MATH-503 – Assignment 3
Sample Solution

1. **Let $X_1, \ldots, X_n$ be iid random variables from**

$$f(x|\theta) = \theta(\theta+1)x^{\theta-1}(1-x), \qquad 0 \le x \le 1, \quad \theta > 0$$

(a) **Find the method of moments estimator of θ.**

$$\begin{aligned}
E[X] &= \int_0^1 x\theta(\theta+1)x^{\theta-1}(1-x)dx = \int_0^1 \theta(\theta+1)x^{\theta}(1-x)dx \\
&= \theta(\theta+1)\int_0^1 x^{\theta} - x^{\theta+1}dx = \theta(\theta+1)\left.\frac{x^{\theta+1}}{\theta+1} - \frac{x^{\theta+2}}{\theta+2}\right|_0^1 \\
&= \theta(\theta+1)\left(\frac{1}{\theta+1} - \frac{1}{\theta+2}\right) = \frac{\theta}{\theta+2}
\end{aligned}$$

Setting the expectation equal to the first sample moment:

$$\frac{\theta}{\theta+2} = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X} \;\Rightarrow\; \theta = (\theta+2)\bar{X} \qquad \Rightarrow \qquad \theta(1-\bar{X}) = 2\bar{X}$$

$$\Rightarrow \; \hat{\theta}_{MOM} = \frac{2\bar{X}}{1-\bar{X}}$$

(b) **Find the maximum likelihood estimator (MLE) of θ.**

$$\begin{aligned}
L(\theta|\boldsymbol{x}) &= \theta^n(\theta+1)^n\prod_{i=1}^n x_i^{\theta-1}\prod_{i=1}^n(1-x_i) \\
\log L(\theta|\boldsymbol{x}) &= n\log\theta + n\log(\theta+1) + (\theta-1)\sum_{i=1}^n \log x_i + \sum_{i=1}^n \log(1-x_i) \\
\frac{d}{d\theta}\log L(\theta|\boldsymbol{x}) &= \frac{n}{\theta} + \frac{n}{\theta+1} + \sum_{i=1}^n \log x_i
\end{aligned}$$

$$\frac{d}{d\theta}\log L(\theta|\boldsymbol{x}) = 0 \;\Rightarrow\; n(2\theta+1) = -\theta(\theta+1)\sum_{i=1}^n \log x_i \;\Rightarrow\; 2\theta+1 = -(\theta^2+\theta)\frac{1}{n}\sum_{i=1}^n \log x_i$$

Let $C = \frac{1}{n}\sum_{i=1}^n \log x_i$. We need to solve the quadratic equation

$$\theta^2 C + \theta(2+C) + 1 = 0 \quad \Rightarrow \quad \theta = \frac{-(2+C) \pm \sqrt{(2+C)^2 - 4C}}{2C} = \frac{-(2+C) \pm \sqrt{4+C^2}}{2C}$$

Since $\theta > 0$, we have to pick which of the solutions is a valid MLE.
Thanks to Brett Newbold for this part to show the valid solution:

$$(2+C)^2 = 4 + 4C + C^2 < 4 + C^2 \qquad \text{since } \; C < 0$$

Taking the square roots, we have

$$\begin{aligned} 2+C &< \sqrt{4+C^2} \qquad \text{or} \qquad 2+C > -\sqrt{4+C^2} \\ \Rightarrow 0 &< -(2+C)+\sqrt{4+C^2} \qquad \text{or} \qquad 0 > -(2+C)-\sqrt{4+C^2} \\ \Rightarrow \frac{-(2+C)+\sqrt{4+C^2}}{2C} &< 0 \qquad \text{or} \qquad \tfrac{-(2+C)-\sqrt{4+C^2}}{2C} > 0 \end{aligned}$$

Thus,

$$\hat{\theta}_{MLE} = \frac{-(2+C)-\sqrt{4+C^2}}{2C} \qquad \text{where } \ C = \frac{1}{n}\sum_{i=1}^{n} \log x_i$$

Check: The second derivative of the log-likelihood is

$$\frac{d^2}{d\theta^2} \log L(\theta|\boldsymbol{x}) = -\frac{n}{\theta^2} - \frac{n}{(\theta+1)^2} < 0$$

therefore, we have found a maximizer of the log-likelihood.

