Math 442 : Solutions to Mid Term II (Take Home)
Spring 2009.

1. Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with mean θ and variance σ^2. Assume that $\theta \neq 0$. Determine the asymptotic distribution of $\overline{X}_n^{\,3}$.

   Solution. Since the sample is taken from a normal distribution, $\overline{X}_n \sim N(0, \sigma^2/n)$ for any sample size. Now, let $g(x) = x^3$. Here we wish to determine the distribution of $g(\overline{X}_n)$ for sufficiently large values of n. Since g is a differentiable function whose derivative is $g'(x) = 3x^2$ nonzero at θ (as $\theta \neq 0$) we can apply the *delta method* which says for large n

$$g(\overline{X}_n) \overset{\text{approx}}{\sim} N(g(\theta), [g'(\theta)]^2\sigma^2/n)$$

   That is, the asymptotic distribution of $\overline{X}_n^{\,3}$ is $N(\theta^3,\ 9\theta^4\ \frac{\sigma^2}{n})$. □

2. (Graduate Students Only.) Let X and Y be independent exponential random variables, with

$$f_X(x|\lambda) = \begin{cases} \frac{1}{\lambda}e^{-x/\lambda} & \text{for } x > 0, \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad f_Y(y|\mu) = \begin{cases} \frac{1}{\mu}e^{-y/\mu} & \text{for } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

   Define Z and W by

$$Z = \min\{X, Y\} \quad \text{and} \quad W = \begin{cases} 1 & \text{if } Z = X, \\ 0 & \text{if } Z = Y. \end{cases}$$

   Suppose we can only observe $(Z_1, W_1), (Z_2, W_2), \ldots, (Z_n, W_n)$. Find the maximum likelihood estimator of λ and μ from the observed data.

   Solution. Let us first find the joint distribution of Z and W. To that end let us compute the following conditional probabilities.

$$P(Z \leq z | W = 1) = \frac{P(Z \leq z, W = 1)}{P(W = 1)} = \frac{P(X \leq z, X \leq Y)}{P(X \leq Y)}.$$

   Now,

$$\begin{aligned} P(Z \leq z, W = 1) &= \int_0^z \int_x^\infty \frac{1}{\lambda}e^{-x/\lambda}\frac{1}{\mu}e^{-y/\mu}dydx \\ &= \int_0^z \frac{1}{\lambda}e^{-x/\lambda}\left[\int_x^\infty \frac{1}{\mu}e^{-y/\mu}dy\right]dx \\ &= \int_0^z \frac{1}{\lambda}e^{-x/\lambda}e^{-x/\mu}dx \\ &= \frac{1}{\lambda}\int_0^z e^{-(1/\lambda+1/\mu)x}dx \\ &= \frac{1/\lambda}{1/\lambda+1/\mu}(1 - e^{-(1/\lambda+1/\mu)z}) \end{aligned}$$

   and

$$\begin{aligned} P(X \leq Y) = \int_0^\infty \int_x^\infty \frac{1}{\lambda}e^{-x/\lambda}\frac{1}{\mu}e^{-y/\mu}dydx &= \lim_{z\to\infty}\int_0^z \int_x^\infty \frac{1}{\lambda}e^{-x/\lambda}\frac{1}{\mu}e^{-y/\mu}dydx \\ &= \frac{1/\lambda}{1/\lambda+1/\mu}\lim_{z\to\infty}(1 - e^{-(1/\lambda+1/\mu)z}) = \frac{1/\lambda}{1/\lambda+1/\mu}. \quad (1) \end{aligned}$$

Hence,

$$P(Z \le z|W=1) = \frac{P(Z \le z, W=1)}{P(W=1)} = \frac{P(X \le z, X \le Y)}{P(X \le Y)} = (1 - e^{-(1/\lambda + 1/\mu)z}).$$

Similar calculations also show that

$$P(Z \le z|W=0) = \frac{P(Z \le z, W=0)}{P(W=0)} = \frac{P(Y \le z, Y \le X)}{P(Y \le X)} = (1 - e^{-(1/\lambda + 1/\mu)z}).$$

