The LLF and MLE for a binomial distribution

Luc Ta

September 2023

1 Introduction

Suppose we have a binomially distributed random variable X, i.e.,

$$X \sim \mathrm{Bin}(n, p),$$

where n is the number of trials and the parameter p is the probability of "success." Recall that the MLE θ of p is the value where, if you plug in θ for p, you get the maximum value of the likelihood function $\mathcal{L}$. We need to show that the MLE θ of p is X/n. Since X is binomially distributed, the binomial theorem gives us the likelihood function

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

2 Blueprints for the proof

We've already shown using the binomial theorem that (1) is the likelihood function of a binomial distribution. So, it only remains to show the MLE θ equals X/n. By definition, the MLE is the value θ of p in (1) that maximizes the likelihood function, for arbitrary n and x.

So, we want to find the maximum of the likelihood function as a function of p. In more detail, we'll begin by differentiating the likelihood function with respect to p (after taking logarithms to lower the exponents and make the calculus easier). Then, we'll set the derivative equal to 0 to find the critical point, and we'll end by arguing why this critical point is a global maximum (rather than a minimum or an inflection point).

3 The proof

Theorem. *Given a binomially distributed random variable $X \sim \mathrm{Bin}(n, p)$, the MLE θ of p is X/n.*

Proof. To find local maxima of the $\mathcal{L}$ given in (1), we'll begin by differentiating. (Even though the binomial distribution is a discrete probability distribution, differentiating $\mathcal{L}$ is possible since p can be *any* value in the interval $[0, 1]$, and $\mathcal{L}$ is a polynomial in p.) To make the calculus easier, we'll begin by taking logarithms of both sides. This suffices since it will preserve all critical points without introducing any new ones. (That is, if the derivative of the likelihood function at a point is 0, then it will stay 0 at that point after taking logarithms; and if is nonzero at a point, then it will stay nonzero after taking logarithms.) So, take logarithms of both sides of (1) and use the logarithm laws to get

$$\begin{aligned}\log(\mathcal{L}(p)) &= \log\left[\binom{n}{x}p^x(1-p)^{n-x}\right] \\ &= \log\binom{n}{x} + \log\left(p^x\right) + \log\left[(1-p)^{n-x}\right] \\ &= \log\binom{n}{x} + x\log p + (n-x)\log(1-p).\end{aligned}$$

Noting that $\log\binom{n}{x}$ is a constant (since n and x are arbitrary/fixed), differentiate with respect with the parameter p to get

$$\begin{aligned}\frac{d}{dp}\log(\mathcal{L}(p)) &= \frac{d}{dp}\log\binom{n}{x} + \frac{d}{dp}x\log p + \frac{d}{dp}(n-x)\log(1-p) \\ &= 0 + \frac{x}{p} - \frac{n-x}{1-p},\end{aligned}$$

where in the last equality we used the chain rule to obtain the third term on the RHS. To find critical points of $\log(\mathcal{L})$, we simply set the RHS equal to 0 and solve for p to get

$$p = \frac{x}{n}.$$

So, $\mathcal{L}$ reaches its only critical point when $p = x/n$. We claim that x/n is in fact a global maximum for $\mathcal{L}$. The "global" part of this claim follows from the fact that this is the *only* critical point of $\mathcal{L}$. Can you justify why this only critical point must be a maximum?

Finally, note that x is simply the desired value for the random variable X; the probability that $X = x$ is precisely what $P\{X = x\}$ is. Hence, we may set x to equal X and set the MLE θ to equal p in the previous equation to get

$$\theta = \frac{X}{n},$$

which is what we needed to show. □