James K. Pringle
140.673 Theory
Dr. Constantine Frangakis
Assignment 3
March 13, 2014

Assignment 3

Problem 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 $\theta_{\mathbf{1}}$ and $\theta_{\mathbf{2}}$) to the sample mean and variance of your data $Y_{\mathbf{i}}$, find estimates of $\theta_{\mathbf{1}}$ and $\theta_{\mathbf{2}}$. This is called a "moment estimation" method.

Proof. Let

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y_n}$$
$$m_2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - m_1)^2 = s_n$$

be the sample mean and variance of the data Y_i. From problem set 2,

$$\mu_1 = E[Y_i|I_i = 1, \theta] = \frac{\theta_2}{\theta_1} + \theta_1$$

and calculating,

$$
\begin{aligned}
\mu_2 &= \mathrm{var}(\mu_2) \\
&= E[Y_i^2 | I_i = 1, \theta] - E[Y_i | I_i = 1, \theta]^2 \\
&= \int E[Y_i|\theta]^{-1} f(Y_i, \theta) Y_i^3 dY_i - \left(\frac{\theta_2}{\theta_1} + \theta_1\right)^2 \\
&= \theta_1^{-1} E[Y_i^3|\theta] - \left(\frac{\theta_2^2}{\theta_1^2} + 2\theta_2 + \theta_1^2\right) \\
&= \theta_1^{-1}\left(\theta_1^3 + 3\theta_1\theta_2 + 2\theta_2^2\theta_1^{-1}\right) - \left(\frac{\theta_2^2}{\theta_1^2} + 2\theta_2 + \theta_1^2\right) \\
&= \theta_1^2 + 3\theta_2 + 2\theta_2^2\theta_1^{-2} - \left(\frac{\theta_2^2}{\theta_1^2} + 2\theta_2 + \theta_1^2\right) \\
&= \frac{\theta_2^2}{\theta_1^2} + \theta_2
\end{aligned}
$$

See `http://bit.ly/N6zu5j` for a derivation of the third moment of the gamma distribution. Therefore,

$$
\begin{aligned}
\frac{\mu_1^2 - \mu_2}{\mu_1} &= \frac{\left(\frac{\theta_2}{\theta_1} + \theta_1\right)^2 - \frac{\theta_2^2}{\theta_1^2} + \theta_2}{\frac{\theta_2}{\theta_1} + \theta_1} \\
&= \frac{\frac{\theta_2^2}{\theta_1^2} + 2\theta_2 + \theta_1^2 - \frac{\theta_2^2}{\theta_1^2} + \theta_2}{\frac{\theta_2}{\theta_1} + \theta_1} \\
&= \frac{\theta_2 + \theta_1^2}{\frac{1}{\theta_1}(\theta_2 + \theta_1^2)} \\
&= \theta_1
\end{aligned}
$$

and

$$
\begin{aligned}
\theta_2 &= \left(\frac{\theta_2}{\theta_1} + \theta_1 - \theta_1\right)\theta_1 \\
&= \left(\mu_1 - \frac{\mu_1^2 - \mu_2}{\mu_1}\right)\left(\frac{\mu_1^2 - \mu_2}{\mu_1}\right) \\
&= \left(\frac{\mu_1^2}{\mu_1} - \frac{\mu_1^2 - \mu_2}{\mu_1}\right)\left(\frac{\mu_1^2 - \mu_2}{\mu_1}\right) \\
&= \left(\frac{\mu_2}{\mu_1}\right)\left(\frac{\mu_1^2 - \mu_2}{\mu_1}\right) \\
&= \frac{\mu_1^2\mu_2 - \mu_2^2}{\mu_1^2}
\end{aligned}
$$

Replacing μ_j by m_j we the method of moments estimators are

$$\theta_1^{MM} = \frac{m_1^2 - m_2}{m_1} = \frac{(\overline{Y}_n)^2 - s_n}{\overline{Y}_n}$$
$$\theta_2^{MM} = \frac{m_1^2 m_2 - m_2^2}{m_1^2} = \frac{(\overline{Y}_n)^2 s_n - s_n^2}{(\overline{Y}_n)^2}$$

