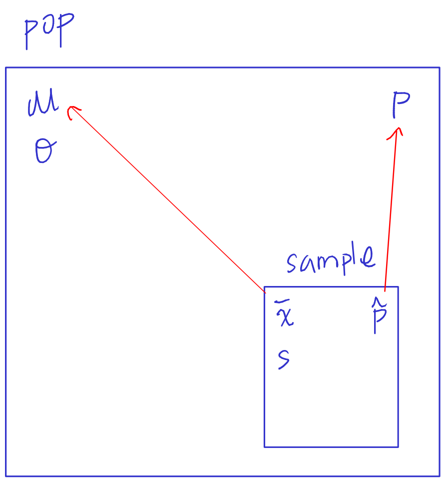
Maximum Liklihood Estimator

David Armstrong

UCI



$$\sum_{i=1}^{n} c = nc$$

$$\sum_{i=1}^{n} (1 - x_i) = n - \sum_{i=1} x_i$$

$$\prod_{i=1}^{n} p = p^n$$

$$\prod_{i=1}^{n} p^{x_i} = p^{\sum_{i=1}^{n} x_i}$$

$$\ln(xy) = \ln x + \ln y$$

$$\ln\left(\prod_{i=1} x_i\right) = \sum_{i=1}^{n} \ln(x_i)$$

$$\ln\left(\frac{x}{y}\right) = \ln x - \ln y$$

$$\ln\left(\frac{x}{yz}\right) = \ln x - \ln y - \ln z$$

$$\ln(x^n) = n \ln x$$

$$\frac{d}{dp} \ln p = \frac{1}{p}$$

$$\frac{d}{dp} \ln(1-p) = -\frac{1}{1-p}$$

$$\frac{d}{dp} p^n = n p^{n-1}$$

$$\frac{d}{dp} p = 1$$

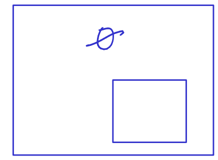
$$\frac{d}{dp} c = 0 \qquad \frac{d}{dp} x_i = 0$$

Estimating Parameters

Suppose $X_1, \ldots, X_n$ is a random sample from a continuous pdf $f_X(x)$ whose unknown parameter is θ.

The question is, how should we use the data to approximate θ?

data



Method of Maximum Likelihood

The likelihood function $L(\theta)$ is the product of the pdf $f_X(x;\theta)$ evaluated at the n data points. That is,

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i;\theta)$$

L is a function of θ; it should not be considered a function of the $x_i's$.

MLE

Let $X_1, \ldots, X_n$ is a random sample from a continuous pdf $f_X(x;\theta)$ and let $L(\theta)$ be the corresponding likelihood function. Suppose $L(\hat{\theta}) \geq L(\theta)$ for all possible values of θ. The $\hat{\theta}$ is called the *maximum likelihood estimate (MLE) for* θ

Note

Finding $\hat{\theta}$ that maximizes a likelihood function is basically an exercise in Calculus. Since $\ln L(\theta)$ increases with $L(\theta)$, the same $\hat{\theta}$ that maximizes the ln of the likelihood function maximizes the likelihood function.

$\log_e = \ln$ $\qquad\qquad$ $\log$

Maximum Likelihood Steps

1. Find $L(\theta)$ Product of the pdfs
2. Find $\ln L(\theta)$ log likelihood
3. Calculate the Score Function $= \dfrac{\partial \ln L(\theta)}{\partial \theta}$ 1st Derivative
4. Find the Score equation by setting the Score Function $\equiv 0$
5. Solve for the parameter
6. Check that the second derivative of the $\ln L(\theta)$ is negative, that is $\dfrac{\partial^2 \ln L(\theta)}{\partial \theta^2} < 0$
7. Check the support

$x \sim \text{Bernoulli}(p) \qquad f(x) = p^{x}(1-p)^{1-x}$

Ex: Suppose $x_1, \ldots, x_n$ is a set of n observations representing a Bernoulli probability model.

- Find the MLE of p.

$n = 10$

- Assume you see the data $\{0, 1, 0, 0, 1, 1, 0, 0, 1, 0\}$ from Bernoulli trials Use the Maximum Liklihood Estimator to estimate the true population parameter p.

$$L\theta = \prod_{i=1}^{n} f_x(x_i ; \theta = p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x}$$

$$= (p^{x_1}(1-p)^{1-x_1})(p^{x_2}(1-p)^{1-x_2})\ldots(p^{x_n}(1-p)^{1-x_n})$$

$$= p^{\sum_{i=1}^{n} x_i}(1-p)^{\sum_{i=1}^{n}(1-x_i)}$$

$$= p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

$$\ln L(\theta) = \ell(p) = \ln\left(p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}\right)$$

$$= \ln p^{\sum_{i=1}^{n} x_i} + \ln(1-p)^{n-\sum_{i=1}^{n} x_i}$$

$$= \sum_{i=1}^{n} x_i \ln p + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p)$$

$$\frac{d}{dp}\ell(p) = \frac{d}{dp}\left(\sum_{i=1}^{n} \ln p + \left(n-\sum_{i=1}^{n} x_i\right)\ln(1-p)\right)$$

$$= \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n-\sum_{i=1}^{n} x_i}{1-p}$$

