STAT 135 Lab 3
Asymptotic MLE and the Method of Moments

Rebecca Barter

February 9, 2015

Maximum likelihood estimation (a reminder)

Maximum likelihood estimation

Suppose that we have a sample, $X_1, X_2, ..., X_n$, where the X_i are IID. Then the

- **Maximum likelihood estimator for** θ: calculate a **single value** which estimates the true value of θ_0 by maximizing the likelihood function with respect to θ
  - i.e. find the value of θ that maximizes the likelihood of observing the data given.

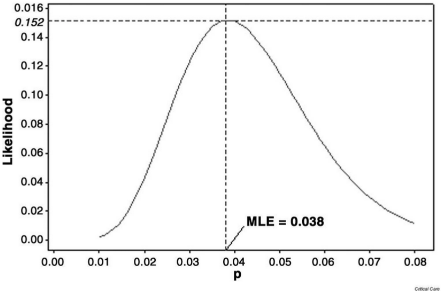
How do we write down the likelihood function? The (non-rigorous) idea:

$$
\begin{aligned}
lik(\theta) &= P(X_1 = x_1, ..., X_n = x_n) \\
&= P(X_1 = x_1)...P(X_n = x_n) \\
&= \prod_{i=1}^{n} f_\theta(X_i)
\end{aligned}
$$

Maximum likelihood estimation

What is the likelihood function?

- The likelihood function, $lik(\theta)$, is a function of θ which corresponds to the probability of observing our sample for various values of θ.



How to find the value of θ that maximizes the likelihood function?

Maximum likelihood estimation: Asymptotic results

Asymptotic results: what happens when our sample size, n, gets really large ($n \to \infty$)

MLE: Asymptotic results

It turns out that the MLE has some very nice asymptotic results

1. **Consistency**: as $n \to \infty$, our ML estimate, $\hat{\theta}_{ML,n}$, gets closer and closer to the true value θ_0.
2. **Normality**: as $n \to \infty$, the distribution of our ML estimate, $\hat{\theta}_{ML,n}$, tends to the normal distribution (with what mean and variance?).

MLE: Asymptotic results

It turns out that the MLE has some very nice asymptotic results

1. **Consistency**: as $n \to \infty$, our ML estimate, $\hat{\theta}_{ML,n}$, gets closer and closer to the true value θ_0.
2. **Normality**: as $n \to \infty$, the distribution of our ML estimate, $\hat{\theta}_{ML,n}$, tends to the normal distribution (with what mean and variance?).

MLE: Asymptotic results
1. Consistency

An estimate, $\hat{\theta}_n$, of θ_0 is called **consistent** if:

$$\hat{\theta}_n \xrightarrow{p} \theta_0 \qquad \text{as} \qquad n \to \infty$$

---

where $\hat{\theta}_n \xrightarrow{p} \theta_0$ *technically means that, for all* $\epsilon > 0$,

$$P(|\hat{\theta}_n - \theta_0| > \epsilon) \to 0 \text{ as } n \to \infty$$

But you don't need to worry about that right now... just think of it as as n gets really large, the probability that $\hat{\theta}_n$ *differs from* θ_0 *becomes increasingly small.*

MLE: Asymptotic results
1. Consistency

The MLE, $\hat{\theta}_{ML,n}$ is a **consistent estimator** for the parameter, θ, that it is estimating, so that

$$\hat{\theta}_{ML,n} \xrightarrow{p} \theta_0 \qquad \text{as} \qquad n \to \infty$$

This nice property also implies that the MLE is **asymptotically unbiased**:

$$E(\hat{\theta}_{ML,n}) \to \theta_0 \quad \text{as } n \to \infty$$

MLE: Asymptotic results

It turns out that the MLE has some very nice asymptotic results

1. **Consistency**: as $n \to \infty$, our ML estimate, $\hat{\theta}_{ML,n}$, gets closer and closer to the true value θ_0.
2. **Normality**: as $n \to \infty$, the distribution of our ML estimate, $\hat{\theta}_{ML,n}$, tends to the normal distribution (with what mean and variance?).