So the conditional distribution Z given $W = w$ does not depend on w which means Z and W are independent and,

$$P(Z \le z|W=w) = P(Z \le z) = 1 - e^{-(1/\lambda + 1/\mu)z}. \tag{2}$$

Let us introduce the following parameterizations:

$$\phi = \left(\frac{1}{\lambda} + \frac{1}{\mu}\right)^{-1} \quad \text{and} \quad \theta = \frac{\frac{1}{\lambda}}{\frac{1}{\mu} + \frac{1}{\mu}}.$$

We recognize the RHS of the second equality in (2) as the CDF of Exponential(ϕ). We know that W has a Bernoulli(θ) distribution (see equation (1) above). Since W and Z are independent, the MLE of ϕ cannot depend on W and that of θ cannot depend on Z. Indeed, we know that the MLE's of ϕ and θ are $\overline{Z}$ and $\overline{W}$, respectively. The MLE's of λ and μ can be obtained by the invariance property of MLE noting that

$$\lambda = \frac{\phi}{\theta} \quad \text{and} \quad \mu = \frac{\phi}{1-\theta}.$$

Hence the MLE of λ and μ are

$$\hat{\lambda} = \frac{\overline{Z}}{\overline{\overline{W}}} \quad \text{and} \quad \hat{\mu} = \frac{\overline{Z}}{1 - \overline{\overline{W}}},$$

respectively. □

3. (Undergraduate Students Only.) Let $X_1, X_2, \ldots, X_n$ be a sample from the *inverse Gaussian* probability density function,

$$f_X(x|\mu, \lambda) = \begin{cases} \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\{-\lambda(x-\mu)^2/(2\mu^2 x)\} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Find the maximum likelihood estimator of λ and μ. Note: $\lambda > 0, \mu > 0$.

Solution. For all observed values $x_1, x_2, \ldots, x_n$, the likelihood function $L(\mu, \lambda|x_1, x_2, \ldots, x_n)$ will be

$$\begin{aligned} L(\mu, \lambda|x_1, x_2, \ldots, x_n) &= \left[\prod_{i=1}^{n} \left(\frac{\lambda}{2\pi {x_i}^3}\right)^{1/2}\right] \exp\left\{\sum_{i=1}^{n} \frac{-\lambda(x_i - \mu)^2}{2\mu^2 x_i}\right\} \\ &= \left(\frac{\lambda}{2\pi}\right)^{n/2} \left[\prod_{i=1}^{n} \left(\frac{1}{{x_i}^3}\right)^{1/2}\right] \exp\left\{\sum_{i=1}^{n} \frac{-\lambda {x_i}^2 + \lambda 2 x_i \mu - \lambda \mu^2}{2\mu^2 x_i}\right\} \\ &= \left(\frac{\lambda}{2\pi}\right)^{n/2} \left[\prod_{i=1}^{n} \left(\frac{1}{{x_i}^3}\right)^{1/2}\right] \exp\left\{\sum_{i=1}^{n} \left(\frac{-\lambda x_i}{2\mu^2} + \frac{\lambda}{\mu} - \frac{\lambda}{2x_i}\right)\right\} \\ &= \left(\frac{\lambda}{2\pi}\right)^{n/2} \left[\prod_{i=1}^{n} \left(\frac{1}{{x_i}^3}\right)^{1/2}\right] \exp\left\{\frac{-\lambda}{2\mu^2} \sum_{i=1}^{n} x_i + \frac{n\lambda}{\mu} - \frac{\lambda}{2} \sum_{i=1}^{n} \frac{1}{x_i}\right\}. \end{aligned}$$

Instead of maximizing the likelihood function directly, it is easier to maximize the natural log which preserves the maximum values as it is a monotonic increasing function. Let

$$\begin{aligned}\ell(\mu,\lambda|x_1,x_2,\ldots,x_n) &= \ln L(\mu,\lambda|x_1,x_2,\ldots,x_n)\\ &= \frac{n}{2}\ \ln\lambda - \frac{n}{2}\ \ln(2\pi) + \ln\left[\prod_{i=1}^{n}\left(\frac{1}{{x_i}^3}\right)^{1/2}\right] - \frac{\lambda}{2\mu^2}\sum_{i=1}^{n} x_i + \frac{n\lambda}{\mu} - \frac{\lambda}{2}\sum_{i=1}^{n}\frac{1}{x_i}.\end{aligned}$$

