4.41. Define the vectors $\boldsymbol{w}_i$ by

- $\boldsymbol{w}_1 = \boldsymbol{u}_1 = \frac{1}{2}\langle 1, 1, 1, 1\rangle$,
- $\boldsymbol{w}_2 = \frac{1}{2}\langle 1, 1, -1, -1\rangle$,
- $\boldsymbol{w}_3 = \frac{1}{2}\langle 1, -1, -1, 1\rangle$,
- $\boldsymbol{w}_4 = \frac{1}{2}\langle 1, -1, 1, -1\rangle$.

a) Show that these vectors form an orthonormal basis of $\mathbb{R}^4$. That is, show that they each have unit length and are pairwise orthogonal.

b) Repeat parts b)–d) of the previous question using $\boldsymbol{w}_i$ in place of $\boldsymbol{u}_i$. What do you notice?

(a)

$\|\vec{w}_1\| = \frac{1}{2}\sqrt{1^2+1^2+1^2+1^2} = \frac{1}{2}\sqrt{4} = 1$

$\|\vec{w}_2\| = \frac{1}{2}\sqrt{1^2+1^2+(-1)^2+(-1)^2} = \frac{1}{2}\sqrt{4} = 1$

$\|\vec{w}_3\| = \frac{1}{2}\sqrt{1^2+(-1)^2+(-1)^2+1^2} = \frac{1}{2}\sqrt{4} = 1$

$\|\vec{w}_4\| = \frac{1}{2}\sqrt{1^2+(-1)^2+1^2+(-1)^2} = \frac{1}{2}\sqrt{4} = 1$

Thus they all have unit length.

w_1 and w_2 : $\frac{1}{2}+\frac{1}{2}-\frac{1}{2}-\frac{1}{2}=0$

w_1 and w_3 : $\frac{1}{2}-\frac{1}{2}-\frac{1}{2}+\frac{1}{2}=0$

w_1 and w_4: $\frac{1}{2}-\frac{1}{2}+\frac{1}{2}-\frac{1}{2}=0$

w_2 and w_3 : $\frac{1}{2}-\frac{1}{2}+\frac{1}{2}-\frac{1}{2}=0$

w_2 and w_4: $\frac{1}{2}-\frac{1}{2}-\frac{1}{2}+\frac{1}{2}=0$

w_3 and w_4: $\frac{1}{2}+\frac{1}{2}-\frac{1}{2}-\frac{1}{2}=0$

Thus they are all pairwise orthogonal.

(b) $x = \langle 3, 4, 5, 8\rangle$

$p_1 = (x \cdot w_1)\, w_1 = \langle 5, 5, 5, 5\rangle$

$p_2 = (x \cdot w_2)\, w_2 = \langle -1.5, -1.5, 1.5, 1.5\rangle$

$p_3 = (x \cdot w_3)\, w_3 = \langle 0.5, -0.5, -0.5, 0.5\rangle$

$p_4 = (x \cdot w_4)\, w_4 = \langle -1, 1, -1, 1\rangle$

$p_1+p_2+p_3+p_4 = \langle 3, 4, 5, 8\rangle = x$

(c) $l_1 = \sqrt{4 \cdot 5^2} = 10$

$l_2 = \sqrt{4 \cdot 1.5^2} = 3$

$l_3 = \sqrt{4 \cdot 0.5^2} = 1$

$l_4 = \sqrt{4 \cdot 1^2} = 2$

(d) $\sum_{i=2}^{4} l_i^2 = 3^2+1^2+2^2 = 9+1+4 = 14$

$s^2 = \frac{(3-5)^2+(4-5)^2+(5-5)^2+(8-5)^2}{4-1} = \frac{14}{3}$

$3s^2 = 3 \times \frac{14}{3} = 14$

Thus $\sum_{i=2}^{4} l_i^2 = 3s^2$

I notice that $(n-1)s^2$ equals to the sum of the last $(n-1)$ terms of l_i^2

2. (1 point) If a 90% confidence interval for σ^2 is reported to be $[51.47, 261.90]$, what is the value of the sample standard deviation? Some comments:

- You may assume that this is an equal tail confidence interval.
- Make a picture of the objective function, namely the equation you are trying to find the root of.
- Include your code and calculations.

