Maximum Likelihood Estimation

Suppose it is known that $X_1, X_2, ..., X_n$ are independent and identically distributed (iid) Poisson random variables with parameter θ, but the value of θ is unknown. We can estimate θ by finding the value of θ which is most compatible with the observed sample, using $P_\theta(\mathbf{X} = \mathbf{x})$ as the measure of compatibility. Viewing $P_\theta(\mathbf{X} = \mathbf{x})$ as a function of θ, it is called the likelihood function, and we seek the value of θ which maximizes it. This value of θ is called the maximum likelihood estimate (mle), and the associated estimator is the maximum likelihood estimator (MLE).

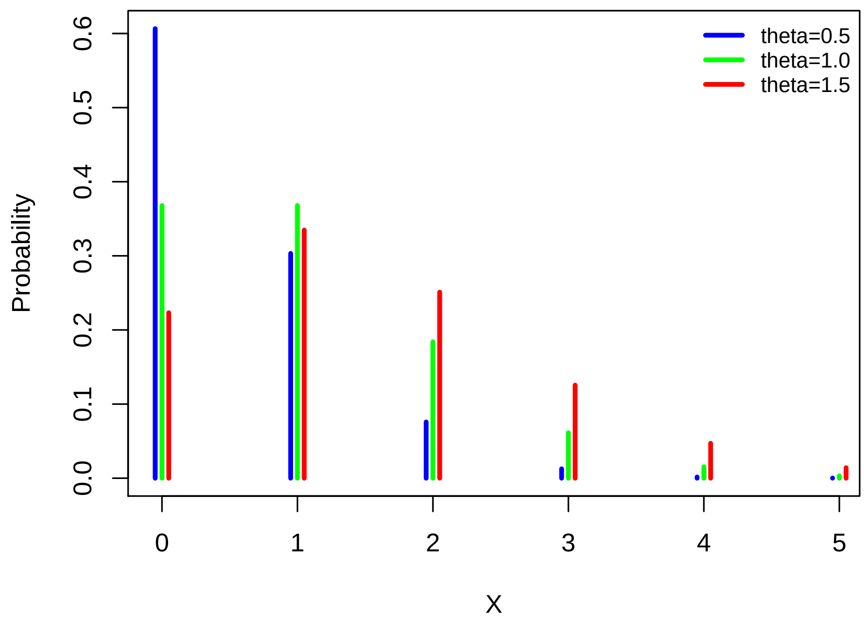

Figure 1: Poisson pmf with different values of θ

Given the observed sample

$$x_1 = 0, x_2 = 1, x_3 = 1, x_4 = 3, x_5 = 0,$$

we have

$$P_{\theta=0.5}(\mathbf{X} = \mathbf{x}) \doteq 0.00043,$$

$$P_{\theta=1.0}(\mathbf{X} = \mathbf{x}) \doteq 0.00112,$$

and

$$P_{\theta=1.5}(\mathbf{X} = \mathbf{x}) \doteq 0.00070.$$

So among the three values of θ considered, putting $\theta = 1.0$ maximizes the likelihood. But what about other values θ could be?

Letting $L(\theta) = P_\theta(\mathbf{X} = \mathbf{x})$, we have

$$L(\theta) = \prod_{i=1}^{n} P_\theta(X_i = x_i)$$

$$= \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!}$$

$$= \frac{\theta^{\Sigma x_i} e^{-n\theta}}{\prod_{i=1}^{n} x_i!}.$$

Solving $L'(\theta) = 0$ and determining that the solution is a maximizing value, it can be concluded that the maximum likelihood estimate is $\bar{x}$, the sample mean.

It is often easier to work with the log-likelihood, $\ell(\theta) = \log L(\theta)$. The value of θ which maximizes $\ell(\theta)$ is the same as the value of θ which maximizes $L(\theta)$. Letting $c = \prod_{i=1}^{n} x_i!$, for the Poisson case we have

$$\ell(\theta) = \sum x_i \log(\theta) - n\theta - \log(c).$$

So

$$\ell'(\theta) = \frac{\Sigma x_i}{\theta} - n,$$

and $\ell'(\theta) = 0$ implies $\theta = \frac{\Sigma x_i}{n} = \bar{x}$. The second derivative test can be used to determine that $\bar{x}$ is a maximizing value.