Score $= 0$ max/min solve for p

$$0 = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n-\sum_{i=1}^{n} x_i\right)}{1-p}$$

$$\frac{\sum_{i=1}^{n} x_i}{p} = \frac{\left(n-\sum_{i=1}^{n} x_i\right)}{1-p}$$

$$\sum_{i=1}^{n} x_i - p\sum_{i=1}^{n} x_i = np - p\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i = np \qquad \text{MLE ?}$$

$$p = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \text{concave down}$$

$$\frac{d^2}{d^2p}\ell(p) = \frac{dp}{dp}\left(\frac{\sum_{i=1}^{n} x_i}{p} - \frac{n-\sum_{i=1}^{n} x_i}{1-p}\right) < 0$$

$$= -\frac{\sum_{i=1}^{n} x_i}{n} - \frac{n-\sum_{i=1}^{n} x_i}{(1-p)^2} < 0$$

$$= -\frac{+}{+} - \frac{0/+}{+}$$

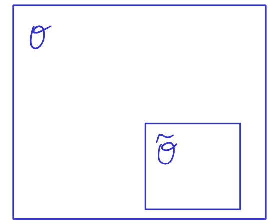
$$= \text{neg} - \text{neg}$$

Thus, $\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$

is the maximum likelihood estimate.

$$\hat{p} = \frac{4}{10} = 0.4$$

independent identitly distributed



EX: Suppose $x_1, \ldots, x_n$ are i.i.d. random variables with density function:

$$f_X(x;\sigma) = \frac{1}{2\sigma} e^{-\dfrac{|x|}{\sigma}}$$

Find the maximum liklihood estimate of σ

$$L(\theta) = f_x(x_i \mid \Theta = \theta) = \prod_{i=1}^{n} \frac{1}{2\theta} e^{-\frac{|x_i|}{\theta}}$$

$$\left(\frac{1}{2\theta} e^{\frac{-|x_1|}{\theta}}\right)\left(\frac{1}{2\theta} e^{\frac{-|x_2|}{\theta}}\right)\left(\frac{1}{2\theta} e^{\frac{-|x_3|}{\theta}}\right)\cdots\left(\frac{1}{2\theta} e^{\frac{-|x_n|}{\theta}}\right)$$

$$= \left(\frac{1}{2\theta}\right)^n e^{-\frac{\sum_{i=1}^{n}|x_i|}{\theta}}$$

$$\ell(\theta) = \ln\left(\left(\frac{1}{2\theta}\right)^n e^{-\frac{\sum_{i=1}^{n}|x_i|}{\theta}}\right)$$

$$= n\ln(1) - n\ln(2) - n\ln(\theta) + \ln\left(e^{-\frac{\sum_{i=1}^{n}|x_i|}{\theta}}\right)$$

$$= 0 - n\ln(2) - n\ln(\theta) - \frac{\sum_{i=1}^{n}|x_i|}{\theta}$$

$$\text{score} = \frac{d}{d\theta}\ell(\theta) = \frac{d}{d\theta}\left(-n\ln 2 - n\ln(\theta) - \frac{\sum_{i=1}^{n}|x_i|}{\theta}\right)$$

$$= -\frac{n}{\theta} + \frac{\sum_{i=1}^{n}|x_i|}{\theta^2}$$

Max or min when score $= 0$

$$-\frac{n}{\theta} + \frac{\sum_{i=1}^{n}|x_i|}{\theta^2} = 0$$

$$\frac{\sum_{i=1}^{n}|x_i|}{\theta^2} = \frac{n}{\theta}$$

$$\theta\sum_{i=1}^{n}|x_i| = n\theta^2$$

$$\sum_{i=1}^{n}|x_i| = n\theta$$

$$\boxed{\frac{\sum_{i=1}^{n}|x_i|}{n} = \theta}$$

we believe this is the maximum likehood estimator

maximums occur when second derivative is concave down

$$\frac{d}{d\theta}\text{score} < 0$$

$$\frac{d}{d\theta}\left(-\frac{n}{\theta} + \frac{\sum_{i=1}^{n}|x_i|}{\theta^2}\right) < 0$$

$$\frac{n}{\theta^2} - \frac{2\sum_{i=1}^{n}|x_i|}{\theta^3} < 0$$

$$\frac{n\theta}{\theta^3} - \frac{2\sum_{i=1}^{n}|x_i|}{\theta^3} < 0$$

$$\frac{n\theta - 2\sum_{i=1}^{n}|x_i|}{\theta^3} < 0$$

$$\frac{n\left(\frac{\sum_{i=1}^{n}|x_i|}{n}\right) - 2\sum_{i=1}^{n}|x_i|}{\theta^3} < 0$$

$$\frac{\sum_{i=1}^{n}|x_i| - 2\sum_{i=1}^{n}|x_i|}{\theta^3} < 0$$

$$-\frac{\sum_{i=1}^{n}|x_i|}{\theta^3} < 0$$

Yes!

Thus $\hat{\theta} = \frac{\sum_{i=1}^{n}|x_i|}{n}$ is the maximum likehood estimator