Background Notes: Lecture 11: Maximum likelihood I

Example Mining disasters
Let X be number of mining disasters (five or more killed) per year in northern England between 1921 and 1925. Observed values[1] of X were 0, 2, 1, 0, 2.

Model disasters as a Poisson process in time, occurring randomly and independently. Number of disasters in any year is $X \sim \text{Poisson}(\mu)$ with $\mathrm{E}[X] = \mu$. so

$$\text{pr}\{X = x; \mu\} = \frac{\mu^x e^{-\mu}}{x!}, \quad x = 0, 1, 2, \ldots.$$

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_5)$ and $\boldsymbol{x} = (0, 2, 1, 0, 2)$. Then, assuming independence, the probability of observing $\boldsymbol{x}$ is

$$\text{pr}\{\boldsymbol{X} = \boldsymbol{x}; \mu\} = \frac{\mu^0 e^{-\mu}}{0!} \times \frac{\mu^2 e^{-\mu}}{2!} \times \frac{\mu^1 e^{-\mu}}{1!} \times \frac{\mu^0 e^{-\mu}}{0!} \times \frac{\mu^2 e^{-\mu}}{2!} = \frac{\mu^{0+2+1+0+2} e^{-5\mu}}{0!\,2!\,1!\,0!\,2!} = \frac{\mu^5 e^{-5\mu}}{4}.$$

For given known μ this gives the joint probability function of observing $\boldsymbol{X} = \boldsymbol{x}$.

If we treat this as a function of μ for fixed $\boldsymbol{x}$, this gives the *likelihood function* of μ given $\boldsymbol{x}$,

$$L(\mu; \boldsymbol{x}) = \frac{\mu^5 e^{-5\mu}}{4}.$$

This is plotted in figure 1. Intuitively we would estimate μ by the value which makes the observed data $\boldsymbol{x}$ "most likely" to occur. The maximum likelihood estimate for μ is here $\mu = 1$.

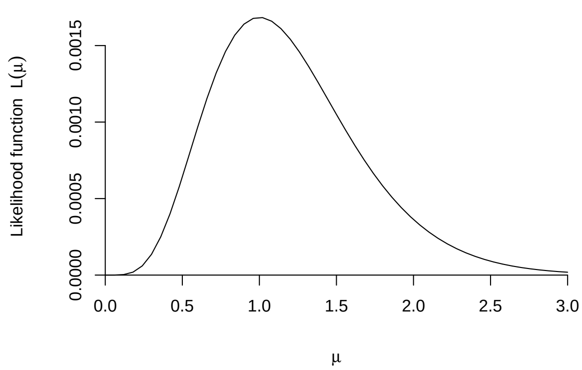

Figure 1: likelihood function $L(\mu; \boldsymbol{x} = (0, 2, 1, 0, 2))$ for mining data.

Example Binomial distribution
Suppose a coin has probability of heads θ. If we toss the coin n times the probability of observing $X = x$ heads is

$$\text{pr}\{X = x;\ \theta\} = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \qquad x = 0, 1, \ldots, n.$$

This is plotted in figure 2(left) in the case $\theta = 0.3$.

[1]Source: Durham Mining Museum web-site `http://www.dmm.org.uk/`

Date of disaster	5 Sep 1922	27 Nov 1922	24 Feb 1923	30 Mar 1925	9 Aug 1925
Killed	39	6	8	38	5

Suppose now that in 10 tosses of the coin we observe 6 heads. This occurs with probability

$$\text{pr}\{X = 6;\ \theta\} = \binom{10}{6}\theta^6(1-\theta)^4 = 210\theta^6(1-\theta)^4.$$

For example, if $\theta = 0.3$, this gives $\text{pr}\{X = 6\} = 0.0368$.

Now consider evaluating this probability for the case $x = 6$ for different values of θ. Treated as a function of θ this gives the likelihood function when $x = 6$:

$$L(\theta;\ x = 6) = \binom{10}{6}\theta^6(1-\theta)^4 = 210\theta^6(1-\theta)^4.$$

This is plotted in figure 2(right). It can be seen that $x = 6$ is not likely to occur if θ is very small or very large. The most likely case is where $\widehat{\theta} = 0.6$. This is the value of θ which maximises the likelihood $L(\theta; x = 6)$.

