Lecture 21: Maximum Likelihood Estimation

16 April, 2018

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

Binomial MLE, unknown number of trials

We flip a 'fair coin' and observe the number of heads but we do not know how many times the coin was flipped. Based on no. of heads reported we want to estimate the no. of coin tosses.

Example 21.1 *Let $X_1, X_2, \cdots, X_n$ be a random sample from a binomial(k, p) population, where p is known and k is unknown. Find the MLE of the parameter k.*

Solution: The likelihood function is

$$\begin{aligned} L(k|x_1, x_2, \cdots, x_n, p) &= \prod_{i=1}^{n} \binom{k}{x_i} p^{x_i}(1-p)^{k-x_i} \\ &= \binom{k}{x_1}\binom{k}{x_2}\cdots\binom{k}{x_n} p^{\sum_{i=1}^n x_i}(1-p)^{nk-\sum_{i=1}^n x_i} \\ &= \frac{k!}{x_1!(k-x_1)!} \times \frac{k!}{x_2!(k-x_2)!} \times \cdots \times \frac{k!}{x_n!(k-x_n)!} p^y (1-p)^{nk-y}, \end{aligned}$$

where $y = \sum_{i=1}^{n} x_i$.

Note that though p is known but no specific value is given, so it is important for us to decide about the possible values of p. Observe that $p \in (0,1)$ (if $p = 0$ or $p = 1$ then we loose the definition of a pmf), therefore $(1-p) \in (0,1) \implies (1-p)^n \in (0,1)$.

If all x_i's are zero then $L(k|0, 0, \cdots, 0, p) = (1-p)^{kn}$. As $(1-p)^n \in (0,1)$, so MLE is going to be $k = 1$. So now we assume that at least one x_i's is ≥ 1. In this case we have $\max\limits_{1 \leq i \leq n} x_i \geq 1$.

Maximizing $L(k|\mathbf{x}, p)$ by differentiation is not legitimate as k varies over set of positive integers. Thus we try a different approach.

Since the number of coin-tosses can not be less than the maximum number of heads observed Thus admissible values of parameter k are all positive integers $\geq \max\limits_{1 \leq i \leq n} x_i$.

Hence MLE is a positive integer $\hat{k} \geq \max_{1\leq i\leq n} x_i$ that satisfies $\dfrac{L(\hat{k}|\mathbf{x},p)}{L(k|\mathbf{x},p)} \geq 1$ for $\max_{1\leq i\leq n} x_i \leq k \leq \hat{k}$ and $\dfrac{L(k|\mathbf{x},p)}{L(\hat{k}|\mathbf{x},p)} \leq 1$ for $k > \hat{k}$. We will show that there is only one such $\hat{k}$.

Claim 21.2 *The equation*

$$(1-p)^n = \prod_{i=1}^{n}(1-x_i z) \tag{21.1}$$

has a unique solution $\hat{z}$ in the interior of interval $\left[0, \dfrac{1}{\max_{1\leq i\leq n} x_i}\right]$.

Proof of the Claim: The right-hand side of the equation (21.1) is the following real n-degree polynomial function

$$f(z) = (1-x_1 z)(1-x_2 z)\cdots(1-x_n z).$$

The derivative of f is

$$\begin{aligned} f'(z) &= -x_1\left[(1-x_2 z)\cdots(1-x_n z)\right] - x_2\left[(1-x_1 z)(1-x_3 z)\cdots(1-x_n z)\right] - \cdots \\ &\quad \cdots - x_n\left[(1-x_1 z)(1-x_2 z)\cdots(1-x_{n-1} z)\right] \\ &< 0, \text{ on } \left[0, \frac{1}{\max_{1\leq i\leq n} x_i}\right) \end{aligned}$$

Because

1. On the interval $\left[0, \dfrac{1}{\max_{1\leq i\leq n} x_i}\right)$, $x_i z < \dfrac{x_i}{\max_{1\leq i\leq n} x_i} \leq 1$ for each i, if all $x_i \geq 1$.
2. If some of x_i's are zero (but not all together) then corresponding to zero x_i's, $(1-x_i z)$ will be one. Still we have $f'(z) < 0$ on the interval $\left[0, \dfrac{1}{\max_{1\leq i\leq n} x_i}\right)$.

Hence f is a strictly decreasing function of z on the interval $\left[0, \dfrac{1}{\max_{1\leq i\leq n} x_i}\right]$ with a value of 1 at $z = 0$ and a value of 0 at $z = \dfrac{1}{\max_{1\leq i\leq n} x_i}$.

Being a polynomial, f is continuous thus by intermediate value theorem, there is a point $\hat{z}$ in the interval $\left(0, \dfrac{1}{\max\limits_{1\le i\le n} x_i}\right)$ that solves the equation (21.1). The point $\hat{z}$ is unique since f is strictly decreasing. This completes the proof of the claim.

