Introduction to Statistics & Maximum Likelihood

SOC 512 & CSSS 505

written by Laina Mercer & Jessica Godwin

Aaron Osgood-Zimmerman

Department of Statistics
University of Washington

March 4, 2021

Outline

- Motivation
- Likelihood
- Maximum Likelihood
- Confidence intervals

Motivation

We have been discussing the *probability* of events given distributions with set parameters.

For example, assume X =the number of heads in two consecutive coin tosses. Then $X \sim$Binomial$(n = 2, p = 0.5)$ and we know the probability of each possible value of X can be calculated with

$$P(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x}$$

resulting in the following probabilities:

X	0	1	2
$P(X = x)$	0.25	0.5	0.25

Likelihood

However, in practice, we will first collect data (or download it from the internet).

Then, based on the nature of the data, the study design, or other factors we will assume the distribution (or family) the data was generated from (Binomial, Guassian, Poisson, etc...).

After which we will use our data to estimate the parameters of the distribution we are assuming. The parameters are estimated to maximize the likelihood function given the data we have observed.

Likelihood

The likelihood of the data will look very similar to the the probability distributions, however the likelihoods are functions of the parameters instead of the random variable.

For example, with a binomial distribution we have the distribution:

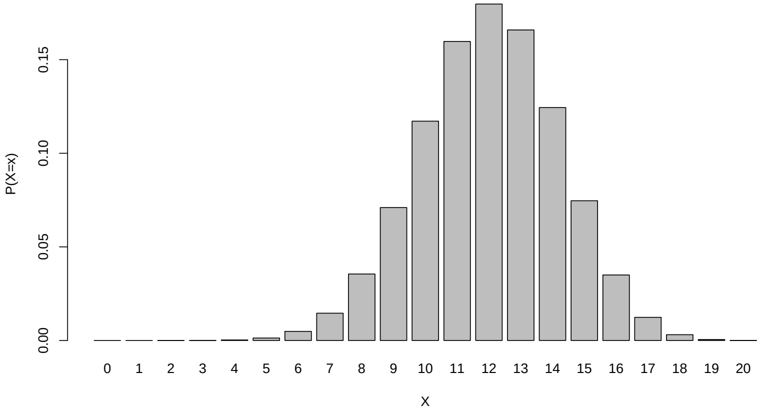
$$P(X = x|p) = \binom{n}{x} p^x (1-p)^{n-x}$$

and the likelihood function

$$L(p|X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

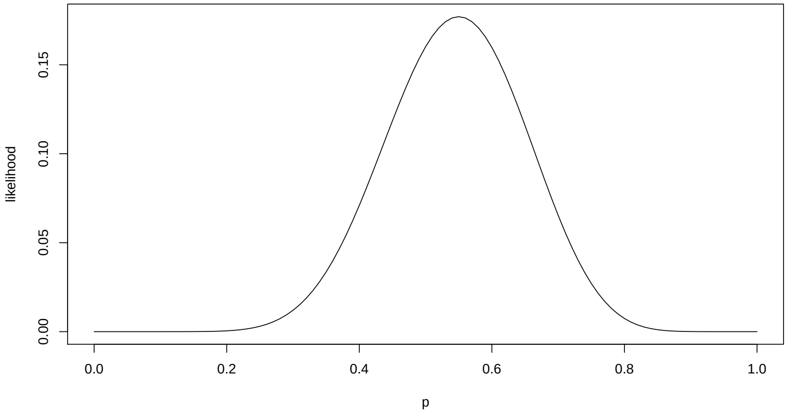
Probability Distribution

$X \sim$Binomial$(n = 20, p = 0.6))$



Data Likelihood based on an experiment

n=20 and $X = 11$



Example

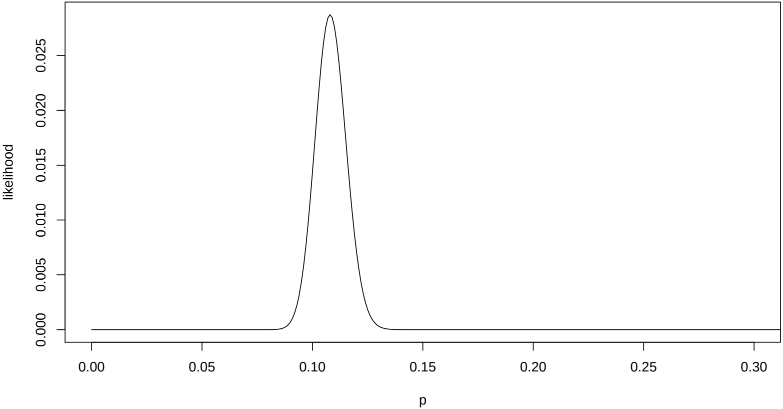
What is the prevalence of HIV in South Africa?