First, for each fixed λ, we shall find the value $\hat{\mu}$ which maximizes this function. To locate the extrema, take the partial derivative with respect to μ, set it equal to zero and solve

$$\frac{\partial}{\partial\mu}\ell(\mu,\lambda|x_1,x_2,\ldots,x_n) = \frac{\lambda}{\mu^3}\sum_{i=1}^{n} x_n - \frac{n\lambda}{\mu^2} = 0 \quad\Rightarrow\quad \mu = \frac{1}{n}\sum_{i=1}^{n} x_n = \overline{x}_n.$$

The second partial derivative of $\ell(\mu,\lambda|x_1,x_2,\ldots,x_n)$ at $\overline{x}_n$ is

$$\frac{\partial^2}{\partial\mu^2}\ell(\mu,\lambda|x_1,x_2,\ldots,x_n)|_{\mu=\overline{x}_n} = -\frac{3\lambda}{{\overline{x}_n}^4}\sum_{i=1}^{n} x_i + \frac{2\lambda n}{{\overline{x}_n}^3} = -\frac{3\lambda n^4}{\sum_{i=1}^{n} x_i} + \frac{2\lambda n^4}{\sum_{i=1}^{n} x_i}.$$

Hence, the second derivative is negative and the function is convex. Therefore $\hat{\mu} = \overline{x}_n$ maximizes the function. Before proceeding, we should check the boundaries of the parameter space; that is, the limits of the likelihood function

$$\lim_{\mu\to 0} L(\mu,\lambda|x_1,x_2,\ldots,x_n) = e^{-\infty} = 0$$

and

$$\lim_{\mu\to\infty} L(\mu,\lambda|x_1,x_2,\ldots,x_n) = \left(\frac{\lambda}{2\pi}\right)^{n/2}\left[\prod_{i=1}^{n}\left(\frac{1}{{x_i}^3}\right)^{1/2}\right]\exp\left\{-\frac{\lambda}{2}\sum_{i=1}^{n}\frac{1}{x_i}\right\} < L(\overline{x}_n,\lambda|x_1,x_2,\ldots,x_n)$$

as

$$\begin{aligned}L(\overline{x}_n,\lambda|x_1,x_2,\ldots,x_n) &= \left(\frac{\lambda}{2\pi}\right)^{n/2}\left[\prod_{i=1}^{n}\left(\frac{1}{{x_i}^3}\right)^{1/2}\right]\exp\left\{\frac{-\lambda}{2(\frac{1}{n}\sum_{i=1}^{n} x_i)^2}\sum_{i=1}^{n} x_i + \frac{n\lambda}{\frac{1}{n}\sum_{i=1}^{n} x_i} - \frac{\lambda}{2}\sum_{i=1}^{n}\frac{1}{x_i}\right\}\\ &= \left(\frac{\lambda}{2\pi}\right)^{n/2}\left[\prod_{i=1}^{n}\left(\frac{1}{{x_i}^3}\right)^{1/2}\right]\exp\left\{\frac{-\lambda n^2}{2(\sum_{i=1}^{n} x_i)} + \frac{n^2\lambda}{\sum_{i=1}^{n} x_i} - \frac{\lambda}{2}\sum_{i=1}^{n}\frac{1}{x_i}\right\}\\ &= \left(\frac{\lambda}{2\pi}\right)^{n/2}\left[\prod_{i=1}^{n}\left(\frac{1}{{x_i}^3}\right)^{1/2}\right]\exp\left\{\frac{n^2\lambda}{2\sum_{i=1}^{n} x_i} - \frac{\lambda}{2}\sum_{i=1}^{n}\frac{1}{x_i}\right\}.\end{aligned}$$

and

$$\frac{n^2\lambda}{2\sum_{i=1}^{n} x_i} > 0$$

for $\lambda > 0$. Next, with $\mu = \overline{x}_n$, we shall find the value of λ which maximizes the profile log-likelihood function of λ, $\ell(\hat{\mu},\lambda|x_1,x_2,\ldots,x_n)$. Again to locate the extrema, take the partial derivative with respect to λ, set it equal to zero and solve

