Methods of Point Estimation. Method of Maximum Likelihood.

Anastasiia Kim

April 20, 2020

Method of Maximum Likelihood

The maximum likelihood estimator (MLE) is the parameter point for which the observed sample is most likely.

- the range of the MLE coincides with the range of the parameter
- drawbacks associated with finding the maximum of a function
  - verifying that global maximum has been found
  - how sensitive is the estimate to small changes in the data? can slightly different samples produce a vastly different ML estimates?

Method of Maximum Likelihood. Intuition

I have a bag that contains 3 balls. Each ball is either red or blue, but I have no information in addition to this. Thus, the number of blue balls, call it θ, might be 0,1,2,3. I am allowed to choose 4 balls at random from the bag with replacement. We define the random variables X_1, X_2, X_3, X_4 as indicator functions: 1 if ith choosen ball is blue and 0 if not.
Note that X_is are i.i.d. and $X_i \sim Bernoulli(\theta/3)$, the pmf is

$$\theta/3 \quad x = 1$$

$$1 - \theta/3 \;\; x = 0$$

After doing my experiment, I observe the following values for X_is:

$$x_1 = 1, x_2 = 0, x_3 = 1, x_4 = 1$$

I observe 3 blue balls and 1 red ball.

- For each possible value of θ, find the probability of the observed sample
- For which value of θ is the probability of the observed sample is the largest?

Method of Maximum Likelihood. Intuition

We define the random variables X_1, X_2, X_3, X_4 as indicator functions: 1 if ith choosen ball is blue and 0 if not. Note that X_is are i.i.d. and $X_i \sim Bernoulli(\theta/3)$:

$$\theta/3 \quad x = 1$$

$$1 - \theta/3 \quad x = 0$$

After doing my experiment, I observe the following values for X_is:

$$x_1 = 1, x_2 = 0, x_3 = 1, x_4 = 1$$

I observe 3 blue balls and 1 red ball.

- For each possible value of θ, find the probability of the observed sample
- For which value of θ is the probability of the observed sample is the largest?

R.v.s are independent so

$$p(x_1, x_2, x_3, x_4) = p(x_1)p(x_2)p(x_3)p(x_4)$$

$$p(1, 0, 1, 1; \theta) = (\theta/3)^3(1 - \theta/3)$$

where θ is the number of blue balls in the bag.

Method of Maximum Likelihood. Intuition

R.v.s are independent so

$$p(x_1, x_2, x_3, x_4) = p(x_1)p(x_2)p(x_3)p(x_4)$$

$$p(1, 0, 1, 1; \theta) = (\theta/3)^3(1 - \theta/3)$$

where θ is the number of blue balls in the bag. For $\theta = 0, 1, 2, 3$ the probabilities obtained from the joint pmf are $0, 0.0247, 0.0988, 0$.

- Why the probability of observed sample for $\theta = 0$ and $\theta = 3$ is zero?
- For which value of θ is the probability of the observed sample is the largest?

Method of Maximum Likelihood. Intuition

R.v.s are independent so

$$p(x_1, x_2, x_3, x_4) = p(x_1)p(x_2)p(x_3)p(x_4)$$

$$p(1, 0, 1, 1; \theta) = (\theta/3)^3(1 - \theta/3)$$

where θ is the number of blue balls in the bag. For $\theta = 0, 1, 2, 3$ the probabilities obtained from the joint pmf are $0, 0.0247, 0.0988, 0$.

- For which value of θ is the probability of the observed sample is the largest?
- The observed data (1,0,1,1) is most likely to occur for $\theta = 2$.
- $\hat{\theta} = 2$ is the maximum likelihood estimate (MLE) of θ: the true number of blue balls in the bag out of total 3 balls.