The quantity $\frac{1}{\hat{z}}$ may not be an integer. Let $\hat{k}$ be the greatest integer less than or equal to $\frac{1}{\hat{z}}$. Then $\hat{k} \le \frac{1}{\hat{z}}$ and $\hat{k}+1 > \frac{1}{\hat{z}}$. Hence

$$
\begin{aligned}
(1-p)^n &= \prod_{i=1}^{n}(1-x_i\hat{z}) \ge \prod_{i=1}^{n}\left(1-\frac{x_i}{\hat{k}}\right) \quad (\because \text{ function is strictly decreasing})\\
&\implies [\hat{k}(1-p)]^n \ge \prod_{i=1}^{n}(\hat{k}-x_i).\\
\text{Similarly}\quad (1-p)^n &= \prod_{i=1}^{n}(1-x_i\hat{z}) < \prod_{i=1}^{n}\left(1-\frac{x_i}{\hat{k}+1}\right)\\
&\implies [(\hat{k}+1)(1-p)]^n < \prod_{i=1}^{n}((\hat{k}+1)-x_i)
\end{aligned}
$$

Claim 21.3 *$\hat{k}$ is the global maximum of $L(k|\mathbf{x},p)$.*

Proof of the Claim: The ratio of likelihoods is

$$
\begin{aligned}
\frac{L(k|\mathbf{x},p)}{L(k-1|\mathbf{x},p)} &= \frac{\frac{k!}{x_1!(k-x_1)!}\times\frac{k!}{x_2!(k-x_2)!}\times\cdots\times\frac{k!}{x_n!(k-x_n)!}\; p^y(1-p)^{nk-y}}{\frac{k-1!}{x_1!(k-1-x_1)!}\times\frac{k-1!}{x_2!(k-1-x_2)!}\times\cdots\times\frac{k-1!}{x_n!(k-1-x_n)!}\; p^y(1-p)^{n(k-1)-y}}\\
&= \frac{k}{k-x_1}\times\frac{k}{k-x_2}\times\cdots\times\frac{k}{k-x_n}\times\frac{1}{(1-p)^{-n}} = \frac{[k(1-p)]^n}{\prod_{i=1}^{n}(k-x_i)}.
\end{aligned}
$$

If $k < \hat{k}$ then $\frac{1}{k} > \frac{1}{\hat{k}}$. Hence

$$
\begin{aligned}
f\left(\frac{1}{k}\right) &< f\left(\frac{1}{\hat{k}}\right)\\
\implies \prod_{i=1}^{n}\left(1-\frac{x_i}{k}\right) &< \prod_{i=1}^{n}\left(1-\frac{x_i}{\hat{k}}\right) \le (1-p)^n\\
\implies \frac{[k(1-p)]^n}{\prod_{i=1}^{n}(k-x_i)} &\ge 1 \implies \frac{L(k|\mathbf{x},p)}{L(k-1|\mathbf{x},p)} \ge 1
\end{aligned}
$$

Therefore $L(k|\mathbf{x},p)$ is an increasing function in k for $k \in \left\{\max\limits_{1\le i\le n} x_i, \max\limits_{1\le i\le n} x_i+1,\cdots,\hat{k}\right\}$. Now let $k>\hat{k}$. then $\frac{1}{k}\le\frac{1}{\hat{k}+1}$. Hence

$$\begin{aligned}
& f\left(\frac{1}{k}\right) \ge f\left(\frac{1}{\hat{k}+1}\right) \\
\implies & \prod_{i=1}^{n}\left(1-\frac{x_i}{k}\right) \ge \prod_{i=1}^{n}\left(1-\frac{x_i}{\hat{k}+1}\right) > (1-p)^n \\
\implies & \frac{[k(1-p)]^n}{\prod_{i=1}^{n}(k-x_i)} < 1 \implies \frac{L(k|\mathbf{x},p)}{L(k-1|\mathbf{x},p)} < 1.
\end{aligned}$$

Therefore $L(k|\mathbf{x},p)$ is a decreasing function in k for $k\in\left\{\hat{k},\hat{k}+1,\cdots\right\}$. This completes the proof of the claim.

Thus, this analysis shows that there is a unique maximum for the likelihood function and it can be found by numerically solving an nth-degree polynomial equality (21.1). ■

Remark 21.4 *If a likelihood function cannot be maximized analytically, it may be possible to use a computer and maximize the likelihood function numerically. In fact, this is one of the most important features of MLEs. If a likelihood can be written down, then there is some hope of maximizing it numerically and, hence, finding MLEs of the parameters. When this is done, there is still always the question of whether a local or global maximum has been found. Thus, it is always important to analyze the likelihood function as much, as possible, to find the number and nature of its local maxima, before using numeric maximization.*

Example 21.5 *Let $X_1, X_2, \cdots, X_n$ be a random sample from a Uniform$(0,\theta)$ distribution, where θ is unknown. Find the maximum likelihood estimator (MLE) of θ based on this random sample.*

Solution: A uniform$(0,\theta)$ random variable has the pdf

$$f(x)=\begin{cases} \frac{1}{\theta} & \text{if } \; 0<x<\theta \\ 0 & \text{otherwise} \end{cases}.$$

The likelihood function is

$$L(\theta|x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n} f(x_i|\theta)=\begin{cases} \dfrac{1}{\theta^n} & \text{if } \; 0<x_1,x_2,\cdots,x_n<\theta \\ 0 & \text{otherwise} \end{cases}.$$

Likelihood function is non-zero if $\theta>\max\limits_{1\le i\le n} x_i$. Also L is a strictly decreasing function of θ, therefore the it attains it's supremum at point $\hat{\theta}=\max\limits_{1\le i\le n} x_i$. Hence MLE is

$$\hat{\theta}=\max\{X_1,X_2,\cdots,X_n\}.$$

■