We know that $\frac{(n-1)s^2}{q_2} \leq \sigma^2 \leq \frac{(n-1)s^2}{q_1}$

Thus we get $\frac{(n-1)s^2}{q_2} = 51.47$ and $\frac{(n-1)s^2}{q_1} = 261.9$

$$\frac{\frac{(n-1)s^2}{q_2}}{\frac{(n-1)s^2}{q_1}} = \frac{51.47}{261.9}$$

$$\frac{q_1}{q_2} = \frac{51.47}{261.9}$$

We can see root $= 9.999$, so we take $n=10$ as our n.

```{r}
uniroot(f=function(n){qchisq(p=0.05,df=n-1)/qchisq(p=0.95,df=n-1)-51.47/261.9}, lower=2,upper=25)
df<-tibble(n=seq(1,50,1), result=qchisq(p=0.05,df=n-1)/qchisq(p=0.95,df=n-1))
ggplot(data=df, mapping=aes(x=n,y=result)) + geom_line() + geom_hline(yintercept=51.47/261.9) +
  ggtitle("The Objevtive Function for a 90% Confidence Interval")
```

R Console

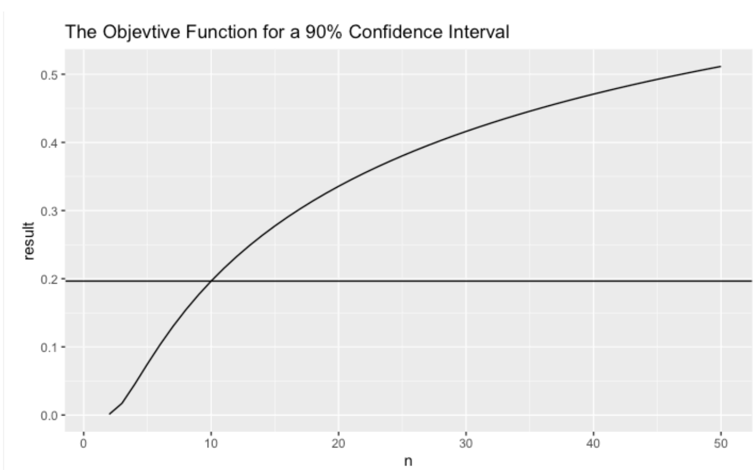
```
$root
[1] 9.999683

$f.root
[1] -8.490925e-09

$iter
[1] 6

$init.it
[1] NA

$estim.prec
[1] 6.103516e-05
```



3. (2 points) This question has three related parts which weave together to tell a story.

(a) Suppose $U \sim Unif(0,1)$. Show, using the CDF method, that $-2\ln(U) \sim Chisq(2)$.

(b) Suppose $X_1 \sim Chisq(2)$ independent of $X_2 \sim Chisq(2)$. Show, using the CDF method, that

$$S = X_1 + X_2 \sim Chisq(4).$$

(We did problems like this in 341.)

(c) Two independent research teams separately conduct a test of a point null hypothesis about the same population parameter. The test is based on a continuous test statistic. One team gets a P-value of 0.06 and the other a P-value of 0.08. Unfortunately, neither team gota significant result at a 5% level of significance. To improve their statistical power, they would really like to combine their P-values but are not sure how to go about it.

It is a fact that when a point null hypothesis is true, P-values based on continuous test statistics have a uniform distribution. Use this fact to help them combine their results. What should they conclude at a 5% level?

Be sure to explain your thought process neatly and clearly.

(a) $F_Y(y) = P(-2\ln(U) \le x)$

$= P(\ln(U) \ge -\frac{x}{2})$

$= 1 - P(\ln(U) \le -\frac{x}{2})$

$= 1 - P(U \le e^{-\frac{x}{2}})$

$= 1 - F(e^{-\frac{x}{2}})$

$= 1 - e^{-\frac{x}{2}}$

$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \frac{1}{2}e^{-\frac{x}{2}} & 0 < x < \infty \\ 0 & \text{otherwise} \end{cases}$

We can see that $\frac{1}{2}e^{-\frac{x}{2}}$ comes from a gamma distribution with $\alpha = 1$ and $\beta = 2$.

