Math 4397/6397
Problem Set 6 due Oct 8

Solutions

Problem 1. Here is an example of a simulation:

```
simData <- matrix(rnorm(1000 * 20, mean = 5, sd = sqrt(2)), 1000, 20)
mns <- apply(simData, 1, mean)
sds <- apply(simData, 1, sd)
tStats <- (mns - 5) / sds * sqrt(20)
RtStats<-rt(1000,df=19)
quantile(tStats, seq(0.1,0.9,0.1))
quantile(RtStats, seq(0.1,0.9,0.1))
```

It gives for `tStats`

```
        10%         20%         30%         40%         50%
-1.336811288 -0.900628675 -0.588846440 -0.272108665  0.006584178
        60%         70%         80%         90%
 0.206184179  0.495108984  0.797292577  1.184275882
```

and for `RtStats`

```
       10%        20%        30%        40%        50%        60%
-1.4379468 -0.8690964 -0.5570161 -0.2898791 -0.0703901  0.1918926
       70%        80%        90%
 0.4616374  0.8005604  1.2163823
```

They do agree to a good accuracy, as they should. To the extent that we believe R's random t-generating function, this simply illustrates the mathematical fact that standardized means of normal random variables follow the t-distribution.

Problem 2. Similarly, we use, with the already generated data:

```
vars <- sds ^ 2
quantile(19*vars/2, seq(0.1,0.9,0.1))
quantile(rchisq(1000,19))
```

This gives

```
     10%      20%      30%      40%      50%      60%      70%      80%
11.87511 13.64634 15.31343 16.63283 18.04428 19.81967 21.69575 23.87361
     90%
27.05320
```

and

```
     10%      20%      30%      40%      50%      60%      70%      80%
11.71071 13.65071 15.44795 16.89407 18.31829 19.92430 21.66645 23.72344
     90%
26.62096
```

The quantiles of appropriately normalized variances from a normal distribution do agree with the theoretical quantiles (as output by R), thus validating that $(n-1)S^2/\sigma^2$ is indeed Chi-squared with $n-1$ degrees of freedom.

Problem 3. If $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ then we know that $(n-1)S^2/\sigma^2$ is chi-squared with $n-1$ degrees of freedom. The expected value is linear, so

$$E[\frac{n-1}{\sigma^2}S^2] = \frac{n-1}{\sigma^2}E[S^2] = \frac{n-1}{\sigma^2}\sigma^2 = n-1$$

Problem 4. Let p denote the unknown proportion of wrinkled peas.

a. For "large" n, the interval is $\hat{p} \pm z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$ with the sample proportion $\hat{p} = 12/20 = 0.6$. For this data set we can use

```
phat + c(-1, 1) * qnorm(.975) * sqrt(phat * (1 - phat) / n)
```

which yields .39 to .81. Therefore we have a 95% confidence interval estimate of values from .39 to .81.

b. The margin of error for a 95% interval is $\epsilon = 1.96\sqrt{p(1-p)/n}$. Solving for n we get that we need

$$n \geq 1.96^2 p(1-p)/\epsilon^2$$

and using the inequality $p(1-p) \leq 1/4$ we know that

$$n \geq \frac{1.96^2}{4\epsilon^2}$$

works regardless of p. This gives $n = 9,604$, a safe bet. On the other hand plugging in $p = .6$ yields $n = 9,220$, a bit of savings over the $9,604$ estimate.

Problem 5. a.

```
n <- 10
p <- .3
x <- rbinom(1000, size = n, prob = p)
phat <- x / n
ul <- phat + qnorm(.975) * sqrt(phat * (1 - phat) / n)
ll <- phat - qnorm(.975) * sqrt(phat * (1 - phat) / n)
mean(ul >= p & ll <= p)
```

No, the interval only contains the true p only .84 percent of the time.

b. With $\tilde{p}$, we get

```
ptilde <- (x + 2) / (n+4)
ul <- phat + qnorm(.975) * sqrt(ptilde * (1 - ptilde) / n)
ll <- phat - qnorm(.975) * sqrt(ptilde * (1 - ptilde) / n)
mean(ul >= p & ll <= p)
```

that the coverage for this interval is .981. This is better than only .84, but now the probability is a bit too large.

c. Adapt the code from the previous part. The adjusted interval performs better.