1 Maximum likelihood estimators

maxlik.tex and maxlik.pdf, March 11, 2003

Simply put, if we know the form of $f_X(x;\theta)$ and have a sample from $f_X(x;\theta)$, not necessarily random, the *ml* estimator of θ, θ_{ml}, is that θ which maximizes

$$f_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n;\theta)$$

Remember that $f_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n;\theta)$ is the joint density function of the sample, written as a function of θ.

In this context, we call the joint density function of the sample, the *likelihood function.*[1] That is

$$L(x_1,x_2,...,x_n;\theta) = f_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n;\theta)$$

θ_{ml} is called the maximum likelihood estimator of θ because it is that estimate of θ that maximizes the likelihood of drawing the given sample, $x_1,x_2,...,x_n$.

Numerous students have used used Maximum likelihood estimation for their projects.

We find θ_{ml} by maximizing $L(x_1,x_2,...,x_n;\theta)$ with respect to θ.

Maximum likelihood estimation is probably the most versatile tool in the econometrician's tool box.

Note that one needs to assume a form for $f_X(x;\theta)$ to get the *ml* estimator.

There are numerous ways to find maximum likelihood estimates.

- One can do it the old-fashioned way: take partial derivatives of L with respect to each element in θ, set all of them equal to zero, solve the system for the θ, and then check second-order conditions for a maximum.
- Let Mathematica or other such program find those values of θ that maximum the likelihood function. These techniques use search algorithms. In *Mathematica* use the command *Min*.Turn it into a maximization command by having it minimize $-\ln L$.

[1]The sample is considered given, and the liklihood function identifies the likelihood of drawing that sample as a function of the parameter values.

1.1 Look what happens when the sample is a random sample

If the sample is a random sample from $f_X(x;\theta)$ then

$$\begin{aligned} L(x_1, x_2, ..., x_n; \theta) &= f_{X_1, X_2, ..., X_n}(x_1, x_2, ..., x_n; \theta) \\ &= f_X(x_1;\theta) f_X(x_2;\theta) ... f_X(x_n;\theta) = \prod_{i=1}^{n} f_X(x_i;\theta) \end{aligned}$$

because each observation is an independent draw from $f_X(x;\theta)$.

That is, θ_{ml} is that θ which maximizes $\prod_{i=1}^{n} f_X(x_i;\theta)$.

Further note that the θ which maximizes $\prod_{i=1}^{n} f_X(x_i;\theta)$ is also the θ that maximizes $\ln[\prod_{i=1}^{n} f_X(x_i;\theta)]$; that is, the θ that maximizes $\ln L$ is θ_{ml}.

And

$$\begin{aligned} \ln L(x_1, x_2, ..., x_n; \theta) &= \ln[\prod_{i=1}^{n} f_X(x_i;\theta)] \\ &= \sum_{i=1}^{n} \ln\left[f_X(x_i;\theta)\right] \end{aligned}$$

1.2 Some examples of maximum likelihood estimates

1.2.1 Assume the rv X has a Bernoulli distribution

That is, assume

$$f(x;p) = \begin{cases} p^x(1-p)^{1-x} & \text{if} \quad x = 0 \\ p^x(1-p)^{1-x} & \text{if} \quad x = 1 \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p \equiv \theta < 1$.

We want p_{ml}. Assume we have a random sample of 10 observations (the sample consists of zeros and ones. In this case

$$\begin{aligned} L(x_1, x_2, ..., x_n; \theta) &= \prod_{i=1}^{10} f_X(x_i;\theta) \\ &= \prod_{i=1}^{10} p^{x_i}(1-p)^{1-x_i} \end{aligned}$$

So, p_{ml} is that p that maximizes p_{ml}. p_{ml} is also that p that maximizes

$$\begin{aligned}
\ln L(x_1, x_2, ..., x_n; \theta) &= \sum_{i=1}^{10} \ln\left[f_X(x_i;\theta)\right]] \\
&= \sum_{i=1}^{10} \ln\left[p^{x_i}(1-p)^{1-x_i}\right] \\
&= \sum_{i=1}^{10} \{x_i \ln p + (1-x_i)\ln(1-p)\} \\
&= \sum_{i=1}^{10} x_i \ln p + \sum_{i=1}^{10}(1-x_i)\ln(1-p) \\
&= \ln p \sum_{i=1}^{10} x_i + \ln(1-p)\sum_{i=1}^{10}(1-x_i)
\end{aligned}$$

Note that

$$\sum_{i=1}^{10} x_i = 10\bar{x}$$

and

$$\sum_{i=1}^{10}(1-x_i) = 10 - \sum_{i=1}^{10} x_i = 10 - 10\bar{x} = 10(1-\bar{x})$$

so

$$\begin{aligned}
&\ln L(x_1, x_2, ..., x_n; \theta) \\
&= \ln p \sum_{i=1}^{10} x_i + \ln(1-p)\sum_{i=1}^{10}(1-x_i) \\
&= 10\bar{x}\ln p + 10(1-\bar{x})\ln(1-p)
\end{aligned}$$

To find p_{ml} we want to maximize $10\bar{x}\ln p + 10(1-\bar{x})\ln(1-p)$ with respect to p.[2]

Make up a sample with 10 observations and use the *Min* command in *Mathematica* to find p_{ml}.

