Departamento de Matemática

Multivariate Analysis

LMAC, MECD, & MMAC

1^{st} Exam
Duration: 2.0 hours

1^{st} Semester – 2022/2023
12/02/2022 – 10:30

Please justify conveniently your answers

Group I **5.5 points**

1. Let $\boldsymbol{X} \sim \mathcal{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (0,0,0)^t$ and (2.5)

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & & \\ 1/2 & 1 & \\ 0 & 0 & 2 \end{pmatrix},$$

Compute $P(X_1 \geq X_2)$.

2. Suppose that we have four observations, for which we compute a dissimilarity matrix, given by (3.0)

$$\boldsymbol{D} = \begin{pmatrix} 0 & & & \\ 0.3 & & & \\ 0.4 & 0.5 & & \\ 0.7 & 0.8 & 0.6 & \end{pmatrix}.$$

On the basis of this dissimilarity matrix, sketch the dendrogram that results from hierarchically clustering these four observations using complete linkage.

Group II **8.0 points**

The examination marks of 88 students in five different mathematical subjects have been recorded. Each examination was marked out of 20 marks. Some summary statistics for these marks are given in the table below.

Examination	mean	sd	median	trimmed	mad	min	max	range
Calculus I	15.79	1.4	15.9	15.78	1.41	12.4	20.0	7.6
Algebra	14.2	1.11	14.15	14.19	1.26	11.9	17.6	5.7
Complex Analysis	16.11	1.32	16.2	16.11	1.33	13.2	19.1	5.9
Calculus II	15.34	1.34	15.2	15.35	1.33	12.7	19.3	6.6
Statistics	16.49	1.07	16.7	16.48	1.26	14.3	18.8	4.5

The correlation matrix for the data is given below.

```
round(cor(x), 2)
                  Calculus I  Algebra Complex Analysis Calculus II  Statistics
Calculus I              1.00     0.51             0.65        0.55        0.43
Algebra                 0.51     1.00             0.66        0.55        0.60
Complex Analysis        0.65     0.66             1.00        0.76        0.70
Calculus II             0.55     0.55             0.76        1.00        0.72
Statistics              0.43     0.60             0.70        0.72        1.00
```

Consider the following R commands and corresponding results:

```
res<-PcaClassic(x, k = 2, kmax = ncol(x),scale=FALSE, signflip=TRUE, crit.pca.distances = 0.975)
summary(res)
Call:
PcaClassic(x = x, k = 2, kmax = ncol(x), scale = FALSE, signflip = TRUE,
crit.pca.distances = 0.975)
Importance of components:
   PC1    PC2
Standard deviation     2.3418 1.0151
Proportion of Variance 0.8418 0.1582
Cumulative Proportion  0.8418 1.0000

round(res@loadings,3)
 PC1    PC2
Calculus I       0.467  0.839
Algebra          0.360 -0.099
Complex Analysis 0.520 -0.077
Calculus II      0.499 -0.336
Statistics       0.364 -0.409
```

1. Before obtaining principal components associated with this dataset, look to the information available and comment on any particularity that can be important for the analysis. Discuss whether it is appropriate to carry out the principal component analysis using the original or standardized variables. (2.0)
2. Interpret the first two principal components. (3.0)
3. The (sorted) score distances (obtained based on the first two principal components) associated with the 88 students are:

```
sort(round(res$sd,2))
[1] 0.04 0.12 0.29 0.32 0.41 0.42 0.43 0.43 0.47 0.50 0.54 0.59 0.64 0.66 0.68 0.68 0.70
[18] 0.70 0.72 0.72 0.75 0.76 0.77 0.78 0.82 0.86 0.87 0.87 0.89 0.97 0.99 1.00 1.01 1.01
[35] 1.03 1.05 1.11 1.14 1.17 1.19 1.19 1.21 1.22 1.23 1.25 1.27 1.28 1.29 1.31 1.35 1.37
[52] 1.37 1.38 1.38 1.42 1.43 1.43 1.44 1.44 1.46 1.47 1.51 1.55 1.61 1.63 1.65 1.66 1.66
[69] 1.70 1.71 1.74 1.74 1.75 1.76 1.83 1.89 1.98 1.99 2.14 2.22 2.24 2.25 2.28 2.46 2.48
[86] 2.62 2.75 3.04
```