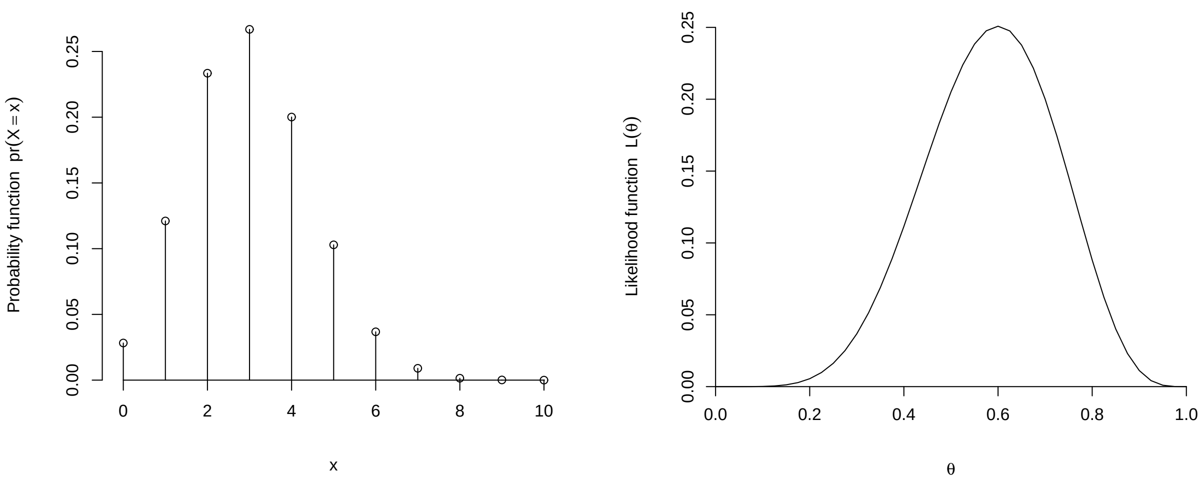

Figure 2: (left) probability function $\text{pr}\{X = x;\ \theta = 0.3\}$; (right) likelihood $L(\theta;\ x = 6)$.

Example Exponential distribution
Consider four observations $\boldsymbol{x} = (0.18, 3.36, 0.95, 7.08)$ from an exponential distribution[2] with probability density function $f_X(x; \lambda) = \lambda e^{-\lambda x}$ for $x > 0$. The joint probability density function is

$$f_{\boldsymbol{X}}(\boldsymbol{x}; \lambda) = \lambda e^{-0.18\lambda} \times \lambda e^{-3.36\lambda} \times \lambda e^{-0.95\lambda} \times \lambda e^{-7.08\lambda} = \lambda^4 e^{-11.57\lambda}.$$

The likelihood function is $L(\lambda; \boldsymbol{x}) = \lambda^4 e^{-11.57\lambda}$ and is plotted in figure 3(left). The log-likelihood function is $l(\lambda; \boldsymbol{x}) = 4\log(\lambda) - 11.57\lambda$ and is plotted in figure 3(right).

Further reading for lectures 12 and 13

Rice, J.A. (1995) *Mathematical Statistics and Data Analysis (2nd edition)*, sections 4.2, 8.5.

Hogg, R.V., McKean, J.W. and Craig, A.T. (2005) *Introduction to Mathematical Statistics (6th edition)*, sections 4.1, 6.1.

Larsen, R.J. and Marx, M.L. (2010) *An Introduction to Mathematical Statistics and its Applications (5th edition)*, sections 5.1, 5.2, 5.4.

Miller, I. and Miller, M. (2004) *John E. Freund's Mathematical Statistics with Applications*, sections 8.8, 10.2, 10.3, 10.8.

[2]Data simulated from an exponential($\lambda = 1/3$) distribution.

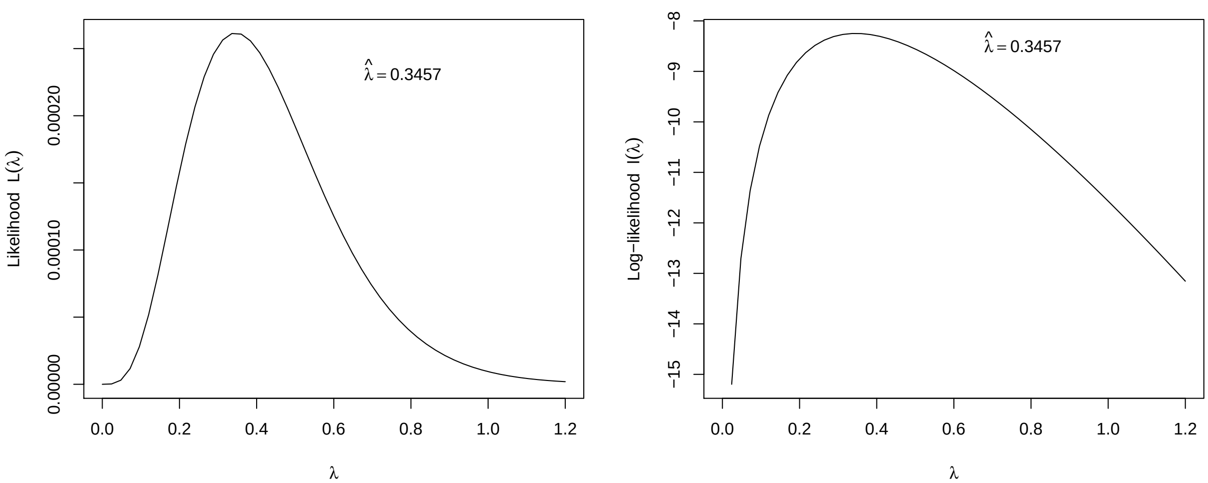

Figure 3: (left) likelihood function $L(\lambda; \boldsymbol{x})$; (right) log-likelihood function $l(\lambda; \boldsymbol{x})$.