MATH 392 Problem Set 4

Sam D. Olson

Time Spent on PS4

Total: 7 hours Exercises: $4\frac{1}{2}$ hours Case Study: $2\frac{1}{2}$ hours

Exercises from the book

8.1 #1

Random Sample $X_1, ..., X_n$ taken from the Uniform distribution on the interval $[0, \theta]$

$\forall \theta$, how large must a random sample be taken in order that:

$Pr\left(\mid max\left(X_1, ..., X_n\right) - \theta \mid \leq 0.1\theta\right) \geq 0.95$

Let $U = max(X_1, ..., X_n)$

Note the c.d.f. of $U = max(X_1, ..., X_n)$ is

$$f(u) = \left\{ \begin{array}{c} 0 \; for \; u \leq \theta \\ \left(\frac{u}{\theta}\right)^n \; for \; 0 < u < \theta \\ 1 \; for \; u \geq \theta \end{array} \right\}$$

Note since $U \leq \theta$ with probability 1, then $\mid U - \theta \mid \leq 0.1\theta$ is the same as the event that $U \geq 0.9\theta$.

The probability of this event is $1 - F(0.9\theta) = 1 - 0.9^n$ To satisfy the condition, we have:

$\rightarrow 0.95 = 1 - 0.9^n \rightarrow 0.95^n = 0.05$

We take the log of each side to solve for n, giving us:

$log(0.95)n = log(0.05) \rightarrow n = \frac{log(0.05)}{log(0.95)}$

Evaluating this gives us $n = 28.43$. Thus we note that as n increases, $1 - 0.9^n$ decreases, such that we need a sample of at least 29 observations to satisfy the condition.

8.1 #3

For the conditions of Exercise 2, how large a random sample must be taken in order that $E_\theta\left(\mid \bar{X}_n - \theta \mid\right) \leq 0.1$ $\forall \theta$

First, let us note the necessary conditions from Exercise 2 for our purposes.

We know a random sample is to be taken from the normal distribution with unknown mean θ and standard deviation 2.

It then follows that $E_\theta\left(\mid \bar{X}_n - \theta \mid^2\right) = Var_\theta\left(\bar{X}_n\right) = \frac{4}{n}$

We then know $\bar{X}_n$ has the normal distribution with mean θ and variance $\frac{4}{n}$

Thus, we may construct the random variable $Z = \frac{\left(\bar{X}_n - \theta\right)}{\frac{2}{\sqrt{n}}}$ that fits the standard normal distribution.

Thus we have:

$$E_\theta \left(\mid \bar{X}_n - \theta \mid \right) = \frac{2}{\sqrt{n}} E_\theta \left(\mid Z \mid \right)$$

$$= \frac{2}{\sqrt{n}} \int\limits_{-\infty}^{\infty} \mid z \mid \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz$$

Taking advantage of symmetry, we can simplify this to:

$$E_\theta \left(\mid \bar{X}_n - \theta \mid \right) = 2\sqrt{\frac{2}{n\pi}} \int\limits_{0}^{\infty} z e^{\frac{-z^2}{2}} dz$$

$$= 2\sqrt{\frac{2}{n\pi}}$$

However, $2\sqrt{\frac{2}{n\pi}} \leq 0.1$ if $n \geq \frac{800}{\pi}$

Evaluating this fraction, we have $n \geq 254.6$, thus we note as n increases the values of $2\sqrt{\frac{2}{n\pi}}$ decreases.

As such, we need a random sample of at least size 255 to satisfy the stated condition.

8.2 #2 Find the mode of the χ^2 distribution with m degrees of freedom $(m = 1, 2, ...)$

The mode will be the value of x at which the p.d.f. $f(x)$ is at a maximum. This applies similarly to the log of $f(x)$.

The p.d.f. of $f(x)$ is given as:

$f(x) = \frac{1}{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)} x^{\frac{m}{2}-1} e^{-\frac{x}{2}}$

Note, the first part of this equation does not include x, thus we may note ϕ as some constant in taking the log, giving us:

Thus $log\left(f(x) \right) = \phi + \left(\frac{m}{2} - 1 \right) log(x) - \frac{x}{2}$

To solve for the mode, we then set $m = 1, 2, 3, 4$, and attempt to maximize (when possible).

Generally, we may maximize by deriving with respect to x and setting the equation equal to 0. This only allows us to solve for $m \geq 2$, as is shown below.