If we take a random sample of 2,000 South Africans and test for HIV. Let X =the number of HIV+ individuals in the sample. Then we assume X follows a binomial distribution with parameters n=2,000 and p=unknown national prevalence.

We will take the p that maximizes the likelihood function as our best estimate for the national HIV prevalence.

Likelihood Example

If 2,000 are sampled and 216 test HIV+.



Maximizing the Likelihood

Finding the Max

How do we find the maximum?

- Take the derivative
- Set it equal to zero
- Solve for the parameter of interest
- Find the second derivative
- Evaluate second derivative at critical value, if negative, we have a max!

For a refresher check out the notes from Lecture 3.

Maximizing the Likelihood

Finding the Max

Most probability likelihood functions are the products of smaller functions of the parameters of interest. For example:

$$L(p|X=x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

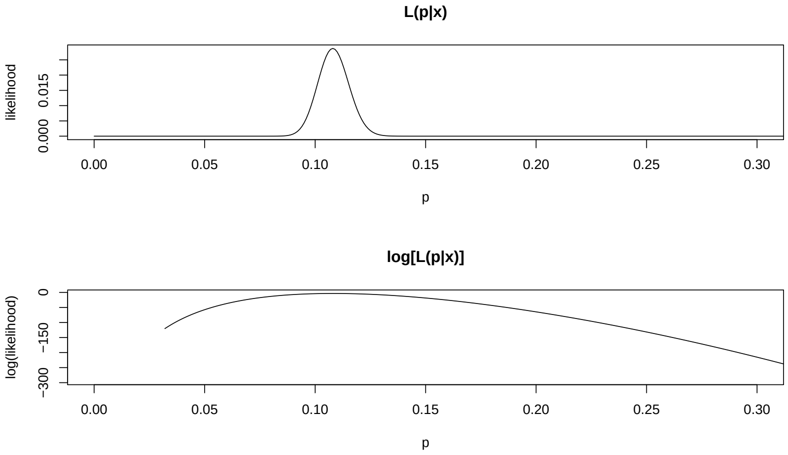
As you know (from Lecture 3) we can find the derivative by using the product rule. However, it is often easier to deal with the log of the likelihood (called the log-likelihood) instead.

Remember from Lecture 1:

$$log(a \cdot b) = log(a) + log(b)$$

The log is a monotone transformation, so it does not affect the location of the maximum.

Likelihood & Log-likelihood



Maximizing the Likelihood

Step 1: Take the log of the likelihood function

Likelihood:

$$L(p|X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Loglikelihood:

$$\begin{aligned} l(p|X = x) &= log\left[\binom{n}{x} p^x (1-p)^{n-x}\right] \\ &= log\left[\binom{n}{x}\right] + log(p^x) + log\left[(1-p)^{n-x}\right] \\ &= log\left[\binom{n}{x}\right] + xlog(p) + (n-x)log(1-p) \end{aligned}$$

Maximizing the Likelihood

Step 2: Take the derivative of the log likelihood with respect to p

Loglikelihood:

$$l(p|X=x) = log\left[\binom{n}{x}\right] + xlog(p) + (n-x)log(1-p)$$

Derivative wrt p:

$$\begin{aligned}
\frac{d}{dp}l(p|X=x) &= \frac{d}{dp}\left[log\left[\binom{n}{x}\right] + xlog(p) + (n-x)log(1-p)\right] \\
&= 0 + \frac{x}{p} + \frac{n-x}{1-p}\cdot(-1) \\
&= \frac{x}{p} - \frac{n-x}{1-p}
\end{aligned}$$

Maximizing the Likelihood

Step 3: Set the derivative equal to zero and solve for $\hat{p}$ (the estimate of p)