MLE: Asymptotic results
2. Normality

An estimate, $\hat{\theta}_n$, of θ is called **asymptotically normal** if, as $n \to \infty$, we have that

$$\hat{\theta}_n \sim N(\mu_{\theta_0}, \sigma^2_{\theta_0})$$

where θ_0 is the true value of the parameter θ.

What else have we seen with this property?

MLE: Asymptotic results
2. Normality

It turns out that our ML estimate, $\hat{\theta}_{ML,n}$, of θ is **asymptotically normal**: as $n \to \infty$, we have that

$$\hat{\theta}_{ML,n} \sim N(\mu_{\theta_0}, \sigma^2_{\theta_0})$$

- **We want to find out, what are μ_{θ_0} and $\sigma^2_{\theta_0}$?**

MLE: Asymptotic results
2. Normality

First, here is a fun definition of **Fisher Information**

$$I(\theta_0) = E\left[\left(\frac{\partial}{\partial\theta}\log(f_\theta(x))\Big|_{\theta_0}\right)^2\right]$$

or alternatively,

$$I(\theta_0) = -E\left[\frac{\partial^2}{\partial^2\theta}\log(f_\theta(x))\Big|_{\theta_0}\right]$$

(we will soon find that the asymptotic variance is related to this quantity)

MLE: Asymptotic results
2. Normality

Fisher Information:

$$I(\theta_0) = -E\left[\frac{\partial^2}{\partial^2\theta}\log(f_\theta(x))\Big|_{\theta_0}\right]$$

Wikipedia says that "*Fisher information is a way of measuring the amount of information that an observable random variable X carries about an unknown parameter θ upon which the probability of X depends*"

MLE: Asymptotic results
2. Normality (example)

Recall last week we showed that if we have a sample $X_1, X_2, ..., X_n$ where $X_i \sim Bernoulli(p_0)$ for each $i = 1, 2, ..., n$, then

$$\hat{p}_{MLE} = \overline{X}_n$$

What is the fisher information for X_i?

$$I(p_0) = -E\left[\frac{\partial^2}{\partial^2 p}\log(f_p(x))\Big|_{p_0}\right]$$

MLE: Asymptotic results
2. Normality (example)

$X_i \sim \mathit{Bernoulli}(p_0)$ for each $i = 1, 2, ..., n$. What is the fisher information for X?

$$I(p_0) = -E\left[\frac{\partial^2}{\partial^2 p}\log(f_p(X))\Big|_{p_0}\right]$$

$$f_p(X) = p^X(1-p)^{1-X}$$

$$I(p_0) = -E\left[\frac{\partial^2}{\partial^2 p}\log(f_p(X))\Big|_{p_0}\right] = \frac{1}{p_0(1-p_0)}$$

MLE: Asymptotic results
2. Normality

It turns out that our ML estimate, $\hat{\theta}_{ML,n}$, of θ is **asymptotically normal**: as $n \to \infty$, we have that

$$\hat{\theta}_{ML,n} \sim N(\mu_{\theta_0}, \sigma^2_{\theta_0})$$

- **We want to find out, what are μ_{θ_0} and $\sigma^2_{\theta_0}$?**

Any ideas as to what μ_{θ_0} might be? (Hint: what is the asymptotic expected value of $\hat{\theta}_{ML,n}$?)

MLE: Asymptotic results
2. Normality

$\hat{\theta}_{ML,n}$, of θ is **asymptotically normal**: as $n \to \infty$, we have that

$$\hat{\theta}_{ML,n} \sim N(\mu_{\theta_0}, \sigma^2_{\theta_0})$$

The *consistency* of $\hat{\theta}_{ML,n}$ tells us that $\hat{\theta}_{ML,n} \xrightarrow{p} \theta_0$, so as $n \to \infty$,

$$E(\hat{\theta}_{ML,n}) \to E(\theta_0) = \theta_0$$

Thus the **asymptotic mean** of the MLE is given by

$$\boxed{\mu_{\theta_0} = \theta_0}$$

MLE: Asymptotic results
2. Normality

$\hat{\theta}_{ML,n}$, of θ is **asymptotically normal**: as $n \to \infty$, we have that

$$\hat{\theta}_{ML,n} \sim N(\mu_{\theta_0}, \sigma^2_{\theta_0})$$

The **asymptotic variance** of the MLE is given by

$$\boxed{\sigma^2_{\theta_0} = \frac{1}{nI(\theta_0)}}$$

MLE: Asymptotic results
2. Normality

So in summary, we have: $\hat{\theta}_{ML,n}$, of θ is **asymptotically normal**: as $n \to \infty$, we have that

$$\boxed{\hat{\theta}_{ML,n} \sim N\left(\theta_0, \frac{1}{nI(\theta_0)}\right)}$$

