Andrew Gard - equitable.equations@gmail.com

Maximum Likelihood Estimation - Example

Suppose we have a set of independent observations $x_1, x_2, \ldots, x_n$ of some phenomenon with an exponential distribution, like the waiting time between orders at a pizza place at dinner time.

Suppose we have a set of independent observations $x_1, x_2, \ldots, x_n$ of some phenomenon with an exponential distribution, like the waiting time between orders at a pizza place at dinner time. Recall that a such random variable has pdf

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \qquad 0 < x < \infty.$$

Suppose we have a set of independent observations $x_1, x_2, \ldots, x_n$ of some phenomenon with an exponential distribution, like the waiting time between orders at a pizza place at dinner time. Recall that a such random variable has pdf

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \qquad 0 < x < \infty.$$

We'd like to use the sample data to obtain the best possible estimate $\hat{\theta}$ of the unknown parameter θ.

Suppose we have a set of independent observations $x_1, x_2, \ldots, x_n$ of some phenomenon with an exponential distribution, like the waiting time between orders at a pizza place at dinner time. Recall that a such random variable has pdf

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \qquad 0 < x < \infty.$$

We'd like to use the sample data to obtain the best possible estimate $\hat{\theta}$ of the unknown parameter θ.

Maximum likelihood estimation is a natural and widely-applicable method for doing just this.

Suppose we have a set of independent observations $x_1, x_2, \ldots, x_n$ of some phenomenon with an exponential distribution, like the waiting time between orders at a pizza place at dinner time. Recall that a such random variable has pdf

$$f(x) = \frac{1}{\theta} e^{-x/\theta}, \qquad 0 < x < \infty.$$

We'd like to use the sample data to obtain the best possible estimate $\hat{\theta}$ of the unknown parameter θ.

Maximum likelihood estimation is a natural and widely-applicable method for doing just this. The basic idea is to write the value of the joint pdf of the random variables $X_1, X_2, \ldots, X_n$ as a function of the parameter θ, then maximize that function.

When viewed as a function of an unknown parameter, the joint pdf of $X_1, X_2, \ldots, X_n$ is known as the **likelihood function**.

$$L(\theta) \quad = \quad f(x_1, x_2, \ldots, x_n; \theta)$$

When viewed as a function of an unknown parameter, the joint pdf of $X_1, X_2, \ldots, X_n$ is known as the **likelihood function**.

$$
\begin{aligned}
L(\theta) &= f(x_1, x_2, \ldots, x_n; \theta) \\
&= \prod_{k=1}^{n} f(x_i; \theta)
\end{aligned}
$$

when the random variables $X_1, X_2, \ldots, X_n$ are identical and independent, as is usually the case.

When viewed as a function of an unknown parameter, the joint pdf of $X_1, X_2, \ldots, X_n$ is known as the **likelihood function**.

$$
\begin{aligned}
L(\theta) &= f(x_1, x_2, \ldots, x_n; \theta) \\
&= \prod_{k=1}^{n} f(x_i; \theta)
\end{aligned}
$$

when the random variables $X_1, X_2, \ldots, X_n$ are identical and independent, as is usually the case. In our example, where each has an exponential distribution,

$$
L(\theta) = \left(\frac{1}{\theta} e^{-x_1/\theta}\right) \left(\frac{1}{\theta} e^{-x_2/\theta}\right) \cdots \left(\frac{1}{\theta} e^{-x_n/\theta}\right)
$$

When viewed as a function of an unknown parameter, the joint pdf of $X_1, X_2, \ldots, X_n$ is known as the **likelihood function**.

$$\begin{aligned} L(\theta) &= f(x_1, x_2, \ldots, x_n; \theta) \\ &= \prod_{k=1}^{n} f(x_i; \theta) \end{aligned}$$

when the random variables $X_1, X_2, \ldots, X_n$ are identical and independent, as is usually the case. In our example, where each has an exponential distribution,

$$\begin{aligned} L(\theta) &= \left(\frac{1}{\theta} e^{-x_1/\theta}\right) \left(\frac{1}{\theta} e^{-x_2/\theta}\right) \cdots \left(\frac{1}{\theta} e^{-x_n/\theta}\right) \\ &= \frac{1}{\theta^n} \exp\left(\frac{-\sum_{k=1}^{n} x_k}{\theta}\right) \end{aligned}$$

This is the function we wish to maximize.

In theory, the value $\theta = \hat{\theta}$ that maximizes $L(\theta)$ can be computed by differentiating with respect to θ and setting the result equal to zero.

$$L(\theta) \;=\; \frac{1}{\theta^n} \exp\left(\frac{-\sum_{k=1}^{n} x_k}{\theta}\right)$$

In theory, the value $\theta = \hat{\theta}$ that maximizes $L(\theta)$ can be computed by differentiating with respect to θ and setting the result equal to zero.

$$L(\theta) \;=\; \frac{1}{\theta^n} \exp\left(\frac{-\sum_{k=1}^{n} x_k}{\theta} \right)$$

But because L is built from products and powers, the derivative will be much easier if we apply a logarithm first.

