14.381 PS#8 SOLUTIONS

FALL 2004
TA: JOSE TESSADA
(tessada@mit.edu)

1. Find the MLE of the following parameters:

(a) λ for an exponential distribution:

$$L(\theta|X) = \prod_{i=1}^{n} \frac{1}{\lambda} \cdot \exp(-x_i/\lambda) = \frac{1}{\lambda^n} \cdot \exp\left(-\sum_{i=1}^{n} x_i/\lambda\right)$$

$$\log L(\theta|X) = -n \cdot \log(\lambda) - \frac{1}{\lambda} \cdot \sum_{i=1}^{n} x_i$$

$$\frac{\partial \log L(\lambda|X)}{\partial \lambda} = -\frac{n}{\lambda} + \frac{1}{\lambda^2} \cdot \sum_{i=1}^{n} x_i = 0$$

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

$$\hat{\lambda}_{ML} = \sum_{i=1}^{n} x_i \Big/ n = \bar{X}$$

$$\frac{d^2 \log L(\lambda|X)}{d\lambda^2} = \frac{n}{\lambda^2} - \frac{2}{\lambda^3} \sum_{i=1}^{n} x_i$$

evaluated at $\lambda = \sum_{i=1}^{n} x_i / n$

$$\frac{n}{\frac{(\sum x_i)^2}{n^2}} - \frac{2}{\frac{(\sum x_i)^3}{n^3}} \cdot \sum x_i$$

$$= \frac{n^3}{(\sum x_i)^2} - \frac{2n^3}{(\sum x_i)^2} = \frac{-n^3}{(\sum x_i)^2} < 0$$

$\Rightarrow$ we found a maximum.

(b) μ and σ for $N(\mu, \sigma^2)$

the likelihood function is

$$L(\theta|x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)$$

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

$$\log L(\theta|x) = -\frac{n}{2}\log(2\cdot\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial \log L(\theta|x)}{\partial\mu} = \frac{1}{2\sigma^2} \cdot 2\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

$$\frac{\partial \log L(\theta|x)}{\partial\sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2} \cdot \frac{1}{(\sigma^2)^2} \cdot \sum_{i=1}^{n}(x_i-\mu)^2$$

so, $\hat{\mu}_{ML}$ and σ^2_{ML} solve the following system of equations

(1) $\sum_{i=1}^{n}(x_i - \hat{\mu}) = 0$

(2) $-\frac{n}{2} + \frac{1}{2} \cdot \frac{1}{\hat{\sigma}^2} \cdot \sum_{i=1}^{n}(x_i - \hat{\mu})^2 = 0$

from (1) we get $\hat{\mu}_{ML} = \sum_{i=1}^{n} x_i / n = \bar{x}$

from (2) (and using the result from (1)) we obtain

$$-n + \frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 = 0$$

$$\Rightarrow \quad \hat{\sigma}^2_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$$

and we know that $\hat{\mu} = \bar{x}$, thus

$$\hat{\sigma}^2_{ML} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Notice that the MLE of σ^2 is biased.
By the invariance property of MLE,

$$\hat{\sigma}_{ML} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Check second order conditions (for an alternative way to check that $\hat{\theta}_{ML}$ is a max see C&B pp. 321-322).

We need at least one second-order partial derivative to be negative, and the Jacobian of the second-order partial derivatives to be positive. (evaluated at $\hat{\mu}_{ML}$ and $\hat{\sigma}^2_{ML}$)

(i) $$\frac{\partial^2 \log L}{\partial \mu^2} = -\frac{n}{\sigma^2}$$

(ii) $$\frac{\partial^2 \log L}{\partial (\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^{n} (x_i - \mu)^2$$

(iii) $$\frac{\partial^2 \log L}{\partial \mu \, \partial \sigma^2} = -\frac{1}{\sigma^4} \sum_{i=1}^{n} (x_i - \mu)$$

It is very easy to check that at least one of the second-order derivatives is <0 (see (i) and evaluate at $\hat{\sigma}^2_{ML}$).
Now, let's deal with the Jacobian

$$J = \left| \begin{matrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4} \sum_{i=1}^{n} (x_i - \mu) \\ -\frac{1}{\sigma^4} \sum_{i=1}^{n} (x_i - \mu) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^{n} (x_i - \mu)^2 \end{matrix} \right|_{\theta = \hat{\theta}}$$

