Jared Yu
Data Mining

Module 2 Assignment

1. Let x have an exponential distribution

$$p(x|\theta) = \begin{cases} \theta e^{-\theta x}, & x \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

   a. Sketch $p(x|\theta)$ versus x for a fixed value of the parameter θ.

Ans:

The plot of $p(x|\theta)$ versus x is shown below in Figure 1 for $\theta = 0.5, 1, 1.5$. It was created in RStudio, where the range of x-values are from -1 to 5. It shows the exponential distribution and how it is nonzero only for values of $x \geq 0$. There is a solid circle at coordinates $(0{,}0.5), (0, 1), (0{,}1.5)$ and an unfilled circle at coordinates $(0, 0)$ to indicate that $p(x|\theta) > 0$ for $x = 0$. The different values for θ for 0.5, 1, and 1.5 are denoted in blue, red, and black respectively. The plot shows how as θ increases, the density increases near zero, but decreases for larger values of x. Also, as x increases, the density tends to decrease.

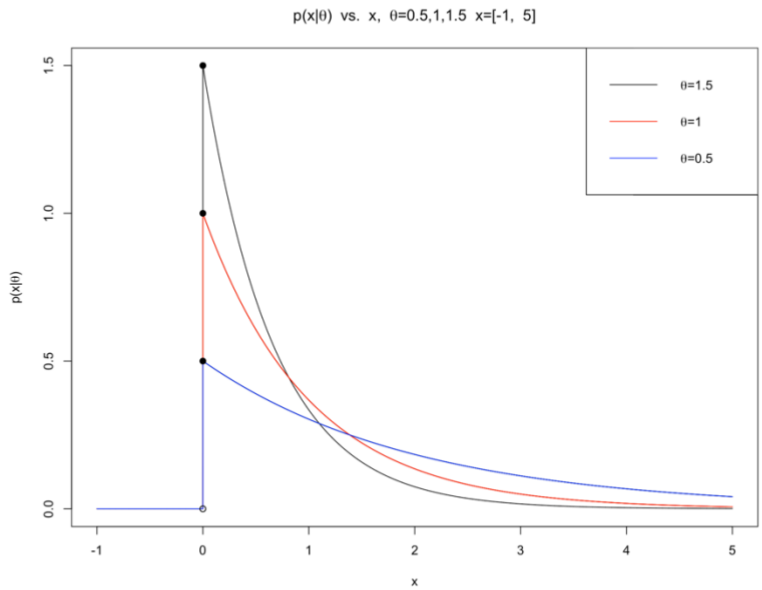

Figure 1 A plot of $p(x|\theta)$ vs. x, where $\theta = 0.5, 1, 1.5$ and $x = [-1, 5]$. At point $(0, 0)$ is an unfilled circle and at points $(0{,}0.5)$, $(0{,}1)$, and $(0{,}1.5)$ there are filled circles. This is to indicate that $p(x|\theta) > 0$ when $x = 0$.

   b. Sketch $p(x|\theta)$ versus θ, $\theta > 0$ for a fixed value of x.

Ans:

The plot of $p(x|\theta)$ versus θ is shown below in Figure 2 for $x = 1{,}2{,}3$. It was created in RStudio, where the range of θ-values are from $(0, 5]$. An unfilled circle at $(0, 0)$ is used to indicate that at that coordinate, the value of $p(x|\theta) = 0$. The different values for x for 1, 2, and 3 are denoted in black, red, and blue respectively. The plot shows how as θ increases, the density tends to peak and the curve down. The point at which it peaks is different depending on the value of x. Furthermore, for larger values of x, the density will be lower for smaller values of θ, before seeming to converge at larger values.