General Form

$0 = \frac{\partial log(f(x))}{\partial x} = \frac{1}{x}\left(\frac{m}{2} - 1 \right) - \frac{1}{2}$

$\rightarrow \frac{1}{2}x = \frac{m}{2} - 1 \rightarrow x = m - 2$

However, the p.d.f. is not defined on negative numbers, such that this formula does not yield answers within the parameter space for $m = 1$, -1.

Specific Commentary on m=1, 2, 3, 4

For $m = 1$, this function is strictly decreasing and increases without bound as x approaches 0. Hence the greatest density will be had at $x = 0$ for $m = 1$.

For $m = 2$, this function is strictly decreases and attains its maximum value when $x = 0$

2

If $m = 3$, the value of x at which the maximum is attained can be found by setting the derivative with respect to x equal to 0. In this instance, $x = m - 2$. This holds for $m \geq 3$

Thus, the mode of the χ^2 distribution, x, with m degrees of freedom, where $m = 1, 2, ...$ is given by the relation:

$$x = \left\{ \begin{array}{l} 0 \ for \ m < 3 \\ m - 2 \ for \ m \geq 3 \end{array} \right\}$$

8.2 #3 Sketch the p.d.f. of the χ^2 distribution with m degrees of freedom for each of the following values of m. Location the mean, the median, and the mode on each sketch.

Note, from the prior Exercise, we are able to calculate the modes of the χ^2 distribution with 1, 2, 3, and 4 degrees of freedom. Namely:

(1): $x_{m=1} = 0$

(2): $x_{m=2} = 0$

(3): $x_{m=3} = 3 - 2 = 1$

(4): $x_{m=4} = 4 - 2 = 2$

Note, the mean of the χ^2 distribution is the degrees of freedom. Hence, we have:

(1): $\bar{x}_{m=1} = 1$

(2): $\bar{x}_{m=2} = 2$

(3): $\bar{x}_{m=3} = 3$

(4): $\bar{x}_{m=4} = 4$

Similarly, to evaluate the medians of each χ^2 distribution, we use the function $qchisq(.5, df = m)$, noting the median is the 50th percentile of the distribution. Thus we have:

```
qchisq(.5, df=1)
```

```
## [1] 0.4549364
```

```
qchisq(.5, df=2)
```

```
## [1] 1.386294
```

```
qchisq(.5, df=3)
```

```
## [1] 2.365974
```

```
qchisq(.5, df=4)
```

```
## [1] 3.356694
```

(1): $\tilde{x}_{m=1} = qchisq(.5, df = 1) = 0.45$

(2): $\tilde{x}_{m=2} = qchisq(.5, df = 2) = 1.39$

(3): $\tilde{x}_{m=3} = qchisq(.5, df = 3) = 2.37$

(4): $\tilde{x}_{m=4} = qchisq(.5, df = 4) = 3.36$

Using the above information, we then make the following plots. Note, though the location of the titles change, red denotes the mode, blue denotes the mean, and green denotes the median.

```
eq = function(x){dchisq(x, df = 1)}
ggplot(data.frame(x=c(-1, 10)), aes(x=x)) +
  stat_function(fun=eq) +
  geom_vline(colour="red", xintercept=0) +
  geom_vline(colour="blue",xintercept=1) +
  geom_vline(colour="green", xintercept=0.45)+
  geom_text(x=-.4, y=0.2, label="mode") +
  geom_text(x=1.35, y=0, label="mean") +
  geom_text(x=.95, y=0.2, label="median") +
  labs(title="Chi-Squared with 1 df.") +
  theme_minimal()
```

Chi–Squared with 1 df.

Chi−Squared with 2 df.

Chi−Squared with 3 df.

