James K. Pringle
140.673 Theory
Dr. Constantine Frangakis
Assignment 3
March 14, 2013

Assignment 3

1. You have been sent by email the data Y_i of 500 persons, which are the lengths of stay described in problem 2, part (ii) of the previous problem set.

   (i) By equating the expressions for $E(Y_i|I_i = 1, \theta)$ and $\text{var}(Y_i|I_i = 1, \theta)$ (in terms of $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$) to the sample mean and variance of your data $Y_\mathbf{i}$, find estimates of $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$. This is called a "moment estimation" method.

   Proof. We use the method of moments as described on Casella and Berger, pg 312.

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \mu_1' = EX = \theta_1 \quad \text{according to problem set 2}$$
$$m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \quad \mu_2' = EX^2 = \theta_2 + \theta_1^2 \quad \text{by problem set 2}$$

   Now we set the moments equal to each other from the sample and the population, then solve the system of equations. Hence

$$\tilde{\theta}_1 = \mu_1 = m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \quad \text{and} \tag{1}$$
$$\tilde{\theta}_2 + \tilde{\theta}_1^{\,2} = \mu_2 = m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \tag{2}$$
$$\tilde{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{3}$$

   These are the estimates we desire. □

   (ii) Consider the likelihood function of your data as in part (ii) of the previous problem set. For the MLEs of that likelihood, there is no known closed form, but the MLEs can be found numerically by maximization algorithms. By using the algorithm "optim" (see `help(optim)`) in the statistical environment `R`, or any other appropriate algorithm and/or programming of your choice, find a stationary point for the mean θ_1 and variance θ_2 of the length of stay in the target population.

Proof. The likelihood function from part (ii) of the last homework is

$$\mathrm{pr}(Y_i \mid I_i = 1, \theta) = \prod_{i=1}^{500} \frac{f(y_i, \theta) y_i}{E_\theta[Y_i]} \tag{4}$$

Since maximizing the likelihood gives the same results as maximizing the log of the likelihood, we maximize

$$\log(\mathrm{pr}(Y_i \mid I_i = 1, \theta)) = \sum_{i=1}^{500} \log(f(Y_i, \theta)) + \log(Y_i) - \log(E_\theta[Y_i]) \tag{5}$$

varying the parameters θ_1 and θ_2 using `optim`. Note that $f(Y_i, \theta)$ is the density of a gamma distribution with mean θ_1 and variance θ_2.

The optimal values are $\hat{\theta}_1 =$ and $\hat{\theta}_2 =$. □

(iii) You may get some warning messages while using the algorithm, which indicate possible numerical instability of the algorithm. To make sure the converged values of the algorithm are a maximum, check that the second derivative matrix of the log-likelihood, also called the Hessian matrix, is negative-definite (as defined in class) when evaluated at the converged values of the algorithm.

Note: you can obtain the Hessian matrix by setting the option `hessian=T` in the algorithm `optim`. You can use the result that the Hessian matrix is negative definite if and only if it satisfies the conditions a.-c. of example 7.2.12 of the text (p. 322).

Proof. The code checks that the Hessian matrix is negative-definite. The theorem is that a symmetric matrix is negative-definite if and only if its eigenvalues are all negative. The code checks that (1) the Hessian is symmetric (which it is by definition since Clairaut's theorem holds) and (2) that the eigenvalues are negative.

This shows that the values we found are a true minimum. □

(iv) Find the MLE of the 95th percentile of the distribution of lengths of stay in the target population. (Hint: you can use the invariance property of MLEs, and a numerical method to do the actual computation. For example, check the function `qgamma()` in `R`).

Proof. We assume that Y_i is distributed as a gamma distribution with mean θ_1 and variance θ_2. Let that distribution have distribution function $F_{\theta_1,\theta_2}(x)$. Thus we find the maximum likelihood estimation for x such that $F_{\theta_1,\theta_2}(x) = 0.95$. By the invariance property of MLEs, we have that the MLE $\hat{x}$ of x is the 95th percentile of the distribution using the MLEs for θ_1 and θ_2. Hence, we seek the value of x such that $F_{\hat{\theta}_1,\hat{\theta}_2}(x) = 0.95$.

By the calculations, $\hat{x} =$.

Each of your colleagues has been given a different independent set of 500 people from the survey. Using each other's ML estimates, report an estimate of the variance of the MLEs $\hat{\theta}_1$ and $\hat{\theta}_2$, and an estimate of the covariance between $\hat{\theta}_1$ and $\hat{\theta}_2$.

Calculating,

So the variance-covariance matrix (using subscripts for row and column number) is

$$\begin{bmatrix} 1 \end{bmatrix} \tag{6}$$

□

2. Problem 2. We wish to study the level of a specific radioactive particle in an environment, using a counter. The number X of particles counted by the counter in a time interval of 1 min. is assumed to follow a Poisson distribution. You have been sent 50 measurements, $X_1, ..., X_{50}$, of counts at different 1 min; assume the 50 measurements are i.i.d. from Poisson(μ).