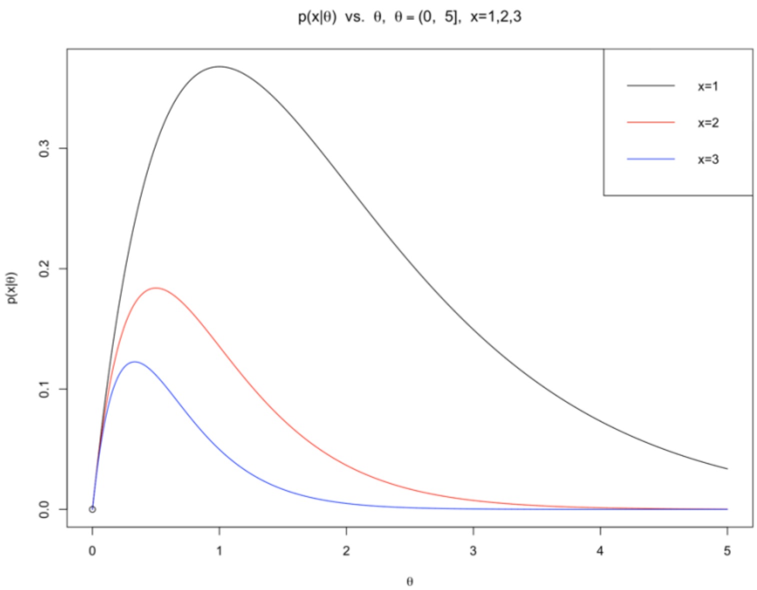

Figure 2 A plot of $p(x|\theta)$ vs. θ, where $x = 1,2,3$ and $\theta = (0, 5]$. An unfilled circle is included at the coordinate (0,0) to indicate that at that point, the value of $p(x|\theta) = 0$.

c. Suppose that n samples $x_1, \cdots, x_n$ are drawn independently according to $p(x|\theta)$. Show that the maximum likelihood estimate for θ is given by

$$\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k}$$

<u>Ans:</u>

Let $X_1, \cdots, X_n \overset{i.i.d.}{\sim} p(x|\theta)$

$$L(\theta) = f(x_1, \cdots, x_n; \theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^{n} x_i} \tag{1.1}$$

Equation (1.1) shows the likelihood formula for the exponential function.

$$l(\theta) = \ln L(\theta) = n \ln \theta - \theta \sum_{i=1}^{n} x_i \tag{1.2}$$

Equation (1.2) shows the log-likelihood version of equation (1).

$$l'(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n} x_i \overset{\text{set to}}{=} 0 \tag{1.3}$$

In equation (1.3), the derivative of $l(\theta)$ is taken and the goal is to find either its minimum or maximum by setting it to 0.

$$\rightarrow \hat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}} \tag{1.4}$$

In equation (1.4), the optimum value of θ is shown to be $\frac{1}{\bar{x}}$. Then, given that at least one X_i is nonzero:

$$l''(\theta) = -\frac{n}{\theta^2} \rightarrow -n\bar{x}^2 < 0 \tag{1.5}$$

In equation (1.5), the second derivative of $l(\theta)$ is taken. This it to confirm whether it is a minimum or maximum. By plugging in the estimate from equation (1.4), it can be seen that this value is always less than zero. This implies that $\hat{\theta}$ is a maximum point and so the MLE is $\boxed{\hat{\theta} = \frac{1}{\bar{x}}}$.

2. Let x have a uniform distribution

$$p(x|\theta) = \begin{cases} \frac{1}{\theta}, & 0 \leq x \leq \theta \\ 0, & \text{otherwise.} \end{cases}$$

   a. Sketch $p(x|\theta)$ versus θ for an arbitrary value of x.

Ans:
The plot of $p(x|\theta)$ versus θ is shown below in Figure 3 for $x = 3$, 5, and 7. It was created in RStudio, where the range of θ-values are from $[2, 10]$. There are solid circles at coordinates $(3, 0.3333)$, $(5,0.2)$, and $(7,0.1428)$. There are unfilled circles at coordinates $(3, 0)$, $(5,0)$, and $(7,0)$. These are to indicate that $p(x|\theta) = \frac{1}{3}, \frac{1}{5}, \frac{1}{7}$ for $x = 3, 5, 7$ respectively. The plot shows that for a fixed x, the density is maximum when $\theta = x$ and starts to decrease as θ increases.