Chi–Squared with 4 df.
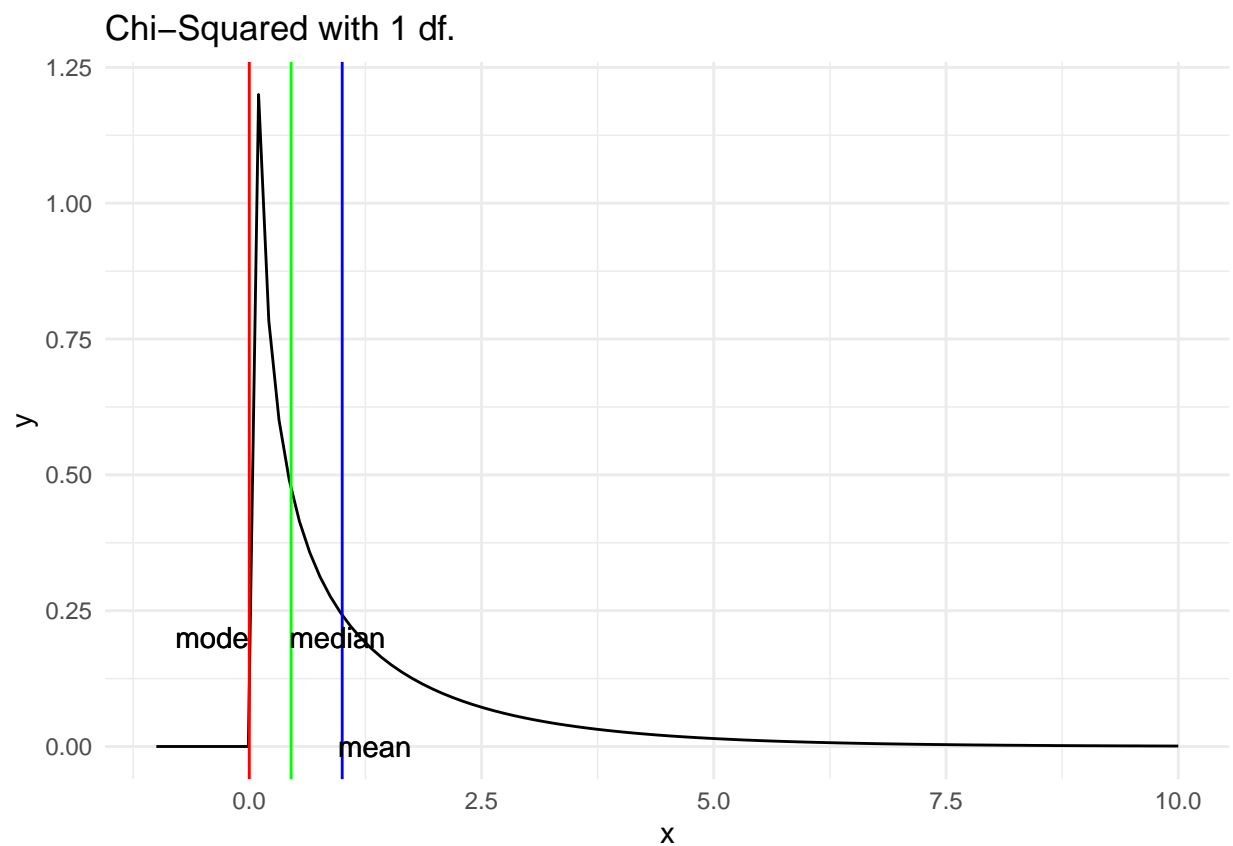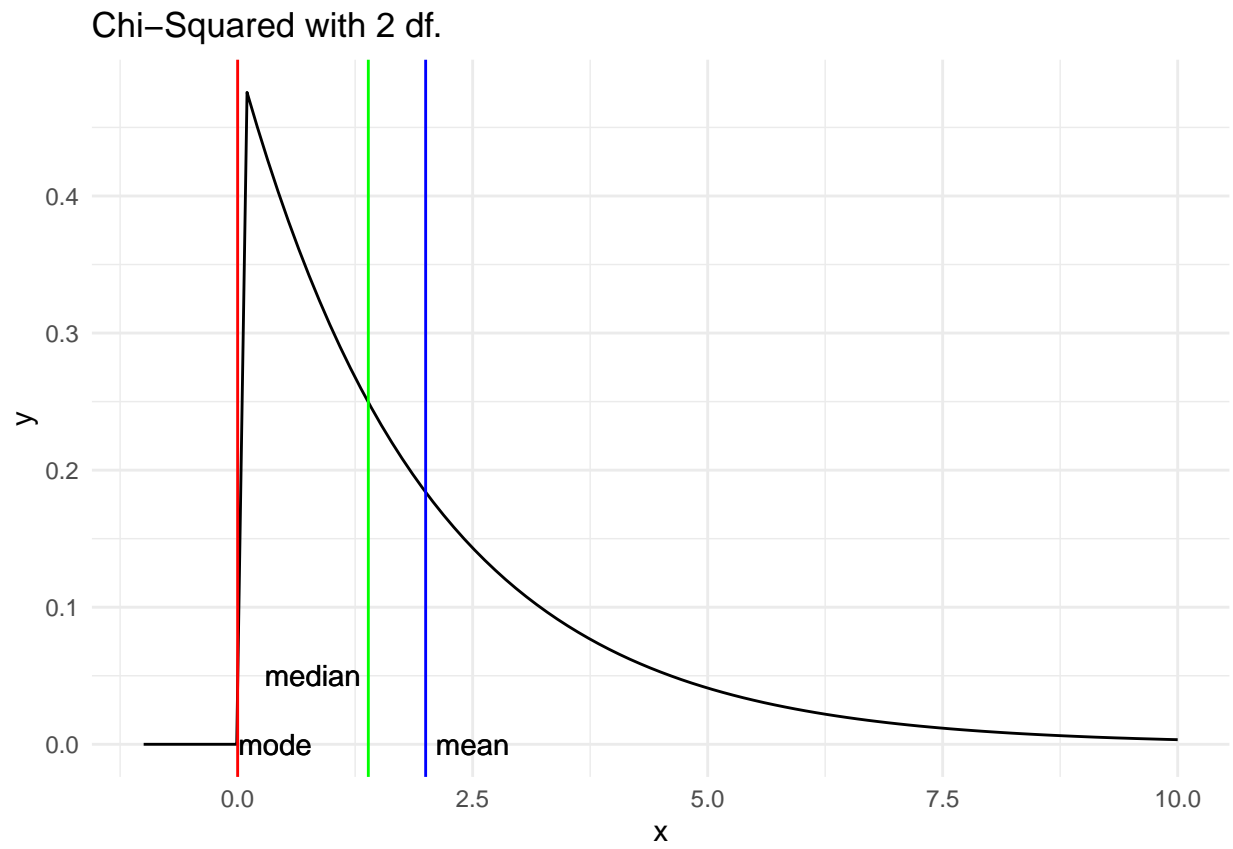