Then do it the old fashion way in terms of any random sample with 10 observations. That is, use calculus to maximize $10\bar{x}\ln p + 10(1-\bar{x})\ln(1-p)$ with respect to p.

$$\begin{aligned}
\frac{d\ln L}{dp} &= 10\bar{x}(\frac{1}{p}) + 10(1-\bar{x})(\frac{1}{1-p})(-1) \\
&= 10\bar{x}(\frac{1}{p}) - 10(1-\bar{x})(\frac{1}{1-p})
\end{aligned}$$

[2] Note that the information in the data required to find the *ml* estimate iscompletely contained by the sample average, $\bar{x}$. $\bar{x}$ is deemed a *sufficient* statistic because it contains sufficient information to estimate the parameter

Set this equal to zero to find the critical point

$$10\bar{x}(\frac{1}{p}) - 10(1-\bar{x})(\frac{1}{1-p}) = 0$$

, Solution is: $\{p = \bar{x}\}$. That is, $\bar{x}$ (the sample mean) is the maximum likelihood estimate of p.

The Bernoulli and Binomial - a simpler way to solve the above problem Our sample of n consists of n repeated Bernoulli trials (draws). It is well known that the number of sucesses (ones) in those n trials has a Binomial distribution. That is, if one has $\sum_{i=1}^{n} x_i$ sucess in n trials

$$f(\sum_{i=1}^{n} x_i) = \left(\begin{array}{c} n \\ \sum_{i=1}^{n} x_i \end{array} \right) p^{(\sum x_i)}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

where $\left(\begin{array}{c} n \\ \sum_{i=1}^{n} x_i \end{array} \right)$ is the binomial coefficient. If $n = 10$

$$f(\sum_{i=1}^{10} x_i) = \left(\begin{array}{c} 10 \\ \sum_{i=1}^{10} x_i \end{array} \right) p^{(\sum x_i)}(1-p)^{n-\sum x_i}$$

Therefore, another way to write the likelihood function (and log likelihood function) for our problem is

$$L(x_1, x_2, ..., x_n; \theta) = \left(\begin{array}{c} 10 \\ \sum_{i=1}^{10} x_i \end{array} \right) p^{(\sum x_i)}(1-p)^{n-\sum x_i}$$

and

$$\ln L(x_1, x_2, ..., x_n; \theta) = \ln \left(\begin{array}{c} 10 \\ \sum_{i=1}^{10} x_i \end{array} \right) + (\sum x_i) \ln p + (^{n-\sum x_i}) \ln(1-p)$$

So

$$\begin{aligned} \frac{d \ln L}{dp} &= (\sum x_i)\frac{1}{p} - (^{n-\sum x_i})(\frac{1}{1-p}) \\ &= 10(\bar{x})\frac{1}{p} - 10(1-\bar{x})(\frac{1}{1-p}) \end{aligned}$$

Set this equal to zero and solve for p to determine that $p_{ml} = \bar{x}$, just what we got when we did it the other way.

1.2.2 Assume the rv X has a Poisson distribution

X has a Poisson distribution if

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x = 0, 1, 2, 3, ...$$

where $\lambda > 0$. The Poisson is a discrete distribution that can take only integer values. It is often the distribution of choice is one wants to *count* something; e.g. the number of times American's get married, or, to make a bad pun, the number of fish caught in a day of fishing. For the Poisson[3]

$$E[x] = \lambda = var[x]$$

Assume that the number of marriages by individuals has a Poisson distribution. This seems reasonable since the number of times one has been married is, hopefully, a nonnegative integer.

Further assume a random sample of 5 observations (0,0,2,2,7).
Write down the likelihood function and the log likelihood function

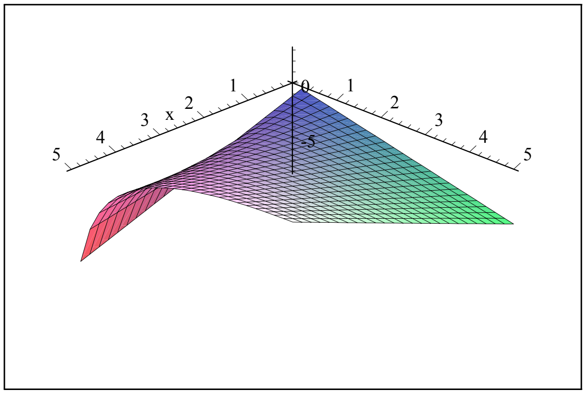
$$L(x_1, x_2, ..., x_5; \lambda) = \prod_{i=1}^{5} f_X(x_i; \lambda) = \prod_{i=1}^{5} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