   (i) Find the MLE of μ. Show that the MLE is a minimal sufficient statistic for μ.

   Proof. The joint density of the fifty random variables is

$$\begin{aligned} f(x_1, \cdots, x_{50} \mid \mu) &= \prod_{i=1}^{50} f(x_i) \quad \text{by independence} && (7) \\ &= \prod_{i=1}^{50} \frac{\mu^{x_i}}{x_i!} e^{-\mu} \quad \text{since all are poisson} && (8) \\ &= \mu^{\sum_{i=1}^{50} x_i} e^{-50\mu} \prod_{i=1}^{50} \frac{1}{x_i!} && (9) \end{aligned}$$

Now we take the derivative of the log of the joint density and set it equal to zero.

$$0 = \frac{d}{d\mu}(\log(f(x_1, \cdots, x_{50} \mid \mu))) \tag{10}$$

$$0 = \frac{d}{d\mu}\left(\sum_{i=1}^{50} x_i \log(\mu) - 50\mu + \sum_{i=1}^{50} \log\left(\frac{1}{x_i!}\right)\right) \tag{11}$$

$$0 = \sum_{i=1}^{50} x_i \mu^{-1} - 50 \tag{12}$$

$$\mu = \frac{1}{50}\sum_{i=1}^{50} x_i \tag{13}$$

$$\mu = \bar{X} \tag{14}$$

Since the second derivative, $-\sum_{i=1}^{50} x_i \mu^{-2}$, is everywhere negative, we have that $\mu = \bar{X}$ is a maximum. Hence it is the MLE.

To show that it is a minimal sufficient statistic for μ, we use theorem 6.2.13 in Casella and Berger, pg. 281. Suppose $Y_1, \cdots, Y_{50}$ had the same MLE as $X_1, \cdots, X_{50}$ with $\bar{Y} = \bar{X}$. Then

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = \prod_{i=1}^{50} \frac{\mu^{x_i}}{x_i!}e^{-\mu} / \prod_{i=1}^{50} \frac{\mu^{y_i}}{y_i!}e^{-\mu} \tag{15}$$

$$= \left(\mu^{\sum_{i=1}^{50} x_i} \prod_{i=1}^{50} \frac{1}{x_i!}e^{-\mu}\right) / \left(\mu^{\sum_{i=1}^{50} y_i} \prod_{i=1}^{50} \frac{1}{y_i!}e^{-\mu}\right) \tag{16}$$

$$= \prod_{i=1}^{50} \frac{1}{x_i!} / \prod_{i=1}^{50} \frac{1}{y_i!} \tag{17}$$

since $\sum_{i=1}^{50} x_i = \sum_{i=1}^{50} y_i$. Thus (15) is constant as a function of μ. Now suppose

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = \left(\mu^{\sum_{i=1}^{50} x_i} \prod_{i=1}^{50} \frac{1}{x_i!}e^{-\mu}\right) / \left(\mu^{\sum_{i=1}^{50} y_i} \prod_{i=1}^{50} \frac{1}{y_i!}e^{-\mu}\right) \tag{18}$$

is constant as a function of μ. Thus

$$\mu^{\sum_{i=1}^{50} x_i} = \mu^{\sum_{i=1}^{50} y_i} \tag{19}$$

for all x_i and y_i. Therefore (19) is equal to a constant, and thus

$$\sum_{i=1}^{50} x_i = \sum_{i=1}^{50} y_i \tag{20}$$

and $\bar{X} = \bar{Y}$. Now by the theorem we conclude that $\bar{X}$ is a minimal sufficient statistic. □

(ii) Suppose we are interested in $g(\mu) = \text{pr}(X = 0 \mid \mu)$. Find the MLE of $g(\mu)$. Do you think this MLE is biased or unbiased for $g(\mu)$ and why?

Proof. Note that $g(\mu) = \text{pr}(X = 0 \mid \mu) = e^{-\mu}$. Thus by the invariance property of MLEs, $e^{-\bar{X}}$ is an MLE of $e^{-\mu}$. Applying Jensen's inequality,

$$E(e^{-\bar{X}}) \geq e^{E(-\bar{X})} = e^{-\mu} \quad \text{since } \bar{X} \text{ is unbiased} \tag{21}$$

Since we are dealing with a strictly convex function, we do not have equality, and thus the MLE for $g(\mu)$ is biased. □

(iii) By considering $X_i^* = 1(X_i = 0), i = 1, \cdots, 50$, where $1()$ is the indicator function, find an unbiased estimator and estimate of $g(\mu)$.