In theory, the value $\theta = \hat{\theta}$ that maximizes $L(\theta)$ can be computed by differentiating with respect to θ and setting the result equal to zero.

$$L(\theta) \;=\; \frac{1}{\theta^n} \exp\left(\frac{-\sum_{k=1}^{n} x_k}{\theta}\right)$$

But because L is built from products and powers, the derivative will be much easier if we apply a logarithm first. The result is called the **log likelihood function**.

$$\ln L(\theta) \;=\; -n \ln \theta - \frac{1}{\theta} \sum_{k=1}^{n} x_k$$

Importantly, a function and its logarithm always have maxima at the same values. We can apply the first derivative test to this easier function instead.

We now maximize $\ln L(\theta)$ over all possible values of θ, namely $0 < \theta < \infty$.

We now maximize $\ln L(\theta)$ over all possible values of θ, namely $0 < \theta < \infty$.

$$\frac{d}{d\theta} \ln L(\theta) \quad = \quad \frac{d}{d\theta}\left(-n \ln \theta - \frac{1}{\theta} \sum_{k=1}^{n} x_k\right)$$

We now maximize $\ln L(\theta)$ over all possible values of θ, namely $0 < \theta < \infty$.

$$\begin{aligned}
\frac{d}{d\theta} \ln L(\theta) &= \frac{d}{d\theta}\left(-n \ln \theta - \frac{1}{\theta}\sum_{k=1}^{n} x_k\right) \\
0 &= -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{k=1}^{n} x_k
\end{aligned}$$

We now maximize $\ln L(\theta)$ over all possible values of θ, namely $0 < \theta < \infty$.

$$
\begin{aligned}
\frac{d}{d\theta} \ln L(\theta) &= \frac{d}{d\theta}\left(-n \ln \theta - \frac{1}{\theta} \sum_{k=1}^{n} x_k\right) \\
0 &= -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{k=1}^{n} x_k \\
\theta &= \frac{1}{n} \sum_{k=1}^{n} x_k = \overline{x}
\end{aligned}
$$

There is only one critical value for the log likelihood function: the sample mean $\overline{x}$.

The first derivative test confirms that $\hat{\theta} = \overline{x}$ is in fact a maximum of $\ln L(\theta)$, as opposed to a minimum or inflection point. The derivative

$$\frac{d}{d\theta} \ln L(\theta) = -\frac{n}{\theta} + \frac{n}{\theta^2}\overline{x}$$

is positive when $\theta < \overline{x}$ and negative when $\theta > \overline{x}$.

The first derivative test confirms that $\hat{\theta} = \overline{x}$ is in fact a maximum of $\ln L(\theta)$, as opposed to a minimum or inflection point. The derivative

$$\frac{d}{d\theta} \ln L(\theta) = -\frac{n}{\theta} + \frac{n}{\theta^2}\overline{x}$$

is positive when $\theta < \overline{x}$ and negative when $\theta > \overline{x}$.

The value $\hat{\theta} = \overline{x}$ makes sense as an estimator since the parameter θ of an exponentially-distributed random variable corresponds to its mean.

Example. Times between orders to a pizza place at dinner time follow an exponential distribution. An employee records the times at which orders are placed, in seconds after opening, for one hour, obtaining the following results.

10	72	114	223	483	494	605	840
842	849	1031	1071	1182	1234	1355	1418
1494	1608	1665	1762	1796	1863	1908	1912
2046	2151	2268	2294	2309	2504	2555	2597
2615	2668	2683	2685	2797	2879	2895	2983
2988	2989	3127	3347	3432	3444	3513	3588

Example. Times between orders to a pizza place at dinner time follow an exponential distribution. An employee records the times at which orders are placed, in seconds after opening, for one hour, obtaining the following results.

10	72	114	223	483	494	605	840
842	849	1031	1071	1182	1234	1355	1418
1494	1608	1665	1762	1796	1863	1908	1912
2046	2151	2268	2294	2309	2504	2555	2597
2615	2668	2683	2685	2797	2879	2895	2983
2988	2989	3127	3347	3432	3444	3513	3588

The maximum likelihood estimator for θ is $\hat{\theta} = \frac{1}{n}\sum x_k$, where x_k is the elapsed time between the k^{th} call and the one previous.

Example. Times between orders to a pizza place at dinner time follow an exponential distribution. An employee records the times at which orders are placed, in seconds after opening, for one hour, obtaining the following results.

10	72	114	223	483	494	605	840
842	849	1031	1071	1182	1234	1355	1418
1494	1608	1665	1762	1796	1863	1908	1912
2046	2151	2268	2294	2309	2504	2555	2597
2615	2668	2683	2685	2797	2879	2895	2983
2988	2989	3127	3347	3432	3444	3513	3588

The maximum likelihood estimator for θ is $\hat{\theta} = \frac{1}{n}\sum x_k$, where x_k is the elapsed time between the k^{th} call and the one previous. Thus $\sum x_k = 3588$ and $\hat{\theta} = \frac{1}{n}\sum x_k = 3588/48 = 74.75$ seconds.

A graph of the likelihood function confirms that the maximum value really does lie at $\theta = 74.75$.