   (a) Assuming that $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, prove that the squared of the Mahalanobis distance between the vector form by the first k principal components, $\boldsymbol{Y}_k = (\gamma_1^t \boldsymbol{X}, \ldots, \gamma_k^t \boldsymbol{X})^t$, and its expected value, $\boldsymbol{\mu}_Y = \mathrm{E}(\boldsymbol{Y}_k)$, follows a chi-squared distribution with k degrees of freedom. (1.5)

   (b) Using the scores distances listed before, determine the number of students assigned as outliers, assuming a $\alpha = 0.025$ false alarm rate. (1.5)

Group III **6.5 points**

An observation x comes from one of the two populations with prior probabilities $P(Y = 0) = P(Y = 1)$ and probability density functions:

$$f_{X|Y=i}(x) = \frac{1}{\lambda_i} \exp\left(-\frac{x}{\lambda_i}\right), \quad x \geq 0$$

with $\lambda_1 > \lambda_0 > 0$, $i = 0, 1$, known as Exponential distribution with parameter λ_i, and $F_{X|Y=j}(x) = 1 - \exp\left(-\frac{x}{\lambda_j}\right)$, $x \geq 0$.

1. Obtain the Bayes classification rule. (3.0)

2. Admitting that $\lambda_1 = 4$, and $\lambda_0 = 1$, calculate the total probability of misclassification. (2.0)

3. Let us admit that the group each observation belongs to, Y, was not observed and $P(Y = 1) = p$ is unknown. Then X can be seen as a mixture of two Exponential distributions. Consider that $\boldsymbol{x} = (x_1, \ldots, x_n)^t$ is a sample of size n from this population. Use, as a given fact, that the complete log-likelihood is:

$$l(\boldsymbol{\lambda}|\boldsymbol{x}, \boldsymbol{y}) = \sum_{j=1}^{n} \ln \left\{ f_{X|Y=y_j}(x_j|\boldsymbol{\lambda}) p^{y_j} (1-p)^{1-y_j} \right\},$$

where $\boldsymbol{\lambda} = (p, \lambda_0, \lambda_1)^t$ and $\boldsymbol{y} = (y_1, \ldots, y_n)^t$ represents the unobserved classes of $\boldsymbol{x}$,

$$\begin{aligned} E\left(l(\boldsymbol{\lambda}|\boldsymbol{X}, \boldsymbol{Y})|\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{\lambda}^{(g)} \right) &= \sum_{j=1}^{n} \ln \left\{ f_{X|Y=1}(x_j) p \right\} P(Y = 1|X = x_j, \boldsymbol{\lambda}^{(g)}) + \\ & \sum_{j=1}^{n} \ln \left\{ f_{X|Y=0}(x_j)(1-p) \right\} P(Y = 0|X = x_j, \boldsymbol{\lambda}^{(g)}), \end{aligned}$$

and the E-step is defined as:

$$\begin{aligned} p_i^{(g+1)} &= \hat{P}(Y_i = 1|\boldsymbol{X} = \boldsymbol{x}_i, \boldsymbol{\lambda}^{(g)}) \\ &= \frac{p_i^{(g)} exp(-x_i/\lambda_1^{(g)})/\lambda_1^{(g)}}{p_i^{(g)} exp(-x_i/\lambda_1^{(g)})/\lambda_1^{(g)} + (1 - p_i^{(g)}) exp(-x_i/\lambda_0^{(g)})/\lambda_0^{(g)}}, \end{aligned}$$

and $p^{(g+1)} = \frac{1}{n} \sum_{i=1}^{n} P(Y_i = 1|\boldsymbol{X} = \boldsymbol{x}_i, \boldsymbol{\lambda}^{(g)})$. Using the EM algorithm, define the associated M-Step that jointly with the E-step will estimate the unknown parameters, p, λ_0, and λ_1. (1.5)

Group I **5.5 points**

1. Let $\boldsymbol{X} \sim \mathcal{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (0,0,0)^t$ and (2.5)

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & & \\ 1/2 & 1 & \\ 0 & 0 & 2 \end{pmatrix},$$

Compute $P(X_1 \geq X_2)$.