2. **Let $X_1, \ldots, X_n$ be discrete iid random variables from**

$$f(x|\alpha,\beta) = (1-\beta)\beta^{x-\alpha} I_{x\in\{\alpha,\alpha+1,\ldots\}}$$

with $-\infty < \alpha < \infty$ and $0 < \beta < 1$. Let $\theta = (\alpha,\beta)$.

(a) **Find a sufficient statistic for θ.**

The joint pmf is given by:

$$f(\boldsymbol{x}|\alpha,\beta) = \frac{(1-\beta)^n}{\beta^{n\alpha}} \beta^{\sum_{i=1}^n x_i} I_{(x_{(1)}\in\{\alpha,\alpha+1,\ldots\})}$$

Using the factorization theorem, the sufficient statistic is $T(\boldsymbol{X}) = \big(X_{(1)}, \sum_{i=1}^n X_i\big)$.

(b) **Find the MLE of θ.**

$$\mathcal{L}(\alpha,\beta|\boldsymbol{X}) = \frac{(1-\beta)^n}{\beta^{n\alpha}} \beta^{\sum_{i=1}^n x_i} I_{\{\alpha \le x_{(1)}\}}$$

For fixed $\beta \in (0,1)$, the function $\frac{C}{\beta^{n\alpha}}$ is an increasing function of α. So, the likelihood achieves its maximum at the largest value of α, i.e.,

$$\hat{\alpha} = X_{(1)}$$

On the interval $\alpha \le X_{(1)}$ over which the likelihood function is defined, the log-likelihood is given by

$$\begin{aligned} \log \mathcal{L}(\alpha,\beta|\boldsymbol{X}) &= n\log(1-\beta) - n\alpha\log\beta + \sum_{i=1}^n x_i \log\beta \\ \frac{\partial}{\partial\beta} \log \mathcal{L}(\alpha,\beta|\boldsymbol{X}) &= -\frac{n}{1-\beta} + \frac{\sum_{i=1}^n x_i - n\alpha}{\beta} \end{aligned}$$

Solving for $\frac{\partial}{\partial\beta}\log\mathcal{L}(\alpha,\beta|\boldsymbol{X})=0$

$$\frac{n}{1-\beta}=\frac{\sum_{i=1}^n x_i-n\alpha}{\beta}\quad\Rightarrow\quad\frac{1-\beta}{\beta}=\frac{n}{\sum_{i=1}^n x_i-n\alpha}\quad\Rightarrow\quad\frac{1}{\beta}-1=\frac{1}{\bar{x}-\alpha}$$
$$\frac{1}{\beta}=\frac{1+\bar{x}-\alpha}{\bar{x}-\alpha}\quad\Rightarrow\quad\hat{\beta}=\frac{\bar{x}-x_{(1)}}{1+\bar{x}-x_{(1)}}$$

Therefore, $\hat{\theta}=\left(X_{(1)},\frac{\bar{X}-X_{(1)}}{1+\bar{X}-X_{(1)}}\right)$.

(c) **Find the MLE of $P(X\leq\alpha+1)$.**

$P(X\leq\alpha+1)=P(X=\alpha)+P(X=\alpha+1)=(1-\beta)\beta^0+(1-\beta)\beta=(1-\beta)(1+\beta)=1-\beta^2$

By the invariance property of the MLE,

$$\widehat{P(X\leq\alpha+1)}=1-\hat{\beta}^2=1-\left(\frac{\bar{X}-X_{(1)}}{1+\bar{X}-X_{(1)}}\right)^2=\frac{1+2(\bar{X}-X_{(1)})}{\left(1+\bar{X}-X_{(1)}\right)^2}$$

3. **Let $(x_1,Y_1),(x_2,Y_2),\ldots,(x_n,Y_n)$ be a set of points satisfying a linear regression model**

$$y_i=\alpha+\beta x_i+\varepsilon_i,\qquad\varepsilon_i\sim N(0,\sigma^2),\quad i=1,\ldots,n$$

That is,

$$Y_i\overset{iid}{\sim}N\left(\alpha+\beta x_i,\sigma^2\right)$$

(a) **Find the maximum likelihood estimators for α,β and σ^2.**

Let $\boldsymbol{\theta}=(\alpha,\beta,\sigma^2)$.

