# Class 1 : Poisson data

$Y_i \equiv$ count of coal mining disaster in year $i$.

$Y_i \in \{0,1,2,\ldots\}$ $\quad i=1,2,\ldots,n$

**Poisson Model** $\quad Y_1,\ldots,Y_n \mid \lambda \overset{iid}{\sim} \text{Poisson}(\lambda)$

$\Rightarrow \Pr(y_i \mid \lambda) = \dfrac{\lambda^{y_i} e^{-\lambda}}{y_i!}$ $\quad \lambda > 0$ occurrence rate

$\Rightarrow E(y_i \mid \lambda) = V(y_i \mid \lambda) = \lambda$

Example:

| $k$ | $\Pr(y_i=k \mid \lambda=1)$ | $\Pr(y_i=k \mid \lambda=2)$ |
|---|---|---|
| 0 | 0.368 | 0.135 |
| 1 | 0.368 | 0.271 |
| 2 | 0.184 | 0.271 |
| 3 | 0.061 | 0.181 |
| SUM | 0.981 | 0.857 |

# Maximum Likelihood Estimation

$$\hat{\lambda}_{MLE} \equiv \arg\max_{\lambda > 0} \; p(y_1, \dots, y_n | \lambda)$$

$$p(y_1, \dots, y_n | \lambda) = \prod_{i=1}^{n} \lambda^{y_i} e^{-\lambda} / y_i! = \frac{\lambda^{S_n} e^{-n\lambda}}{\prod_{i=1}^{n} y_i!}$$

$$S_n = \sum_{i=1}^{n} y_i$$

As a function of $\lambda$, $p(y_1 \dots y_n | \lambda)$ is usually rewritten as

$$L(\lambda | y_1 \dots y_n) = c \, \lambda^{S_n} e^{-n\lambda}$$

where $c^{-1} = \prod_{i=1}^{n} y_i!$ is constant.

The Likelihood function $L(\lambda | y_1 \dots y_n)$ is a function of the data only through $(S_n, n)$.

$\Rightarrow S_n$ is a sufficient statistic for $\lambda$.

$\hat{\lambda}_{MLE}$ is obtained by finding the maximum of $L(\lambda|y_1 \dots y_n)$ or its log:

$$\ell(\lambda) \equiv \log L(\lambda|y_1 \dots y_n) = \log c + S_n \log \lambda - n\lambda$$

$$\Rightarrow \frac{\partial \ell(\lambda)}{\partial \lambda} = \frac{S_n}{\lambda} - n \text{ and } \frac{\partial \ell(\lambda)}{\partial \lambda} = 0 \Leftrightarrow$$

$$\Leftrightarrow \hat{\lambda}_{MLE} = \frac{S_n}{n} = \frac{\sum_{i=1}^{n} y_i}{n} \text{ (sample mean)}$$

Is it a maximum?

$$\frac{\partial^2 \ell(\lambda)}{\partial \lambda^2} = -\frac{S_n}{\lambda^2} < 0 \Rightarrow$$

$\Rightarrow \hat{\lambda}_{MLE}$ is a point of maximum.

Also,

$$E(\hat{\lambda}_{MLE}|\lambda) = \frac{1}{n} \sum_{i=1}^{n} E(y_i|\lambda) = \lambda$$

$\hat{\lambda}_{MLE}$ is unbiased.

## Consistancy

$$V(\hat{\lambda}_{MLE}|\lambda) = \frac{1}{n^2}\sum_{i=1}^{n} V(y_i|\lambda) = \frac{\lambda}{n}$$

$$\xrightarrow{n\to\infty} 0$$

$$\Rightarrow \quad \hat{\lambda}_{MLE} \xrightarrow{n\to\infty} \lambda$$

$\hat{\lambda}_{MLE}$ is a consistent estimator.

## Efficiency

An estimator is efficient if its variance equals the Cramér-Rao lower boundary:

$$CRLB = \frac{1}{I(\lambda)} \qquad I(\lambda) \equiv \text{Expected Fisher Information}$$

$$I(\lambda) = n E_{y|\lambda}\left\{\left(\frac{\partial \log L(\lambda|y)}{\partial\lambda}\right)^2\right\}$$

$$\Rightarrow \; I(\lambda) = n\, E_{y|\lambda}\left\{\left(\frac{y}{\lambda}-1\right)^2\right\}$$

$$= n\, E_{y|\lambda}\left\{\frac{y^2}{\lambda^2} - 2\frac{y}{\lambda} + 1\right\}$$

$$= n\left[\frac{1}{\lambda^2}E_{y|\lambda}(y^2) - \frac{2}{\lambda}E_{y|\lambda}(y) + 1\right]$$

$$= n\left[\frac{\lambda+1}{\lambda} - 1\right] = n/\lambda = \frac{1}{V(\hat{\lambda}_{MLE}|\lambda)}$$

$$E_{y|\lambda}(y^2) = V_{y|\lambda}(y) + \left[E_{y|\lambda}(y)\right]^2 = \lambda + \lambda^2 = \lambda(\lambda+1).$$

Efficiency can be computed as

$$e(\hat{\lambda}_{MLE}) = \frac{\left\{I(\hat{\lambda}_{MLE})\right\}^{-1}}{V(\hat{\lambda}_{MLE}|\lambda)}$$