

MATH 392 Problem Set 4

Sam D. Olson

Exercises from the book

8.1 #1

Random Sample X_1, \dots, 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:

$$PR(|\max(X_1, \dots, X_n) - \theta| \leq 0.1\theta) \geq 0.95$$

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

$$f(u) = \begin{cases} 0 & \text{for } u \leq 0 \\ \left(\frac{u}{\theta}\right)^n & \text{for } 0 < u < \theta \\ 1 & \text{for } u \geq \theta \end{cases}$$

Note since $U \leq \theta$ with probability 1, then $|U - \theta| \leq 0.1\theta$ 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 $E_\theta(|\bar{X}_n - \theta|) \leq 0.1$ $\forall \theta$

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

We know \bar{X}_n has the normal distribution with mean θ and variance $\frac{\sigma^2}{n}$

Thus, we have the random variable $Z = \frac{(\bar{X}_n - \theta)}{\frac{\sigma}{\sqrt{n}}}$ has the standard normal distribution.

Thus we have:

$$\begin{aligned} & E_\theta(|\bar{X}_n - \theta|) \\ &= \frac{\sigma}{\sqrt{n}} E_\theta(|Z|) \\ &= \frac{\sigma}{\sqrt{n}} \int_{-\infty}^{\infty} |z| \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \end{aligned}$$

Taking advantage of symmetry, we can simplify this to:

$$= 2\sqrt{\frac{\sigma^2}{n}} \int_0^{\infty} z e^{-\frac{z^2}{2}} dz$$

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

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 ($m = 1, 2, \dots$)

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 $f(x)$.

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

Where ϕ is some constant.

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

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

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 $m = 1, 2, \dots$ is given by the relation:

$$x = \begin{cases} 0 & \text{for } m < 3 \\ m - 2 & \text{for } m \geq 3 \end{cases}$$

8.2 #3 (use R) 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:

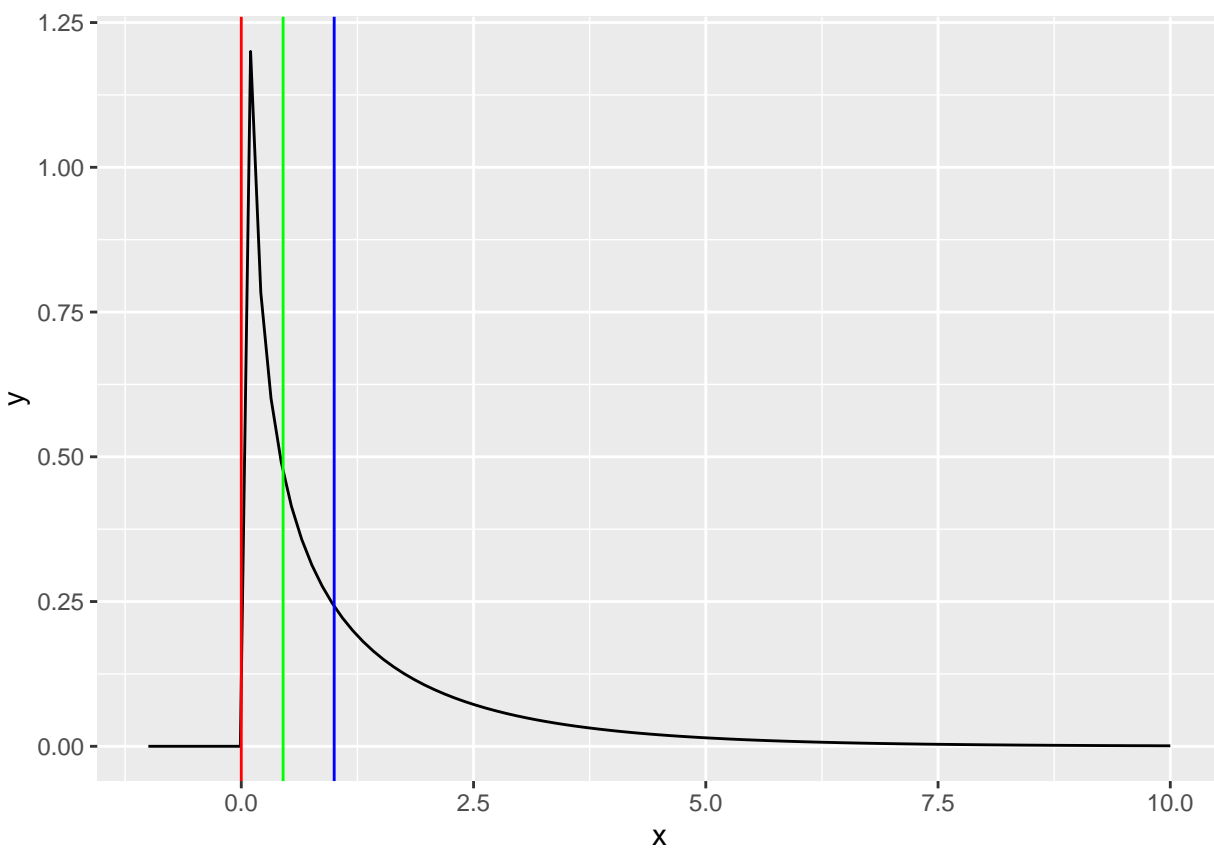
$$(1): \tilde{x}_{m=1} = qchisq(.5, df = 1) = 0.45$$