For continuous random variables, the joint pdf, instead of the joint pmf, is used for the likelihood function.

Suppose $X_1, X_2, ..., X_n$ are independent and identically distributed Normal random variables with mean μ and variance 1 and we want to obtain the mle of μ. We have

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

$$= (2\pi)^{-n/2} \exp\left(-\tfrac{1}{2}\textstyle\sum_{i=1}^{n}[x_i - \mu]^2\right).$$

So

$$\ell(\mu) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}[x_i - \mu]^2,$$

and

$$\ell'(\mu) = \sum_{i=1}^{n}[x_i - \mu] = \sum_{i=1}^{n} x_i - n\mu.$$

Setting $\ell'(\mu)$ equal to 0, it can be found that $\hat{\mu}_{mle} = \bar{x}$.

Now consider iid $N(\mu, \sigma^2)$ random variables, $X_1, X_2, ..., X_n$, where both μ and σ^2 are unknown.

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[x_i - \mu]^2\right)$$

and

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}[x_i - \mu]^2.$$

To maximize the log-likelihood, it is clear that we need to minimize $\sum_{i=1}^{n}[x_i - \mu]^2$, which again gives us that $\hat{\mu}_{mle} = \bar{x}$.

With the time series model considered in E&T, writing down a likelihood in terms of the Y_t or the Z_t variables may at first be difficult due to the lack of independence. But the disturbance random variables are independent, and so we have

$$f(\mathbf{e}) = \prod_{t=u}^{v} \frac{1}{\sqrt{2\pi}\sigma} \exp(-e_t^2/(2\sigma^2))$$

$$= (2\pi\sigma^2)^{-(v-u+1)/2} \exp\left(-\tfrac{1}{2\sigma^2}\textstyle\sum_{t=u}^{v} e_t^2\right).$$

Now using (8.15) from p. 93 of E&T we can put this in terms of the z_t:

$$f(\mathbf{z}) = (2\pi\sigma^2)^{-(v-u+1)/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=u}^{v}[z_t - \beta z_{t-1}]^2\right).$$

Taking the estimated z_t from (8.19) on p. 94, we can write a log-likelihood in terms of the z_t:

$$\ell(\beta, \sigma^2) = -\frac{(v-u+1)}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=u}^{v}[z_t - \beta z_{t-1}]^2.$$

To maximize the log-likelihood we minimize $\sum_{t=u}^{v}[z_t - \beta z_{t-1}]^2$. We have

$$\frac{d}{d\beta}\sum_{t=u}^{v}[z_t - \beta z_{t-1}]^2 = -2\sum_{t=u}^{v}\{z_{t-1}[z_t - \beta z_{t-1}]\}.$$

Setting this to 0 we get

$$\sum_{t=u}^{v} z_{t-1} z_t = \beta \sum_{t=u}^{v} z_{t-1}^2.$$

So the estimate of β is

$$\hat{\beta} = \frac{\sum_{t=u}^{v} z_{t-1} z_t}{\sum_{t=u}^{v} z_{t-1}^2}.$$

Below is the R code to obtain the plot of the three Poisson probability mass functions:

```
> pr1=numeric(6)
> pr2=numeric(6)
> pr3=numeric(6)
> for (i in 1:6) pr1[i]=dpois(i-1,0.5)
> for (i in 1:6) pr2[i]=dpois(i-1,1)
> for (i in 1:6) pr3[i]=dpois(i-1,1.5)
> x1=c(-0.05, 0.95, 1.95, 2.95, 3.95, 4.95)
> x2=c(0:5)
> x3=c(0.05, 1.05, 2.05, 3.05, 4.05, 5.05)
> plot(x1, pr1, type="h", col="4", ylab="Probability", xlab="X",
+ lwd="3")
> lines(x2,pr2, type="h", col="green", lwd="3")
> lines(x3, pr3, type="h", col="10", lwd="3")
> legend( x="topright", cex=0.8, bty="n", legend=c("theta=0.5",
+ "theta=1.0", "theta=1.5"), col = c(4,"green",10),lwd=c(3,3,3))
```