MLE: Asymptotic results (example)

For large samples, the ML estimate of θ is approximately normally distributed:

$$\boxed{\hat{\theta}_{ML,n} \sim N\left(\theta_0, \frac{1}{nI(\theta_0)}\right)}$$

For our $X_i \sim Bernoulli(p_0)$, $i = 1, ..., n$ example. Recall:

$$\hat{p}_{ML} = \overline{X}_n$$

$$I(p_0) = \frac{1}{p_0(1-p_0)}$$

Thus, when $X_i \sim Bernoulli(p_0)$, for large n

$$\boxed{\hat{p}_{ML} = \overline{X}_n \sim N\left(p_0, \frac{p_0(1-p_0)}{n}\right)}$$

Why do we believe this result? How else could we have obtained it?

Exercise

MLE: Asymptotic results (exercise)

In class, you showed that if we have a sample $X_i \sim Poisson(\lambda_0)$, the MLE of λ is

$$\hat{\lambda}_{ML} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

1. What is the asymptotic distribution of $\hat{\lambda}_{ML}$ (You will need to calculate the asymptotic mean and variance of $\hat{\lambda}_{ML}$)?
2. Generate $N = 10000$ samples, $X_1, X_2, ..., X_{1000}$ of size $n = 1000$ from the $Poisson(3)$ distribution.
3. For each sample, calculate the ML estimate of λ. Plot a histogram of the ML estimates
4. Calculate the variance of your ML estimate, and show that this is close to the asymptotic value derived in part 1

Method of Moments (MOM)

(An alternative to MLE)

Method of Moments

- **Maximum likelihood estimator for θ**: calculate a **single value** which estimates the true value of θ by maximizing the likelihood function with respect to θ.
- **Method of moments estimator for θ**: By equating the theoretical moments to the empirical (sample) moments, derive equations that relate the theoretical moments to θ. The equations are then solved for θ.

Suppose X follows some distribution. The kth **moment of the distribution** is defined to be

$$\mu_k = E[X^k] = g_k(\theta)$$

which will be some function of θ.

Method of Moments

MOM works by equating the theoretical moments (which will be a function of θ) to the empirical moments.

Moment	Theoretical Moment	Empirical Moment
first moment	$E[X]$	$\frac{\sum_{i=1}^{n} X_i}{n}$
second moment	$E[X^2]$	$\frac{\sum_{i=1}^{n} X_i^2}{n}$
third moment	$E[X^3]$	$\frac{\sum_{i=1}^{n} X_i^3}{n}$

Method of Moments

MOM is perhaps best described by example.
Suppose that $X \sim Bernoulli(p)$. Then the first moment is given by

$$E[X] = 0 \times P(X = 0) + 1 \times P(X = 1) = p$$

Moreover, we can estimate the E[X] by taking a sample $X_1, ..., X_n$ and calculating the sample mean :

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

We approximate the first theoretical moment, $E[X]$, by the first empirical moment, $\overline{X}$, i.e.

$$\hat{p}_{MOM} = \overline{X}$$

which is the same as the MLE estimator! (note that this is not always the case...)

Exercise

Exercise – Question 43, Chapter 8 (page 320) from Rice

The file `gamma-arrivals` contains a set of gamma-ray data consisting of the times between arrivals (interarrival times) of 3,935 photons (units are seconds)

The gamma distribution can be written as

$$f_{\alpha,\beta}(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

1. Make a histogram of the interarrival times. Does it appear that a gamma distribution would be a plausible model?
2. Fit the parameters by the method of moments and by maximum likelihood. How do the estimates compare? *(Hint: the MLE for α has no closed-form solution use $\hat{\alpha}_{MLE} = 1$)*
3. Plot the two fitted gamma densities on top of the histogram. Do the fits look reasonable?