④

$$= \frac{1}{\sigma^6}\left[-\frac{n^2}{2} + \frac{n}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu)\right)^2 \right]\Bigg|_{\theta=\hat{\theta}}$$

$$= \frac{1}{\hat{\sigma}^6}\left[-\frac{n^2}{2} + \frac{n^2}{\hat{\sigma}^2}\hat{\sigma}^2 - \frac{1}{\hat{\sigma}^2}\left(\sum_{i=1}^{n}(x_i-\bar{x})\right)^2 \right]$$

$$= \frac{1}{\hat{\sigma}^6}\left[\frac{n^2}{2} - \frac{1}{\hat{\sigma}^2}\Bigg(\underbrace{\sum_{i=1}^{n}(x_i-\bar{x})}_{=0}\Bigg)^2 \right]$$

$$= \frac{1}{\hat{\sigma}^6}\,\frac{n^2}{2} > 0$$

(c) θ if pdf is

$$f(x|\theta) = \theta x^{-2} \qquad 0 < \theta \leq x < \infty$$

the likelihood is then

$$L(\theta|x) = \prod_{i=1}^{n} \theta x_i^{-2} I_{[\theta,\infty)}(x_i)$$

and we can rewrite it as

$$L(\theta|x) = \theta^n \cdot \prod_{i=1}^{n} x_i^{-2} \quad I_{[\theta,\infty)}(x_{(1)})$$

where $x_{(1)}$ is the $\min\{x_i\}$

Now we face a situation similar to the case where we wanted to estimate the parameters of the uniform distribution.

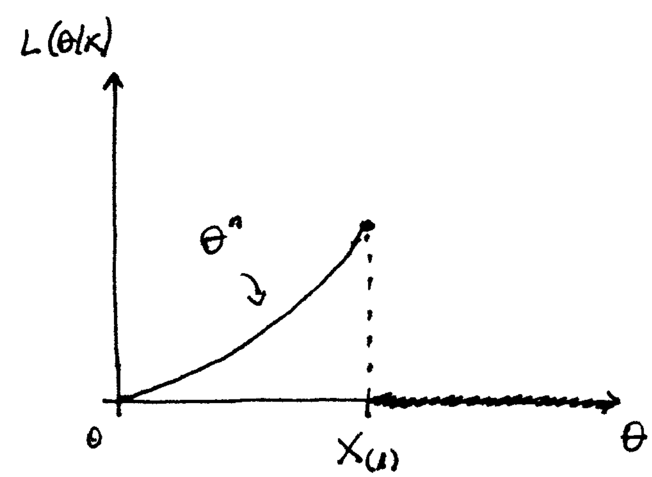
$\prod_{i=1}^{n} x_i^{-2}$ does not depend on θ so we do not look at it.

θ^n is increasing

The indicator function $I_{[\theta,\infty)}(x_{(1)})$ puts a limit, so in fact if we set $\hat{\theta}$ too high we are not maximizing the likelihood.

$$\Rightarrow \hat{\theta}_{ML} = \min\{x_i\} = x_{(1)}$$

Graphically:



2. C&B 7.7

We first write the likelihood function

$$L(\theta|x) = \begin{cases} 1 & \text{if } \theta = 0, \quad 0 < x_i < 1 \\ \prod_{i=1}^{n} \frac{1}{2\sqrt{x_i}} & \text{if } \theta = 1, \quad 0 < x_i < 1 \end{cases}$$

Thus the MLE of θ is

$$\hat{\theta}_{ML} = \begin{cases} 0 & \text{if } 1 \geq \prod_{i=1}^{n} \frac{1}{2\sqrt{x_i}} \\ 1 & \text{if } 1 < \prod_{i=1}^{n} \frac{1}{2\sqrt{x_i}} \end{cases}$$

3. C&B 7.12

$X_i \sim$ iid Bernoulli (θ), $0 \leq \theta \leq 1/2$

(a) method of moments

$$E(X) = \theta = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \quad \Rightarrow \hat{\theta}_{MM} = \bar{X}$$

ML:

$$L(\theta|x) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}$$

$$\log L(\theta|x) = \Sigma x_i \cdot \log(\theta) + (n - \Sigma x_i)\cdot \log(1-\theta)$$

$$\frac{d \log L(\theta|x)}{d\theta} = \frac{\Sigma x_i}{\theta} - \frac{(n-\Sigma x_i)}{1-\theta} = 0$$

$$(1-\theta)\,\Sigma x_i = (n - \Sigma x_i)\,\theta$$

$$\Sigma x_i = n\theta$$

$$\hat{\theta}_{ML} = \frac{\Sigma x_i}{n} = \bar{X}$$

But we have that $\bar{X}$ can take any value between 0 and 1. So the MLE after taking into account the constraints is

$$\hat{\theta}_{ML} = \begin{cases} \bar{X} & \text{if } \bar{X} \le 1/2 \\ 1/2 & \text{if } \bar{X} > 1/2 \end{cases}$$

$$\hat{\theta}_{ML} = \min\{\bar{X}, 1/2\}$$

(b) The MSE of $\hat{\theta}_{MM} = Var(\hat{\theta}_{MM}) + Bias(\hat{\theta}_{MM})^2$

$$= \left(\theta(1-\theta)/n\right) + 0^2 = \frac{\theta(1-\theta)}{n}$$

We can write the MSE of $\hat{\theta}_{ML}$ in the following way:

$$MSE(\hat{\theta}_{ML}) = E(\hat{\theta}_{ML} - \theta)^2 = \sum_{j=0}^{n} (\hat{\theta}_{ML} - \theta)^2 \binom{n}{j} \theta^j (1-\theta)^{n-j}$$

$$= \sum_{j=0}^{[n/2]} \left(\frac{j}{n} - \theta\right)^2 \binom{n}{j} \theta^j (1-\theta)^{n-j}$$

$$+ \sum_{j=[n/2]+1}^{n} \left(\frac{1}{2} - \theta\right)^2 \binom{n}{j} \theta^j (1-\theta)^{n-j}$$

where we use the fact that $Y = \sum_i X_i \sim Binomial(n, \theta)$ and $[n/2] = n/2$, if n is even, and $[n/2] = (n-1)/2$, if n is odd.

