2. The method of maximum likelihood (ML)

Motivation

Suppose that we have a box with two balls in it, each of which is either black or white. But we have no idea how many of these two balls are black.

We randomly draw a ball from the box and find that it is black.

Is the other ball also black?

If the other ball is black then then the probability of us having drawn a black is 100%. If the other ball is white then the probability of us having drawn a black is only 50%.

Because 100% > 50%, it's reasonable to conclude that the other ball is black, although we can't be sure.

What we have done is an example of estimation based on the principle that we should choose the value which maximises the likelihood of what has happened.

Suppose that Y is an observation from some probability distribution that depends on an unknown parameter θ.

The **likelihood function** is defined to be the pdf of Y, $f(y)$,
but considered as a function of θ.

We denote the likelihood by $L(\theta)$ or $L(\theta; y)$.

The **maximum likelihood estimate (MLE)** of θ is the value of θ which maximises the likelihood $L(\theta)$.

Note: The MLE may not be unique.

Example 6 (Formalisation of our motivating example)

We have a box with 2 balls in it, each of which is either black or white. We randomly draw a ball from the box and find that it is black.

What is the MLE of the number of black balls originally in the box?

Let θ be the number of black balls originally in the box.
Also let Y be the number of black balls in our sample of one.

Then $Y \sim \text{Bern}(\theta/2)$, and $p(y) = \begin{cases} \theta/2, & y=1 \\ 1-\theta/2, & y=0 \end{cases}$

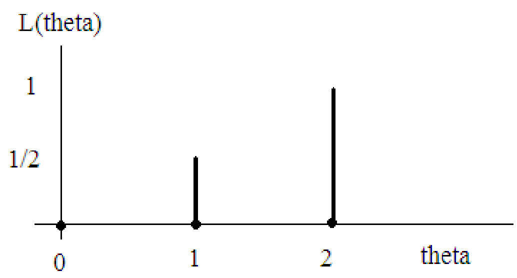
where $\theta = 0,1,2$.

But we actually observed $y = 1$ blacks.
Thus $p(y) = \theta/2$.

So the likelihood is $L(\theta) = \dfrac{\theta}{2} = \begin{cases} 0/2 = 0, & \theta = 0 \\ 1/2 = 1/2, & \theta = 1 \\ 2/2 = 1, & \theta = 2 \end{cases}$

But max(0,1/2,1) = 1.

So the MLE of θ is $\hat{\theta} = 2$.



What if the chosen ball were white?
Then $y = 0$ and so $p(y) = 1-\theta/2$.

So $L(\theta) = 1 - \dfrac{\theta}{2} = \begin{cases} 1-0/2 = 1, & \theta = 0 \\ 1-1/2 = 1/2, & \theta = 1 \\ 1-2/2 = 0, & \theta = 2 \end{cases}$

So the MLE of θ is now 0.

In conclusion we may write the MLE generally as $\hat{\theta} = 2y$, y = 0,1.
We may also write the MLE as a random variable: $\hat{\theta} = 2Y$.

Note that the method of moments leads to the same estimate as the method of maximum likelihood: equating $\mu_1' = EY = \theta / 2$ with m_1' = y, we get $\hat{\theta} = 2y$.

Example 7 A bent coin is tosses 5 times and heads come up twice.

Find the MLE of the probability of heads coming up on a single toss.

Let Y be the number of heads that come up, and let p be the probability of interest.

Then Y ~ Bin(n,p), where n = 5, and $p(y) = \binom{n}{y} p^y (1-p)^{n-y}$, y = 0,...,n (0 < p < 1).

So the likelihood is $L(p) = \binom{n}{y} p^y (1-p)^{n-y}$, 0 < p < 1 (y = 0,...,n).

We now calculate $L'(p) = \binom{n}{y} \left\{ p^y (n-y)(1-p)^{n-y-1}(-1) + yp^{y-1}(1-p)^{n-y} \right\}$.
Setting this to zero yields $\hat{p} = \dfrac{y}{n} = \dfrac{2}{5}$.

Alternatively, we can work on the log scale, using the result that the number which maximises a function also maximises the natural logarithm of that function.

The **loglikelihood function** is $l(p) = \log L(p) = \log \binom{n}{y} + y \log p + (n-y)\log(1-p)$.
Then $l'(p) = 0 + \dfrac{y}{p} - \dfrac{n-y}{1-p}$.
Setting $l'(p)$ to zero yields the MLE, $\hat{p} = \dfrac{y}{n} = \dfrac{2}{5}$.

(Note that this is the same as the MOME. Also note that

$$l''(p) = -\frac{y}{p^2} - \frac{n-y}{(1-p)^2} < 0,$$

which confirms that $\hat{p}$ = y/n maximises L.)

Simplification

Observe that the $\binom{n}{y}$ term could have been left out, with no change to the result.