1. $P(X_1 \geq X_2) = P(X_1 - X_2 \geq 0) = P(\underset{\sim}{a}^t \underset{\sim}{X} \geq 0)$

where $\underset{\sim}{a} = (1, -1, 0)^t$. Como $\underset{\sim}{X} \sim \mathcal{N}_3(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$

then $\underset{\sim}{a}^t \underset{\sim}{X} \sim \mathcal{N}_1(\ E(\underset{\sim}{a}^t \underset{\sim}{X}) = \mu_1 - \mu_2 = 0,$

$\qquad\qquad \mathrm{Var}(a^t \underset{\sim}{X}) = a^t \underset{\sim}{\Sigma}\, a = 1)$

$$\mathrm{Var}(a^t \underset{\sim}{X}) = a^t \underset{\sim}{\Sigma}\, a = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & 0 \\ 1/2 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix} = \frac{1}{2} + \frac{1}{2} = 1$$

thus $\underset{\sim}{a}^t \underset{\sim}{X} \sim N(0, 1)$

$P(X_1 \geq X_2) = P(\underset{\sim}{a}^t X \geq 0) = \frac{1}{2}$ (symmetry of N(0,1) dist.)

2. Suppose that we have four observations, for which we compute a dissimilarity matrix, given by (3.0)

$$\boldsymbol{D} = \begin{pmatrix} 0 & & & \\ 0.3 & & & \\ 0.4 & 0.5 & & \\ 0.7 & 0.8 & 0.6 & \end{pmatrix}.$$

On the basis of this dissimilarity matrix, sketch the dendrogram that results from hierarchically clustering these four observations using complete linkage.

Step 1:

$\min d_{ij} = \boxed{0.3 = d_{12}}$

$d_{3(12)} = \max(d_{13}, d_{23}) = \max(0.4, 0.5) = 0.5$

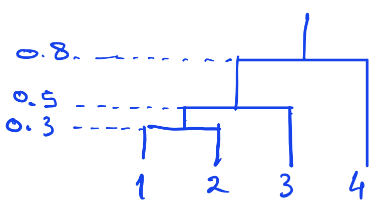
$d_{4(12)} = \max(d_{14}, d_{24}) = \max(0.7, 0.8) = 0.8$

Step 2:

$$D = \begin{array}{c} 12 \\ 3 \\ 4 \end{array} \begin{bmatrix} 0 & & \\ 0.5 & 0 & \\ 0.8 & 0.6 & 0 \end{bmatrix} \qquad \min d_{ij} = \boxed{0.5 = d_{3(12)}}$$

$d_{4(123)} = \max(d_{34}, d_{4(12)}) = \boxed{\max(0.6, 0.8) = 0.8}$

Dendrogram



Group II **8.0 points**

The examination marks of 88 students in five different mathematical subjects have been recorded. Each examination was marked out of 20 marks. Some summary statistics for these marks are given in the table below.

1. Before obtaining principal components associated with this dataset, look to the information available and comment on any particularity that can be important for the analysis. Discuss whether it is appropriate to carry out the principal component analysis using the original or standardized variables. (2.0)

- the variables are highly correlated so summarize them using PCA makes all sense
- the variances are not so different, so the use of PCA based on the original variables seems interesting for the analysis. In practice the two analysis should have been made and analysed.

2. Interpret the first two principal components. (3.0)

```
round(res@loadings,3)
 PC1    PC2
Calculus I        0.467  0.839
Algebra           0.360 -0.099
Complex Analysis  0.520 -0.077
Calculus II       0.499 -0.336
Statistics        0.364 -0.409
```

1st PC: weigh mean of all courses
- global measure of performance of the students.