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{Y}) &= \prod_{i=1}^n\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(y_i-[\alpha+\beta x_i])^2\right\}\right]\\
&= (2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-[\alpha+\beta x_i])^2\right\}\\
\log\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{Y}) &= -\frac{n}{2}\log(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^n(y_i-[\alpha+\beta x_i])^2\\
\frac{\partial}{\partial\alpha}\log\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{Y}) &= \frac{1}{\sigma^2}\sum_{i=1}^n(y_i-[\alpha+\beta x_i])\\
\frac{\partial}{\partial\beta}\log\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{Y}) &= \frac{1}{\sigma^2}\sum_{i=1}^n x_i(y_i-[\alpha+\beta x_i])\\
\frac{\partial}{\partial\sigma^2}\log\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{X},\boldsymbol{Y}) &= -\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^n(y_i-[\alpha+\beta x_i])^2
\end{aligned}$$

Setting the partial derivatives to 0:

$$(1)\quad\sum_{i=1}^n(y_i-[\alpha+\beta x_i])=0\quad\Rightarrow\quad\sum_{i=1}^n y_i-n\alpha-\beta\sum_{i=1}^n x_i=0$$

$$(2) \quad \sum_{i=1}^{n} x_i(y_i - [\alpha + \beta x_i]) = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} x_i y_i - \alpha \sum_{i=1}^{n} x_i - \beta \sum_{i=1}^{n} x_i^2 = 0$$

$$(3) \quad \frac{1}{2\sigma^2}\left[-n + \frac{\sum_{i=1}^{n}(y_i - [\alpha + \beta x_i])^2}{\sigma^2}\right] = 0 \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - [\hat{\alpha} + \hat{\beta} x_i])^2}{n}$$

$$(1) \quad \hat{\alpha} = \frac{\sum_{i=1}^{n} y_i - \beta \sum_{i=1}^{n} x_i}{n} \qquad \Rightarrow \qquad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$$

$$(2) \qquad \sum_{i=1}^{n} x_i y_i - \left(\frac{\sum_{i=1}^{n} y_i - \beta \sum_{i=1}^{n} x_i}{n}\right) \sum_{i=1}^{n} x_i - \beta \sum_{i=1}^{n} x_i^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i + \frac{\beta}{n}\left(\left(\sum_{i=1}^{n} x_i\right)^2 - n \sum_{i=1}^{n} x_i^2\right) = 0$$

$$\Rightarrow \hat{\beta} = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i - n \sum_{i=1}^{n} x_i y_i}{(\sum_{i=1}^{n} x_i)^2 - n \sum_{i=1}^{n} x_i^2}$$

(b) **The following table shows Starbuck's annual net revenue (x) and the cost of operating the stores (y) that generated these revenues (in millions of dollars).**

Year	Net revenue (x)	Cost of sales (y)
1999	1687	748
2000	2178	962
2001	2649	1113
2002	3289	1350
2003	4076	1686
2004	5294	2199
2005	6369	2605
2006	7787	3179
2007	9411	3999

i. **Provide a scatterplot of the data (y versus x) and describe the relationship.**

```
x = c(1687, 2178, 2649, 3289, 4076, 5294, 6369, 7787, 9411)
y = c(748, 962, 1113, 1350, 1686, 2199, 2605, 3179, 3999)
plot(x, y)
```

We note a strong positive linear association between net revenue and cost of sales.