and

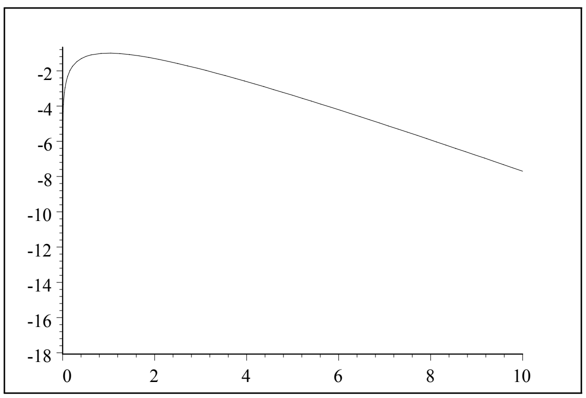
$$\begin{aligned}
\ln L(x_1, x_2, ..., x_5; \lambda) &= \sum_{i=1}^{5} \ln f_X(x_i; \lambda) \\
&= \sum_{i=1}^{5} \ln\left[\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}\right] \\
&= \sum_{i=1}^{5} \ln(e^{-\lambda}\lambda^{x_i}) - \ln(x_i!) \\
&= \sum_{i=1}^{5} \ln(e^{-\lambda}) + \ln(\lambda^{x_i}) - \ln(x_i!) \\
& \sum_{i=1}^{5} -\lambda + x_i \ln(\lambda) - \ln(x_i!) \\
&= -5\lambda + \ln(\lambda)\sum_{i=1}^{5} x_i - \sum_{i=1}^{5} \ln(x_i!) \\
&= -5\lambda + \ln(\lambda)5\bar{x} - \ln(x_i!)
\end{aligned}$$

[3]It is obviously restrictive to assume that the mean and variance are equal. This restriction can be relaxed by, for example, assuming a negative binomial distribution. See, for example, Green page xxx

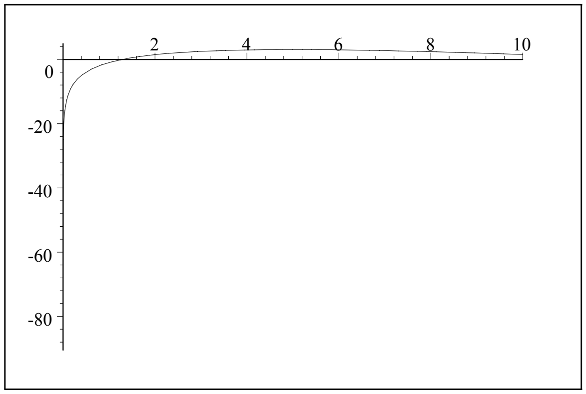
Since the term containing $x_i!$ does not contain λ, $\bar{x}$ is a sufficient statistic. The following is a graph of the liklihood function with $\bar{x}$ and λ on the horizontal plane and $\ln L$ on the vertical axis. One could use this graph to find λ_{ml} for any given value of $\bar{x}$. Note how $\ln L$ is always negative.



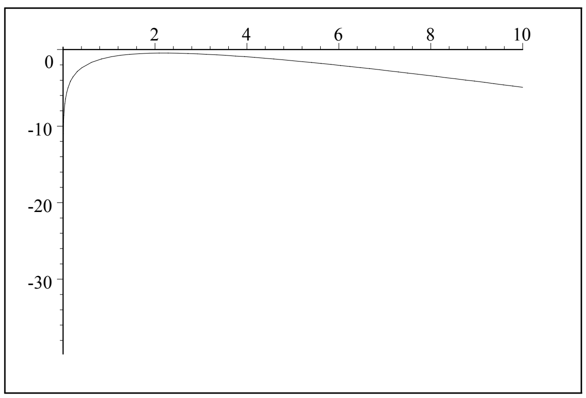
Now graph some slices of the above. Assuming $\bar{x} = 1$



Assuming $\bar{x} = 5$



Assuming $\bar{x} = 2.2$



Note that $\ln(x_i!)$ is not a function of λ, so maximizing $(-\lambda + \ln(\lambda)\bar{x})$ is equivalent to maximizing $\ln L(x_1, x_2, ..., x_5; \lambda)$.

$$\frac{d \ln L}{d\lambda} = -5 + 5\bar{x}\frac{1}{\lambda}$$

Set this equal to zero and solve for λ

$$-5 + 5\bar{x}\frac{1}{\lambda} = 0$$

, Solution is: $\{\lambda_{ml} = \bar{x}\}$

In the numerical example, $\bar{x} = 2.2$ (the average of $0, 0, 2, 2, 7$). Wow.

Then I let Maple find the answer.

$-\lambda + \ln(\lambda)(2.2)$ Candidate(s) for extrema: $\{-0.465\,39\}$, at $\{\{\lambda = 2.\,2\}\}$

Let's get the probability associated with the first eight integer values

$$\text{PoissonDen}\,(0; 2.2)$$

0.110 8; that is, there is a 11% chance one will not get married

$$\text{PoissonDen}\,(1; 2.2)$$

0.243 77; that is, there is a 24% change one will marry once

$$\text{PoissonDen}\,(2; 2.2)$$

: 0.268 14; that is, there is a 26% chance one will marry twice

$$\text{PoissonDen}\,(3; 2.2)$$

: 0.196 64; a 19% chance one will marry thrice

$$\text{PoissonDen}\,(4; 2.2)$$

: 0.108 15; that is, there is a 10% change one will marry four times

$$\text{PoissonDen}\,(5; 2.2)$$

: 4.7587×10^{-2};; that is, there is a 4% chance one will marry five times

$$\text{PoissonDen}\,(6; 2.2)$$

: 1.7448×10^{-2}; a 1% chance one will marry six times

$$\text{PoissonDen}\,(7; 2.2)$$

: 5.4838×10^{-3}; a .5% chance that one will marry seven times.