$$\begin{aligned}
\frac{\partial}{\partial\lambda}\ell(\overline{x}_n,\lambda|x_1,x_2,\ldots,x_n) &= \frac{n}{2\lambda}-\frac{n^2}{2\sum_{i=1}^n x_i}+\frac{n^2}{\sum_{i=1}^n x_i}-\frac{1}{2}\sum_{i=1}^n\frac{1}{x_i}=0\\
\Rightarrow\quad & \frac{n}{\lambda}-\frac{n^2}{\sum_{i=1}^n x_i}+\frac{2n^2}{\sum_{i=1}^n x_i}-\sum_{i=1}^n\frac{1}{x_i}=0\\
\Leftrightarrow\quad & \frac{n}{\lambda}+\frac{n^2}{\sum_{i=1}^n x_i}-\sum_{i=1}^n\frac{1}{x_i}=0\\
\Leftrightarrow\quad & \frac{1}{\lambda}+\frac{n}{\sum_{i=1}^n x_i}-\frac{1}{n}\sum_{i=1}^n\frac{1}{x_i}=0\\
\Leftrightarrow\quad & \frac{1}{\lambda}=\frac{1}{n}\sum_{i=1}^n\frac{1}{x_i}-\frac{n}{\sum_{i=1}^n x_i}\\
\Leftrightarrow\quad & \lambda=\left(\frac{1}{n}\sum_{i=1}^n\frac{1}{x_i}-\frac{1}{\overline{x}_n}\right)^{-1}
\end{aligned}$$

The second partial derivative of $\ell(\overline{x}_n,\lambda|x_1,x_2,\ldots,x_n)$ is

$$\frac{\partial^2}{\partial\lambda^2}\ \ell(\overline{x}_n,\lambda|x_1,x_2,\ldots,x_n)\ =\ -\frac{n}{2\lambda^2}<0\quad\text{for all }\lambda>0.$$

Hence, the function is convex and therefore $\lambda=\left(\frac{1}{n}\sum_{i=1}^n\frac{1}{x_i}-\frac{1}{\overline{x}_n}\right)^{-1}$ maximizes the function. Before reaching a conclusion, we should check the boundaries of the parameter space; that is, the limits of the likelihood function at $\mu=\hat{\mu}$. The limits can be easily observed in the unsimplified likelihood function; that is,

$$L((\hat{\mu},\lambda))=\left[\prod_{i=1}^n\left(\frac{\lambda}{2\pi {x_i}^3}\right)^{1/2}\right]\exp\left\{-\lambda\sum_{i=1}^n\frac{(x_i-\hat{\mu})^2}{2\hat{\mu}^2x_i}\right\}.$$

Thus

$$\lim_{\lambda\to0}L(\overline{x}_n,\lambda|x_1,x_2,\ldots,x_n)=0$$

and

$$\lim_{\lambda\to\infty}L(\overline{x}_n,\lambda|x_1,x_2,\ldots,x_n)=e^{-\infty}=0.$$

Therefore the maximum likelihood estimator $(\hat{\mu},\hat{\lambda})$ of (μ,λ) is $\left(\overline{X}_n,\left(\frac{1}{n}\sum_{i=1}^n\frac{1}{X_i}-\frac{1}{\overline{X}_n}\right)^{-1}\right)$.

□

4. Let $X_1,X_2,\ldots,X_n$ be a random sample from

$$f_X(x|\theta)=\theta(1-\theta)^xI_{\{0,1,2,\ldots\}}(x),\quad 0<\theta<1$$

where $I_A(x)$ is defined as

$$I_A(x)=\begin{cases}1 & \text{if } x\in A,\\ 0 & \text{if } x\in A^c.\end{cases}$$

Let $g(\theta)=(1-\theta)/\theta$.

(a) Find the method of moments estimator and the maximum likelihood estimator of $g(\theta)$.

(b) (Graduate students only.) Assume a Uniform$(0,1)$ prior for θ and find the Bayes estimator of $g(\theta)$.

(c) Is the maximum likelihood estimator unbiased? Why?