ii. **Provide maximum likelihood estimates of α, β and σ^2.**

```
beta.hat = (sum(y)*sum(x) - n*sum(x*y)) / (sum(x)^2- n*sum(x^2))
alpha.hat = mean(y)-beta.hat*mean(x)
sigma.hat = sum((y-(alpha.hat+beta.hat*x))^2)/n

> beta.hat
[1] 0.4135198
```

```
> alpha.hat
[1] 18.57375
> sigma.hat
[1] 1971.674
```

iii. **Fit the linear regression model in R and compare the regression coefficient estimates to those you obtained in (b.ii). [Hint: use `lm(y ~ x)` in R].**

```
> lm(y ~ x)

Coefficients:
(Intercept)            x
    18.5737       0.4135
```

We note that the MLEs of α and β match those we found in (b.ii).

4. **Let $(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n)$ be independent observations such that**

$$Y_i = \begin{cases} 1 & \textbf{if subject } i \textbf{ experienced event} \\ 0 & \textbf{otherwise} \end{cases} \qquad Y_i \sim \textbf{ Bernoulli}(p_i)$$

We fit a logistic regression model on the fixed covariate x_i $(i = 1, \dots, n)$

$$\log \frac{p_i}{1 - p_i} = \alpha + \beta x_i$$

(a) **Use Newton-Raphson algorithm to find the MLE of $\theta = (\alpha, \beta)$.**

i. **Provide the likelihood and log-likelihood functions of θ.**

$$f(y_i|p_i) = p_i^{y_i}(1-p_i)^{1-y_i} = \left(\frac{p_i}{1-p_i}\right)^{y_i}(1-p_i) \qquad p_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}$$

$$\begin{aligned}
\mathcal{L}(\theta|\boldsymbol{X},\boldsymbol{Y}) &= \prod_{i=1}^{n}\left[(e^{\alpha+\beta x_i})^{y_i}\frac{1}{1+e^{\alpha+\beta x_i}}\right] = \frac{e^{\sum_{i=1}^{n} y_i(\alpha+\beta x_i)}}{\prod_{i=1}^{n}(1+e^{\alpha+\beta x_i})} \\
\log \mathcal{L}(\theta|\boldsymbol{X},\boldsymbol{Y}) &= \sum_{i=1}^{n} y_i(\alpha+\beta x_i) - \sum_{i=1}^{n}\log\left(1+e^{\alpha+\beta x_i}\right)
\end{aligned}$$

ii. **Evaluate the partial derivatives of the log-likelihood and the Hessian matrix.**

$$\begin{aligned}
\frac{\partial}{\partial\alpha}\log \mathcal{L}(\theta|\boldsymbol{X},\boldsymbol{Y}) &= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} \\
\frac{\partial}{\partial\beta}\log \mathcal{L}(\theta|\boldsymbol{X},\boldsymbol{Y}) &= \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n}\frac{x_i e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial^2}{\partial \alpha^2} \log \mathcal{L}(\theta|\boldsymbol{X},\boldsymbol{Y}) &= -\sum_{i=1}^{n} \frac{e^{\alpha+\beta x_i}(1+e^{\alpha+\beta x_i}) - (e^{\alpha+\beta x_i})^2}{(1+e^{\alpha+\beta x_i})^2} = -\sum_{i=1}^{n} \frac{e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2} \\
\frac{\partial^2}{\partial \alpha \partial \beta} \log \mathcal{L}(\theta|\boldsymbol{X},\boldsymbol{Y}) &= -\sum_{i=1}^{n} \frac{x_i e^{\alpha+\beta x_i}(1+e^{\alpha+\beta x_i}) - x_i(e^{\alpha+\beta x_i})^2}{(1+e^{\alpha+\beta x_i})^2} = -\sum_{i=1}^{n} \frac{x_i e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2} \\
\frac{\partial^2}{\partial \beta^2} \log \mathcal{L}(\theta|\boldsymbol{X},\boldsymbol{Y}) &= -\sum_{i=1}^{n} \frac{x_i^2 e^{\alpha+\beta x_i}(1+e^{\alpha+\beta x_i}) - (x_i e^{\alpha+\beta x_i})^2}{(1+e^{\alpha+\beta x_i})^2} = -\sum_{i=1}^{n} \frac{x_i^2 e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2}
\end{aligned}
$$