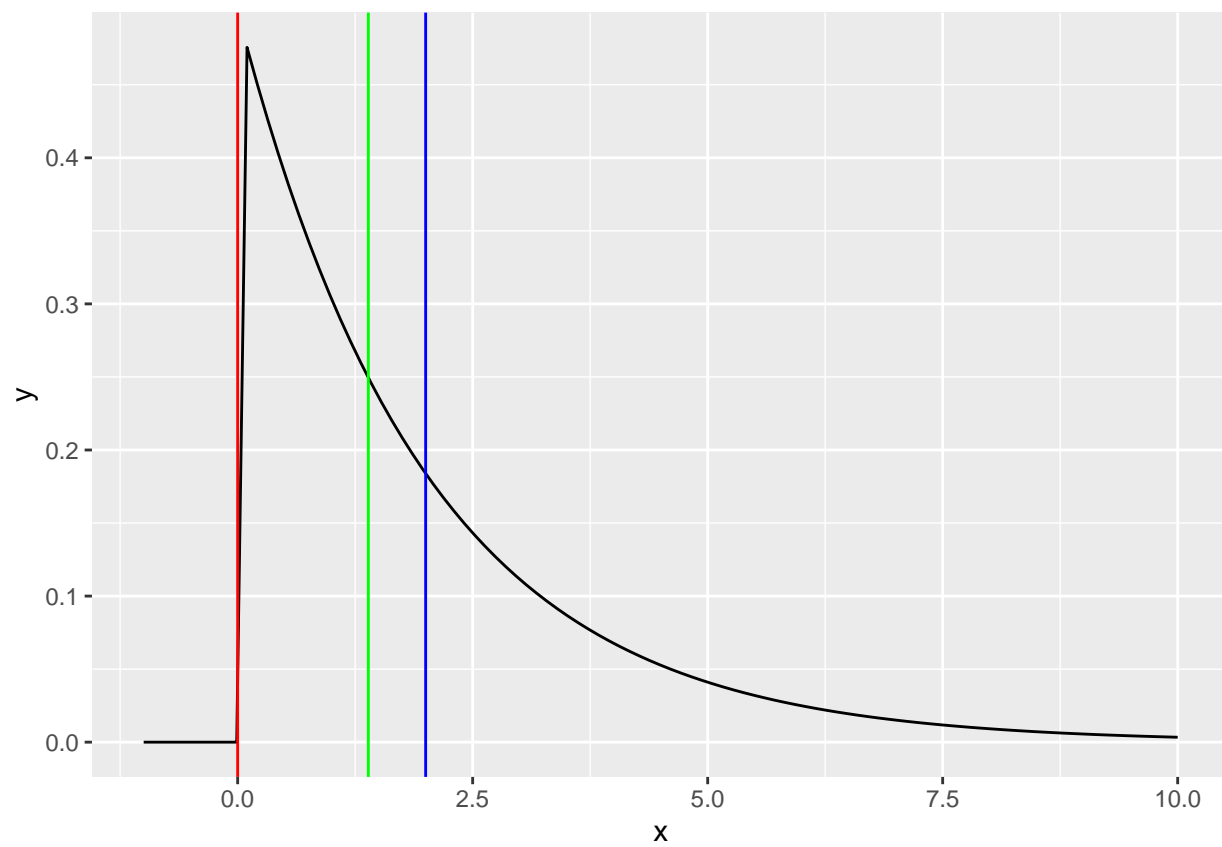
$$(2): \tilde{x}_{m=2} = qchisq(.5, df = 2) = 1.39$$

