STA113 (Spring 2002) - Homework 11

April 25, 2002

Solution (6.22)

a. We need the expectation of X,

$$E(X) = \int_0^1 (\theta+1)x^{\theta+1}dx = \frac{\theta+1}{\theta+2}.$$

The method of moments estimator $\hat{\theta}^{mme}$ of θ, arises from the solution (in θ) of the equation

$$\bar{X} = E(X) = \frac{\theta+1}{\theta+2},$$

resulting in $\hat{\theta}^{mme} = (1-2\bar{X})/(\bar{X}-1)$.
Based on the data given, $\hat{\theta}^{mme} = 3$.

b. The likelihood function is

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

and hence the log likelihood is given by

$$\log L(\theta) = n\log(\theta+1) + \theta \sum_{i=1}^n \log x_i.$$

Solving (in θ) the equation $\frac{d\log L(\theta)}{d\theta} = 0$, we obtain

$$\hat{\theta}^{mle} = -\frac{n + \sum_{i=1}^n \log x_i}{\sum_{i=1}^n \log x_i}.$$

This is indeed the maximum likelihood estimator of θ since it gives a negative value to the second derivative with respect to θ of $\log L(\theta)$.
Based on the data given, $\hat{\theta}^{mle} = 3.12$.

Solution (6.23)

We are given $X_1, ..., X_n$ i.i.d. $Poisson(\lambda_1)$ and $Y_1, ..., Y_n$ i.i.d. $Poisson(\lambda_2)$. The additional assumption we need (and is not clearly stated in the problem) is that of independence between the two random samples. This assumption provides the joint likelihood function as follows

$$L(\lambda_1, \lambda_2) = \prod_{i=1}^{n} \frac{e^{-\lambda_1}\lambda_1^{x_i}}{x_i!} \prod_{j=1}^{n} \frac{e^{-\lambda_2}\lambda_2^{y_j}}{y_j!}.$$

The maximum likelihood estimators $\hat{\lambda_1}^{mle}$ and $\hat{\lambda_2}^{mle}$ of λ_1 and λ_2, respectively, are obtained from the solution of the system of equations

$$\frac{d\log L(\lambda_1, \lambda_2)}{d\lambda_1} = 0, \frac{d\log L(\lambda_1, \lambda_2)}{d\lambda_2} = 0,$$

yielding $\hat{\lambda_1}^{mle} = \bar{X}$ and $\hat{\lambda_2}^{mle} = \bar{Y}$. (One can check that the pair $(\hat{\lambda_1}^{mle}, \hat{\lambda_2}^{mle})$ indeed maximizes $\log L(\lambda_1, \lambda_2)$.) Finally, the maximum likelihood estimator of $\lambda_1 - \lambda_2$ is given by

$$\hat{\lambda_1}^{mle} - \hat{\lambda_2}^{mle} = \bar{X} - \bar{Y},$$

using the invariance principle for maximum likelihood estimators.

Solution (6.29)

a. This is an example where maximizing the likelihood function cannot be done by simply taking derivatives since the maximum for one of the parameters is obtained at the boundary.

The likelihood function is

$$L(\lambda, \theta) = \lambda^n e^{-\lambda(-n\theta + \sum_{i=1}^n x_i)}, \theta \leq \min_i x_i$$

and hence the log likelihood becomes

$$\log L(\lambda, \theta) = n \log(\lambda) - \lambda \sum_{i=1}^n x_i + n\lambda\theta, \theta \leq \min_i x_i.$$

Note that the constraint $\theta \leq \min_i x_i$ is very important and has to be included. This is true for any problem where the set of possible values for the random variable depends on the parameters of the distribution (in our case, $x \geq \theta$ in the definition of the density function).

Now let's consider maximizing $\log L(\lambda, \theta)$ with respect to θ (that is keeping λ fixed). We note that $\frac{d \log L(\lambda,\theta)}{d\theta} = n\lambda > 0$. This means that $\log L(\lambda, \theta)$ as a function of θ (with λ fixed) is increasing. This, along with the fact that $\theta \leq \min_i x_i$, implies that the maximum of $\log L(\lambda, \theta)$ (with λ fixed) is obtained at $\theta = \min_i x_i$. Since this value doesn't depend on λ (is the same for any value of λ we fix) the maximum likelihood estimator of θ is

$$\hat{\theta}^{mle} = \min_i x_i.$$

Having found the maximum with respect to θ, we can substitute this value in $\log L(\lambda, \theta)$ leading to

$$\log L(\lambda, \hat{\theta}^{mle}) = \log L(\lambda, \min_i x_i) = n \log(\lambda) - \lambda \sum_{i=1}^n x_i + n\lambda \min_i x_i$$

which needs to be maximized with respect to λ in order to get the maximum likelihood estimator of λ. This is now a standard maximization that involves derivatives. Solving

$$\frac{d \log L(\lambda, \min_i x_i)}{d\lambda} = 0$$

we obtain

$$\hat{\lambda}^{mle} = \frac{n}{-n \min_i x_i + \sum_{i=1}^n x_i}.$$

(Again, it can, and should, be checked through the second derivative that $\hat{\lambda}^{mle}$ indeed maximizes $\log L(\lambda, \min_i x_i)$).

b. For the particular data set, we get $\hat{\theta}^{mle} = 0.64$ and $\hat{\lambda}^{mle} = 0.202$.