STAT40180 — Stochastic Models

Brendan Murphy

Week 1

Inference Example

Models

- We will consider the crime statistics example.
- We will establish how we can fit the posited model to the data.

Crime Statistics

- We have ten observation $x_1, x_2, \ldots, x_{10}$ that we are proposing to model as Binomial(n, p) with n and p unknown.
- This is an unusual binomial problem because n is unknown.
- Further, n is a discrete quantity so we can't use calculus based methods.
- Also, $0 < p < 1$ which means it is bounded; this may (or may not) be problematic.

Crime Statistics: Method of Moments

- We could try to use method of moments to estimate the model.
- We have two unknown parameters (n, p), so we will need two equations to uniquely identify them.
- We know that under the binomial model

$$\mathbb{E}(X_i) = np \text{ and } \mathbb{V}\text{ar}(X_i) = np(1-p).$$

- If we replace the expected values by the sample moments, we get

$$np = \overline{x} \text{ and } np(1-p) = s^2.$$

Crime Statistics: Method of Moments

- Thus,

$$\begin{aligned} & \overline{x}(1-p) = s^2 \\ \Rightarrow \quad & (1-p) = \frac{s^2}{\overline{x}} \\ \Rightarrow \quad & p = 1 - \frac{s^2}{\overline{x}} \end{aligned}$$

  and

$$n = \frac{\overline{x}}{p}$$

Crime Statistics: Estimates

- For the given data, we get:
  $\hat{p} = 0.61$ and $\hat{n} = 54$.
- Thus, we estimate that there are 54 crimes per week and 61% of crimes occuring are reported.
- The following code can be used:

```
x <- scan()
38 34 32 34 32 27 28 36 37 33

xbar <- mean(x)
s2 <- var(x)

phat <- 1-s2/xbar
nhat <- xbar/phat

phat
nhat
```

Crime Statistics: Likelihood

- We could try to estimate (n, p) using maximum likelihood.
- It turns out to be non-trivial, but it is perfectly managable.
- For the observed data, we get the following likelihood function:

$$L(n,p) = \prod_{i=1}^{m} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i}.$$

- The log-likelihood is:

$$\ell(n,p) = \sum_{i=1}^{m} \log \binom{n}{x_i} + \left(\sum_{i=1}^{m} x_i\right) \log p + \left(nm - \sum_{i=1}^{m} x_i\right) \log(1-p).$$

- We want to maximize this, with respect to (n, p).

Crime Statistics: Likelihood

- Suppose, for a moment, that n is known.
- We could maximize the likelihood with respect to p to find that

$$\hat{p}(n) = \frac{\sum_{i=1}^{m} x_i}{nm} \qquad \text{Check!}$$

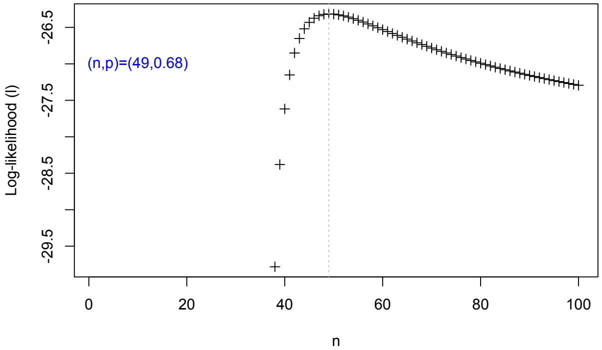
- *I have written it as a function of n because the calculation assumed n known.*
- We could replace p in the likelihood by $\hat{p}(n)$ to get

$$\ell(n, \hat{p}(n)) = \sum_{i=1}^{m} \log \binom{n}{x_i} + \left(\sum_{i=1}^{m} x_i\right) \log \hat{p}(n) + \left(nm - \sum_{i=1}^{m} x_i\right) \log(1 - \hat{p}(n))$$

- There is no straightforward way to maximize the resulting function with respect to n.

Crime Statistics: Likelihood

- However, because n is a whole number, we can evaluate $\ell(n, \hat{p}(n))$ for a range of values of n.
- The resulting plot is as follows:



- In this case, we got a lower value for the number of crimes but a higher percentage being reported.

Crime Statistics: Likelihood Code

- The code for doing the maximum likelihood estimation.

```
l<-function(n,p,x)
{
sum(dbinom(x,n,p,log=TRUE))
}

phat<-function(x,n)
{
m<-length(x)
sum(x)/(n*m)
}

l2<-function(n,x)
{
l(n,phat(x,n),x)
}

nvec<-1:100
lvec<-rep(NA,length(nvec))

for (n in nvec)
{
lvec[n]<-l2(nvec[n],x)
}
plot(nvec,lvec,pch=3,xlab="n",ylab="Log-likelihood (l)")

abline(v=49,col="gray",lty=3)
text(10,-27,"(n,p)=(49,0.68)",col="blue")
```