Estimating Beta-Binomial Parameters

STA1403

February 22, 2017

INTRODUCTION

In the "gender in families" case study, we compared Geissler's data to two different binomial distributions and a beta-binomial distribution. The latter was characterized by two *shape* parameters which determine the mean and variance of this distribution.

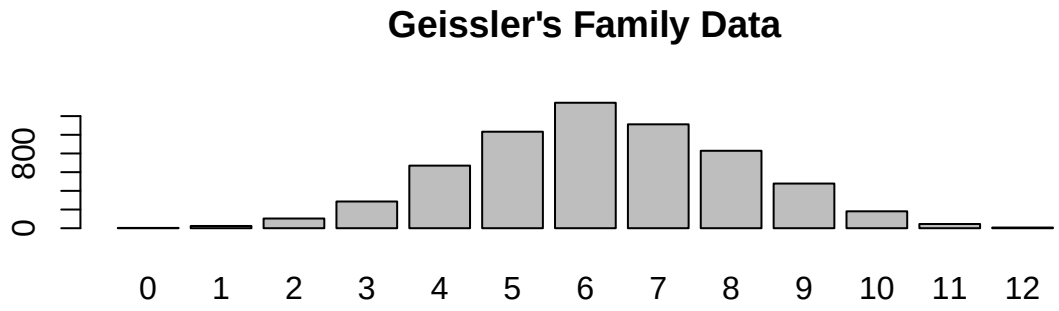
$$E[X] = \frac{n\alpha}{\alpha+\beta} \qquad Var[X] = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

where n is the number of "trials" in the modeled experiment. In the case of the number of boys in families with twelve children, $n = 12$. The challenge here is to calculate the two shape parameters α and β, knowing only n and the sample mean and variance. This technique is called the *method of moments*.

THE DATA

Get Geissler's family data, and calculate the mean number of boys and the variance.

```
boy.data <- read.table("http://tinyurl.com/zgwaep4/GENDER.txt", header=TRUE)
attach(boy.data)
barplot(families, names.arg=boys, main="Geissler's Family Data" )
```




```
# find mean and variance of the number of boys
Nf      <- sum(families)
X.sum   <- sum(boys*families)
X.mean  <- X.sum / Nf
X.sumsq <- sum(families*boys^2)
X.var   <- (X.sumsq - Nf*X.mean^2) / (Nf-1)
detach(boy.data)
c(X.mean, X.var)
```

```
## [1] 6.230581 3.489840
```

COMPUTATION

A Non-Linear Equation Solver

The **nleqslv** package enables R to solve systems of non-linear equations. Each equation is written with all the terms on the right-hand side, and the solver attempts to find solutions that will set each equation to zero.

```
# install.packages("nleqslv")
library(nleqslv)
```

Define the System of Equations...

Set up the equations to be solved as a single function:

```
MoMEstimates <- function(ab) {
  y <- numeric(2)
  y[1] <- X.mean - 12* ab[1] / (ab[1] + ab[2])
  y[2] <- X.var - 12*ab[1]*ab[2]*(ab[1]+ab[2]+12)/((ab[1]+ab[2])^2*(ab[1]+ab[2]+1))
  y
}
```

...and Solve Them

Then give it a set of starting values and find the solution:

```
ab <- c(30,30)
solution <- nleqslv(ab, MoMEstimates, method="Newton", control=list(allowSingular=TRUE))
solution$x
```

```
## [1] 34.09501 31.57144
```