STAT 703/J703
January 18th, 2007

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

Today

- Sections 8.1-8.3: Estimation and Sampling Distributions
- Section 8.4: Method of Moments

One of the primary goals in statistics is to model a population based on a sample.

Given the data set, what distribution and what parameters seem to explain it the best.

Examples:

- Time until product failure
- Number of occurrences in a certain amount of time
- Distribution of lengths in a species
- Value of the stock market on a given day

Consider a sample of heights X_1, X_2, X_3, …. X_n that you believe are normally distributed.

What parameters do you need to estimate?

What is the easiest way to do it?

Now consider applying this to a Poisson distribution.

What parameters do we need to estimate?

What are two obvious estimates?

```
x<-rpois(500,2)
mean(x)
var(x)
```


How do we decide which estimate is best?

A sampling distribution is the probability distribution of a statistic.

In general we want a sampling distribution that is as close as possible to the corresponding parameter.

Consider a sample of size two from a population with probability distribution:

x	0	2	8
p(x)	0.25	0.5	0.25

For some other distributions we can also get the sampling distributions from our previous results.

Consider X_1, X_2, X_3, …. X_n from a normal distribution. How do $\bar{x}$ and s^2 behave?

One way of avoiding the complex mathematics for other distributions is to use simulation to estimate the sampling distribution based on our current parameters.

This is the idea behind the parametric bootstrap.

8.4 – Method of Moments

One of the major ways of getting estimates of parameters is related to what we proposed with the normal distribution.

Use a number of moments equal to the number of parameters that need to be estimated, and set the sample moments equal to the distribution's moments.

STAT 702/J702 B.Habing Univ. of S.C. 13

Recall that…

$\mu_1 = E(X) = \mu$

$\mu_2 = E(X^2) = Var(X)+(E(X))^2 = \sigma^2 + \mu^2$

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \qquad \hat{\mu}_2 = \frac{\sum_{i=1}^{n} x_i^2}{n} = \frac{\sum_{i=1}^{n}(x_i - \bar{x} + \bar{x})^2}{n}$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 + n\bar{x}^2}{n} = \hat{\sigma}^2 + \bar{x}^2$$

STAT 702/J702 B.Habing Univ. of S.C. 14

In this case though we get a slightly different answer for the normal distribution.

STAT 702/J702 B.Habing Univ. of S.C. 15

Now consider a Gamma distribution.

$\mu = \alpha/\lambda$

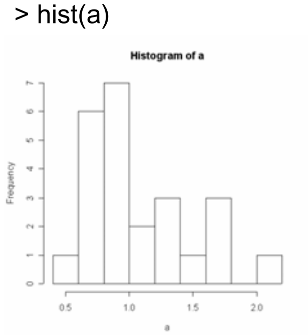
$\sigma^2 = \alpha/\lambda^2$

But how would these estimates behave?

Example: Discrimination parameters for a law school admissions test.

```
0.52208 0.61226 0.61651
0.67259 0.68124 0.70027
0.79531 0.80179 0.85638
0.87090 0.88407 0.90651
0.95291 0.99212 1.08418
1.09365 1.23861 1.36625
1.36719 1.57871 1.61840
1.67781 1.77927 2.02504
```


> hist(a)



> mean(a)
[1] 1.070585

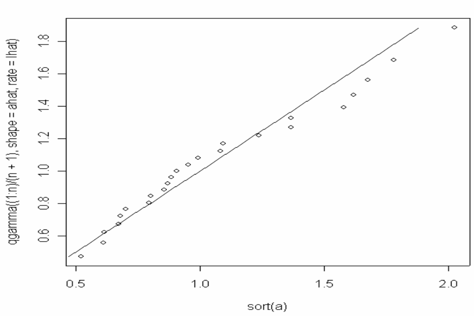
> (n-1)/n*var(a)
[1] 0.1691372

Question 1: Does the gamma with these parameters seem to match our data?

A quantile-quantile plot of our data against $F^{-1}(i/(n+1))$ could be used to see.


```
n<-length(a)
xbar<-mean(a)
sigma2hat<-(n-1)/n*var(a)
lhat<-xbar/sigma2hat
ahat<-xbar^2/sigma2hat


plot(sort(a),qgamma((1:n)/(n+1),
  shape=ahat,rate=lhat))
lines(qgamma((1:n)/(n+1),shape=ahat,
  rate=lhat),qgamma((1:n)/(n+1),
  shape=ahat,rate=lhat))
```

Say we repeated the plot with

```
a<-rnorm(24,1.07,sqrt(.1677))
```