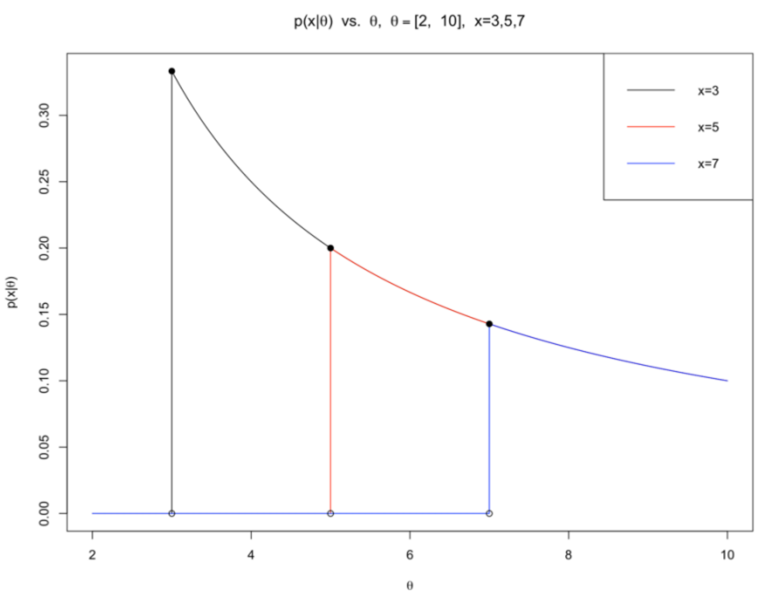

Figure 3 A plot of $p(x|\theta)$ vs. θ, where $x = 3,5,7$ and $\theta = [2, 7]$. There are filled circles at coordinates $(3,0.3333)$, $(5,0.2)$, and $(7,0.1428)$. There are unfilled circles at coordinates $(3,0)$, $(5,0)$, and $(7,0)$. This is to indicate that $p(x|\theta) = \frac{1}{3}, \frac{1}{5}, \frac{1}{7}$ when $x = 3,5,7$ respectively.

   b. Suppose that n samples $x_1, \cdots, x_n$ are drawn independently according to $p(x|\theta)$. Show that the maximum likelihood estimate for θ is $\max_k x_k$.

Ans: (used reference [1])
Let $X_1, \cdots, X_n \overset{i.i.d.}{\sim} p(x|\theta)$.
The support of $p(x|\theta)$, $0 \leq x \leq \theta$, contains the parameter θ, therefore the method of finding the log likelihood will not work. Instead, it is possible to look at the following joint probability density function (p.d.f.),

$$f_X(x_1, \cdots, x_n; \theta) = \prod_{i=1}^{n} \frac{1}{\theta} 1_{\{x_i \le \theta\}} = \frac{1}{\theta^n} \prod_{i=1}^{n} 1_{\{x_i \le \theta\}},$$

where $1_{\{x_i \le \theta\}}$ is an indicator function defined as follows,

$$1_{\{0 \le x_i \le \theta\}} = \begin{cases} 1, & x_i \le \theta \\ 0, & \text{otherwise.} \end{cases}$$

When $\theta < x_{(n)}$, where $x_{(n)}$ is the $n'th$ order statistic from the sample, then $f_X = 0$ because that would violate the rules of the support. Therefore, the only case considered is $\theta \ge x_{(n)}$. Then the joint p.d.f. will become,

$$f_X(x_1, \cdots, x_n; \theta) = \frac{1}{\theta^n} 1_{\{x_{(n)} \le \theta\}}.$$

It can be seen that when θ decreases, $f_X(x_1, \cdots, x_n; \theta)$ increases, therefore to maximize $f_X(x_1, \cdots, x_n; \theta)$ requires choosing the minimum value for θ, where $\theta \in [x_{(n)}, \infty)$. $\therefore$ The MLE is $\boxed{\hat{\theta} = x_{(n)}}$, where $x_{(n)}$ is the largest value in the sample, or $\boxed{\max_k x_k}$.

c. Find the method of moments estimator for θ.

