0.9 \leq S_5^2 \leq 1.2) &= \mathbb{P}(4 \cdot 0.9 \leq 4S_5^2 \leq 4 \cdot 1.2) \\ &= \mathbb{P}(3.6 \leq 4S_5^2 \leq 4.8) \\ &= F_{\chi^2_4}(4.8) - F_{\chi^2_4}(3.6)\end{aligned}$$

  where $F_{\chi^2_\nu}(x)$ denotes the CDF of the χ^2_ν distribution, evaluated at x.
- Using a computer software, we can find this to be around 0.1544.

t-Distribution

Definition (Student's $t-$distribution)

Given $Z \sim \mathcal{N}(0, 1)$ and $W \sim \chi^2_\nu$ with $Z \perp W$, then

$$T := \frac{Z}{\sqrt{W/\nu}}$$

follows the so-called **$t-$distribution** with ν degrees of freedom, notated $T \sim t_\nu$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

t-Distribution

Theorem ($t-$distribution Density)

If $T \sim t_\nu$, then T has density given by

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \cdot \Gamma\left(\frac{\nu}{2}\right)} \cdot \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Proof.
This can be proven using the CDF method, and the definition of the $t-$distribution on the previous slide. □

Result

- Let's quickly recall why we even started talking about the sample variance.
- We know that $\sqrt{n}(\overline{Y}_n - \mu)/\sigma$ follows a standard normal distribution; however, sometimes σ is unknown.
- So, a natural question arises - what happens if we replace σ with $S_n := \sqrt{S_n^2}$? I.e., what happens if we replace the population standard deviation with the sample standard deviation

$$S_n := \sqrt{S_n^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2}$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Result

Theorem (Modified Standardized Sample Mean)

Given $Y_1, Y_2, \cdots \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, $\overline{Y}_n := n^{-1} \sum_{i=1}^n Y_i$, and $S_n := \sqrt{(n-1)^{-1} \sum_{i=1}^n (Y_i - \overline{Y}_n)^2}$,

$$U := \sqrt{n} \left(\frac{\overline{Y}_n - \mu}{S_n} \right) \sim t_{n-1}$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Result

- In other words, replacing the population standard deviation with the sample standard deviation breaks the normality of the sample mean.
- Intuitively, this makes sense! When we replace σ (which is deterministic) with S_n (which is random), we're essentially increasing the randomness of our overall statistic [we'll define a *statistic* more rigorously next lecture]. It makes sense, then, that we would need to use a distribution that has *wider* tails than the normal distribution (to allow for greater uncertainty). Indeed, the t distribution has wider tails than the normal distribution!

Example

Example

Let $Y_1, Y_2, Y_3 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$, and set

$$U := \sqrt{3} \cdot \frac{\frac{1}{3}(Y_1 + Y_2 + Y_3)}{\sqrt{\frac{1}{2}\left[\left(Y_1 - \frac{Y_1+Y_2+Y_3}{3}\right)^2 + \left(Y_2 - \frac{Y_1+Y_2+Y_3}{3}\right)^2 + \left(Y_3 - \frac{Y_1+Y_2+Y_3}{3}\right)^2\right]}}$$

What is the distribution of U? Include both the distribution's name as well as any/all relevant parameter(s)!

UC SANTA BARBARA
Department of Statistics and Applied Probability

Solution

- So, I guess we should start with the CDF method, right?
- Ew, no. (Unless you want to - in which case, go ahead...?)
- I think it would be a better idea to use our previously-derived results!
- That is, once we stare at U a bit more, we start to recognize some familiar quantities.

Solution

- For example, $\frac{1}{3}(Y_1 + Y_2 + Y_3) = \frac{Y_1+Y_2+Y_3}{3}$ is just another name for $\overline{Y}_3$, the sample mean of $\{Y_i\}_{i=1}^3$!
- Additionally, take a look at the denominator:

$$\sqrt{\frac{1}{2}\left[\left(Y_1 - \frac{Y_1+Y_2+Y_3}{3}\right)^2 + \left(Y_2 - \frac{Y_1+Y_2+Y_3}{3}\right)^2 + \left(Y_3 - \frac{Y_1+Y_2+Y_3}{3}\right)^2\right]}$$

- We can rewrite this as

$$\sqrt{\frac{1}{3-1} \cdot \left[(Y_1 - \overline{Y}_3)^2 + (Y_2 - \overline{Y}_3)^2 + (Y_3 - \overline{Y}_3)^2\right]}$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Solution

$$\sqrt{\frac{1}{3-1} \cdot \left[(Y_1 - \overline{Y}_3)^2 + (Y_2 - \overline{Y}_3)^2 + (Y_3 - \overline{Y}_3)^2 \right]}$$

$$= \sqrt{\frac{1}{3-1} \sum_{i=1}^{3} (Y_i - \overline{Y}_3)^2} = \sqrt{S_3^2} = S_3$$

- In other words, the denominator is just the sample standard deviation of $\{Y_i\}_{i=1}^3$!

Solution

- So, once the dust clears, we have

$$U := \sqrt{3} \cdot \frac{\frac{1}{3}(Y_1 + Y_2 + Y_3)}{\sqrt{\frac{1}{2}\left[\left(Y_1 - \frac{Y_1+Y_2+Y_3}{3}\right)^2 + \left(Y_2 - \frac{Y_1+Y_2+Y_3}{3}\right)^2 + \left(Y_3 - \frac{Y_1+Y_2+Y_3}{3}\right)^2\right]}}$$

$$= \sqrt{3}\left(\frac{\overline{Y}_3 - 0}{S_3}\right)$$

- Hence, by our result (Modified Standardized Sample Mean) with $\mu = 0$, we can immediately conclude that $U \sim t_{3-1}$; i.e. $U \sim t_2$.

UC SANTA BARBARA
Department of Statistics and Applied Probability