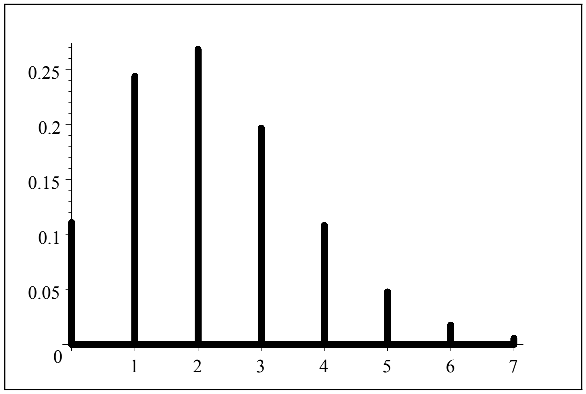
The probability that one will marry twelve times is

$$\text{PoissonDen}\,(12; 2.2)$$

: 2.9736×10^{-6}, which is not much.

Graphing this Poisson for 0 through 7 $\text{PoissonDen}\,(k; 2.2) = \frac{2.2^k e^{-2.2}}{k!}$
$(0, 0, 0, 0.1108, 0, 0, 1, 0, 1, 0.24377, 1, 0, 2, 0, 2, 0.26814, 2, 0, 3, 0, 3, 0.19664, 3, 0, 4, 0, 4, 0.10815, 4, 0, 5, 0, 5,$

$4.7587 \times 10^{-2}, 5, 0, 6, 0, 6, 1.7448 \times 10^{-2}, 6, 0, 7, 0, 7, 5.4838 \times 10^{-3}, 7, 0)$



Poisson distribution with $\lambda = 2.2$

That is, $E[x] = 2.2$. What is the maximum likelihood estimate of the variance? 2.2

Now let's make the problem a little more interesting. Assume

$married = x_i$	age_i
0	12
0	50
2	30
2	36
7	97

That is, we know each individuals age, and suspect that there might be a relationship between how many times one has been married and one's age.

How would you change the above Poisson model to take this into account?

One could assume that λ is a function of *age*; e.g.[4]

$$\lambda = \lambda_0 age_i$$

[4] It is also common to assume a nonlinear function such as $\lambda = \exp(\lambda_0 age)$. See, for example, Green page xxx.

In which case,

$$\begin{aligned}
\ln L(x_1, x_2, ..., x_5; \lambda_0, \lambda_1) &= \sum_{i=1}^{5} \ln f_X(x_i; \lambda_0 age_i) \\
&= \sum_{i=1}^{5} \ln \left[\frac{e^{-(\lambda_0 age)} (\lambda_0 age_i)^{x_i}}{x_i!} \right] \\
&= \sum_{i=1}^{5} \ln(e^{-(\lambda_0 age_i)} (\lambda_0 age_i)^{x_i}) - \ln(x_i!) \\
&= \sum_{i=1}^{5} \ln(e^{-(\lambda_{01} age)}) + \ln((\lambda_0 age_i)^{x_i}) - \ln(x_i!) \\
&= \sum_{i=1}^{5} -(\lambda_0 age_i) + x_i \ln(\lambda_0 age_i) - \ln(x_i!) \\
&= \left[-\lambda_1 \sum_{i=1}^{5} age_i + \sum_{i=1}^{5} x_i \ln(\lambda_0 age_i) - \sum_{i=1}^{5} \ln(x_i!) \right]
\end{aligned}$$

Now let's take the partial with respect to λ_o

$$\begin{aligned}
\frac{\partial \ln L(x_1, x_2, ..., x_5; \lambda_0)}{\partial \lambda_0} &= \frac{\partial \left[-\lambda_0 \sum_{i=1}^{5} age_i + \sum_{i=1}^{5} x_i \ln(\lambda_0 age_i) - \sum_{i=1}^{5} \ln(x_i!) \right]}{\partial \lambda_1} \\
&= -\sum_{i=1}^{5} age_i + \sum_{i=1}^{5} x_i \frac{\partial \ln(\lambda_0 age_i)}{\partial \lambda_0} \\
&= -\sum_{i=1}^{5} age_i + \sum_{i=1}^{5} x_i \frac{age_i}{(\lambda_0 age_i)} \\
&= -\sum_{i=1}^{5} age_i + \sum_{i=1}^{5} x_i \frac{1}{\lambda_0} \\
&= -\sum_{i=1}^{5} age_i + \frac{1}{\lambda_0} \sum_{i=1}^{5} x_i \\
&= -5ag\bar{e} + \frac{5\bar{x}}{\lambda_0}
\end{aligned}$$

Set this equal to zero and solve for λ_0.

$$-5ag\bar{e} + \frac{5\bar{x}}{\lambda_0} = 0$$

, Solution is $\frac{\bar{x}}{ag\bar{e}}$, average number of marriages divided by average age, and

$$\lambda_{ml} = \frac{age_i \bar{x}}{ag\bar{e}}$$

This is interesting, our expectation of one's number of marriages is the sample average, weighted by $\frac{age_i}{ag\bar{e}}$ (the individual's age a a proportion of the average age in the sample.

The ml estimate of λ_0 for the sample at hand is something like .0488

Now make the problem more interesting by assuming.

$$\lambda = \lambda_0 + \lambda_1 age_i$$

That is, estimate a slope and an intercept.

1.2.3 Assume some random variable X has a normal distribution, that is

$$f_X(x; \mu_x, \sigma_x^2) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(\frac{1}{2\sigma_x^2})(x-\mu_x)^2}$$

We draw a random sample from of n observations from this distribution.