Ans: (used reference [2])

Let $X_1, \cdots, X_n \overset{i.i.d.}{\sim} p(x|\theta)$.

$$\alpha_1 = \alpha_1(\theta) = E_\theta(X^1) = \int_0^\theta x^1 dF_\theta(x) = \int_0^\theta x f_\theta(x) dx = \int_0^\theta \frac{x}{\theta} dx = \frac{1}{2\theta} x^2 |_0^\theta = \frac{\theta}{2} \qquad (2.1)$$

In equation (2.1), the first theoretical moment is derived. Only one is needed, since $p(x|\theta)$ only has a single parameter.

$$\hat{\alpha}_1 = \frac{1}{n} \sum_{i=1}^{n} X_i^1 \qquad (2.2)$$

In equation (2.2), the first sample moment is calculated.

$$E_\theta(X^1) = \frac{\theta}{2} = \frac{1}{n} \sum_{i=1}^{n} X_i^1 \qquad (2.3)$$

Equation (2.3) comes from equation the first theoretical moment about the origin with the first sample moment.

$$\boxed{\hat{\theta} = \frac{2}{n} \sum_{i=1}^{n} X_i = 2\bar{X}_n} \qquad (2.4)$$

In Equation (2.4), it shows that the method of moments estimator for θ is $2\bar{X}_n$.

3. Let $\mathbf{x}$ be a binary $(0,1)$ vector with multivariate Bernoulli distribution

$$p(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i} (1 - \theta_i)^{1 - x_i},$$

where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d)^\top$ is an unknown parameter vector, θ_i being the probability that $x_i = 1$. Show that the maximum likelihood estimate for $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k.$$

Ans: (used reference [3])

Let $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n \overset{i.i.d.}{\sim} p(\mathbf{x}|\boldsymbol{\theta})$.

$$L(\boldsymbol{\theta}) = f_{\boldsymbol{X}}(\mathbf{x}_1, \cdots, \mathbf{x}_n; \boldsymbol{\theta}) = \prod_{j=1}^{n} \prod_{i=1}^{d} \theta_i^{x_{ij}} (1-\theta_i)^{1-x_{ij}} = \prod_{i=1}^{d} \theta_i^{\sum_{j=1}^{n} x_{ij}} (1-\theta_i)^{n-\sum_{j=1}^{n} x_{ij}} \quad (3.1)$$

Equation (3.1) shows the likelihood function for $p(\mathbf{x}|\boldsymbol{\theta})$.

$$l(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) = \ln\left[\prod_{i=1}^{d} \theta_i^{\sum_{j=1}^{n} x_{ij}} (1-\theta_i)^{n-\sum_{j=1}^{n} x_{ij}}\right] = \ln\left[\prod_{i=1}^{d} \theta_i^{\sum_{j=1}^{n} x_{ij}} \prod_{i=1}^{d} (1-\theta_i)^{n-\sum_{j=1}^{n} x_{ij}}\right] (3.2)$$

$$= \ln\left[\prod_{i=1}^{d} \theta_i^{\sum_{j=1}^{n} x_{ij}}\right] + \ln\left[\prod_{i=1}^{d} (1-\theta_i)^{n-\sum_{j=1}^{n} x_{ij}}\right] = \sum_{i=1}^{d} \ln\left[\theta_i^{\sum_{j=1}^{n} x_{ij}}\right] + \sum_{i=1}^{d} \ln\left[(1-\theta_i)^{n-\sum_{j=1}^{n} x_{ij}}\right] (3.3)$$

$$= \sum_{i=1}^{d}\left[\sum_{j=1}^{n} x_{ij} \ln(\theta_i)\right] + \sum_{i=1}^{d}\left[\left(n - \sum_{j=1}^{n} x_{ij}\right) \ln(1-\theta_i)\right] \quad (3.4)$$

Equations (3.2) to (3.4) show the log-likelihood of equation (3.1).