Maximum Likelihood Estimator

Suppose that $X_1, ..., X_n$ are i.i.d. random variables with probability distribution $f(x;\theta)$ where θ is a single unknown parameter. Let $x_1, x_2, \ldots, x_n$ be the obserbed values in a random sample size n. Then the likelihood function of the sample is

$$L(\theta) = f(x_1;\theta) \cdot f(x_2;\theta) \cdot \cdots \cdot f(x_n;\theta).$$

- Note that likelihood function is now a funciton of only the unknown parameter θ
- The maximum likelihood estimator (MLE) of θ is the value of θ that maximizes the likelihood function $L(\theta)$
- In the case of a discrete random variable: the likelihood function of the sample $L(\theta)$ is just a probability

$$P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n).$$

- In a discrete case, the MLE is an estimator that maximizes the probability of occurance of the sample values.

MLE. Bernoulli Example

Suppose that an experiment consists of $n = 5$ independent Bernoulli trials each having probability of success p. Let X be the total number of successes in the trials, so that $X \sim \mathrm{Bin}(5, p)$. If the outcome is $X = 3$, the likelihood

$$
\begin{aligned}
L(p; x) &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \frac{5!}{5!(5-3)!} p^3 (1-p)^{5-3}
\end{aligned}
$$

$$\frac{d \log L(p; x)}{dp} = 3p^2 - 8p^3 + 5p^4$$

$$3p^2 - 8p^3 + 5p^4 = 0$$

$$\hat{p} = 0.6$$

If we observe X=3 successes in n=5 trials, a reasonable estimate of the long-run proportion of successes p is 0.6.

MLE. Poisson Example

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}.$$

For $X_1, X_2, \ldots, X_n$ iid Poisson random variables with have a joint frequency function that is a product of the marginal frequency functions, the log likelihood will thus be:

$$\begin{aligned}\log L(\lambda) &= \sum_{i=1}^{n} (X_i \log \lambda - \lambda - \log X_i!) \\ &= \log \lambda \sum_{i=1}^{n} X_i \log \lambda - n\lambda - \sum_{i=1}^{n} \log X_i!.\end{aligned}$$

We need to find the maximum by finding the derivative and set it to 0:

$$\hat{\lambda} = \bar{X}$$

MLE. Normal Example

Let X be normally distributed with unknown μ and known variance σ^2. The likelihood function of a random sample of size n, say $X_1, X_2, \ldots, X_n$ is

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}.$$

Now log likelihood will thus be

$$\log L(\mu) = -(n/2)\log(2\pi\sigma^2) - (2\sigma^2)^{-1}\sum_{i=1}^{n}(x_i-\mu)^2$$

with the derivative

$$\frac{d\log L(\mu)}{d\mu} = (\sigma^2)^{-1}\sum_{i=1}^{n}(x_i-\mu).$$

Equating derivative to 0 and solving for μ yields

$$\hat{\mu} = \frac{\sum_{i=1}^{2} X_i}{n} = \overline{X}.$$

Conclusion: The sample mean is the maximum likelihood estimator of μ. Notice that this is identical to the moment estimator.

MLE. Exponential Example

Let X be exponentially distributed with parameter λ. The likelihood function of a random sample of size n, say $X_1, X_2, X_3, \ldots, X_n$, is

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}.$$

The log likelihood is

$$\log L(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} x_i \tag{1}$$

and its derivative is

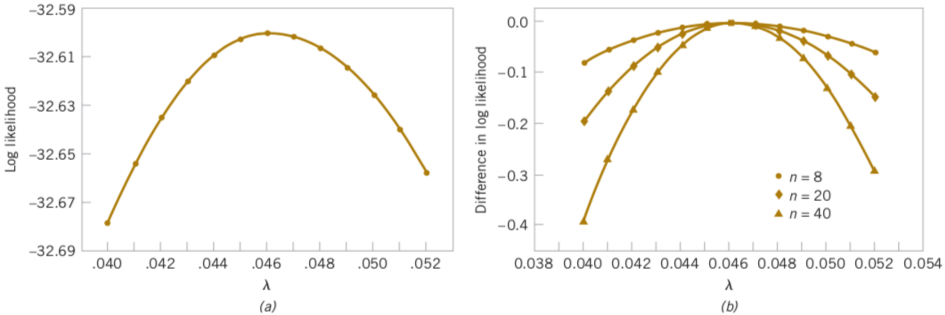
$$\frac{d \log L(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i.$$

Equating derivative to 0 and solving for λ yields

$$\hat{\lambda} = n / \sum_{i=1}^{n} X_i = 1/\overline{X}.$$

Conclusion: The reciprocal of the sample mean is the maximum likelihood estimator of λ. Notice that this is identical to the moment estimator.