Calculating,

```
dat_file <- "data9.txt"
dat_str <- paste(readLines(dat_file), collapse = "")
dat <- eval(parse(text = dat_str))

head(dat)

## [1] 231 497 366 674  85 372

summary(dat)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       4      88     175     213     293    1010

length(dat)

## [1] 500

m1 <- mean(dat)
m2 <- var(dat)
mm1 <- (m1^2 - m2)/m1
mm2 <- (m1^2 * m2 - m2^2)/(m1^2)
theta_mm <- c(mm1, mm2)

theta_mm

## [1]    86.68 10968.35
```

Thus the estimates from the method of moments are

$$\begin{bmatrix} \theta_1^{MM} \\ \theta_2^{MM} \end{bmatrix} = \begin{bmatrix} 86.6767 \\ 1.0968 \times 10^4 \end{bmatrix}$$

Another Method of Moments

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

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} Y_i, \quad \mu_1' = E[Y_i|I_i = 1, \theta] = \frac{\theta_2}{\theta_1} + \theta_1 \quad \text{by problem set 2, problem (iii)}$$

Calculating,

$$\begin{aligned} m_2 = \frac{1}{n}\sum_{i=1}^{n} Y_i^2, \quad \mu_2' = E[Y_i^2|I_i = 1, \theta] &= \int E[Y_i|\theta]^{-1} f(Y_i, \theta) Y_i^3 dY_i \\ &= \theta_1^{-1} E[Y_i^3|\theta] \\ &= \theta_1^{-1}\left(\theta_1^3 + 3\theta_1\theta_2 + 2\theta_2^2\theta_1^{-1}\right) \\ &= \theta_1^2 + 3\theta_2 + 2\theta_2^2\theta_1^{-2} \end{aligned}$$

Let $\theta_1 = k\gamma$ where $k > 0$ is the shape and $\gamma > 0$ is the scale. Then $\theta_2 = k\gamma^2$. Therefore,

$$\mu_1' = \frac{k\gamma^2}{k\gamma} + k\gamma = \gamma(k+1)$$

and

$$\mu_2' = k^2\gamma^2 + 3k\gamma^2 + 2\gamma^2 = \gamma^2(k+1) + \gamma^2(k+1)^2$$

Calculating,

$$\begin{aligned} \gamma &= \frac{\gamma^2(k+1) + \gamma^2(k+1)^2 - (\gamma(k+1))^2}{\gamma(k+1)} \\ &= \frac{\mu_2' - (\mu_1')^2}{\mu_1'} \end{aligned}$$

Furthermore,

$$\begin{aligned} k &= \frac{\gamma(k+1)}{\gamma} - 1 \\ &= \frac{\mu_1'}{\frac{\mu_2' - (\mu_1')^2}{\mu_1'}} - 1 \\ &= \frac{(\mu_1')^2}{\mu_2' - (\mu_1')^2} - 1 \\ &= \frac{2(\mu_1')^2 - \mu_2'}{\mu_2' - (\mu_1')^2} \end{aligned}$$

Combining these definitions of γ and k, it follows

$$\begin{aligned}\theta_1 &= k\gamma \\ &= \left(\frac{2(\mu_1')^2 - \mu_2'}{\mu_2' - (\mu_1')^2}\right)\left(\frac{\mu_2' - (\mu_1')^2}{\mu_1'}\right) \\ &= \frac{2(\mu_1')^2 - \mu_2'}{\mu_1'}\end{aligned}$$

and

$$\begin{aligned}\theta_2 &= k\gamma^2 \\ &= \left(\frac{2(\mu_1')^2 - \mu_2'}{\mu_2' - (\mu_1')^2}\right)\left(\frac{\mu_2' - (\mu_1')^2}{\mu_1'}\right)^2 \\ &= \left(\frac{2(\mu_1')^2 - \mu_2'}{\mu_1'}\right)\left(\frac{\mu_2' - (\mu_1')^2}{\mu_1'}\right) \\ &= \frac{-2(\mu_1')^4 - (\mu_2')^2 + 3(\mu_1')^2\mu_2'}{(\mu_1')^2}\end{aligned}$$

