Maximum Likelihood Estimation (MLE)

Suppose you make some observations $y_1, \ldots, y_n$ and you strongly believe that they're coming from a particular distribution (e.g., a Bernoulli) but that some or all of the parameters of that distribution are unknown (e.g., you don't know p). You might want to use your data $y_1, \ldots, y_n$ to estimate the distribution parameters (e.g., p).

If the data are from a discrete distribution, the *likelihood function* L is the conditional probability that you would observe the data you actually observed given the parameter values, treated as a function of the distribution parameters $\theta_1, \theta_2, \ldots, \theta_k$:

$$\begin{aligned} L(y_1, y_2, \ldots, y_n \quad &| \quad \theta_1, \theta_2, \ldots, \theta_k) \\ &= \quad P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n | \theta_1, \theta_2, \ldots, \theta_k). \end{aligned}$$

For example, if the y_i values came from a Bernoulli(p) distribution, the likelihood function is

$$L(y_1, y_2, \ldots, y_n | p) \quad = \quad P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n | p).$$

Furthermore, if those observations are taken to be independent, this likelihood function becomes

$$L(y_1, y_2, \ldots, y_n | p) \quad = \quad \prod_{i=1}^{n} P(Y_i = y_i | p) = p^{\sum_{i=1}^{n} y_i} (1-p)^{n - \sum_{i=1}^{n} y_i}.$$

Then it might make sense to try to maximize this likelihood function over all values of p. We could try to differentiate twice and check for concavity, then set the first derivative equal to 0. Remember to also check boundary points. Or, we can attempt to maximize the natural log of the likelihood function. Because the log function is monotone and increasing, we have that

$$argmax(f(x)) = argmax(\log(f(x)),$$

and logs of products result in sums, and it's often easier to differentiate sums than products.

If the data values are assumed to be the result of independent observations from a particular continuous probability distribution with probability density function $f(\cdot)$ and unknown parameters $\theta_1, \ldots, \theta_k$, the likelihood function is just the product of the density functions:

$$L(y_1, y_2, \ldots, y_n | \theta_1, \ldots, \theta_k) = \prod_{i=1}^{n} f(y_i | \theta_1, \ldots, \theta_k).$$

The *maximum likelihood estimator* (MLE) of a distribution parameter is taken to be the value of that parameter that maximizes the likelihood function L (or the log-likelihood function).

Example 1 Suppose $Y_1, \ldots, Y_n$ are independent observations from a Bernoulli(p) distribution. You wish to estimate p. Of course, you could use the unbiased estimator $\hat{p} = n^{-1} \sum_{i=1}^{n} Y_i$. Let's try the MLE approach.

Example 2 Let r be a non-negative constant. Let $Y_1, \ldots, Y_n$ be an independent random sample from a distribution with the density function

$$f(y|\theta) = \theta^{-1} r y^{r-1} e^{-y^r/\theta} \mathbb{1}\{\theta > 0,\, y > 0\}.$$

Find the MLE of θ.

Example 3 Show that when making random, independent observations $Y_1, \dots, Y_n$ from a normal distribution with unknown mean μ and unknown variance σ^2 that the MLE for μ is $\overline{Y} := n^{-1} \sum_{i=1}^{n} Y_i$ but that the MLE for σ^2 is

$$\widehat{\sigma^2_{MLE}} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \frac{n-1}{n} S^2.$$

Homework

1. The random variable Y has a geometric(p) distribution if it models the number of successive independent trials until the first success (p = probability of success on any individual trial). In this case,

$$P(Y = k) = p(1-p)^{k-1}, \;\; k \in \{1, 2, 3, \ldots\}.$$

   Let $Y_1, \ldots, Y_n$ be an independent random sample from a geometric(p) distribution. Find the MLE of p.

2. The random variable N has a Poisson(λ) distribution if

$$P(N = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \;\; k \in \{0, 1, 2, 3, \ldots\}.$$

   It turns out that
$$E[N] = Var(N) = \lambda.$$

   Show that when making random, independent observations $Y_1, \ldots, Y_n$ from a Poisson distribution with mean λ that the maximum likelihood estimator for λ is the sample mean: $\widehat{\lambda_{MLE}} = n^{-1}\sum_{i=1}^{n} Y_i$.

3. The random variable U has a continuous uniform distribution on the interval $[a, b]$ if its density function is given by

$$f_U(u) = \mathbb{1}\{a \leq u \leq b\}\frac{1}{b-a}, \;\; u \in \mathbb{R}.$$

   (a) Show that when making random, independent observations $X_1, \ldots, X_n$ from a continuous uniform distribution on $[0, b]$ that the maximum likelihood estimator for b is $\widehat{b_{MLE}} = max_i(X_i)$.

   (b) Show that $2\overline{X}$ is an unbiased estimator for b.

   (c) It turns out that $\frac{n+1}{n}\max_i(X_i)$ is also an unbiased estimator for b. Which of the three estimators from parts (a), (b), and (c) do you think is best and why?