(d) Calculate the mean squared error of the maximum likelihood estimator.

(e) (Graduate students only.) Which of the three estimators is preferable?

Solution. (a) We can recognize that $X \sim$ Geometric(θ) where $X =$ the number of failures before the 1st success, so that $E(X) = (1-\theta)/\theta = g(\theta)$. Equating this with the first sample moment $\overline{X}_n$, we get the MOM $\widetilde{g(\theta)}$ of $g(\theta)$ to be $\overline{X}_n$.

The log-likelihood function is given by:

$$\ell(\theta|x_1, x_2, \ldots, x_n) \;=\; n\log\theta + \sum_{i=1}^{n} x_i \cdot \log(1-\theta).$$

Differentiating with respect to θ and setting to zero gives:

$$\begin{aligned} \frac{d}{d\theta}\ell(\theta|x_1, x_2, \ldots, x_n) &= \frac{n}{\theta} - \frac{\sum x_i}{1-\theta} = 0 \;\Rightarrow\; \frac{n}{\theta} = \frac{\sum_{i=1}^{n} x_i}{1-\theta} \\ &\Rightarrow\; \theta = \frac{n}{n + \sum_{i=1}^{n} x_i}. \end{aligned}$$

Since

$$\frac{d^2}{d\theta^2}\ell(\theta|x_1, x_2, \ldots, x_n) = -\frac{n}{\theta^2} - \frac{\sum_{i=1}^{n} x_i}{(1-\theta)^2} < 0$$

$\forall\, \theta \in (0,1)$, then this is the global max of $\ell(\theta|x_1, x_2, \ldots, x_n)$ in $(0,1)$. Now checking the boundary points,

$$\lim_{\theta \to 0} \ell(\theta|x_1, x_2, \ldots, x_n) = \lim_{\theta \to 1} \ell(\theta|x_1, x_2, \ldots, x_n) = -\infty.$$

Hence the MLE of θ is

$$\hat{\theta} = \frac{1}{1 + \overline{X}_n}.$$

By the Invariance Property of MLEs, the MLE for $g(\theta) = (1-\theta)/\theta$ is $g(\hat{\theta})$

$$\widehat{g(\theta)} = g(\hat{\theta}) = \frac{1 - \frac{1}{1+\overline{X}_n}}{\frac{1}{1+\overline{X}_n}} = \overline{X}_n.$$

(b) The posterior distribution of $\theta|\boldsymbol{x}$ is computed as:

$$f(\theta|\boldsymbol{x}) \;\propto\; f_{\mathbf{X}}(\boldsymbol{x}|\theta)\pi(\theta) = \theta^n (1-\theta)^{\sum_{i=1}^{n} x_i} \cdot 1 = \theta^n (1-\theta)^{\sum_{i=1}^{n} x_i}.$$

Recognizing this as a Beta$(n+1, \sum_{i=1}^{n} x_i + 1)$ kernel, we conclude that the posterior distribution of θ given $x_1, x_2, \cdots, x_n$ is Beta$(n+1, \sum_{i=1}^{n} x_i + 1)$. The Bayes' estimator of $g(\theta)$

(for squared error loss) is thus:

$$\widehat{g(\theta)}_B = E\left[g(\theta)|x_1, x_2, \cdots, x_n\right] = \int_0^1 \frac{1-\theta}{\theta} \cdot \frac{\Gamma(n+\sum_{i=1}^n x_i + 2)}{\Gamma(n+1)\Gamma(\sum_{i=1}^n x_i + 1)} \theta^n (1-\theta)^{\sum_{i=1}^n x_i} d\theta$$

$$= \frac{\Gamma(n+\sum_{i=1}^n x_i + 2)}{\Gamma(n+1)\Gamma(\sum_{i=1}^n x_i + 1)} \cdot \frac{\Gamma(n)\Gamma(\sum_{i=1}^n x_i + 2)}{\Gamma(n+\sum_{i=1}^n x_i + 2)} = \frac{(\sum_{i=1}^n x_i + 1)}{n} = \overline{X} + 1/n.$$

(c) Yes. Since $\overline{X}_n$ is unbiased estimator of $\mu = E(X) = (1-\theta)/\theta = g(\theta)$.