Thus $-2\ln(U) \sim$ Gamma $(\alpha = 1, \beta = 2)$

Thus $-2\ln(U) \sim$ Chisq (df=2

(b) $P(S \le s) = P(X_1 + X_2 \le s)$

$= \int_0^s \int_0^{s-x_1} f_{x_2} f_{x_1} \, dx_2 \, dx_1$

$= \int_0^s \int_0^{s-x_1} \frac{1}{4} e^{-\frac{x_2}{2}} e^{-\frac{x_1}{2}} \, dx_2 \, dx_1$

$= \int_0^s \frac{1}{4} e^{-\frac{x_1}{2}} (-2) \left[e^{-\frac{x_2}{2}}\right]_0^{s-x_1} dx_1$

$= -\frac{1}{2} \int_0^s e^{-\frac{x_1}{2} - \frac{s-x_1}{2}} - e^{-\frac{x_1}{2}} \, dx_1$

$= 1 - e^{-\frac{s}{2}} - \frac{1}{2} e^{-\frac{s}{2}} s$

$f_S(s) = \frac{d}{ds} \left(1 - e^{-\frac{s}{2}} - \frac{1}{2} e^{-\frac{s}{2}} s\right)$

$= \frac{1}{2} e^{-\frac{s}{2}} - \frac{1}{2}(-\frac{1}{2}) e^{-\frac{s}{2}} s - \frac{1}{2} e^{-\frac{s}{2}}$

$= \frac{1}{4} s e^{-\frac{s}{2}} \quad 0 < s < \infty$

Thus $S \sim$ Chisq (4)

(c) From part a, we know that $-2\ln(U) \sim$ Chisq(2). And from part (b), we know that Chisq(2) + Chisq(2) $\sim$ Chisq(4), which meas $-2\ln(U) - 2\ln(U) \sim$ Chisq(4).

$-2\ln(0.06) - 2\ln(0.08) = 10.67828$

$P(x \geq 10.6738) = 1 - P(x \leq 10.6738) = 1 - \text{pchisq}(10.6738, 4) = 0.0304$

We see that $0.0304 < 0.05$, so we can conclud it is significant at 5%.

4. (1 point) An Arterisonde machine prints blood-pressure readings on a tape so that the measurement can be read rather than heard. A major argument for using such a machine is that the variability of measurements obtained by different observers on the same person will be lower than the variability with a standard blood-pressure cuff. From previously published work, the variance with a standard blood pressure cuff is $\sigma_0^2 = 35$.

   Suppose we have data consisting of systolic blood pressure (SBP) measurements obtained on 10 people and read by two observers. We use the difference, X, between the first and second observers to assess *inter-observer* variability. In particular, if we assume

   $$X \sim N(\mu, \sigma),$$

   then it is of primary interest to test $H_0 : \sigma^2 = 35$.

   The data is in the file `systolic.csv`. Download it and create a vector of 10 differences and then calculate a 95% confidence interval for σ^2. (Even though investigators think the variability of the new method will be lower, they want to conduct a two sided test as the observers are less experienced in using it and this might result in an increase in the variability.)

   Create a brief report (of sorts) where you include a description of the data and scientific problem, model/assumptions, the hypothesis, the formula for the confidence interval, and a conclusion.

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

We assume $H_0: \sigma^2 = 35$, $H_1: \sigma^2 \neq 35$

In the data set we see there are two groups of data with each of ten. So the degree of freedom $= 10-1 = 9$. From the calculation, we get the $\bar{X}_n$ is -0.2, and $S^2 = 8.17778$. And $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = 0.06$. Then the confidence interval is

$$P\left(q_1 \leq \frac{(n-1)S^2}{\sigma^2} \leq q_2\right) = 0.95$$

$$P\left(\frac{(n-1)S^2}{q_2} \leq \sigma^2 \leq \frac{(n-1)S^2}{q_1}\right) = 0.95$$

$$P\left(\frac{9 \times 8.17778}{19.02278} \leq \sigma^2 \leq \frac{9 \times 8.17778}{2.70039}\right) = 0.95$$

$\therefore$ $[3.869, 27.256]$ is our 95% confidence interval. However 35 is not in the range. Thus we reject the H_0, and conclude that $\sigma^2 = 35$ is not reasonable.

```r
```{r}
library(tidyverse)
#read csv file
systolic<-read_csv(file="systolic.csv")
#rename variable called observer 1 as observer1
systolic <- systolic %>% rename(observer1 = "observer 1")
#calculate the difference
systolic <- systolic %>% mutate(diff=observer1-observer2)

diff_mean <- mean(systolic$diff)
S2 <- sum((systolic$diff-diff_mean)^2)/(10-1)
X2 <- 9*S2/35^2

print(diff_mean)
print(S2)
print(X2)
```
```

```
── Column specification ──────────────────────────────
cols(
  patient = col_double(),
  `observer 1` = col_double(),
  observer2 = col_double()
)

[1] -0.2
[1] 8.177778
[1] 0.06008163
```