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_i} = \sum_{j=1}^{n} \frac{x_{ij}}{\theta_i} + \left(n - \sum_{j=1}^{n} x_{ij}\right)\left(-\frac{1}{1-\theta_i}\right) = \frac{n\bar{x}_i}{\theta_i} - \frac{n(1-\bar{x}_i)}{1-\theta_i}, \text{for } i = 1, \cdots, d, \quad (3.5)$$

where $\bar{x}_i = \frac{\sum_{j=1}^{n} x_{ij}}{n}$. Equation (3.5) shows the partial derivative of equation (3.4) w.r.t. a single θ_i, for $i = 1, \cdots, d$. Then, setting $\frac{\partial l(\boldsymbol{\theta})}{\partial \theta_i} = 0$, for $i = 1, \cdots, d$, it is possible to find the optimum value for each θ_i.

$$\frac{n\,\bar{x}_i}{\theta_i} - \frac{\mathrm{n}(1-\bar{x}_i)}{1-\theta_i} = 0 \quad (3.6)$$

$$\frac{n\,\bar{x}_i}{\theta_i} = \frac{\mathrm{n}(1-\bar{x}_i)}{1-\theta_i} \quad (3.7)$$

$$n\,\bar{x}_i(1-\theta_i) = \mathrm{n}(1-\bar{x}_i)\theta_i \quad (3.8)$$

$$\bar{x}_i - \bar{x}_i\theta_i = \theta_i - \bar{x}_i\theta_i \quad (3.9)$$

$$\hat{\theta}_i = \bar{x}_i, \text{for } i = 1, \cdots, d \quad (3.10)$$

Equations (3.6) to (3.10) show that the optimum value, $\hat{\theta}_i = \bar{x}_i$ for $i = 1, \cdots, d$. To prove that it is a maximum, the Hessian matrix must be shown to be negative definite.

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_i^2} = -\frac{n\bar{x}_i}{\theta_i^2} - \frac{n(1-\bar{x}_i)}{(1-\theta_i)^2} \quad (3.11)$$

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_i} = \frac{\partial}{\partial \theta_j}\left[\frac{n\bar{x}_i}{\theta_i} - \frac{n(1-\bar{x}_i)}{1-\theta_i}\right] = 0 \quad (3.12)$$

Equation (3.11) and (3.12) show the second derivatives for the diagonal and off-diagonal elements.

$$\boldsymbol{H} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top \partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_1^2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_2 \partial \theta_1} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_d \partial \theta_1} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_2^2} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_d \partial \theta_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_d} & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_2 \partial \theta_d} & \cdots & \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \theta_d^2} \end{bmatrix} \tag{3.13}$$

$$= \begin{bmatrix} -\frac{n\bar{x}_1}{\theta_1^2} - \frac{n(1-\bar{x}_1)}{(1-\theta_1)^2} & 0 & \cdots & 0 \\ 0 & -\frac{n\bar{x}_2}{\theta_2^2} - \frac{n(1-\bar{x}_2)}{(1-\theta_2)^2} \quad 0 & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\frac{n\bar{x}_d}{\theta_d^2} - \frac{n(1-\bar{x}_d)}{(1-\theta_d)^2} \end{bmatrix} \tag{3.14}$$

In equations (3.13), the second derivatives are seen in the Hessian matrix denoted $\boldsymbol{H}$. In equation (3.14), the values are inserted, where only the diagonal elements have distinctly nonzero values.

Let the vector $\boldsymbol{z} = [z_1 \cdots z_d]^\top$, $\boldsymbol{z} \neq \boldsymbol{0}$ and $z_i \in \mathbb{R}$.

$$\boldsymbol{z}^\top \boldsymbol{H} \boldsymbol{z} = \sum_{i=1}^{d} \left(-\frac{n\bar{x}_i}{\theta_i^2} - \frac{n(1-\bar{x}_i)}{(1-\theta_i)^2} \right) z_i^2 = -\sum_{i=1}^{d} \left(\frac{n\bar{x}_i}{\theta_i^2} + \frac{n(1-\bar{x}_i)}{(1-\theta_i)^2} \right) z_i^2 < 0 \tag{3.15}$$

Equation (3.15) then shows that $\boldsymbol{z}^T \boldsymbol{H} \boldsymbol{z} < 0$, which means that the Hessian $\boldsymbol{H}$ is negative-definite. So, the MLE of θ_i is $\hat{\theta}_i = \bar{x}_i = \frac{\sum_{j=1}^{n} x_{ij}}{n}$, for $i = 1, \cdots, d$. In other words, the MLE of $\boldsymbol{\theta}$ is as follows,