(c) Rewrite $MSE(\hat{\theta}_{MM})$:

$$MSE(\hat{\theta}_{MM}) = \sum_{j=0}^{n} \left(\frac{j}{n} - \theta\right)^2 \binom{n}{j} \theta^j (1-\theta)^{n-j}$$

Therefore,

$$MSE(\hat{\theta}_{MM}) - MSE(\hat{\theta}_{ML}) = \sum_{j=[n/2]+1}^{n} \left[\left(\frac{j}{n} - \theta\right)^2 - \left(\frac{1}{2} - \theta\right)^2 \right] \binom{n}{j} \theta^j (1-\theta)^{n-j}$$
$$= \sum_{j=[n/2]+1}^{n} \left(\frac{j}{n} + \frac{1}{2} - 2\theta\right)\left(\frac{j}{n} - \frac{1}{2}\right) \binom{n}{j} \theta^j (1-\theta)^{n-j}$$

The facts that $j/n > 1/2$ in the sum and $\theta \le 1/2$ imply that every term in the sum is positive. Therefore $MSE(\hat{\theta}_{ML}) < MSE(\hat{\theta}_{MM})$ for every $\theta \in (0, 1/2]$. For $\theta = 0$, then $MSE(\hat{\theta}_{ML}) = MSE(\hat{\theta}_{MM})$.

If you take the restriction $\theta \in [0, 1/2]$ as valid, then you could have argued the following:
for any $\theta \in (0, 1/2]$, $\hat{\theta}_{ML} \le \hat{\theta}_{MM}$
then $\hat{\theta}_{ML} - \theta \le \hat{\theta}_{MM} - \theta$
implying that
$$E_\theta\left[(\hat{\theta}_{ML} - \theta)^2\right] \le E_\theta\left[(\hat{\theta}_{MM} - \theta)^2\right]$$

But you have to be careful. As in the previous argument, this is valid iff the restriction is true. Furthermore, it is not evident that for $\theta = 0$ both MSE are the same, even if they are!

Another issue arises when we think that the restriction $\theta \in [0, 1/2]$ is not valid. If that's the case, then we need to allow for $\theta \in [0, 1]$ in the analysis. In this case you will see that truncating the estimator at $1/2$ will lead you to a trade-off.

4.

(a) Write $\varepsilon_i = y_i - \beta x_i$ and write down the likelihood using $(y_i - \beta x_i)$ instead of ε_i.

$$L(\theta|y) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_i (y_i^2 - 2\beta x_i y_i + \beta^2 x_i^2)\right)$$

$$\log L = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_i y_i^2 + \frac{\beta}{\sigma^2}\sum_i x_i y_i - \frac{\beta^2}{2\sigma^2}\sum_i x_i^2$$

$$\frac{\partial \log L}{\partial \beta} = \frac{1}{\sigma^2}\sum_i x_i y_i - \frac{\beta}{\sigma^2}\sum_i x_i^2$$

set the derivative equal to 0 and obtain

$$\hat{\beta} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

and $$\frac{\partial^2 \log L}{\partial \beta^2} = -\frac{1}{\sigma^2}\sum_i x_i^2 < 0$$

so $\hat{\beta}$ is a maximum.

Note: we can ignore the role of σ^2 because $\hat{\beta}$ does not depend on σ^2 and the second order condition holds for any $\sigma^2 > 0$.
This is not true for $\hat{\sigma}^2$, which depends on β.

Now

$$\hat{\beta} = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\sum_i x_i \cdot (\beta x_i + \varepsilon_i)}{\sum_i x_i^2} = \beta + \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2}$$

$$E(\hat{\beta}) = \beta + \frac{1}{\sum_i x_i^2} \cdot \cancel{\sum_i x_i E(\varepsilon_i)}^{\,0} = \beta$$

$\therefore$ $E(\hat{\beta}) = \beta$ so $\hat{\beta}$ is unbiased.

(b)

y_i is a linear transformation of a normal variable, ε_i.

$\Rightarrow y_i \sim N(\beta x_i, \sigma^2)$

And $\hat{\beta}$ is a linear combination of the y_i, so it must have a normal distribution too.
We already know that $E(\hat{\beta}) = \beta$, we just need to calculate the variance.

$$Var(\hat{\beta}) = E\left[(\hat{\beta}-\beta)^2\right] = E\left[\left(\frac{\sum x_i \varepsilon_i}{\sum x_i^2}\right)^2\right] = \frac{1}{(\sum x_i^2)^2} \cdot E\left[\left(\sum x_i \varepsilon_i\right)^2\right]$$

$$= \frac{\sigma^2 \sum x_i^2}{(\sum x_i^2)^2} = \frac{\sigma^2}{\sum x_i^2}$$