Now replacing μ_j' with m_j, we get the moment of methods estimator.

$$\begin{aligned}\theta_1^{MM} &= \frac{2(\overline{Y_n})^2 - \overline{Y_n^2}}{\overline{Y_n}} \\ \theta_2^{MM} &= \frac{-2(\overline{Y_n})^4 - (\overline{Y_n^2})^2 + 3(\overline{Y_n})^2\overline{Y_n^2}}{(\overline{Y_n})^2}\end{aligned}$$

Using the data Y_i of 500 persons, the method of moments gives

```
m1 <- mean(dat)
m2 <- mean(dat * dat)
mm1 <- (2 * m1^2 - m2)/m1
mm2 <- (-2 * m1^4 - m2^2 + 3 * m1^2 * m2)/(m1^2)
theta_mm <- c(mm1, mm2)

theta_mm

## [1]    86.93 10978.38
```

Thus the estimates from the method of moments are

$$\begin{bmatrix}\theta_1^{MM} \\ \theta_2^{MM}\end{bmatrix} = \begin{bmatrix}86.9297 \\ 1.0978 \times 10^4\end{bmatrix}$$

$\square$

(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

$$\text{pr}(Y^{obs} \mid I_i = 1, \theta) = E[Y|\theta]^{-500} \left(\prod_{i=1}^{500} f(Y_i, \theta) Y_i \right)$$

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

$$\ell(Y^{obs}; \theta) = \log(\text{pr}(Y^{obs} \mid I_i = 1, \theta)) = -500 \log(E[Y|\theta]) + \sum_{i=1}^{500} (\log(f(Y_i, \theta)) + \log(Y_i))$$

by 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. Therefore the shape is $\alpha = \theta_1^2/\theta_2$ and the rate is $\beta = \theta_1/\theta_2$. Numerical optimization gives values for θ_1 and θ_2 that maximize the likelihood.

```r
get_log_lik <- function(theta, dat = dat) {
    alpha <- theta[1]^2/theta[2]
    beta <- theta[1]/theta[2]

    log_pr <- dgamma(x = dat, shape = alpha, rate = beta, log = TRUE)
    log_Y <- log(dat)
    log_EY <- log(theta[1])

    log_lik_each <- log_pr + log_Y - log_EY
    log_lik <- sum(log_lik_each)
    return(log_lik)
}

optim_out <- suppressWarnings(optim(par = theta_mm, fn = get_log_lik, dat = dat,
    control = list(fnscale = -1), hessian = TRUE))

optim_out$convergence

## [1] 0
```

```
theta_mle <- optim_out$par
theta_mle

## [1]    80.36 10679.92
```

The convergence code 0 means that `optim` successfully completed. The values that maximize the likelihood are

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \begin{bmatrix} 80.3553 \\ 1.068 \times 10^4 \end{bmatrix}$$

□

(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 Hessian is

```
hess <- optim_out$hessian
hess

##              [,1]       [,2]
## [1,] -0.0338227  2.422e-04
## [2,]  0.0002422 -3.070e-06
```

and by

```
isSymmetric(hess)

## [1] TRUE
```

`R` says that the Hessian is symmetric. Since the Hessian is symmetric, it can be diagonalized. Its eigenvalues and eigenvectors are

```
eigen_out <- eigen(hess)
eigen_out
```

```
## $values
## [1] -1.335e-06 -3.382e-02
##
## $vectors
##           [,1]      [,2]
## [1,] -0.007161 -0.999974
## [2,] -0.999974  0.007161
```