Thus: $L(p) = p^y(1-p)^{n-y}$

$l(p) = y\log p + (n-y)\log(1-p)$

$l'(p) = \dfrac{y}{p} - \dfrac{n-y}{1-p}$

etc.

Therefore we redefine the likelihood as **any constant multiple of *Y*'s pdf**.

This means that we can safely ignore multiplicative constants when writing down the likelihood function $L(\theta)$, and we can ignore any additive constants when writing down the loglikeliohood function $l(\theta)$.

The case of several observations

The method of ML can also be used when there are several sample observations, say $Y_1,\ldots,Y_n$, whose distribution depends on θ.

Example 8 Suppose that 1.2, 2.4 and 1.8 are a random sample from an exponential distribution with unknown mean.

Find the MLE of that mean.

The sample observations $Y_1,\ldots,Y_n$ have joint pdf

$$\begin{aligned} f(y_1,\ldots,y_n) &= \prod_{i=1}^n f(y_i) \quad \text{(by independence)} \\ &= \prod_{i=1}^n \frac{1}{b}e^{-y_i/b} \quad \text{(where } b = EY_i \text{ is the unknown mean)} \\ &= b^{-n}e^{-\frac{1}{b}\sum_{i=1}^n y_i} \\ &= b^{-n}e^{-\dot{y}/b}, \end{aligned}$$

where $\dot{y} = y_1 + \ldots + y_n$.

So the likelihood is $L(b) = b^{-n}e^{-\dot{y}/b}$, $b > 0$. (NB: There are no constants to ignore.)
So the loglikelihood is $l(b) = -n\log b - \dot{y}/b$.

Then $l'(b) = -\frac{n}{b} + \frac{\dot{y}}{b^2}$.

Solving $l'(b) = 0$ leads to the MLE of *b*, namely $\hat{b} = \frac{\dot{y}}{n} = \bar{y} = 1.8$.

The case of several parameters

ML estimation also works when there are two or more unknown parameters.

Example 9 $Y_1, \dots, Y_n \sim iid\ N(a, b^2)$.

Find the MLE's of *a* and b^2.

$$f(y_1, \dots, y_n) = \prod_{i=1}^{n} \frac{1}{b\sqrt{2\pi}} \exp\left\{-\frac{1}{2b^2}(y_i - a)^2\right\}$$

$$= b^{-n}(2\pi)^{-n/2} \exp\left\{-\frac{1}{2b^2}\sum_{i=1}^{n}(y_i - a)^2\right\}.$$

So $L(a, b^2) = (b^2)^{-n/2} \exp\left\{-\frac{1}{2b^2}\sum_{i=1}^{n}(y_i - a)^2\right\}$, $-\infty < a < \infty$, $b^2 > 0$.

$$l(a, b^2) = -\frac{n}{2}\log b^2 - \frac{1}{2b^2}\sum_{i=1}^{n}(y_i - a)^2 .$$

$$\frac{\partial l(a, b^2)}{\partial a} = -\frac{1}{2b^2}\sum_{i=1}^{n} 2(y_i - a)^1(-1) = \frac{1}{b^2}\sum_{i=1}^{n}(y_i - a) = \frac{1}{b^2}(\dot{y} - na). \qquad (1)$$

$$\frac{\partial l(a, b^2)}{\partial b^2} = -\frac{n}{2b^2} + \frac{1}{2b^4}\sum_{i=1}^{n}(y_i - a)^2 . \qquad (2)$$

Setting (1) to 0 yields the MLE of *a*, namely $\hat{a} = \dot{y}/n = \bar{y}$.

Substituting $\bar{y}$ for *a* in (2) and then setting (2) to zero yields

$$\hat{b}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{n-1}{n}\left\{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2\right\} = \left(\frac{n-1}{n}\right)s^2 .$$

(Note that these MLE's are exactly the same as the MOME's mentioned earlier.)

Example 10 A partly melted die is rolled repeatedly until the first 6 comes up. Then it is rolled again the same number of times.

We are interested in p, the probability of 6 coming up on a single roll,

Suppose that the first 6 comes up on the third roll, and the numbers which then come up are 6,2,6.

Find the MLE of p.

Let X = number of rolls until first 6, and Y = number of 6's on last half of rolls.

Then $X \sim \text{Geo}(p)$ (with $x = 3$) and $(Y \mid X = x) \sim \text{Bin}(x,p)$ (with $y = 2$).

So $p(x,y) = p(x)p(y \mid x) = (1-p)^{x-1} p \times \binom{x}{y} p^y (1-p)^{x-y}$; $x = 1,2,\ldots;\ y = 0,\ldots,x$.

So $L(p) = (1-p)^{x-1+x-y} p^{1+y} = (1-p)^a p^b$, where $a = 2x - y - 1$ and $b = 1 + y$.