(d) Since the MLE is unbiased

$$\text{MSE}_{\overline{X}_n}(\theta) = \text{Var}_\theta(\overline{X}_n) = \sigma^2/n = (1/n)(1-\theta)/\theta^2.$$

(e) Since the MLE and MOM are the same for this problem, it suffices to compare the MLE and the Bayes Estimate. Further,

$$\text{Var}_\theta(\overline{X}_n) = \text{Var}_\theta(\overline{X}_n + 1/n).$$

also,

$$\text{Bias}_{\overline{X}_n+1/n}(\theta) = E_\theta(\overline{X}_n + 1/n) - \theta = E_\theta(\overline{X}_n) + 1/n - \theta = \theta + 1/n - \theta = 1/n.$$

Thus,

$$\begin{aligned}\text{MSE}_{\overline{X}_n+1/n}(\theta) &= \text{Var}_\theta(\overline{X}_n + 1/n) + [\text{Bias}_{\overline{X}_n+1/n}(\theta)]^2 \\ &= \text{Var}_\theta(\overline{X}_n) + (1/n)^2 = \text{MSE}_{\overline{X}_n}(\theta) + (1/n)^2 > \text{MSE}_{\overline{X}_n}(\theta).\end{aligned}$$

Hence the MLE is preferable.

□

5. (Computational Problem.) Suppose the proportion of survivors of melanoma in 10 randomly selected communities of Montana were:

   0.7013, 0.2661, 0.5751, 0.1433, 0.1929, 0.4917, 0.7922, 0.4220, 0.1841 and 0.3096.

   (a) Choose an appropriate model for this data and plot the likelihood function for the parameter of the model and comment on the plot. (attach your R-code)
   (b) Find the method of moments estimator of the model. (Give an explicit analytical formula for the method of moments estimators and the numerical estimates.)
   (c) Find the maximum likelihood estimator of the parameters and the chance of survivor-ship for a randomly selected community is less than 50%. (attach your R-code and output with a brief discussion)
   (d) (Graduate students only.) Conduct a small scale simulation to see if the maximum likelihood estimator of the parameters are unbiased and/or consistent.

Solutions.

(a) Here the observed random variable X is a proportion, hence $0 \leq X \leq 1$. Thus a beta distribution provides an appropriate model for the data. For all observed values of $x_1, x_2, \ldots, x_n$ the likelihood function $L(\alpha, \beta)$ has the form

$$L(\alpha,\beta) = \prod_{i=1}^{n} f_X(x_i|\alpha,\beta) = \prod_{i=1}^{n} \frac{1}{\mathrm{B}(\alpha,\beta)} x_i^{\alpha-1}(1-x_i)^{\beta-1} = \frac{1}{\mathrm{B}(\alpha,\beta)^n}\left(\prod_{i=1}^{n} x_i\right)^{\alpha-1}\left(\prod_{i=1}^{n}(1-x_i)\right)^{\beta-1}.$$

The plot of this likelihood for the given data is given in Figure 1. We clearly see that a global maximum exists. More specifically, the likelihood function appears to take on a maximum value when $\alpha \approx 2.5$ and $\beta \approx 4$.

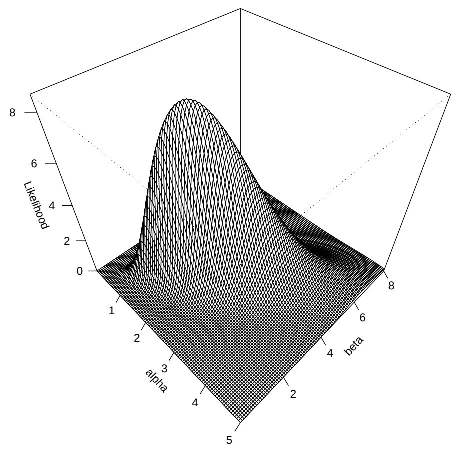

Figure 1: The plot of the Likelihood the Data assuming Beta pdf as the population.