We want to find the ml estimates of μ_x and σ_x^2.

This is the most famous ml problem

In this case,

$$L(x_1, x_2, ..., x_n; \mu_x, \sigma_x^2) = \prod_{i=1}^{n} f_X(x; \mu_x, \sigma_x^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(\frac{1}{2\sigma_x^2})(x_i-\mu_x)^2}$$

and the ln of the likelihood function is

$$\begin{aligned}
\ln L(x_1, x_2, ..., x_n; \mu_x, \sigma_x^2) &= \ln \prod_{i=1}^{n} f_X(x; \mu_x, \sigma_x^2) \\
&= \ln \prod_{i=1}^{n} \ln \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(\frac{1}{2\sigma_x^2})(x_i-\mu_x)^2} \\
&= \ln \prod_{i=1}^{n} (2\pi)^{1/2} \sigma_x^{-1} e^{-(\frac{1}{2\sigma_x^2})(x_i-\mu_x)^2} \\
&= \sum_{i=1}^{n} \left\{ -\frac{1}{2} \ln(2\pi) - \ln \sigma_x - (\frac{1}{2\sigma_x^2})(x_i - \mu_x)^2 \right\} \\
&= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma_x^2 - (\frac{1}{2\sigma_x^2}) \sum_{i=1}^{n} (x_i - \mu_x)^2
\end{aligned}$$

We want to maximize this with respect to μ_x and σ_x^2. Take the partials

$$\frac{\partial[-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma_x^2-(\frac{1}{2\sigma_x^2})\sum_{i=1}^n(x_i-\mu_x)^2]}{\partial\mu_x}=\left[-\frac{n\mu_x-\sum_{i=1}^n x_i}{\sigma_x^2}\right]$$

and

$$\frac{\partial[-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma_x^2-(\frac{1}{2\sigma_x^2})\sum_{i=1}^n(x_i-\mu_x)^2]}{\partial\sigma_x^2}=\frac{\sum_{i=1}^n(x_i-\mu_x)^2-n\sigma_x^2}{2\sigma_x^4}$$

Set these both equal to zero and solve for μ_x and σ_x^2. Start with the first equation

$$\left[-\frac{n\mu_x-\sum_{i=1}^n x_i}{\sigma_x^2}\right]=0$$

Note that the μ_x that solves this is $\mu_x=\frac{1}{n}\sum_{i=1}^n x_i$. That is, the maximum likelihood estimate of μ_x is $\frac{1}{n}\sum_{i=1}^n x_i=\bar{x}$

Plug this into the second partial, set equal to zero, and solve for the maximum likelihood estimate of σ_x^2

$$\frac{\sum_{i=1}^n(x_i-\bar{x})^2-n\sigma_x^2}{2\sigma_x^4}=0$$

That is, solve $\sum_{i=1}^n(x_i-\bar{x})^2-n\sigma_x^2$ for σ_x^2, which is

$$\hat{\sigma}_x^2=\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})^2$$

So the maximum likelihood estimate of μ_x is $\bar{x}$ and the maximum likelihood estimate of σ_x^2 is $\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})^2$. Note that the first is unbiased, the second is not - both are asymtoticallly unbiased.

Make up some data - maybe $4, 7, 1$ and find the max likelihood estimates.

The log likelihood function with unnecesary terms removed is.
$-\frac{3}{2}\ln\sigma_x^2-(\frac{1}{2\sigma_x^2})[(4-\mu_x)^2+(7-\mu_x)^2+(1-\mu_x)^2]$

Note that we maximize LnL by taking its derivative with respect to the parameter and searching for a local interior maximum. We could also have used a computer search algorithm such as MIN in Mathematica to find $\widehat{\mu}_x$ and $\widehat{\sigma}_x^2$. That is,

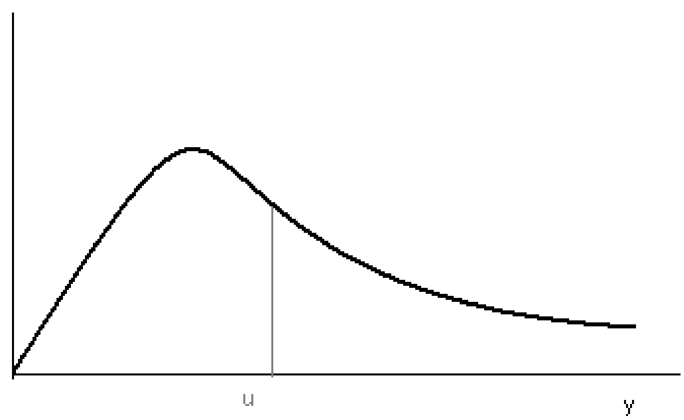
$$Minimize\ \ -Ln\ L$$

1.3 A more general max lik problem

Consider the following problem. Assume that the i^{th} random variable Y_i is distributed[5]

$$f_{Y_i}\left(y_i, \mu_{y_i}, \sigma_y^2\right) \text{ where } i = 1, 2, ..., n$$

Note that, for now, we are not assuming a specific density for Y_i such as normal or Poisson, only that it has some known density. It might look as follows (the subscrips are supressed in the example density).