So high (low) scores of the students on the 1st PC means high (low) performance of the students at math courses

2^{nd} PC: Is nearly a contrast between calculus I vs (calculus II and statistics)

Thus, high (low) scores on the 2^{nd} PC characterizes a student with high mark on calculus I and low (high) mark on calculus II and statistics.

3. The (sorted) score distances (obtained based on the first two principal components) associated with the 88 students are:

```
sort(round(res$sd,2))
[1]  0.04 0.12 0.29 0.32 0.41 0.42 0.43 0.43 0.47 0.50 0.54 0.59 0.64 0.66 0.68 0.68 0.70
[18] 0.70 0.72 0.72 0.75 0.76 0.77 0.78 0.82 0.86 0.87 0.87 0.89 0.97 0.99 1.00 1.01 1.01
[35] 1.03 1.05 1.11 1.14 1.17 1.19 1.19 1.21 1.22 1.23 1.25 1.27 1.28 1.29 1.31 1.35 1.37
[52] 1.37 1.38 1.38 1.42 1.43 1.43 1.44 1.44 1.46 1.47 1.51 1.55 1.61 1.63 1.65 1.66 1.66
[69] 1.70 1.71 1.74 1.74 1.75 1.76 1.83 1.89 1.98 1.99 2.14 2.22 2.24 2.25 2.28 2.46 2.48
[86] 2.62 2.75 3.04
```

(a) Assuming that $\boldsymbol{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, prove that the squared of the Mahalanobis distance between the vector form by the first k principal components, $\boldsymbol{Y}_k = (\gamma_1^t \boldsymbol{X}, \ldots, \gamma_k^t \boldsymbol{X})^t$, and its expected value, $\boldsymbol{\mu}_Y = \mathrm{E}(\boldsymbol{Y}_k)$, follows a chi-squared distribution with k degrees of freedom. (1.5)

(b) Using the scores distances listed before, determine the number of students assigned as outliers, assuming a $\alpha = 0.025$ false alarm rate. (1.5)

Let $y_{ij} = \gamma_i^t x_j$ the score of the j-th student on the i-th PC. Let $\underset{\sim}{y}_i = (y_{i1}, y_{ij})^t$ then

$$SD^2(\underset{\sim}{y}_i) = (\underset{\sim}{y}_i - \Gamma_2^t \underset{\sim}{\mu})^t \underset{\sim}{\Lambda}^{-1} (\underset{\sim}{y}_i - \Gamma_2^t \underset{\sim}{\mu})$$

where $\Gamma_2 = [\underset{\sim}{\gamma}_1, \underset{\sim}{\gamma}_2]$ and $\underset{\sim}{\mu} = (\mu_1, \mu_2)^t$

If we can assume that $\underset{\sim}{X} \sim \mathcal{N}_p(\underset{\sim}{\mu}, \underset{\sim}{\Sigma})$ then

$$\underset{\sim}{Y} = \Gamma_2^t \underset{\sim}{X} \sim \mathcal{N}_2(\Gamma_2^t \underset{\sim}{\mu}, \Gamma_2^t \underset{\sim}{\Sigma} \underset{\sim}{\Gamma}_2)$$

where by construction $Var(\underset{\sim}{Y}) = Diag(d_1, d_2)$

And so,

$$SD^2(\underset{\sim}{y}_i) = \sum_{i=1}^{2} \frac{(y_i - \bar{y}_i)^2}{d_i}$$

where $\bar{y}_i = \underset{\sim}{v}_i^t \underset{\sim}{\bar{x}}$, $i=1,2$. Then we can prove that

$SD^2(\underset{\sim}{Y})$ is the Mahalanobis distance and if $X \sim N_p(\underset{=}{\mu}, \underset{=}{\Sigma})$ then $SD^2(Y) \sim \chi^2_{(2)}$

thus, the cutting value, τ is estimated as $\alpha = P(SD^2(Y) \geq \tau \mid SD^2(\underset{\sim}{Y}) \sim \chi^2_{(2)})$

$$\Leftrightarrow \quad \tau = F^{-1}_{\chi^2_{(2)}}(1-0.025)$$

$$\Leftrightarrow \quad \tau = F^{-1}_{\chi^2_{(2)}}(0.975) = 7.3778$$

thus, the decision rule is:

If $\hat{SD}^2(\underset{\sim}{y}_i) \geq 7.3778 \Rightarrow$ Assign $\underset{\sim}{x}_i$ as an outlier

otherwise $\Rightarrow$ Assign $\underset{\sim}{x}_i$ as a regular obs.

Moreover, $\sqrt{\tau} = \tau^* = 2.7162$, thus

if $SD^2(\underset{\sim}{y}_i) \geq 2.7162 \Rightarrow$ Assign i-th obs as outlier. In our case we have two outliers ([86] 2.62 2.75 3.04)

Group III **6.5 points**

An observation x comes from one of the two populations with prior probabilities $P(Y=0)=P(Y=1)$ and probability density functions:

$$f_{X|Y=i}(x)=\frac{1}{\lambda_i}\exp\left(-\frac{x}{\lambda_i}\right),\quad x\geq 0$$

with $\lambda_1>\lambda_0>0$, $i=0,1$, known as Exponential distribution with parameter λ_i, and $F_{X|Y=j}(x)=1-\exp\left(-\frac{x}{\lambda_j}\right)$, $x\geq 0$.

1. Obtain the Bayes classification rule. (3.0)

According to Bayes rule:

Assign x to $\{Y=1\}$ iff

$$P(Y=1|X=x)\geq P(Y=0|X=x)$$

$$\Leftrightarrow \quad \frac{f_{X|Y=1}(x)\,P(Y=1)}{f_X(x)}\geq\frac{f_{X|Y=0}(x)\,P(Y=0)}{f_X(x)}$$

$P(Y=0)=P(Y=1)$

$$\Leftrightarrow \quad \frac{f_{X|Y=1}(x)}{f_{X|Y=0}(x)}\geq 1$$

$$\Leftrightarrow \quad \frac{1}{d_1}\,\frac{\exp(-x/d_1)}{\exp(-x/d_0)\,\frac{1}{d_0}}\geq 1$$

$$\Leftrightarrow \quad \exp\left\{-x\left(\frac{1}{d_1}-\frac{1}{d_0}\right)\right\}\geq\frac{d_1}{d_0}$$

$$\Leftrightarrow \quad -x\left(\frac{1}{d_1}-\frac{1}{d_0}\right)\geq\log\left(\frac{d_1}{d_0}\right)$$

Since $\lambda_1 > \lambda_0 \iff \frac{1}{\lambda_1} < \frac{1}{\lambda_0}$

and $\frac{1}{\lambda_1} - \frac{1}{\lambda_0} < 0$ so

$$x\left(\frac{1}{\lambda_0} - \frac{1}{\lambda_1}\right) \geq \log \frac{\lambda_1}{\lambda_0}$$

$$\iff \quad x \geq \frac{\lambda_0 \lambda_1}{\lambda_1 - \lambda_0} \log \frac{\lambda_1}{\lambda_0}$$

thus

<u>Classification Rule</u> is:

$$\begin{cases} \text{If } x \geq \dfrac{\lambda_0 \lambda_1}{\lambda_1 - \lambda_0} \log \dfrac{\lambda_1}{\lambda_0} \text{ then} \\ \text{assign } x \text{ to } \{Y=1\}. \\ \text{Otherwise assign } x \text{ to } \{Y=0\} \end{cases}$$

2. Admitting that $\lambda_1 = 4$, and $\lambda_0 = 1$, calculate the total probability of misclassification. (2.0)

then $\tau = \frac{\lambda_1 \lambda_0}{\lambda_1 - \lambda_0} \log\left(\frac{\lambda_1}{\lambda_0}\right) = \frac{4 \times 1}{4 - 1} \log\left(\frac{4}{1}\right)$