An *R*-script to plot the likelihood is

```
xx<-c(0.7012873,0.2661434,0.5750884,0.1432547,0.1929323,
0.4916651,0.7922205,0.4219847,0.1840799,0.3095818)
BetaL<-function(a,b,x)
{
n<-length(x)
exp(-n*log(beta(a,b))+(a-1)*sum(log(x))+(b-1)*sum(log(1-x)))
}

a<-seq(0.01,5,length=100)
b<-seq(0.01,8,length=100)
Z<-outer(a,b,BetaL,xx)
persp(a,b,Z,theta=45,phi=45,ticktype="detailed",xlab=expression(alpha),
ylab=expression(beta),zlab="Likelihood")
```

(b) To find the method of moments estimator we equate the distribution moments with the sample moments until we can solve for α and β. To that end let us first reparametrize as follows:

$$\phi := \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \theta := \alpha + \beta.$$

Expressing α and β interms of ϕ and θ:

$$\alpha = \theta\phi \quad \text{and} \quad \beta = \theta(1-\phi).$$

Now,

$$E(X) = \frac{\alpha}{\alpha+\beta} = \phi \quad \text{and} \quad \text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\phi(1-\phi)}{\theta+1}.$$

Then equating the population mean and variance with their sample counterparts,

$$\overline{X} = \tilde{\phi} \quad \text{and} \quad S^2 = \frac{\tilde{\phi}(1-\tilde{\phi})}{(\tilde{\theta}+1)},$$

where $S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/n$, and solving for $\tilde{\phi}$ and $\tilde{\theta}$,

$$\tilde{\phi} = \overline{X} \quad \text{and} \quad \tilde{\theta} = \frac{\overline{X}(1-\overline{X})}{S^2} - 1.$$

Finally the MOM estimators for α and β are

$$\tilde{\alpha} = \tilde{\theta}\tilde{\phi} = \left(\frac{\overline{X}(1-\overline{X})}{S^2} - 1\right)\overline{X} \quad \text{and} \quad \tilde{\beta} = \left(\frac{\overline{X}(1-\overline{X})}{S^2} - 1\right)(1-\overline{X}).$$

Using the sample data $\overline{X} = 0.4078$ and $S^2 = 0.0465$, the MOM estimates are

$$\tilde{\alpha} = 1.71 \quad \text{and} \quad \tilde{\beta} = 2.4830.$$

(c) To determine the maximum likelihood estimator of the parameters, we used the `nlm` command for this problem (`optim` command can also be used). First define a function to compute the value of $-\ell(\boldsymbol{\theta}|\mathbf{x})$. As the command preforms non-linear *minimization*, to locate the maximum the negative value of the likelihood function should be returned. We use the MOM estimators as starting values for numerical maximization. The *R*-code to maximize the log-likelihood is

```
##########################
# Beta MLE using nlm()
#####


beta.mlen<- function(xx,a0,b0)
{
negLL<-function(p,x)
{
a<-p[1]
b<-p[2]
n<-length(x)
n*log(beta(a,b))-(a-1)*sum(log(x))-(b-1)*sum(log(1-x))
}
nlm(negLL, p = c(a0, b0), hessian = T, x = xx)
}
beta.mlen(xx,1.71,2.4830)
```

The output from the above code is:

```
 $minimum
[1] -2.154904
$estimate
[1] 1.998044 2.844340
$gradient
[1] -1.814325e-07 -7.597013e-08
$hessian
          [,1]      [,2]
[1,]  4.163664 -2.292760
[2,] -2.292760  1.911319
$code
[1] 1
$iterations
[1] 8
```

The output is rather self explanatory, *in this case* the estimated maximum is -2.155046 at $\hat{\alpha} = 1.9981722$ and $\hat{\beta} = 2.844455$. Gradient is the relative gradient at the estimated maximum of the function. Here the value of code means that relative gradient is close to zero.

Notice that the probability that the survival rate is less than 50% is

$$P(X < 0.5) = g(\alpha, \beta)$$

whose MLE can be obtained by the invariance property the MLE. Thus

$$P(\widehat{X < 0}.5) = g(\hat{\alpha}, \hat{\beta}) = g(1.998172, 2.844455) = 0.6631215.$$

The *R*-code to compute the estimated probability is:

```
pbeta(0.5, 1.998172, 2.844455)
```

(d) Answers vary.

□