We know the form of $f_{Y_i}\left(y_i, \mu_{y_i}, \sigma_y^2\right)$ but not the specific value of σ_y^2 or values of $\mu_{y_1}, \mu_{y2}, ...\mu_{y_n}$. We want to estimate them. Further assume

$$\mu_{y_i} = \alpha + \beta x_i \text{ where } i = 1, 2, ..., n$$

where the x_i are observed. In other words, the x_i are not random variables from our perspective. Instead, from our perspective, they are known constants. In which case, $f_{y_i}\left(y_i, \alpha + \beta x_i, \sigma_y^2\right)$ and the parameters are α, β, and σ_y^2. Note that μ_{y_i} is a linear function of x_i and $\sigma_{y_i}^2 = \sigma_y^2 \ \forall \ i$.

Imagine a random sample of n observations of (y_i, x_i), $i = 1, 2, ..., n$ and we want the maximum likelihood estimates of α, β, and σ_y^2.

$$L\left(y_1, y_2, ..., y_n, x_1, x_2, ..., x_n; \alpha, \beta, \sigma_y^2\right) = \prod_{i=1}^{n} f_{y_i}\left(y_i, \alpha + \beta x_i, \sigma_y^2\right)$$

[5] Note that I am now naming the random variable Y rather than X. This is more conventional when one assumes that the expected value of Y varys across observations as a function of one of more explanatory variables. Denoting the dependent variable Y is the convention in regression analysis.

and

$$\ln L = \sum_{i=1}^{n} \ln f_{y_i}\left(y_i, \alpha + \beta x_i, \sigma_y^2\right)$$

We would get the maximum likelihood estimates of α, β, and σ_y^2 by maximizing $\ln L$ with respect to these parameters.

For example, if one assumes a normal distribution

$$f_{y_i}\left(y_i, \alpha + \beta x_i, \sigma_y^2\right) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\left(\frac{1}{2\sigma_y^2}\right)[yi-(\alpha+\beta x_i)]^2}$$

That is,

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

where

$$\varepsilon \tilde{\ } N\left(0, \sigma_y^2\right)$$

This is the classical linear regression model (CLR model). In which case,

$$\begin{aligned}
\ln L() &= \sum_{i=1}^{n} \ln f_{y_i}\left(y_i, \alpha + \beta x_i, \sigma_y^2\right) \\
&= \sum_{i=1}^{n} \ln\left[(2\pi)^{-\frac{1}{2}}\left(\sigma_y^2\right)^{-\frac{1}{2}} e^{-\left(\frac{1}{2\sigma_y^2}\right)[yi-(\alpha+\beta x_i)]^2}\right] \\
&= \sum_{i=1}^{n}\left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\sigma_y^2\right) - \left(\frac{1}{2\sigma_y^2}\right)\left(y_i - [\alpha + \beta x_i]\right)^2\right] \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma_y^2\right) - \left(\frac{1}{2\sigma_y^2}\right)\sum_{i=1}^{n}\left(y_i - [\alpha + \beta x_i]\right)^2
\end{aligned}$$

The maximum likelihood estimates of α, β, and σ_y^2 are those values of α, β, and σ_y^2 that maximize $\ln L()$. Lets find them.

$$\begin{aligned}
\frac{d\ln L}{d\alpha} &= 2\left(-\frac{1}{2\sigma_y^2}\right)\sum_{i=1}^{n}\left(y_i - \alpha - \beta x_i\right)(-1) \\
&= \left(\frac{1}{\sigma_y^2}\right)\sum_{i=1}^{n}\left(y_i - \alpha - \beta x_i\right) \;\; set = 0 \qquad (1)
\end{aligned}$$

$$\begin{aligned}
\frac{d\ln L}{d\beta} &= 2\left(-\frac{1}{2\sigma_y^2}\right)\sum_{i=1}^{n}\left(y_i - \alpha - \beta x_i\right)(-x_i) \\
&= \frac{1}{\sigma_y^2}\sum_{i=1}^{n}\left(y_i - \alpha - \beta x_i\right)(x_i) \\
&= \frac{1}{\sigma_y^2}\sum_{i=1}^{n}\left(y_i x_i - \alpha x_i - \beta x_i^2\right) \;\; set = 0 \qquad (2)
\end{aligned}$$

$$
\begin{aligned}
\frac{d\ln L}{d\sigma_y^2} &= -\left(\frac{n}{2}\right)\frac{1}{\sigma_y^2} + \left(\frac{1}{2\sigma_y^4}\right)\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2 \\
&= \left(\frac{1}{2\sigma_y^2}\right)\left[-n + \frac{1}{\sigma_y^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2\right] \\
&= \left(\frac{1}{2\sigma_y^2}\right)\left[\frac{1}{\sigma_y^2}\sum_{i=1}^{n}(y_i - \alpha - \beta x_i)^2 - n\right] \; set = 0 \qquad (3)
\end{aligned}
$$

There are three equations in three unknowns $\left(\alpha, \beta, \sigma_y^2\right)$. Solve for $\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma}_y^2$. Assuming $\sigma_y^2 > 0$, from the first equation we know that

$$
\sum_{i=1}^{n}(y_i - \alpha - \beta x_i) = 0
$$

but

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i - \alpha - \beta x_i) &= -n\alpha + \sum_{i=1}^{n}(y_i - \beta x_i) \\
&= -n\alpha + \sum_{i=1}^{n} y_i - \beta \sum_{i=1}^{n} x_i
\end{aligned}
$$