Let $H(\hat{\theta_1}, \hat{\theta_2})$ denote the Hessian evaluated at the MLE (what was returned by `optim`). Then, calculating,

$$\begin{aligned} H(\hat{\theta_1}, \hat{\theta_2}) &= \begin{bmatrix} -0.033823 & 0.000242 \\ 0.000242 & -0.000003 \end{bmatrix} \\ &= \begin{bmatrix} -0.007161 & -0.999974 \\ -0.999974 & 0.007161 \end{bmatrix} \begin{bmatrix} -0.000001 & 0.000000 \\ 0.000000 & -0.033824 \end{bmatrix} \begin{bmatrix} -0.007161 & -0.999974 \\ -0.999974 & 0.007161 \end{bmatrix} \end{aligned}$$

Let V be the matrix of eigenvectors. Let D be the diagonal matrix of eigenvalues. Then

$$H(\hat{\theta_1}, \hat{\theta_2}) = V'DV$$

Notice V has non-zero determinant:

```
det(eigen_out$vectors)

## [1] -1
```

Given any non-zero $x \in \mathbf{R}^2$, let $Vx = y$. It follows that since V is invertible, y is also non-zero. Therefore,

$$\begin{aligned} x'H(\hat{\theta_1}, \hat{\theta_2})x &= x'V'DVx \\ &= y'Dy \\ &= y_1^2 D_{11} + y_2^2 D_{22} < 0 \end{aligned}$$

since the y_j^2 are non-negative, with at least one of the y_j^2 non-zero, and the D_{jj} are both negative. Therefore, the Hessian evaluated at the MLE is negative-definite. □

(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 the corresponding distribution function be $F(x|\theta_1, \theta_2)$. Thus we find the maximum likelihood estimation for x such that $F(x|\theta_1, \theta_2) = 0.95$. Equivalently, we find

$x = F^{-1}(0.95|\theta_1, \theta_2)$. By the invariance property of MLEs, we have that the MLE $\hat{x}$ of x is the 95th percentile of the distribution defined by the MLEs for θ_1 and θ_2. Hence, we seek $\hat{x} = F^{-1}(0.95|\hat{\theta}_1, \hat{\theta}_2)$.

```
# Shape
alpha <- theta_mle[1]^2/theta_mle[2]
# Rate
beta <- theta_mle[1]/theta_mle[2]
p95 <- qgamma(0.95, shape = alpha, rate = beta)
p95

## [1] 288.4
```

By these calculations, $\hat{x} = 288.354$. $\square$

(v) 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 $\hat{\theta}_2$.

Proof. Calculating,

```
mle <- data.frame(theta1 = c(80.36865, 109.7853, 90.99945, 81.44357, 101.9241,
    92.92737, 96.05893, 70.32257, 85.84821, 73.89659), theta2 = c(10675.78343,
    13356.19, 12486.82, 12821.99356, 13084.32, 12286.38729, 12657.37306, 12421.78847,
    11932.33, 11354.32))
nrow(mle)

## [1] 10

t1t1 <- cov(mle$theta1, mle$theta1)
t1t2 <- cov(mle$theta1, mle$theta2)
t2t2 <- cov(mle$theta2, mle$theta2)
t1t1

## [1] 153.6

t1t2

## [1] 6469

t2t2

## [1] 649753
```

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

$$\begin{bmatrix} 153.6307 & 6469.3515 \\ 6469.3515 & 6.4975 \times 10^5 \end{bmatrix}$$

□

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

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

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

$$\begin{aligned} 0 &= \frac{d}{d\mu}(\log(f(x_1, \cdots, x_{50} \mid \mu))) \\ 0 &= \frac{d}{d\mu}\left(\sum_{i=1}^{50} x_i \log(\mu) - 50\mu + \sum_{i=1}^{50} \log\left(\frac{1}{x_i!}\right)\right) \\ 0 &= \sum_{i=1}^{50} x_i \mu^{-1} - 50 \\ \mu &= \frac{1}{50} \sum_{i=1}^{50} x_i \\ \mu &= \bar{X} \end{aligned}$$