Derivative wrt p:

$$\frac{d}{dp} l(p|X = x) = \frac{x}{p} - \frac{n-x}{1-p}$$

Set equal to 0 and solve for $\hat{p}$:

$$\begin{aligned}
0 &= \frac{x}{\hat{p}} - \frac{n-x}{1-\hat{p}} \quad \text{[add } \frac{n-x}{1-\hat{p}} \text{ to both sides]} \\
\frac{n-x}{1-\hat{p}} &= \frac{x}{\hat{p}} \quad \text{[multiply by common denominator } \hat{p}(1-\hat{p})\text{]} \\
(n-x)\hat{p} &= x(1-\hat{p}) \quad \text{[distribute]} \\
n\hat{p} - x\hat{p} &= x - x\hat{p} \quad \text{[add } x\hat{p} \text{ to both sides]} \\
n\hat{p} &= x \quad \text{[divide by n]} \\
\hat{p} &= \frac{x}{n}
\end{aligned}$$

Maximizing the Likelihood

Step 4: Take the second derivative of the log likelihood.

Derivative wrt p:

$$\frac{d}{dp} l(p|X = x) = \frac{x}{p} - \frac{n-x}{1-p}$$

Second derivative:

$$\begin{aligned}
\frac{d^2}{dp^2} l(p|X = x) &= \frac{d}{dp}\left[\frac{x}{p} - \frac{n-x}{1-p}\right] \\
&= \frac{x}{p^2}(-1) - \frac{n-x}{(1-p)^2}(-1)(-1) \\
&= \frac{-x}{p^2} - \frac{n-x}{(1-p)^2}
\end{aligned}$$

Maximizing the Likelihood

Step 5: Plug $\hat{p}$ into the second derivative. Is it a maximum?

Second derivative:

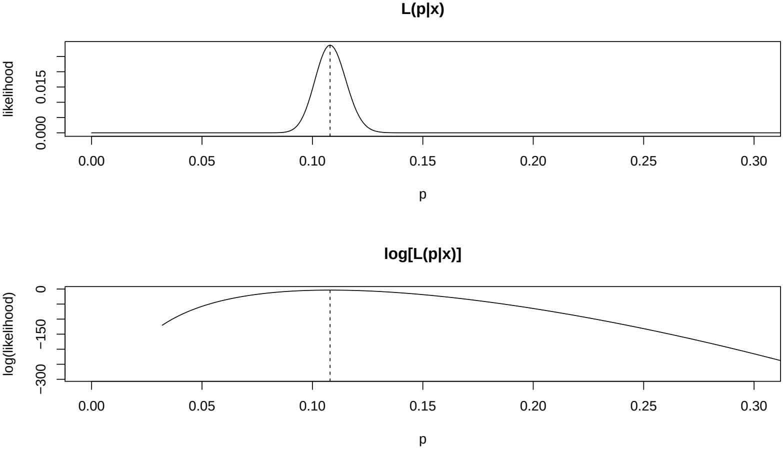
$$\frac{d^2}{dp^2} l(p|X = x) = \frac{-x}{p^2} - \frac{n - x}{(1 - p)^2}$$

Second derivative evaluated at $\hat{p}$:

$$\begin{aligned}
\frac{d^2}{dp^2} l(\hat{p}|X = x) &= \frac{-x}{\hat{p}^2} - \frac{n - x}{(1 - \hat{p})^2} \\
&= \frac{-x}{(x/n)^2} - \frac{n - x}{(1 - (x/n))^2} \\
&= \frac{-n^2}{x} - \frac{n^2}{n - x} \\
&= \frac{-n^3}{x(n - x)} \\
0 &> \frac{-n^3}{x(n - x)}, \text{ for all } 0 \leq x \leq n
\end{aligned}$$

Maximum Likelihood Estimator

If X=216, the MLE is 216/2000=0.108



Maximum Likelihood Estimator

Distribution of the MLE

If we want to know how well we are estimating our parameter (calculate the uncertainty), we need to think about the distribution of $\hat{p}$.

From Lecture 6 (slide 22) we know how to find the mean and variance of a random variable multiplied by a constant.

If $X \sim$Binomial(n, p), then $E[X] = np$ and $Var[X] = np(1-p)$. Thus,

$$E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{E[X]}{n} = \frac{np}{n} = p$$

and

$$Var[\hat{p}] = Var\left[\frac{X}{n}\right] = \frac{Var[X]}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

Maximum Likelihood Estimator

Distribution of the MLE

There are some nice results in statistics that state as $n \to \infty$ this standardization of the MLE $\sqrt{n}[\hat{p} - p] \to N(0, p(1-p)$.

