STAT 703/J703
April 19th, 2005
-Lecture 25-

Instructor: Brian Habing
Department of Statistics
LeConte 203
Telephone: 803-777-3578
E-mail: habing@stat.sc.edu

Today

- Practice Problems
- Introducing Bayesian Statistics

Chapter 8:

1) It can be shown (pg. 376) that the variance of the sample median of a continuous random variable with median γ is approximately $1/\,4nf^2(\gamma)$. The variance of the sample mean on the other hand is always σ^2/n.

a) Consider trying to estimate the center of a normal distribution with mean μ and variance σ^2. What is the efficiency of the mean relative to the median?

b) What condition must a distribution satisfy for the median to be more efficient than the mean for estimating the center?

2a) Show that the gamma distribution is an exponential family.

b) Find the sufficient statistic for (α,β) for a gamma distribution.

Chapter 10: The given code estimates the *F* distribution using MoM, the gamma using both MoM and MLE, and the log-normal by transforming to a normal and using the standard estimates. It then calculates the Kolmogorov-Smirnov test statistic and p-value

1) Imagine that we just used the part of the code for the MoM estimator for the gamma and its test. Why isn't the p-value testing the null hypothesis "the distribution of the data is gamma"?


```
> whichdist(x)   par1  par2     D  pval
f distribution   0.100 0.100 0.482 0.000
gamma (moments)  3.039 4.984 0.024 0.611
gamma (mle)      3.095 5.075 0.021 0.753
lognormal       -0.665 0.613 0.055 0.005
```

2) What is with looking at the four tests here and concluding "we accept the null hypothesis that the data comes from an gamma distribution with parameters 3.095 and 5.075 with a p-value of 0.753."

3) If you try this with an *F* distribution, say using x<-rf(1000,3,5), several times you will find that the *F* doesn't always seem to work well. On one run I got:

```
> whichdist(x)   par1  par2     D  pval
f distribution   4.040 5.594 0.049 0.017
gamma (moments)  0.422 0.271 0.216 0.000
gamma (mle)      0.972 0.624 0.055 0.005
lognormal       -0.154 1.176 0.047 0.023
```

Any idea what could be going on with the part that checks the F? (Yes, the formula for the MoM estimator is correct).

4) For a sample of size 5 I got that all 4 distributions were accepted! What is going on here?

```
> whichdist(x)
                   par1  par2     D  pval
f distribution   17.128 5.757 0.373 0.123
gamma (moments)   0.800 0.522 0.175 0.919
gamma (mle)       0.607 0.396 0.125 0.998
lognormal        -0.590 1.809 0.154 0.971
```


5) If the sample size is really huge and you are using it on real data, why does it make sense to simply ignore the p-values and take the one with the smallest D?

Concepts for Bayes…

1) A player recently promoted to the major leagues has had 1 hit in his first 25 at bats. What do you estimate his batting average to be? (Batting average = % of times a hit is gotten in an at bat).

2) Consider your answer in 1. You are then told that the batting averages of professional major league players has a mean of around 0.266 and a standard deviation of around 0.026. What do you think about your estimate in 1 now?

3) Bayes Rule can be written

$$f(\theta \mid x) = \frac{f(x \mid \theta) g(\theta)}{\int f(x \mid \theta) g(\theta) d\theta}$$

Imagine that we knew $f(x|\theta)$ and $g(\theta)$ and wanted to find the maximum likelihood estimate of $f(\theta|x)$. Why can we just find the value of θ that maximizes $f(x|\theta)\, g(\theta)$ and not have to worry about the integral in the bottom?