Thus,

$$
U(\theta) = \left(\begin{array}{c} \sum_{i=1}^n y_i - \sum_{i=1}^n \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} \\ \sum_{i=1}^n x_i y_i - \sum_{i=1}^n \frac{x_i e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} \end{array} \right) \qquad H(\theta) = -\left[\begin{array}{cc} \sum_{i=1}^n \frac{e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2} & \sum_{i=1}^n \frac{x_i e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2} \\ \sum_{i=1}^n \frac{x_i e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2} & \sum_{i=1}^n \frac{x_i^2 e^{\alpha+\beta x_i}}{(1+e^{\alpha+\beta x_i})^2} \end{array} \right]
$$

iii. **Provide the iterative equation for the Newton-Raphson procedure.**

$$
\begin{aligned}
\theta^{(j+1)} &= \theta^{(j)} - H^{-1}(\theta^{(j)})U(\theta^{(j)}) \\
&= \left(\begin{array}{c} \alpha^{(j)} \\ \beta^{(j)} \end{array} \right) + \left[\begin{array}{cc} \sum_{i=1}^n \frac{e^{\eta_i^{(j)}}}{\left(1+e^{\eta_i^{(j)}}\right)^2} & \sum_{i=1}^n \frac{x_i e^{\eta_i^{(j)}}}{\left(1+e^{\eta_i^{(j)}}\right)^2} \\ \sum_{i=1}^n \frac{x_i e^{\eta_i^{(j)} i}}{\left(1+e^{\eta_i^{(j)}}\right)^2} & \sum_{i=1}^n \frac{x_i^2 e^{\eta_i^{(j)}}}{(1+e^{\eta_i^{(j)}})^2} \end{array} \right]^{-1} \left(\begin{array}{c} \sum_{i=1}^n y_i - \sum_{i=1}^n \frac{e^{\eta_i^{(j)}}}{1+e^{\eta_i^{(j)}}} \\ \sum_{i=1}^n x_i y_i - \sum_{i=1}^n \frac{x_i e^{\eta_i^{(j)}}}{1+e^{\eta_i^{(j)}}} \end{array} \right)
\end{aligned}
$$

where $\eta_i^{(j)} = \alpha^{(j)} + \beta^{(j)} x_i$.

iv. **Write an R function that implements the Newton-Raphson algorithm.**

```
newtraph = function(x, y, theta, eps=1e-6) {
      delta = 1; theta.hat = theta
      while(delta > eps) {
            eta = theta[1]+theta[2]*x
            U1 = sum(y)-sum(exp(eta)/(1+exp(eta)))
            U2 = sum(x*y) - sum(x*exp(eta)/(1+exp(eta)))
            U = c(U1, U2)
            H11 = sum(exp(eta)/(1+exp(eta))^2)
            H12 = sum(x*exp(eta)/(1+exp(eta))^2)
            H22 = sum(x^2*exp(eta)/(1+exp(eta))^2)
            H = matrix(c(H11, H12, H12, H22), 2, 2)
            thetanew = theta + solve(H)%*%U
            delta = sqrt(sum((thetanew-theta)^2))
            theta = thetanew; theta.hat = rbind(theta.hat, t(theta))
     }
    return(theta.hat)
}
```

v. **Use your R function to fit the logistic regression model on the `O-ring.txt` data.**

These data provide information for the 23 space shuttle flights that occurred before the Challenger mission disaster in 1986. The event of interest is whether at least one primary O-ring suffered thermal distress (TD = 1). We want to investigate if this event is associated with the temperature (°F) at the time of the flight (Temp).

```
dat = read.table("O-ring.txt", header=T)
x = dat$Temp; y = dat$TD
newtraph(x, y, c(0,0))
```

```
                [,1]       [,2]
theta.hat  0.000000  0.0000000
           9.619048 -0.1495238
          13.655738 -0.2112470
          14.938289 -0.2306001
          15.042291 -0.2321537
          15.042902 -0.2321627
          15.042902 -0.2321627
```