Proof. Intuition would say that the proportion of the sample equal to zero, $\frac{1}{50}\sum_{i=1}^{50} X_i^*$, would be a good guess for an unbiased estimator of $g(\mu)$. Calculating,

$$E\left[\frac{1}{50}\sum_{i=1}^{50} X_i^*\right] = \frac{1}{50}\sum_{i=1}^{50} E[X_i^*] \tag{22}$$

$$= \frac{1}{50}\sum_{i=1}^{50} \mathcal{P}(X_i = 0 \mid \mu) \quad \text{since } X_i^* \text{ is an indicator} \tag{23}$$

$$= g(\mu) \tag{24}$$

we see that our guess is indeed unbiased. Calculating from the data provided,

We get that an unbiased estimate of $g(\mu)$ is □

(iv) Find the distribution of $\text{pr}(X_1 \mid \bar{X}, \mu)$. (Here, X_1 indicates the first measurement as given to you in random order, and is not necessarily the smallest measurement).

Proof. Note that since we are dealing with 50 observations,

$$\text{pr}(X_1 \mid \bar{X} = \bar{x}, \mu) = \text{pr}(X_1 \mid \sum_{i=1}^{50} X_i = \sum_{i=1}^{50} x_i, \mu) \tag{25}$$

Define a new random variables

$$S = \sum_{i=1}^{50} X_i, \quad \text{and} \quad S_{-1} = \sum_{i=2}^{50} X_i \tag{26}$$

Since the X_i are independent, we have X_1 and S_{-1} are also independent. Notice that S_{-1} is the sum of 49 Poisson(μ) which is a poisson distribution with parameter 49μ. These two facts give

$$\text{pr}(X_1 = x_1, S_{-1} = s_{-1} \mid \mu) = \frac{\mu^{x_1}}{x_1!}e^{-\mu}\frac{(49\mu)^{s_{-1}}}{s_{-1}!}e^{-49\mu} \tag{27}$$

Now we wish to calculate the joint density of X_1 and S by the change of variable method (for reference, see pg. 108 of Grimmett and Stirzaker). Let

$$U = S = S_{-1} + X_1, \quad \text{and} \quad V = X_1 \tag{28}$$

If follows that $S_{-1} = U - V$. Thus the Jacobian, J, is 1 everywhere. Now the joint density of X_1 and S is as follows

$$\begin{aligned} f_{X_1,S}(X_1 = x_1, S = s) &= f_{X_1,S_{-1}}(x_1, s - x_1)|J(x_1, s)| && (29)\\ &= f_{X_1,S_{-1}}(x_1, s - x_1) && (30)\\ &= \frac{\mu^{x_1}}{x_1!}e^{-\mu}\frac{(49\mu)^{s-x_1}}{(s-x_1)!}e^{-49\mu} && (31) \end{aligned}$$

Hence, by conditional probability,

$$\begin{aligned} \mathrm{pr}(X_1 \mid \bar{X}, \mu) &= \mathrm{pr}(X_1 \mid S, \mu) && (32)\\ &= \mathrm{pr}(X_1, S \mid \mu)/\mathrm{pr}(S \mid \mu) && (33)\\ &= \left(\frac{\mu^{x_1}}{x_1!}e^{-\mu}\frac{(49\mu)^{s-x_1}}{(s-x_1)!}e^{-49\mu}\right) \Big/ \left(\frac{(50\mu)^s}{s!}e^{-50\mu}\right) && (34)\\ &= \frac{\mu^{x_1}e^{-50\mu}}{x_1!}\frac{(49\mu)^{s-x_1}}{(s-x_1)!}\frac{s!}{e^{50\mu}(-50\mu)^s} && (35)\\ &= \binom{s}{x_1}\left(\frac{49}{50}\right)^s\left(\frac{1}{49}\right)^{-x_1} && (36) \end{aligned}$$

as desired. □

(v) Use your estimator in (iii), your result in (iv) and "Blackwellization" to obtain an unbiased estimator (and estimate) for $g(\mu)$ that has smaller variance than the one in (iii). Is this the minimum unbiased estimator for $g(\mu)$, and why or why not?

Proof. We have an estimate for $g(\mu)$. It is $(1/50)\sum_{i=1}^{50} X_i^*$. Since the distribution of the data given the statistic $\bar{X}$ does not depend on μ, as we found in (iv) we

Blackwellize by finding

$$
\begin{aligned}
E\left[(1/50)\sum_{i=1}^{50} X_i^* \mid \bar{X}\right] &= (1/50)\sum_{i=1}^{50} E[X_i^*|\bar{X}] && (37)\\
&= (1/50)\sum_{i=1}^{50} \mathrm{pr}(X_i = 0|\bar{X}, \mu) && (38)\\
&= (1/50)\sum_{i=1}^{50} \mathrm{pr}(X_i = 0|S, \mu) && (39)\\
&= (1/50)\sum_{i=1}^{50} \binom{s}{0}\left(\frac{49}{50}\right)^s\left(\frac{1}{49}\right)^0 && (40)\\
&= \left(\frac{49}{50}\right)^s && (41)
\end{aligned}
$$

It is not quite clear if this is the minimum unbiased estimator for $g(\mu)$. □