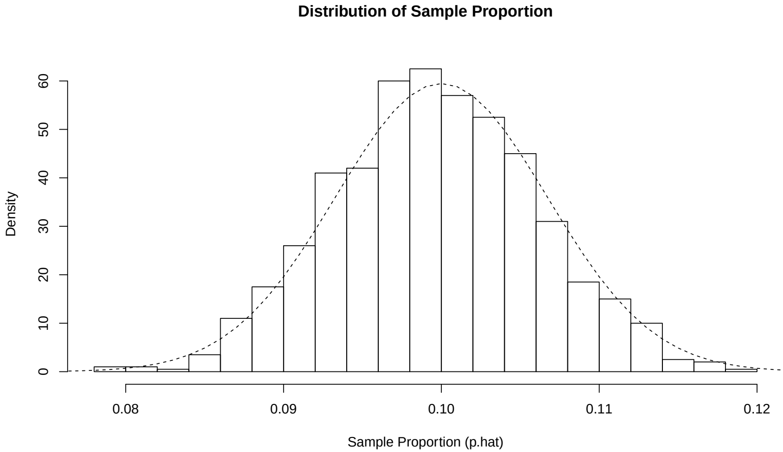
In practice we are always dealing with finite samples, so we use this asymptotic result to say that $\hat{p}$ is approximately distributed as $N\left(p, \frac{p(1-p)}{n}\right)$.

How well does this approximation hold? The approximation is better for larger n.

The next three slides will show examples of taking 1,000 random samples from Binomial distributions and calculating the MLE for each sample and comparing it to the Normal distribution based on the parameters.

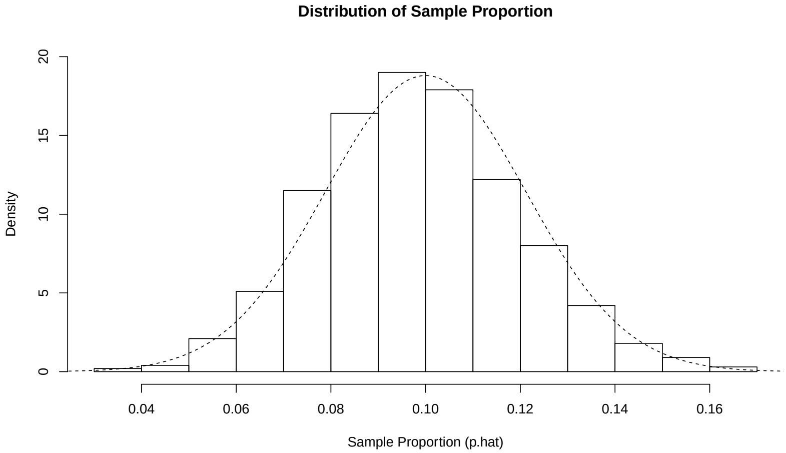
Asymptotic Normality of MLE

p=0.1, n=2,000, $\hat{p} \sim N(p, p(1-p)/n)$ is a good approximation



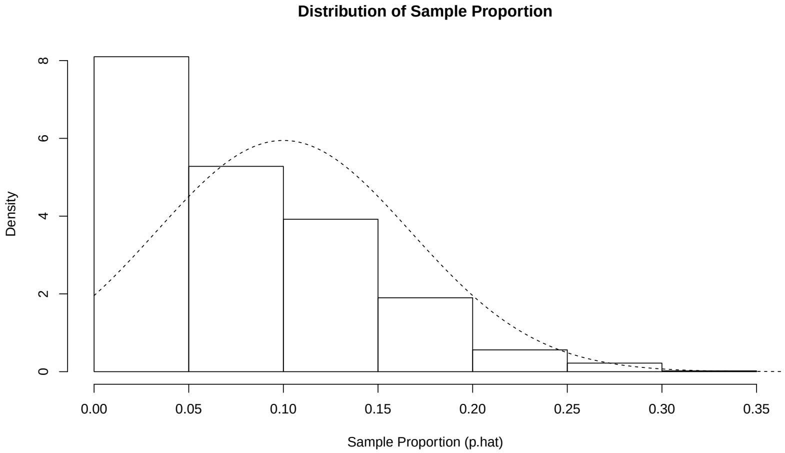
Asymptotic Normality of MLE

p=0.1, n=200, $\hat{p} \sim N(p, p(1-p)/n)$ is still a pretty good approximation



Asymptotic Normality of MLE

p=0.1, n=20, $\hat{p} \sim N(p, p(1-p)/n)$ is NOT a good approximation



Maximum Likelihood Estimator

Uncertainty

Why do we care about the distribution of the MLE? This is how we will generate uncertainty estimates (confidence intervals).