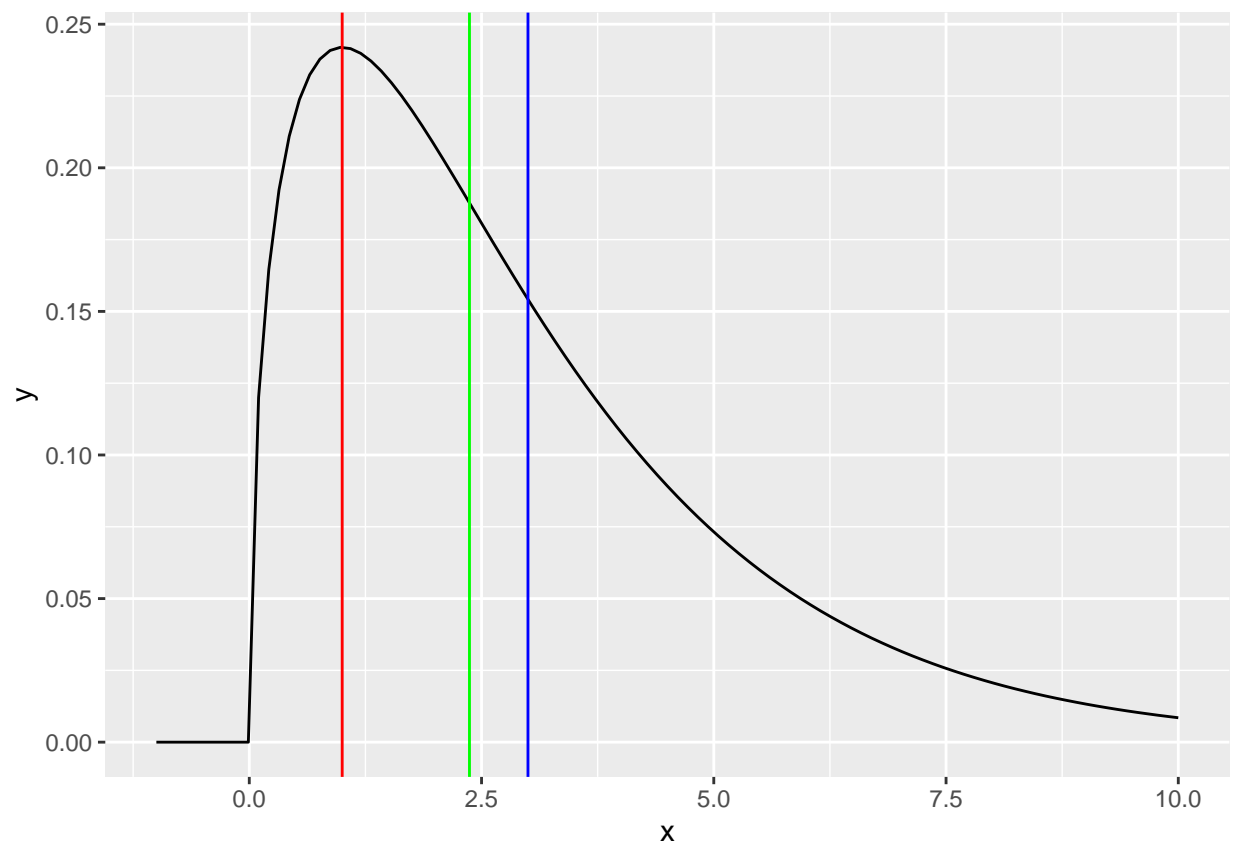
$$(3): \tilde{x}_{m=3} = qchisq(.5, df = 3) = 2.37$$