Then: $l(p) = a\log(1-p) + b\log p$

$$l'(p) = -\frac{a}{1-p} + \frac{b}{p} = 0 \Rightarrow p = \frac{b}{a+b}.$$

So the MLE of p is

$$\hat{p} = \frac{b}{a+b} = \frac{1+y}{(2x-y-1)+(y+1)} = \frac{1+y}{2x} = \frac{1+2}{2(3)} = \frac{3}{6} = \frac{1}{2}.$$

(This makes sense, since 6 came up half the time: 1 + 2 = 3 times in 3 + 3 = 6 rolls.)

Note: We see that ML estimation works when the sample observations are not iid. The same cannot be said for the MOM. We could not use the MOM directly here.

Example 11 Suppose that 3.6 and 5.4 are two numbers chosen randomly and independently from between 0 and c. Find the MLE of c.

Let X and Y denote the two numbers as random variables. Then $X,Y \sim iid\ U(0,c)$.
So $p(x,y) = p(x)p(y) = (1/c)(1/c) = 1/c^2$, $0 < x < c$, $0 < y < c$ $(c > 0)$.
So $L(c) = c^{-2}$, $c > 0$.
So $l(c) = \log L(c) = -2\log c$. So $l'(c) = -2/c$. So $l'(c) = 0 \Rightarrow c = \infty$.

But this is wrong. First observe that $c > x$ and $c > y$. Thus $c > u$, where $u = \max(x,y)$.
So the likelihood function is in fact $L(c) = c^{-2}$, $c > u$ (actually, $c \geq u$).

Now $L(c)$ is a strictly decreasing function. So it has a maximum at $c = u$ (not $c = \infty$).
Therefore the MLE of c is $\hat{c} = u = \max(3.6, 5.4) = 5.4$.

The invariance property of MLEs

If $\hat{\theta}$ is an MLE of θ and g is a function, then an MLE of $\phi = g(\theta)$ is $\hat{\phi} = g(\hat{\theta})$.

Example 12 Suppose that $Y \sim Bin(n, p)$. What is the MLE of $r = p^2$?
Is this MLE unbiased? If not, find an unbiased estimator of r.

The MLE of p is $\hat{p} = Y/n$. Therefore the MLE of r is $\hat{r} = \hat{p}^2 = Y^2/n^2$. Now

$$E\hat{r} = E\hat{p}^2 = V\hat{p} + (E\hat{p})^2 = \frac{p(1-p)}{n} + p^2 = \left(\frac{n-1}{n}\right)p^2 + \frac{p}{n} = \left(\frac{n-1}{n}\right)r + \frac{p}{n} \neq r.$$

So $\hat{r}$ is biased with bias $B(\tilde{r}) = E\tilde{r} - r = \left\{\left(\frac{n-1}{n}\right)r + \frac{p}{n}\right\} - r = \frac{p-r}{n} = \frac{p(1-p)}{n} = V\hat{p}$.

Next observe that $E\left\{\left(\frac{n}{n-1}\right)\hat{r}\right\} = r + \left(\frac{n}{n-1}\right)\frac{p}{n} = r + \frac{p}{n-1}$ and $E\left(\frac{\hat{p}}{n-1}\right) = \frac{p}{n-1}$.

So an unbiased estimator of r is

$$\tilde{r} = \left(\frac{n}{n-1}\right)\hat{r} - \frac{\hat{p}}{n-1} = \frac{n\hat{p}^2 - \hat{p}}{n-1} = \frac{\hat{p}(n\hat{p}-1)}{n-1} = \frac{Y}{n}\left(\frac{Y-1}{n-1}\right).$$

As a check that this estimator is unbiased, consider the case $n = 2$ and $p = 1/2$.
In this case $r = 1/4$ and the possible values of Y are:

0 with probability 1/4 and corresponding value of $\tilde{r}$ $\frac{0}{2}\left(\frac{0-1}{2-1}\right) = 0$

1 with probability 1/2 and corresponding value of $\tilde{r}$ $\frac{1}{2}\left(\frac{1-1}{2-1}\right) = 0$

2 with probability 1/4 and corresponding value of $\tilde{r}$ $\frac{2}{2}\left(\frac{2-1}{2-1}\right) = 1$.

Thus $f(\tilde{r}) = \begin{cases} 3/4, & \tilde{r} = 0 \\ 1/4, & \tilde{r} = 1 \end{cases}$, and so $E\tilde{r} = 0\left(\frac{3}{4}\right) + 1\left(\frac{1}{4}\right) = \frac{1}{4} = r$. I.e. $\tilde{r}$ is unbiased.

(*Exercise*: In a similar way, check that $\tilde{r}$ is unbiased for the case $n = 3$ and $p = 1/3$.)