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_d \end{bmatrix} = \begin{bmatrix} \frac{\sum_{j=1}^{n} x_{1j}}{n} \\ \frac{\sum_{j=1}^{n} x_{2j}}{n} \\ \vdots \\ \frac{\sum_{j=1}^{n} x_{dj}}{n} \end{bmatrix} = \frac{1}{n} \sum_{j=1}^{n} \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{dj} \end{bmatrix} = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k \, . \tag{3.16}$$

This completes the proof of this problem. ■

4. Let x have a Gamma distribution

$$p(x|\alpha,\beta) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & x > 0 \text{ and } \alpha, \beta > 0 \\ 0, & \text{otherwise.} \end{cases}$$

   a. Suppose that n samples $x_1, \cdots, x_n$ are drawn independently according to $p(x|\alpha,\beta)$. Find the method of moments estimator for α and β.

Ans: (used reference [4], [5], [6], [7])

The first step will be to find the moment generating function (M.G.F.) for the Gamma distribution.

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_{-\infty}^{\infty} x^{\alpha-1} e^{-x\left(\frac{1}{\beta}-t\right)} dx \quad (4.1)$$

Equation (4.1) shows the formula for the M.G.F. along with plugging in the p.d.f. of the Gamma distribution. It is important to note that the integral is finite only when $\frac{1}{\beta} - t > 0$, otherwise the integrand will increase towards infinity.

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_{-\infty}^{\infty} x^{\alpha-1} e^{-\frac{x}{\lambda}} dx \text{ where } \frac{1}{\lambda} = \left(\frac{1}{\beta} - t\right) \quad (4.2)$$

In equation (4.2), the $\frac{1}{\lambda}$ is used to replace the part in the exponential function, so it can be seen that the integrand is similar to another $Gamma(\alpha, \lambda)$ p.d.f.

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha} [\Gamma(\alpha)\lambda^\alpha] = \frac{\lambda^\alpha}{\beta^\alpha} = \left(\frac{\lambda}{\beta}\right)^\alpha \quad (4.3)$$

In equation (4.3), the previous $\int_{-\infty}^{\infty} x^{\alpha-1} e^{-\frac{x}{\lambda}} dx$ is integrated to 1, however since it lacks the constant term $\frac{1}{\Gamma(\alpha)\lambda^\alpha}$, $\Gamma(\alpha)\lambda^\alpha$ will be left as a constant.

$$\frac{1}{\lambda} = \frac{1}{\beta} - t = \frac{1 - \beta t}{\beta}; \lambda = \frac{\beta}{1 - \beta t} \quad (4.4)$$

In equation (4.4), the $\frac{1}{\lambda}$ term is rewritten so that it is more convenient to substitute back into equation (4.3).

$$\left(\frac{\lambda}{\beta}\right)^\alpha = \left(\frac{\beta}{1 - \beta t} \cdot \frac{1}{\beta}\right)^\alpha = \left(\frac{1}{1 - \beta t}\right)^\alpha, \forall\, t < \frac{1}{\beta} \quad (4.5)$$

In equation (4.5), the λ term is substituted back into equation (4.4). It is then seen that the M.G.F. for the $Gamma(\alpha, \beta) = \left(\frac{1}{1-\beta t}\right)^\alpha, \forall\, t < \frac{1}{\beta}$.