Log likelihood for the exponential distribution. Example



Properties of MLE

Under very general and not restrictive conditions when the sample size n is large and if $\hat{\Theta}$ is the MLE of the parameter θ.

(1) $\hat{\Theta}$ is an approximately unbiased estimator for θ: $E(\hat{\Theta}) \simeq \theta$.

(2) The variance of $\hat{\Theta}$ is nearly as small as the variance that could be obtained with any other estimator.

(3) $\hat{\Theta}$ has an approximate normal distribution.

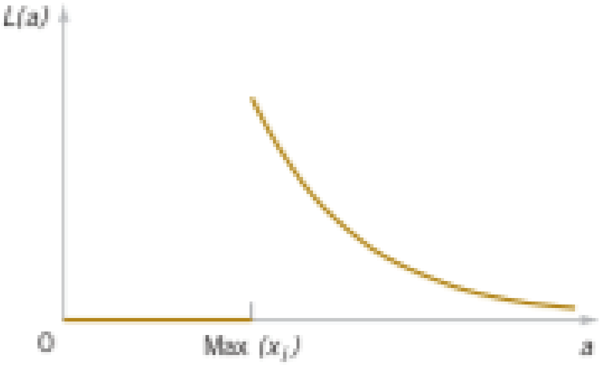
Complications in using Maximum Likelihood Estimation

- It is not always easy to maximize the likelihood function because the equations obtained from $dL(\theta)/d\theta$ may be difficult to solve. Furthermore, it may not always be possible to use calculus methods directly to determine maximum $L(\theta)$.
- Uniform distribution MLE. Let X be uniformly distributed on the interval $[0, a]$. Because the density function is $f(x) = 1/a$ for $0 \leq x \leq a$ and zero otherwise, the likelihood function of a random sample of size n is

  $$L(a) = \prod_{i=1}^{n} \frac{1}{a} = \frac{1}{a^n},$$

  for $0 \leq x_1 \leq a, 0 \leq x_2 \leq a, \ldots, 0 \leq x_n \leq a$.
  We could maximize $L(a)$ by setting $\hat{a}$ equal to the smallest value it could logically take on, which is $\max(x_i)$. This is because $a \geq x_1, a \geq x_2, ...$ for all x, so I can write $a \geq max(x_1, x_2, ...)$

Uniform MLE



MLE. Gamma Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from the gamma distribution. The log likelihood function is

$$\log L(r, \lambda) = \log \left(\prod_{i=1}^{n} \frac{\lambda^r x_i^{r-1} e^{-\lambda x_i}}{\Gamma(r)} \right)$$

$$= nr \log(\lambda) + (r-1) \sum_{i=1}^{n} \log(x_i) - n \log[\Gamma(r)] - \lambda \sum_{i=1}^{n} x_i$$

with partial derivatives

$$\frac{\partial \log L(r, \lambda)}{\partial r} = n \log(\lambda) + \sum_{i=1}^{n} \log(x_i) - n \frac{\Gamma'(r)}{\Gamma(r)}, \qquad \frac{\partial \log L(r, \lambda)}{\partial \lambda} = \frac{nr}{\lambda} - \sum_{i=1}^{n} x_i.$$

By equating these to 0 we get the equations that must be solved to find the maximum likelihood estimators r and λ:

$$\hat{\lambda} = \frac{\hat{r}}{\bar{x}}, \qquad n \log(\hat{\lambda}) + \sum_{i=1}^{n} \log(x_i) = n \frac{\Gamma'(\hat{r})}{\Gamma(\hat{r})}. \tag{2}$$

There is no closed form solution to these equations.