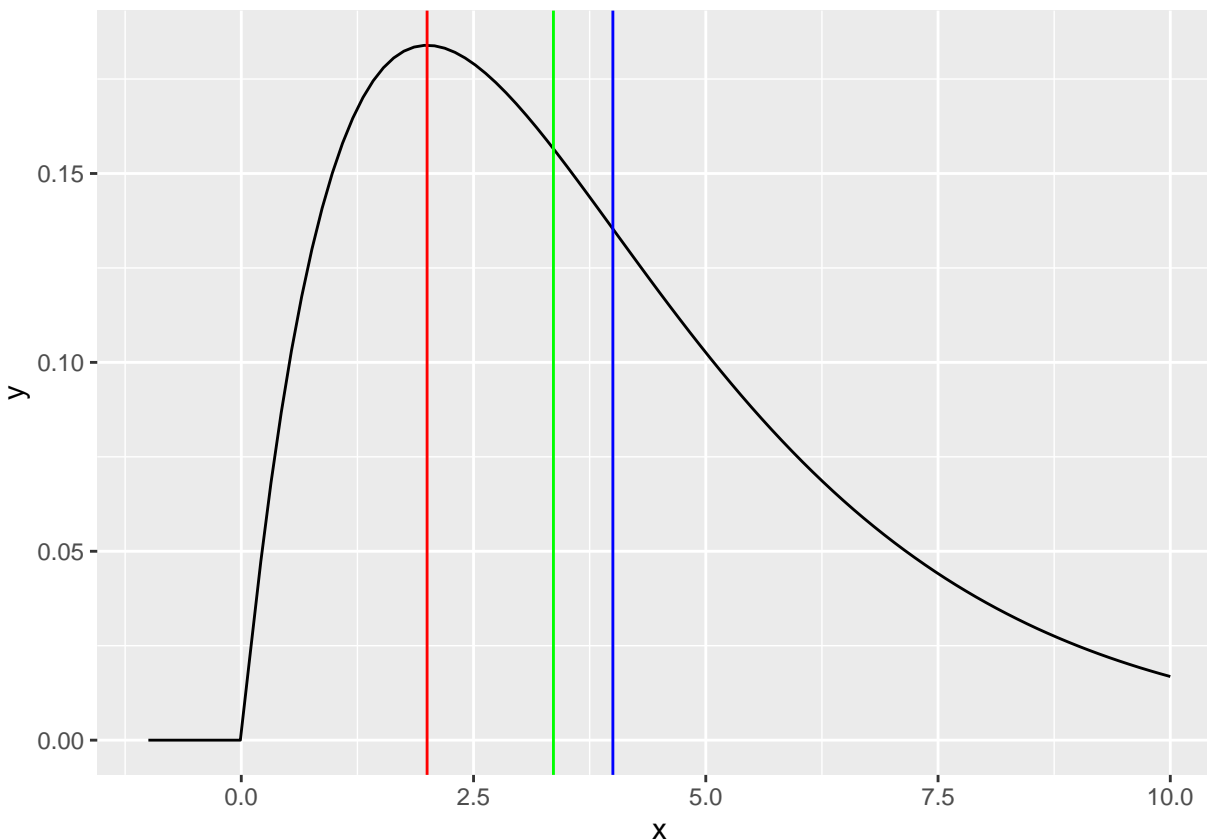
$$(4): \tilde{x}_{m=4} = qchisq(.5, df = 4) = 3.36$$

Using the above information, we then make the following plots. Note, red denotes the mode, green denotes the median, and blue denotes the mean. Full disclosure: I couldn't find a consistent working (and aesthetically pleasing) way to display labels on the below vertical lines. Very open to your thoughts.









For 8.2.3, some useful functions include: `stat_function()`, `dchisq()`, `geom_hline()`, and `qchisq()` (to find the medians).

8.2 #4 (can use a cdf for the probability calculation) 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.

Using the table at the end of the book, we have:

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

Therefore, we must have $r^2 \geq 9.210$

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, the distribution of $W = \frac{n\hat{\sigma}^2}{\sigma^2}$ is the χ^2 distribution with n degrees of freedom.

Note, this is also the gamma distribution with parameters $\frac{n}{2}$ and $\frac{1}{2}$.

If we multiple a gamma random variable by a constant, we change its distribution to another gamma distribution with the same first parameters and a modified second parameter which is the initial second parameter divided by some constant.

Since $\hat{\sigma}^2 = \frac{\sigma^2}{n}W$, we see that 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, \dots, X_n are i.i.d. and that each has the standard normal distribution. Determine a constant c such that the random variable

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

will have a t distribution

Note $X_1 + X_2$ has the normal distribution with mean 0 and variance 2.

Therefore, $Y = \frac{(X_1+X_2)}{\sqrt{2}}$ 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, noting Y and Z are independent.

Therefore, $U = \frac{Y}{(\frac{Z}{3})^{\frac{1}{2}}}$ has the t distribution with 3 degrees of freedom.

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

Case Study: German Tank Problem

Let's pick up the example that we began in class but make the simplifying assumption that we're studying a process where our sample is drawn from the continuous distribution, $X_1, X_2, \dots, X_n \sim \text{Unif}(0, \theta)$, but we're still interested in estimating θ . The MLE and Method of Moments estimators are the same:

$$\begin{aligned}\hat{\theta}_{MLE} &= \max(X_1, X_2, \dots, X_n) = X_{max} \\ \hat{\theta}_{MOM} &= 2\bar{X}\end{aligned}$$

1. Calculate the Bias What happens in the asymptote?

The pdf of X_{max} is $nF^{n-1}(x)f(x)$.

2. Calculate the Variance What happens in the asymptote?
3. Simulate and do ggridges plot to show convergence/consistency