$= \frac{8}{3} \log 2 \approx 1.848$

$$TPM = P(\text{classif } X \text{ in } Y=1, Y=0) + $$
$$+ P(\text{classif } X \text{ in } Y=0, Y=1)$$
$$= P(X \geq \tau \mid Y=0)\frac{1}{2} + P(X < \tau \mid Y=1)\frac{1}{2}$$
$$= \left[1 - F_{Exp(0)}(\tau)\right]\frac{1}{2} + F_{Exp(1)}(\tau)\frac{1}{2}$$
$$= \frac{1}{2}\left[e^{-\tau}\right] + \frac{1}{2}\left[1 - e^{-\tau/4}\right]$$
$$\simeq 0.2638$$

3. Let us admit that the group each observation belongs to, Y, was not observed and $P(Y=1)=p$ is unknown. Then X can be seen as a mixture of two Exponential distributions. Consider that $\boldsymbol{x}=(x_1,\ldots,x_n)^t$ is a sample of size n from this population. Use, as a given fact, that the complete log-likelihood is:

$$l(\boldsymbol{\lambda}|\boldsymbol{x},\boldsymbol{y}) = \sum_{j=1}^{n} \ln\left\{f_{X|Y=y_j}(x_j|\boldsymbol{\lambda})p^{y_j}(1-p)^{1-y_j}\right\},$$

where $\boldsymbol{\lambda}=(p,\lambda_0,\lambda_1)^t$ and $\boldsymbol{y}=(y_1,\ldots,y_n)^t$ represents the unobserved classes of $\boldsymbol{x}$,

$$\begin{aligned} E\left(l(\boldsymbol{\lambda}|\boldsymbol{X},\boldsymbol{Y})|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{\lambda}^{(g)}\right) &= \sum_{j=1}^{n} \ln\left\{f_{X|Y=1}(x_j)p\right\}P(Y=1|X=x_j,\boldsymbol{\lambda}^{(g)}) + \\ & \sum_{j=1}^{n} \ln\left\{f_{X|Y=0}(x_j)(1-p)\right\}P(Y=0|X=x_j,\boldsymbol{\lambda}^{(g)}), \end{aligned}$$

and the E-step is defined as:

$$\begin{aligned} p_i^{(g+1)} &= \hat{P}(Y_i=1|\boldsymbol{X}=\boldsymbol{x}_i,\boldsymbol{\lambda}^{(g)}) \\ &= \frac{p_i^{(g)}exp(-x_i/\lambda_1^{(g)})/\lambda_1^{(g)}}{p_i^{(g)}exp(-x_i/\lambda_1^{(g)})/\lambda_1^{(g)} + (1-p_i^{(g)})exp(-x_i/\lambda_0^{(g)})/\lambda_0^{(g)}}, \end{aligned}$$

and $p^{(g+1)} = \frac{1}{n}\sum_{i=1}^{n} P(Y_i=1|\boldsymbol{X}=\boldsymbol{x}_i,\boldsymbol{\lambda}^{(g)})$. Using the EM algorithm, define the associated M-Step that jointly with the E-step will estimate the unknown parameters, p, λ_0, and λ_1. (1.5)

Being so,

$$E\left(\ell(d \mid \underset{\sim}{X}, \underset{\sim}{Y}) \mid \underset{\sim}{X} = \underset{\sim}{x}, \underset{\sim}{d}^{(g)}\right) = \sum_{j=1}^{n} \log\left(p \frac{1}{d_1} \exp\left(-\frac{x_j}{d_1}\right)\right) *$$

$$P(Y_j = 1 \mid \underset{\sim}{X} = \underset{\sim}{x}_j, \underset{\sim}{d}^{(g)}) + \sum_{j=1}^{n} \log\left((1-p) \frac{1}{d_0} \exp\left(-\frac{x_j}{d_0}\right)\right) *$$

$$* \left(1 - P(Y_j = 1 \mid \underset{\sim}{X} = \underset{\sim}{x}_j, \underset{\sim}{d}^{(g)})\right)$$