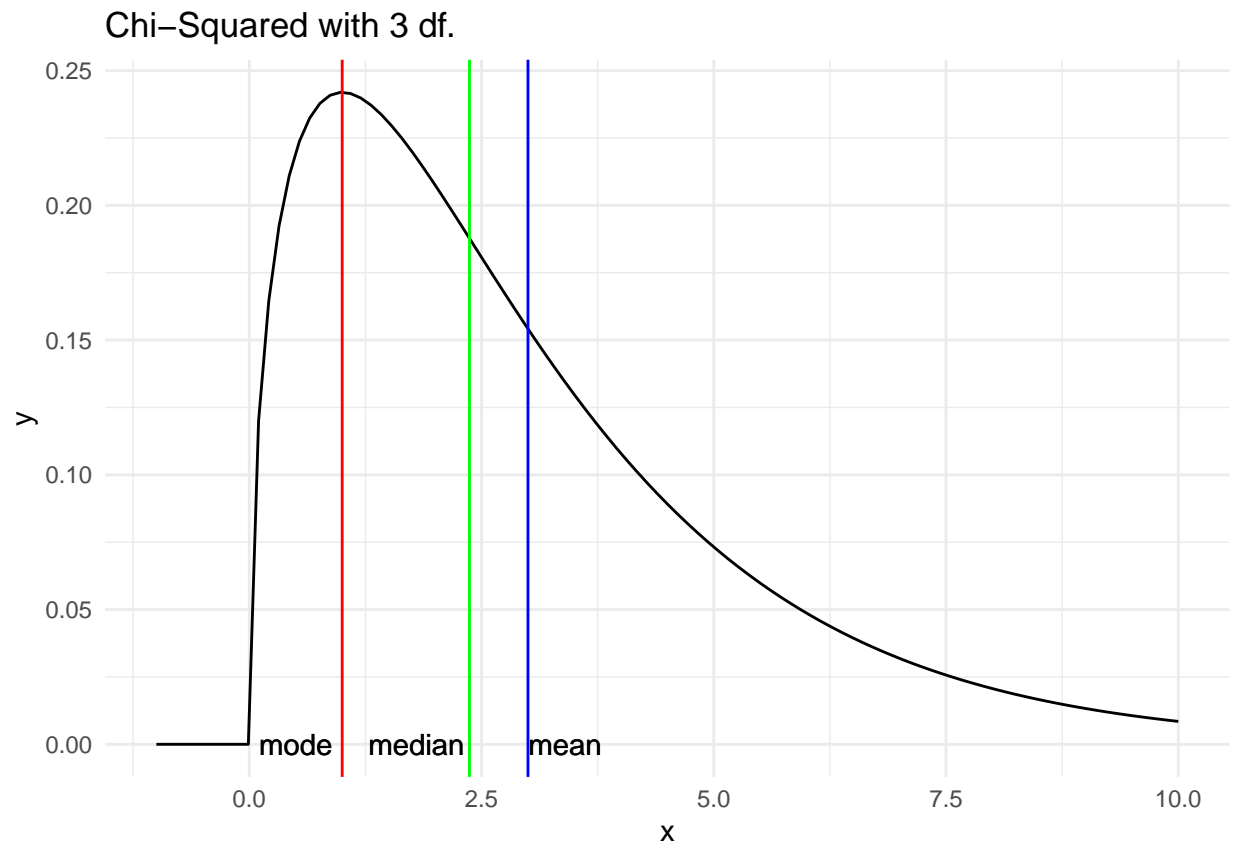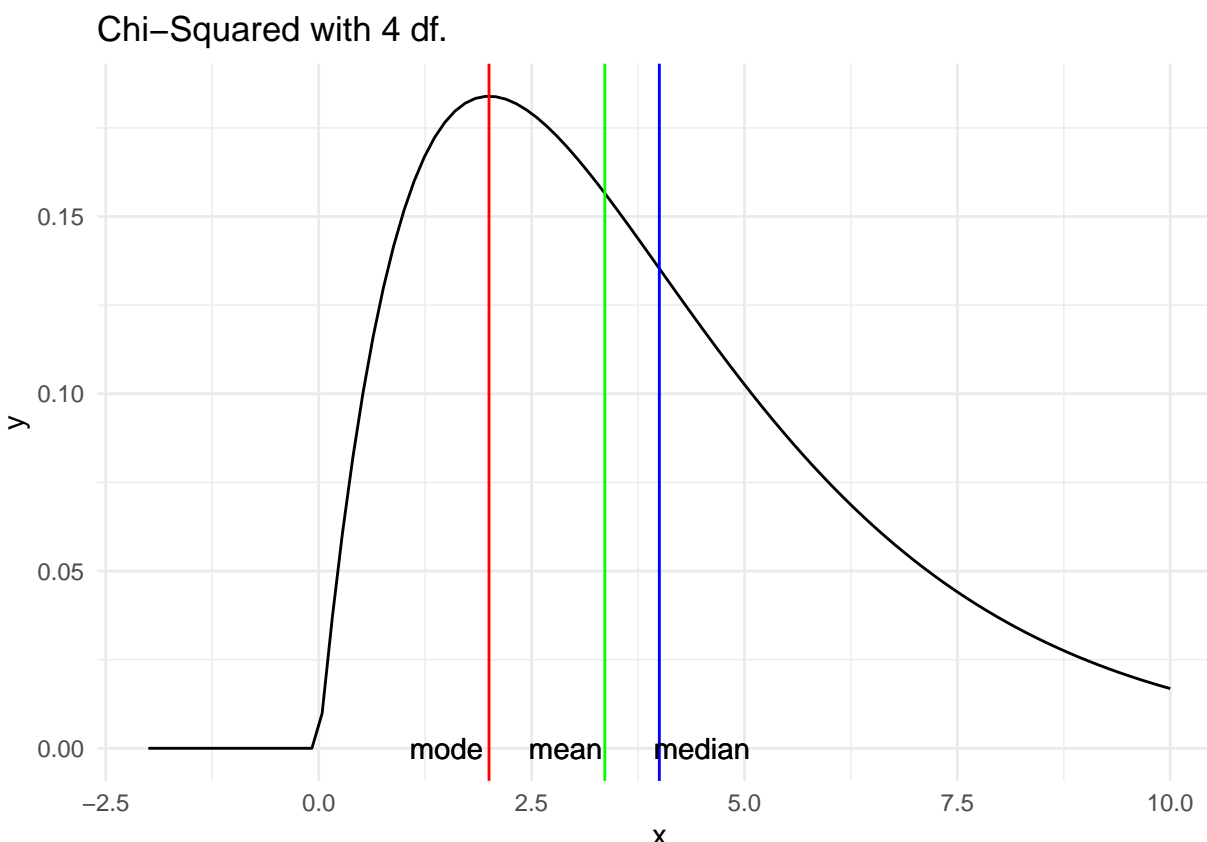
8.2 #4 Suppose that a point (X, Y) is to be chosen at random in the xy-plane, where X and Y are independent random variables and each has the standard normal distribution.