Based on our approximate distribution $Z = \frac{p - \hat{p}}{\sqrt{\hat{p}(1-\hat{p})/n}}$ is approximately $N(0,1)$.

The standardized version of our MLE provides a framework for generating an interval that is expected to include the true parameter value a certain percentage of the time. Generally this is set at 95%.

NOTE: Once you have collected data and created a confidence interval your interval either will or will not include the true value. It is not appropriate to say that you are 95% sure that your interval includes the true value.

Maximum Likelihood Estimator

95% Confidence Intervals

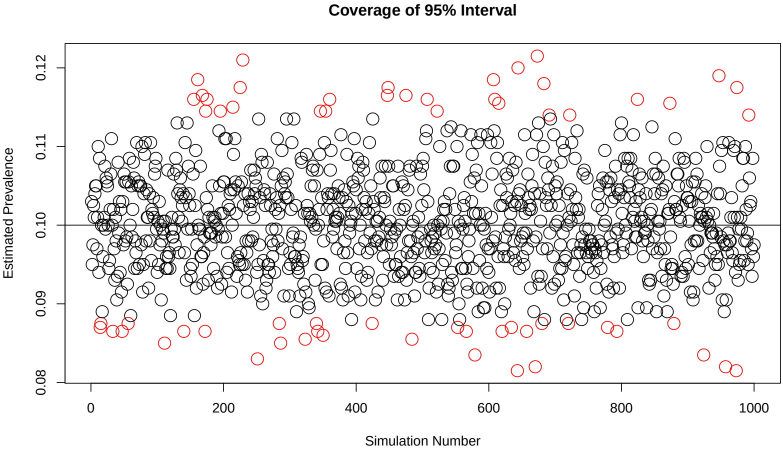
To generate a 95% confidence interval

$$\begin{aligned}
0.95 &= P\left(Z_{0.025} < Z < Z_{0.975}\right) = P\left(-1.96 < Z < 1.96\right) \\
&= P\left(-1.96 < \frac{p - \hat{p}}{\sqrt{\hat{p}(1-\hat{p})/n}} < 1.96\right) \\
&= P\left(-1.96\sqrt{\hat{p}(1-\hat{p})/n} < p - \hat{p} < 1.96\sqrt{\hat{p}(1-\hat{p})/n}\right) \\
&= P\left(\hat{p} - 1.96\sqrt{\hat{p}(1-\hat{p})/n} < p < \hat{p} + 1.96\sqrt{\hat{p}(1-\hat{p})/n}\right)
\end{aligned}$$

Thus, the interval $[\hat{p} - 1.96\sqrt{\hat{p}(1-\hat{p})/n}, \hat{p} + 1.96\sqrt{\hat{p}(1-\hat{p})/n}]$ will give us an interval that should include the true parameter, p, 95% of the time.

95% Confidence Intervals

1,000 random samples from a Binomial(n=2,000,p=0.1). Red points had confidence intervals that did not include the true value (63/1000).



Maximum Likelihood Estimator

95% Confidence Intervals

Back to our example. We sampled 2,000 individuals in South Africa and found $X = 216$ to be HIV+.

What is our MLE? $\hat{p} = X/n = 216/2000 = 0.108$

What is our 95% Confidence Interval?

$$[\hat{p} - 1.96\sqrt{\hat{p}(1-\hat{p})/n}, \hat{p} + 1.96\sqrt{\hat{p}(1-\hat{p})/n}] = [0.094, 0.122]$$

If the true prevalence was 0.1, what is the probability that it is contained in this confidence interval? 1

.

If the true prevalence was 0.13, what is the probability that it is contained in this confidence interval? 0

.

If we always calculate confidence intervals in this fashion, what percentage of our intervals should include the true parameter value? 95%

The End

Questions?