The next step is to find the first and second theoretical moments of $p(x|\alpha, \beta)$ by using the M.G.F. derived in equation (4.5).

$$M_X'(t) = \frac{d}{dt}\left(\frac{1}{1 - \beta t}\right)^\alpha = \frac{d}{dt}(1 - \beta t)^{-\alpha} = -\alpha(1 - \beta t)^{-(\alpha+1)}(-\beta) = \alpha\beta(1 - \beta t)^{-(\alpha+1)} \quad (4.6)$$

$$M_X'(0) = \alpha\beta\big(1 - \beta(0)\big)^{-(\alpha+1)} = \alpha\beta \quad (4.7)$$

Equation (4.6) shows the first derivative of the M.G.F. Equation (4.7) shows that substituting $t = 0$ evaluates to the first theoretical moment being $\mu_1 = E(X) = \alpha\beta$.

$$M_X''(t) = \frac{d}{dt}\alpha\beta(1 - \beta t)^{-(\alpha+1)} = -(\alpha + 1)\alpha\beta(1 - \beta t)^{-(\alpha+2)}(-\beta) = (\alpha + 1)\alpha\beta^2(1 - \beta t)^{-(\alpha+2)} \quad (4.8)$$

$$M_X''(0) = (\alpha + 1)\alpha\beta^2\big(1 - \beta(0)\big)^{-(\alpha+2)} = (\alpha + 1)\alpha\beta^2 \quad (4.9)$$

Equation (4.8) shows the second derivative of the M.G.F. Equation (4.9) shows that substituting $t = 0$ evaluates to the second theoretical moment being $\mu_2 = E(X^2) = (\alpha + 1)\alpha\beta^2$.

The next steps are to find values for $\hat{\alpha}$ and $\hat{\beta}$ based on the sample moments.

$$\mu_1 = \alpha\beta \to \beta = \frac{\mu_1}{\alpha} \quad (4.10)$$

$$\mu_2 = (\alpha + 1)\alpha\left(\frac{\mu_1}{\alpha}\right)^2 = \frac{\alpha + 1}{\alpha}\mu_1^2 \quad (4.11)$$

$$\rightarrow 1 + \frac{1}{\alpha} = \frac{\mu_2}{\mu_1^2} \rightarrow \frac{1}{\alpha} = \frac{\mu_2 - \mu_1^2}{\mu_1^2} \rightarrow \alpha = \frac{\mu_1^2}{\mu_2 - \mu_1^2} \tag{4.12}$$

$$\rightarrow \beta = \mu_1 \left(\frac{\mu_1^2}{\mu_2 - \mu_1^2} \right) = \frac{\mu_2 - \mu_1^2}{\mu_1} \tag{4.13}$$

$$\boxed{\hat{\alpha} = \frac{\bar{X}^2}{\hat{\sigma}^2}, \hat{\beta} = \frac{\hat{\sigma}^2}{\bar{X}}} \tag{4.14}$$

where

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

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\bar{X}^2 + \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 = \hat{\mu}_2 - \hat{\mu}_1^2 \tag{4.15}$$

In equations (4.10) to (4.13), the goal is to solve for α and β based on μ_1 and μ_2. Equation (4.15) is to show that $\hat{\sigma}^2 = \hat{\mu}_2 - \hat{\mu}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Where $\hat{\mu}_1$ and $\hat{\mu}_2$ are the first and second sample moments. These values were also plugged in at the end in equation (4.14). This is analogous to showing that $Var(X) = E[X^2] - (E[X])^2$.

b. Show that the exponential distribution is $\Gamma(1, 1/\theta)$.

Ans:

The Gamma distribution has the following p.d.f.,

$$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, x > 0 \text{ and } \alpha, \beta > 0.$$

Plugging in $\alpha = 1$ and $\beta = \frac{1}{\theta}$ leads to the following,

$$\frac{1}{\Gamma(1)\left(\frac{1}{\theta}\right)^{(1)}} x^{(1)-1} e^{-\frac{x}{\left(\frac{1}{\theta}\right)}} = \theta e^{-x\theta},$$

which is equal to the exponential distribution. The constraints of $x \geq 0$ are slightly different, but the probability is still 0 at $x = 0$ so it does not make a significant difference.

