Assignment 1 – Questions and answers

- Question 1 – Likelihood of binomial model

- Answer 1

$$l(\theta) \sim \theta^{21}(1-\theta)^{173}$$

or

$$\log l(\theta) = 21log\theta + 173\log(1-\theta)$$

Questions and answers

- Question 2 – Plot the likelihood of θ from 0 to 1

- Answer 2

-if plot likelihood:

curve(x^21 * (1-x)^173, from=0, to=1,xlab=expression(theta),
ylab="Likelihood", main="Likelihood of theta between 0 and 1",las=1)

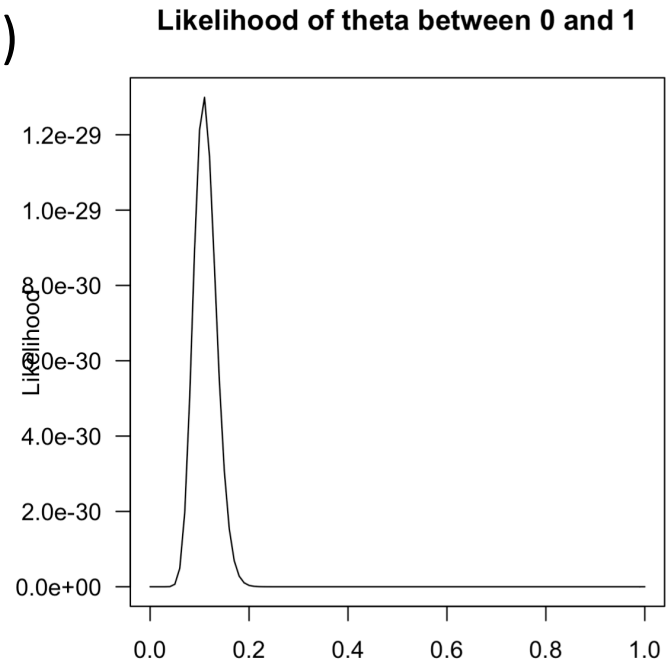
-if plot log-likelihood:

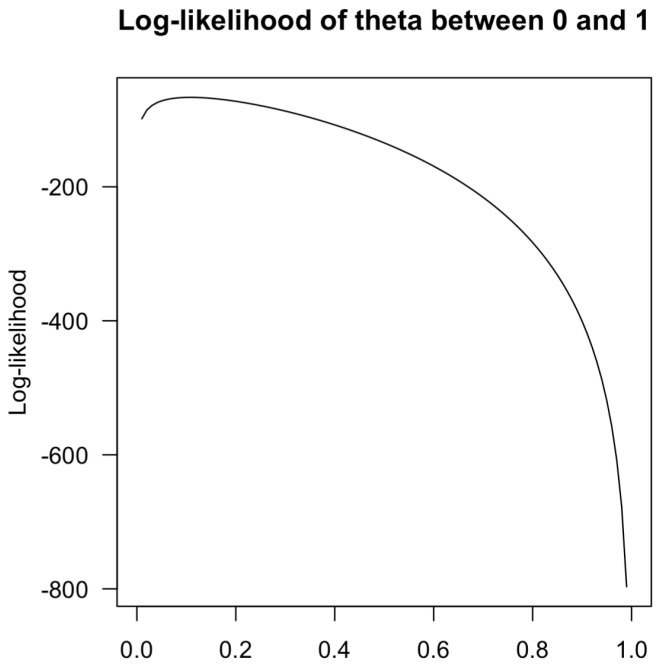
curve(21*log(x)+173*log(1-x),0,1,xlab=expression(theta),ylab="Log-likelihood",main ="Log-likelihood of theta between 0 and 1",type="l",las=1)

Questions and answers

- Question 2 – Plot the likelihood of θ from 0 to 1

- Answer 2 (continued)





Questions and answers

- Question 3 – Find the MLE of θ

- Answer 3 – Can use optimize(function(x) x^21 *
  (1-x)^173, interval=c(0,1), maximum=TRUE)
  to get around 0.11, or 21/194

  or

```
loglik <- function(x) 21*log(x) + 173*log(1-x)
optim(par=0.01, loglik, lower=0.01, upper=0.2,
method="L-BFGS-B", control=list(fnscale=-1))
```

Questions and answers

- Question 4 – 95% confidence interval of θ

- Answer 4 – Can use normal approximation of

$n\hat{\theta} \sim N(n\theta, n\theta(1-\theta))$ to obtain CI of

$$21 \pm 1.96 \times \sqrt{194 \times \frac{21}{194} \times (1 - \frac{21}{194})} = (12.5, 29.5), \text{ then}$$

dividing n = 194 to get (0.06, 0.15)

- Alternative answer 4 – Using likelihood ratio

$$\left|\log l\left(\frac{21}{194}\right) - \log l(\theta)\right| < 1.92 \rightarrow (0.06, 0.15)$$

- Alternative answer 4 – using bootstrap

Questions and answers

- Question 5 – p-value of MLE

- Answer 5 – Need to specify null hypothesis.
- Suppose H_0: $\mu = 0.10$, then use

  binom.test(21, 194, 0.10)

  p-value is 0.72 in this case, hence cannot reject the null hypothesis

- Remark – this function also returns the 95% CI

Questions and answers

- Question 6 – Comment on findings

- Answer 6 – How would you interpret and communicate the CI?

  95% confident prevalence is between 6% and 15%. Result depends on binomial model being correct, which assumes independence between students, and that our sample is representative, etc.