(b) **Fit the logistic regression model using R and compare to your ML estimates in (a.iii). [Hint: use `glm(y ~ x, family=binomial")` in R.]**

```
 glm(y ~ x, family="binomial")

 Call:  glm(formula = y ~ x, family = "binomial")

Coefficients:
(Intercept)            x
    15.0429      -0.2322

Degrees of Freedom: 22 Total (i.e. Null);  21 Residual
Null Deviance:     28.27
Residual Deviance: 20.32  AIC: 24.32
```

We note that the MLEs of α and *beta* match those we found with our implementation of the Newton-Raphson algorithm.

5. **Suppose $X_1, \ldots, X_n$ and $Z_1, \ldots, Z_m$ are random samples from a Normal$(\theta, 1)$ distribution. The n observations $X_1, \ldots, X_n$ are observed, but $Z_1, \ldots, Z_m$ are missing.**

(a) **Provide the complete data likelihood.**

The complete data likelihood is given by the joint pdf of the observed and missing data:

$$\mathcal{L}(\theta|\boldsymbol{X}, \boldsymbol{Z}) \quad = \quad \prod_{i=1}^{n} f(x_i|\theta) \cdot \prod_{i=1}^{m} f(z_i|\theta)$$

$$= \quad (2\pi)^{-n/2} \exp\left\{ -\frac{1}{2}\sum_{i=1}^{n}(x_i - \theta)^2 \right\} \cdot (2\pi)^{-m/2} \exp\left\{ -\frac{1}{2}\sum_{i=1}^{n}(z_i - \theta)^2 \right\}$$

(b) **Show that the expectation of the complete data log-likelihood for the E-step of the EM algorithm is given by**

$$J(\theta|\theta^{(j)}) = -\frac{n+m}{2}\log(2\pi) - \frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 + m\cdot\nu\right) + \theta^{(j)}\left(\sum_{i=1}^{n} x_i + m\cdot\mu\right) - \frac{n+m}{2}\theta^{(j)^2}$$

where

$$\mu = \theta^{(j)} \qquad \nu = 1 + \theta^{(j)^2}$$

The log-likelihood is given by

$$\begin{aligned}
\log\mathcal{L}(\theta|\boldsymbol{X},\boldsymbol{Z}) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2 - \frac{m}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{m}(z_i-\theta)^2 \\
&= -\frac{n+m}{2}\log(2\pi) - \frac{1}{2}\left[\sum_{i=1}^{n} x_i^2 - 2\theta\sum_{i=1}^{n} x_i + n\theta^2 + \sum_{i=1}^{m} z_i^2 - 2\theta\sum_{i=1}^{m} z_i + m\theta^2\right] \\
&= -\frac{n+m}{2}\log(2\pi) - \frac{1}{2}\left[\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{m} z_i^2 - 2\theta(\sum_{i=1}^{n} x_i + \sum_{i=1}^{m} z_i) + (n+m)\theta^2\right] \\
&= -\frac{n+m}{2}\log(2\pi) - \frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{m} z_i^2\right) + \theta\left(\sum_{i=1}^{n} x_i + \sum_{i=1}^{m} z_i\right) - \frac{n+m}{2}\theta^2
\end{aligned}$$