The second derivative of the likelihood is $-\sum_{i=1}^{50} x_i \mu^{-2}$. Since it is negative for all values of $\mu > 0$, 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 apply the theorem from class notes chapter 2, slide 7. Let $\mathbf{Y} \in \mathbf{R}^{50}$ have the same distribution as $\mathbf{X} \in \mathbf{R}^{50}$. Let $T(\mathbf{Y}) = \bar{\mathbf{Y}}$, the MLE calculated above. Now suppose $T(\mathbf{X}) = T(\mathbf{Y})$. Then

$$\begin{aligned}\frac{1}{50}\sum_{i=1}^{50} X_i &= \frac{1}{50}\sum_{i=1}^{50} Y_i \\ \sum_{i=1}^{50} X_i &= \sum_{i=1}^{50} Y_i\end{aligned}$$

and

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

which is free of μ. Now suppose that

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

is free of μ. Then it is the case that

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

is constant with respect to μ, so that

$$\begin{aligned}0 &= \sum_{i=1}^{50} x_i - \sum_{i=1}^{50} y_i \\ \sum_{i=1}^{50} x_i &= \sum_{i=1}^{50} y_i \\ \frac{1}{50}\sum_{i=1}^{50} x_i &= \frac{1}{50}\sum_{i=1}^{50} y_i \\ T(\mathbf{x}) &= T(\mathbf{y})\end{aligned}$$

Now by the theorem we conclude that $\bar{X}$ is a minimal sufficient statistic for μ. $\square$

(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, $\widehat{g(\mu)} = e^{-\bar{X}}$ is the MLE of $e^{-\mu}$. Applying Jensen's inequality,

$$E(\widehat{g(\mu)}) = E(g(\hat{\mu})) = E(e^{-\bar{X}}) \geq e^{E(-\bar{X})} = e^{-\mu} \quad \text{since } \bar{X} \text{ is unbiased for } \mu$$

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

$$E(\widehat{g(\mu)}) > e^{-\mu}$$

□

(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,

$$\begin{aligned} E\left[\frac{1}{50}\sum_{i=1}^{50} X_i^*\right] &= \frac{1}{50}\sum_{i=1}^{50} E[X_i^*] \\ &= \frac{1}{50}\sum_{i=1}^{50} P(X_i = 0 \mid \mu) \quad \text{since } X_i^* \text{ is an indicator} \\ &= g(\mu) \end{aligned}$$

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

```
dat2 <- c(3, 4, 4, 3, 6, 0, 2, 4, 4, 3, 3, 5, 2, 2, 4, 3, 5, 4, 1, 2, 3, 3,
    6, 0, 5, 6, 5, 2, 4, 4, 0, 4, 4, 1, 2, 8, 4, 7, 3, 5, 3, 3, 2, 4, 5, 7,
    3, 2, 4, 5)
est <- mean(dat2 == 0)
est

## [1] 0.06
```

We get that an unbiased estimate of $g(\mu)$ is 0.06 □

(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,

$$\begin{aligned}
\text{pr}(X_1 \mid \bar{X} = \bar{x}, \mu) &= \text{pr}(X_1 \mid \frac{1}{50}\sum_{i=1}^{50} X_i = \frac{1}{50}\sum_{i=1}^{50} x_i, \mu) \\
&= \text{pr}(X_1 \mid \sum_{i=1}^{50} X_i = \sum_{i=1}^{50} x_i, \mu)
\end{aligned}$$

Define new random variables

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

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

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

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

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

If follows that $S_{-1} = U - V$. Thus the Jacobian is

$$\begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$$

Now the joint density of X_1 and S is as follows

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

Hence, by conditional probability,

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

as desired. This is recognized as a binomial distribution with s trials, with x_1 successes, and with probability of success $1/50$. □

(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 $(1/50)\sum_{i=1}^{50} X_i^*$. In problem (i), we found that $\bar{X}$ is a minimal sufficient statistic for μ. By the Blackwell-Rao Theorem,

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

is the "Blackwellized" estimator. Since we conditioned on a minimal sufficient statistic in the Blackwellization process, $\left(\frac{49}{50}\right)^s$ is the minimum variance unbiased estimator for $g(\mu)$.

Calculating,

```
s <- sum(dat2)
blac <- (49/50)^s
blac

## [1] 0.02743
```

we see that our Blackwellized estimator is 0.0274. □