1. Calculate the asymptotic relative efficiency for the sign statistic relative to the one-sample t-test (which you should approximate using the one-sample z-test). Do this for observations from the

a. uniform distribution, on \([-1/2, 1/2]\) with variance \(1/12\) and mean under the null hypothesis of 0, and

\[ T = \sqrt{12} \sum_{j=1}^{n} X_j / n. \]

Then \(\mu(\theta) = \theta \sqrt{12}\), and \(\mu'(0) = \sqrt{12}\). Also, \(\sigma(0) = \sqrt{n \text{Var}[T]} = 1\), and the efficiency is \(e_1 = \sqrt{12}\). As we noted in class, \(\mu'(0)\) is the density of the observations at 0, which is 1. Furthermore, \(\sigma(0) = 1/2\). So the efficacy is 2. The asymptotic relative efficiency is \((2/\sqrt{12})^2 = 1/3\).

b. the logistic distribution, symmetric about 0, with variance \(\pi^2/3\) and density \(\exp(x)/(1 + \exp(x))^2\).

As above, the T statistic has \(\mu'(0) = 1/(\pi/\sqrt{3})\), and \(\sigma(0) = 1\), with efficacy \(\sqrt{3}/\pi\). The sign statistic has mean derivative \(\mu'_2(0) = 1/4\), and \(\sigma_2(0) = 1/2\).

Hence the efficacy is \((1/4)/(1/2) = 1/2\), and the asymptotic relative efficiency is \(((\sqrt{3}/\pi)/(1/2))^{-2} = (2\sqrt{3}/\pi)^{-2} = 0.82\), indicating that the sign is 20% less efficient than the T test for the logistic distribution.

2. The data set

http://ftp.uni-bayreuth.de/math/statlib/datasets/lupus

gives data on 87 lupus patients. The third column gives duration, and the fourth column gives transformed disease duration. Give a 90% confidence interval for the median duration, through inverting the sign test, and compare this to the normal theory interval for the mean. Keep in mind that the normal theory and sign test approach are only comparable if you can argue that the mean and the median for the distribution are plausibly the same.

Comment on this.

Here is the R code to read in the data. The skip command tells R to skip the first six lines of the input file, which has a text header, and the nmax tells R to stop reading after 87 observations.

```r
lupus<-as.data.frame(scan("lupus",skip=6,nmax=87,
what=list(tim=0,status=0,duration=0,tduration=0)))
library("BSDA")#We need this for the SIGN.test function
SIGN.test(lupus[,3],conf.level=.9)
```

The appropriate interval is what R reports as the Upper Achieved CI, \((1, 4)\). To compare this to the normal theory result, do

```r
t.test(lupus[,3],conf.level=.9)
```
which gives a confidence interval of (6.65, 14.30). These are quite different, since the
distribution is heavily skewed, and the median and mean are quite different, as the histogram in Fig
1 of the transformed duration values shows:

One could have done this calculation on the transformed data:

```R
lupus<-as.data.frame(scan("lupus",skip=6,nmax=87,
what=list(tim=0,status=0,duration=0,tduration=0)))
library("BSDA")#We need this for the SIGN.test function
SIGN.test(lupus[,4],conf.level=.9)
t.test(lupus[,4],conf.level=.9)
```

which gives intervals (0.69, 1.61) and (1.18, 1.65) respectively.

3. The data set
http://lib.stat.cmu.edu/datasets/bodyfat

gives data on body fat in 252 men. The second column gives proportion of lean body
tissue. Give a 95% confidence interval for upper quartile proportion of lean body tissue.
Note that the first 116 lines and last 10 lines are data set description, and should be deleted.
(Line 117 is blank, and should also be deleted).

Calculate the positions of the bounds via

```R
a<-qbinom(0.025,21,.75);b<-21+1-qbinom(0.025,21,1-0.75)
plbt<-scan("bodyfat",skip=117,nlines=252,what=list(a=0,b=0),flush=TRUE)$b
sort(plbt)[c(a,b)]
```

Then $a = 175$, and $b = 203$. Here is the R code to read in the data. The skip command tells
R to skip the first 117 lines of the input file, which has a text header, and the nlines tells R to stop
reading after 252 observations.
The appropriate interval is
sort(plbt)[c(a,b)]

which gives a confidence interval of (23.6, 26.7).

4. Suppose 49 observations are drawn from a Cauchy distribution, displaced to have location parameter 1. What is the power of the sign test at level 0.05 to test the null hypothesis of expectation zero for these observations?

Note that $\sigma(0) = 0.5$, $\mu'(0) = 1/\pi$, and $\mu(\theta^A) = 0.75$. Use the power formula

$1 - \Phi(\sqrt{N}(\sigma(0)z_\alpha/\sqrt{N} - \mu'(0)\theta^A)) = 1 - \Phi(7(0.5 \times 1.64/7 - 1/\pi \times 1)) = 1 - \Phi(-1.408) = 0.92$, or

$1 - \Phi(\sqrt{N}(\mu(0) - \mu(\theta^A))/\sigma(0) - z_\alpha) = 1 - \Phi(7 \times (0.5 - 0.75)/0.5 + 1.64)1 - \Phi(-1.84) = 0.969$. 