Noting that $\sum\limits_{i=1}^{n} y_i = n\bar{y}$ and $\sum\limits_{i=1}^{n} x_i = n\bar{x}$,

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i - \alpha - \beta x_i) &= -n\alpha + n\overline{Y} - \beta n\overline{x} = 0 \\
&= -\alpha + \bar{y} - \beta\bar{x} = 0 \\
\alpha &= \bar{y} - \beta\bar{x} \qquad (4)
\end{aligned}
$$

Plug this result into the second equation

$$
\frac{d\ln L}{d\beta} = \frac{1}{\sigma_y^2}\sum_{i=1}^{n}\left(y_i x_i - \alpha x_i - \beta x_i^2\right) = 0 \qquad (5)
$$

to obtain

$$\begin{aligned}
\frac{1}{\sigma_y^2}\sum_{i=1}^{n}\left(y_i x_i - (\bar{y}-\beta\bar{x})\,x_i - \beta x_i^2\right) &= 0 \\
\sum_{i=1}^{n}\left(y_i x_i - \bar{y}x_i + \beta\bar{x}x_i - \beta x_i^2\right) &= 0 \\
\sum_{i=1}^{n} y_i x_i - \bar{y}\sum_{i=1}^{n} x_i + \beta\bar{x}\sum_{i=1}^{n} x_i - \beta\sum_{i=1}^{n} x_i^2 &= 0 \\
\sum_{i=1}^{n} y_i x_i - \bar{y}n\bar{x} + \beta\bar{x}n\bar{x} - \beta\sum_{i=1}^{n} x_i^2 &= 0 \\
\beta n\bar{x}^2 - \beta\sum_{i=1}^{n} x_i^2 &= n\bar{y}\bar{x} - \sum_{i=1}^{n} y_i x_i \\
\beta\left(n\bar{x}^2 - \sum_{i=1}^{n} x_i^2\right) &= n\bar{y}\bar{x} - \sum_{i=1}^{n} y_i x_i \\
\widehat{\beta}_{ml} &= \frac{n\bar{y}\bar{x} - \sum_{i=1}^{n} y_i x_i}{n\bar{x}^2 - \sum_{i=1}^{n} x_i^2} \qquad (6)
\end{aligned}$$

There are many common ways of expressing this result. If one multiplies the numerator and denominator by -1, one obtains

$$\widehat{\beta}_{ml} = \frac{\sum_{i=1}^{n} y_i x_i - n\bar{y}\bar{x}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

(G, p. 139). To obtain another common form of $\widehat{\beta}_{ml}$ note that

$$\begin{aligned}
\sum (y_i - \bar{y})(x_i - \bar{x}) &= \sum \left[y_i x_i - y_i\bar{x} - \bar{y}x_i + \bar{x}\bar{y}\right] \\
&= \sum y_i x_i - \bar{x}\sum y_i - \bar{y}\sum x_i + n\bar{x}\bar{y}\bar{x} \\
&= \sum y_i x_i - \bar{x}n\bar{y} - \bar{y}\sum x_i + n\bar{x}\bar{y} \\
&= \sum y_i x_i - \bar{y}\sum x_i \\
&= \sum y_i x_i - n\bar{y}\bar{x}
\end{aligned}$$

and

$$\begin{aligned}\sum (x_i - \bar{x})^2 &= \sum \left(x_i^2 - x_i\bar{x} - \bar{x}x_i + \bar{x}^2\right) \\ &= \sum x_i^2 - \bar{x}\sum x_i - \bar{x}\sum x_i + n\bar{x}^2 \\ &= \sum x_i^2 - \bar{x}n\bar{x} - \bar{x}n\bar{x} + n\bar{x}^2 \\ &= \sum x_i^2 - n\bar{x}^2 - n\bar{x}^2 + n\bar{x}^2 \\ &= \sum x_i^2 - n\bar{x}^2\end{aligned}$$

So,

$$\widehat{\beta}_{ml} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

(MGB, p. 499 for LS estimate of β; G p. 139). Or in terms of the deviations around the means

$$\widetilde{y}_i \equiv y_i - \overline{Y}$$

and

$$\widetilde{x}_i \equiv x_i - \overline{x}$$

$$\widehat{\beta}_{ml} = \frac{\sum \widetilde{y}_i\widetilde{x}_i}{\sum \widetilde{x}_i^2}$$

Now lets calculate $\widehat{\alpha}_{ml}$. Recall that

$$\alpha = \bar{y} - \beta\bar{x}$$

Plug in $\widehat{\beta}_{ml}$.

$$\widehat{\alpha}_{ml} = \bar{y} - \widehat{\beta}_{ml}\bar{x}$$

to obtain

$$\widehat{\alpha}_{ml} = \bar{y} - \frac{\bar{x}\sum \widetilde{y}_i\widetilde{x}_i}{\sum \widetilde{x}_i^2}$$

Looking ahead, the maximum likelihood estimates of α and β, assuming

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

where

$$\varepsilon_i\, \tilde{}\, N\left(0, \sigma_\varepsilon^2\right)$$

are also the least square estimates. That is, for the Classical Linear Regression Model, the maximum likelihood estimates of α and β are equivalent to the least squares estimates.