References

[1] http://www2.stat.duke.edu/~banks/111-lectures.dir/lect10.pdf

[2] https://online.stat.psu.edu/stat415/lesson/1/1.4

[3] https://en.wikipedia.org/wiki/Definite_symmetric_matrix

[4] https://www.youtube.com/watch?v=TePh29vzVEk

[5] https://www.youtube.com/watch?v=-elod4SsOts

[6] https://www.stat.berkeley.edu/~vigre/activities/bootstrap/2006/wickham_stati.pdf

[7] http://www2.econ.iastate.edu/classes/econ500/hallam/documents/Sample_Moments.pdf

Code Appendix

```
library(latex2exp)
### 1
### a
exponential_distribution <- function(x, theta) {
  ifelse(x >= 0,
         theta * exp(-theta * x),
         0)
}

xs <- seq(-1, 5, length.out = 1e4)
ys <- exponential_distribution(x = xs, theta = 1)

thetas <- seq(0.5, 1.5, length.out = 3)
y_vec <- sapply(X = thetas, FUN = function(x)
  exponential_distribution(x = xs, theta = x))

plot(xs, y_vec[,3], type = 'l',
     main = TeX('$p(x|\\theta)\\;vs.\\;x,\\;\\theta =0.5,1,1.5\\;x=\\[-1,\\;5\\]$'),
     xlab = TeX('$x$'), ylab = TeX('$p(x|\\theta)$'))
lines(xs, y_vec[,2], col = 'red'); lines(xs, y_vec[,1], col = 'blue')
points(0, 0, pch=1)
points(0, 0.5, pch=19)
points(0, 1, pch=19)
points(0, 1.5, pch=19)
legend("topright",
       legend = c(TeX('$\\theta =1.5$'), TeX('$\\theta =1$'), TeX('$\\theta =0.5$')),
       col = c("black", "red", "blue"), lty = rep(1,3))
### b
thetas <- seq(0, 5, length.out = 1e4)
exp_dist_vec <- Vectorize(exponential_distribution, vectorize.args = "theta")

y_vec <- sapply(X = seq(1,3,length.out = 3), FUN = function(x)
  exp_dist_vec(x = x, theta = thetas))

plot(thetas, y_vec[,1], type = 'l',
     main = TeX('$p(x|\\theta)\\;vs.\\;\\theta,\\;\\theta = (0,\\;5\\],\\;x=1, 2, 3$'),
     xlab = TeX('$\\theta$'), ylab = TeX('$p(x|\\theta)$'))
lines(thetas, y_vec[,2], col = 'red')
lines(thetas, y_vec[,3], col = 'blue')
points(0, 0, pch=1)
legend("topright", legend = c(TeX('$x=1$'), TeX('$x=2$'), TeX('$x=3$')),
       col = c('black', 'red', 'blue'), lty = rep(1,3))

### 2
### a
uniform_distribution <- function(x, theta) {
  ifelse((0<=x) & (x<=theta),
         1 / theta,
         0)
}

# xs <- seq(-1, 6, length.out = 1e4)
# ys <- uniform_distribution(x = xs, theta = 5)
# plot(xs, ys, type = 'l')
# points(0, 0, pch=1); points(0, 0.2, pch=16)
# points(5, 0, pch=1); points(5, 0.2, pch=16)

thetas <- seq(2, 10, length.out = 1e4)
ys1 <- uniform_distribution(x = 3, theta = thetas)
ys2 <- uniform_distribution(x = 5, theta = thetas)
ys3 <- uniform_distribution(x = 7, theta = thetas)

y_vec <- sapply(X = seq(1,3,length.out = 3), FUN = function(x)
  uniform_distribution(x = x, theta = thetas))

plot(thetas, ys1, type = 'l',
     main = TeX('$p(x|\\theta)\\;vs.\\;\\theta,\\;\\theta = \\[2,\\;10\\],\\;x=3,5,7$'),
```

```r
     xlab = TeX('$\\theta$'), ylab = TeX('$p(x|\\theta)$'))
lines(thetas, ys2, col = 'red')
lines(thetas, ys3, col = 'blue')
points(3, 0, pch=1); points(3, max(ys1), pch=19)
points(5, 0, pch=1); points(5, max(ys2), pch=19)
points(7, 0, pch=1); points(7, max(ys3), pch=19)
legend("topright", legend = c(TeX('$x=3$'), TeX('$x=5$'), TeX('$x=7$')),
       col = c('black', 'red', 'blue'), lty = rep(1,3))
```