The expectation of the complete data log-likelihood is

$$\begin{aligned}
J(\theta|\theta^{(j)}) &= E\left[\log\mathcal{L}(\theta|\boldsymbol{X},\boldsymbol{Z})|\boldsymbol{X}\theta^{(j)}\right] \\
&= -\frac{n+m}{2}\log(2\pi) - \frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{m} E[Z_i^2|\boldsymbol{X},\theta^{(j)}]\right) \\
&+ \theta\left(\sum_{i=1}^{n} x_i + \sum_{i=1}^{m} E[Z_i|\boldsymbol{X},\theta^{(j)}]\right) - \frac{n+m}{2}\theta^2 \\
E[Z_i^2|\boldsymbol{X},\theta^{(j)}] &= \mathrm{Var}(Z_i|\boldsymbol{X},\theta^{(j)}) + \left(E[Z_i|\boldsymbol{X},\theta^{(j)}]\right)^2 = 1 + \theta^{(j)2} = \nu \\
E[Z_i|\boldsymbol{X},\theta^{(j)}] &= \theta^{(j)} = \mu
\end{aligned}$$

Therefore,

$$J(\theta|\theta^{(j)}) = -\frac{n+m}{2}\log(2\pi) - \frac{1}{2}\left(\sum_{i=1}^{n} x_i^2 + m\cdot\nu\right) + \theta^{(j)}\left(\sum_{i=1}^{n} x_i + m\cdot\mu\right) - \frac{n+m}{2}\theta^{(j)2}$$

(c) **Show that the M-step of the EM algorithm updates θ by**

$$\theta^{(j+1)} = \frac{n\cdot\bar{x} + m\cdot\mu}{n+m} = \frac{n\cdot\bar{x} + m\cdot\theta^{(j)}}{n+m}$$

For the M-step, we maximize $J(\theta|\theta^{(j)})$. Taking the derivative with respect to θ, we have

$$\frac{\partial}{\partial\theta}J(\theta|\theta^{(j)}) = \left(\sum_{i=1}^{n} x_i + m\mu\right) - (n+m)\theta$$

setting this to 0 and solving for θ, we get

$$\sum_{i=1}^{n} x_i + m\mu = (n+m)\theta \quad \Rightarrow \quad \theta = \frac{n\bar{x} + m\mu}{n+m}$$

that is, the M-step update θ using the iterative equation

$$\theta^{(j+1)} = \frac{n\bar{x} + m\theta^{(j)}}{n+m}$$

6. **Let $X_1, \ldots, X_n$ be iid geometric random variables where**

$$f(x|\theta) = P(X = x|\theta) = (1-\theta)^{x-1}\theta, \qquad x = 1, 2, \ldots, \quad 0 < \theta < 1$$

Suppose we consider a Beta(α, β) prior for θ

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \qquad 0 < \theta < 1, \quad \alpha > 0, \quad \beta > 0$$

(a) **Find the posterior distribution of θ.**

$$\begin{aligned}
\pi(\theta|\boldsymbol{x}) &= \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{\int_0^1 f(\boldsymbol{x}|\theta)\pi(\theta)d\theta} = \frac{(1-\theta)^{\sum_i (x_i-1)}\theta^n \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}}{\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \underbrace{\theta^{n+\alpha-1}(1-\theta)^{\sum_i x_i - n+\beta-1}}_{\text{kernel of Beta}(n+\alpha, \sum_i x_i - n+\beta)} d\theta} \\
&= \frac{\theta^{n+\alpha-1}(1-\theta)^{\sum_i x_i-n+\beta-1}}{\frac{\Gamma(n+\alpha)\Gamma(\sum_i x_i - n+\beta)}{\Gamma(\sum_i x_i+\alpha+\beta)}} = \frac{\Gamma(\sum_i x_i + \alpha+\beta)}{\Gamma(n+\alpha)\Gamma(\sum_i x_i - n+\beta)}\theta^{n+\alpha-1}(1-\theta)^{\sum_i x_i-n+\beta-1}
\end{aligned}$$

Therefore, the posterior distribution of θ is

$$\theta|\boldsymbol{x} \sim \text{ Beta}(n+\alpha, \sum_{i=1}^{n} x_i - n + \beta)$$

(b) **Provide the Bayes estimator of θ.**

The expectation of a Beta(α, β) distribution is given by $\frac{\alpha}{\alpha+\beta}$, thus

$$E[\theta|\boldsymbol{x}] = \frac{n+\alpha}{\sum_{i=1}^{n} x_i + \alpha + \beta}$$