Now find the maximum likelihood estimates of σ_y^2. Recall the third first-order conditon

$$\begin{aligned}
\left(\frac{1}{2\sigma_y^2}\right)\left[\frac{1}{\sigma_y^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2-n\right] &= 0 \\
\left[\frac{1}{\sigma_y^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2-n\right] &= 0 \\
\frac{1}{\sigma_y^2}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2 &= n \\
\frac{1}{\sigma_y^2} &= \frac{n}{\sum\limits_{i=1}^{n}(y_i-\alpha-\beta x_i)^2} \\
\sigma_y^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i-\alpha-\beta x_i)^2
\end{aligned}$$

So the maximum likelihood estimate of σ_y^2 is

$$\widehat{\sigma}_y^2=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\widehat{\alpha}_{ml}-\widehat{\beta}_{ml}x_i\right)^2$$

In summary, we have just derived the maximum likelihood estimates of α, β, and σ_y^2 with a random sample of size n assuming in the population

$$y_i=\alpha+\beta x_i+\varepsilon_i$$

where

$$\varepsilon_i\tilde{\ }N\left(0,\sigma_\varepsilon^2\right)$$

That is, we have derived the maximum likelihood estimators for α, β, and σ_y^2 for the classical linear regression model.

1.3.1 Lets do one more maximum likelihood problem: return to the Bernoulli problem

Assume two alternatives and the probability that individual i chooses alternative 1 on any trial, t, is p_i. That is,

$$\begin{aligned}
f_{X_{it}}(x_{it},p_i) &= p_i^{x_{it}}(1-p_i)^{1-x_{it}} \quad \text{for } x_{it}=0 \text{ or } 1 \\
&= 0 \text{ otherwise}
\end{aligned}$$

where $i=1,2,....,n$. $x_{it}=1$ if individual chooses alternative 1 on trial t, and zero otherwise, $t=1,2,.....,T$.

Let x_i be the number of times individual i chooses alternative 1 in T trials.

$$x_i=\sum_{t=1}^{T}x_{it}$$

In which case, we can (have) shown that

$$f_{X_i}(x_i, p_i, T) = \begin{pmatrix} T \\ x_i \end{pmatrix} p_i^{x_i}(1-p_i)^{T-x_i} \qquad i = 1, 2, ..., n$$

Further assume

$$p_i = \alpha + \beta G_i$$

where

$$\begin{aligned} G_1 &= male \\ G_0 &= female \end{aligned}$$

Note that the variable G_i, which can only take one of two values, 0 or 1. Variables with this property are typically referred to as dummy variables (G., chap 9). Lets says we know that $\alpha = .10$, implying that $P_i = .1$ if female. Also, β equals either $0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$, or 0.8.

We have a random sample of n individuals. That is, for n independent individuals we observe the choices each makes on T independent trials. What is the maximum likelihood estimate of β? First note that

$$\begin{aligned} f_{X_i}(x_i, p_i, T) &= \begin{pmatrix} T \\ x_i \end{pmatrix} p_i^{x_i}(1-p_i)^{T-x_i} \\ &= \begin{pmatrix} T \\ x_i \end{pmatrix} (.1+\beta G_i)^{x_i}(1-(.1+\beta G_i))^{T-x_i} \end{aligned}$$

Therefore,

$$L = \prod_{i=1}^{n} \begin{pmatrix} T \\ x_i \end{pmatrix} (.1+\beta G_i)^{x_i}(1-(.1+\beta G_i))^{T-x_i}$$

So,

$$\begin{aligned} \ln L &= \sum_{i=1}^{n} \ln\left[\begin{pmatrix} T \\ x_i \end{pmatrix} (.1+\beta G_i)^{x_i}(1-(.1+\beta G_i))^{T-x_i}\right] \\ &= \sum_{i=1}^{n}\left[\ln\begin{pmatrix} T \\ x_i \end{pmatrix} + x_i \ln(.1+\beta G_i) + (T-x_i)\ln(1-0.1-\beta G_i)\right] \\ &= \sum_{i=1}^{n}\left[\ln\begin{pmatrix} T \\ x_i \end{pmatrix} + x_i \ln(.1+\beta G_i) + (T-x_i)\ln(0.9-\beta G_i)\right] \end{aligned}$$

The maximum likelihood estimate of β is the β that maximizes the above equation but it is also the β that maximizes

$$\sum_{i=1}^{n}\left[x_i \ln(.1+\beta G_i) + (T-x_i)\ln(0.9-\beta G_i)\right]$$

for a given random sample. One would calculate this for $\beta = 0.1, 0.2, ...0.8$ and the $\widehat{\beta}_{ml}$ is the one that maximizes the function. For example, assume $T = 2$ and $n = 3$ such that

$$\begin{aligned} x_{11} &= 1 \\ x_{12} &= 1 \\ x_{21} &= 1 \\ x_{22} &= 0 \\ x_{31} &= 0 \\ x_{32} &= 0 \end{aligned}$$

where individuals 1 and 2 are males and individual 3 is a female. What is the maximum likelihood estimate of β? Remember that for maximum likelihood estimation, one needs to know the form of $f_x(x; \theta)$

1.4 why we like max lik technique

1. very general
2. Estimates have desireable properties under very general conditions.
3. ML estimates most asymtotically efficient
4. doesn't require a random sample
5. classical linear regression model is just a very special case
6. Easy to do hypothesis testing and tests of significance

then discuss lik ratio tests in a general way. MGB page 441