Stat 290: Lab 4

Binomials and Proportions

Lab Objective

1. To study binomials and proportions

Assignment

There is nothing to be turned in for this lab.

Getting the Data

You probably still have the VA data from lab 2, but if not, you will need to get it again from the course web page (the 1992 and 1993 datasets). Follow those instructions to create the objects `y92`, `n92`, and `f92`.

Intervals for Proportions

Let's make some basic intervals for the proportions in the VA data. First plot the data:

```
plot(n92, f92, ylim=c(0,1))
```

Next compute the mean proportion observed in 1992, `p92_mean(f92)`, and create a sorted vector of the counts, `z_sort(n92)`. We will first make pointwise intervals assuming that all hospitals have the same true proportion. At each datapoint, we construct an interval for the proportion based on the binomial density function:

```
abline(h=p92, col=2)
lines(z, qbinom( 0.005, z, p92 )/z, col=3)
lines(z, qbinom( 0.995, z, p92 )/z, col=3)
```

(the colors are optional)
Make sure you understand what these lines are (read the code carefully until you figure it out). Note that a lot of the points fall outside the lines. What is the significance of this?

Next we try intervals centered at the observed proportions for each hospital. We compute the lower bound of the interval as `v92` using the asymptotic approximation, and use the function `error.bar` (see the on-line help for more info on this function):

```
v92_1.96*sqrt(f92*(1-f92)/n92)
error.bar( n92, f92, lower=v92, bar.end=F,  ylim=c(0,1), ylab="f92")
```

(Note that `error.bar` returns `NULL` at the S prompt because all functions must return something, but this function merely exists to produce a graph, so it returns the empty set at the S prompt.) Notice how the lengths of the intervals change with the size.

Log Quadratic Approximations

The log quadratic approximation to the relative likelihood is discussed in Section 3.3 of the Problems of Proportions handout. Here we'll fit that approximation to a few of the hospitals. First we look at hospital 1 in 1992:

```
f92[1]; n92[1]; y92[1]
```

We'll calculate and plot the relative likelihood over a sequence of 400 evenly spaced points:

```
theta_seq(0,1,length=400)
rlike_dbinom(y92[1],n92[1],theta)/dbinom(y92[1],n92[1],f92[1])
plot(theta,rlike,type='l')
```

Finally we compute the asymptotic variance, τ^2, and compute and plot the log quadratic approximation:

```
tau2_f92[1]*(1-f92[1])/n92[1]
lines(theta, exp(-(theta-f92[1])^2/(2*tau2)), col=3)
```

Alternatively, instead of using color (remember to change your color scheme under the motif options menu if you are using color), you could use a different line type:

```
lines(theta, exp(-(theta-f92[1])^2/(2*tau2)), lty=3)
```

Either way, you should see that the lines are nearly identical, so the approximation is excellent in this case. Now repeat this process for hospital 6, and you should see that while the approximation is still good, it is not as good as in the first case.

Finishing Up

There is no assignment for this lab. When you are done in S, the command to quit is `q()`. Make sure you quit S before logging off so that you don't spawn a runaway process.