If a circle is drawn in the xy-plane with its center at the origin, what is the radius of the smallest circle that can be chosen in order for there to be probability 0.99 that the point (X,Y) will lie inside the circle?

Let r denote the radius of the circle. Thus, the point (X,Y) will lie inside the circle if $X^2 + Y^2 < r^2$

Also, $X^2 + Y^2$ has a χ^2 distribution with 2 degrees of freedom.

We may then note the c.d.f. of the χ^2 distribution with 2 degrees of freedom as follows:

$F(x, m) = \frac{\gamma\left(\frac{m}{2}, \frac{x}{2}\right)}{\Gamma\left(\frac{m}{2}\right)}$

Where $\gamma(s, x)$ is the lower incomplete gamma function given as:

$\gamma(s, x) = \int\limits_0^x t^{s-1} e^{-t} dt$

We may set the c.d.f. equal to 0.99 and solve for the bound. This is, frankly, a bit hectic (though I would be happy to screencap Wolfram Alpha if need be).

Instead, we may use the above formula to check our work, but in the meantime I highly recommend turning to page 859 and looking at $p = 0.99$ for the χ^2 distribution with 2 degrees of freedom, which gives us a value of 9.210.

Thus, we have:

$Pr\left(X^2 + Y^2 \leq 9.210\right) = 0.99$

Therefore, we substitute r^2 into the above equation to give us $r^2 \geq 9.210$, noting that r can be as small as $r = \sqrt{9.210} = 3.03$ units.

8.2 #13

Prove that the distribution of $\hat{\sigma}_0^2$ in Examples 8.2.1 and 8.2.2 is the gamma distribution with parameters $\frac{n}{2}$ and $\frac{n}{2\sigma^2}$

From the prior examples noted, let $W = \frac{n\hat{\sigma}^2}{\sigma^2}$. Then it follows that W has the χ^2 distribution with n degrees of freedom.

Note, the χ^2 distribution with n degrees of freedom is the same as the gamma distribution with parameters $\frac{n}{2}$ and $\frac{1}{2}$.

Thus, if we multiply a gamma random variable by a constant, we change its distribution to another gamma distribution. Of note, this distribution will have the same first parameter as the initial gamma distribution (in our case $\frac{n}{2}$) and a modified second parameter, $\beta_{new} = \frac{\beta_{old}}{c}$, with some constant c.

Since $\hat{\sigma}^2 = \frac{\sigma^2}{n}W$, we see that the constant c is given by:

$c = \frac{\sigma^2}{n}$

Thus, we may say the distribution of $\hat{\sigma}^2$ is the gamma distribution with parameters $\frac{n}{2}$ and $\frac{n}{2\sigma^2}$

8.4 #3 Suppose that five random variables, $X_1, ..., X_n$ are i.i.d. and that each has the standard normal distribution. Determine a constant c such that the random variable

$\frac{c(X_1+X_2)}{\left(X_3^2+X_4^2+X_5^2\right)^{\frac{1}{2}}}$

will have a t distribution.

As each $X_1, ..., X_n$ has the standard normal, $X_1 + X_2$ has the normal distribution with mean 0 and variance 2.

Therefore, $Y = \frac{(X_1+X_2)}{\sqrt{2}}$ also has a standard normal distribution.

Additionally, $Z = X_3^2 + X_4^2 + X_5^2$ has the χ^2 distribution with 3 degrees of freedom.

By their construction, Y and Z are independent as they do not share X_i terms.

Therefore, we create a variable U using a product (a division in this case) of Y and Z, noted:

$U = \frac{Y}{\left(\frac{Z}{3}\right)^{\frac{1}{2}}} = \frac{(X_1+X_2)}{\sqrt{2}\left(\frac{\left(X_3^2+X_4^2+X_5^2\right)}{3}\right)^{\frac{1}{2}}}$ has the t distribution with 3 degrees of freedom.

Thus, we may choose a constant $c = \sqrt{\frac{3}{2}}$, noting $\frac{1}{\frac{\sqrt{2}}{\sqrt{3}}} = \frac{\sqrt{3}}{\sqrt{2}}$, then the given condition will be satisfied.