where $P(Y_j = 1 \mid X_j = x_j, \underset{\sim}{d}^{(g)}) =$

$$= \frac{p_j^{(g)} f_{X \mid Y=1}(x_j \mid \underset{\sim}{d}^{(g)})}{p_j^{(g)} f_{X \mid Y=1}(x_j \mid \underset{\sim}{d}^{(g)}) + (1-p_j^{(g)}) f_{X \mid Y=0}(x_j \mid \underset{\sim}{d}^{(g)})}$$

$$= \frac{p_j^{(g)} \frac{1}{d_1^{(g)}} \exp\left(-\frac{x_j}{d_1^{(g)}}\right)}{p_j^{(g)} \frac{1}{d_1^{(g)}} \exp\left(-\frac{x_j}{d_1^{(g)}}\right) + \frac{(1-p_j^{(g)})}{d_0^{(g)}} \exp\left(-\frac{x_j}{d_0^{(g)}}\right)}$$

$$= \xi_j(\underset{\sim}{d}^{(g)}) \equiv \xi_j^{(g)}$$

thus,

$$E\left(\ell(d \mid \underset{\sim}{X}, \underset{\sim}{Y}) \mid \underset{\sim}{X} = \underset{\sim}{x}, \underset{\sim}{d}^{(g)}\right) =$$

$$= \sum_{j=1}^{n} \left[\log\left(\frac{p}{d_1}\right) - \frac{x_j}{d_1}\right] \xi_j^{(g)}$$

$$+ (1 - \xi^{(g)}) \sum_{j=1}^{p} \left[\log\left(\frac{1-p}{d_0}\right) - \frac{x_j}{d_0}\right]$$

$$= Q(\underset{\sim}{d} \mid \underset{\sim}{d}^{(g)})$$

$$\frac{\partial Q(\underset{\sim}{d} \mid \underset{\sim}{d}^{(g)})}{\partial p} = 0 \iff$$

$$\iff \sum_{j=1}^{n} \frac{1}{d_1} \frac{1}{p/d_1} * \xi_j^{(g)} + \sum_{j=1}^{n} (1-\xi_j^{(g)}) \left[-\frac{1}{d_0} \frac{1}{(1-p)/d_0} \right] = 0$$

$$\iff \sum_{j=1}^{n} \frac{1}{p} \xi_j^{(g)} - \sum_{j=1}^{n} (1-\xi_j^{(g)}) \frac{1}{1-p} = 0$$

$$\iff (1-p) \sum_{j=1}^{n} \xi_j^{(g)} = p \sum_{j=1}^{n} (1-\xi_j^{(g)})$$

$$\iff p \left(\sum_{j=1}^{n} (1-\xi_j^{(g)}) + \sum_{j=1}^{n} \xi_j^{(g)} \right) = \sum_{j=1}^{n} \xi_j^{(g)}$$

$$\iff p = \frac{1}{n} \sum_{j=1}^{n} \xi_j^{(g)}$$

Thus, $$\boxed{p^{(g+1)} = \frac{1}{n} \sum_{j=1}^{n} \xi_j^{(g)}}$$

this is the E-step.

<u>M-step:</u>

$$\frac{\partial Q(\underset{\sim}{d} \mid \underset{\sim}{d}^{(g)})}{\partial d_1} = 0$$

$$\Leftrightarrow \sum_{j=1}^{n} \left[+ \frac{x_j}{d_1^2} - \frac{1}{d_1} \right] \xi_j^{(g)} +$$

$$\Leftrightarrow \sum_{j=1}^{n} x_j \, \xi_j^{(g)} = d_1 \sum_{j=1}^{n} \xi_j^{(g)}$$

$$\Leftrightarrow \boxed{d_1 = \frac{\sum_{j=1}^{n} x_j \, \xi_j^{(g)}}{\sum_{j=1}^{n} \xi_j^{(g)}}}$$

In an analogous way:

$$\frac{\partial Q(d \mid d^{(g)})}{\partial d_0} = 0 \Leftrightarrow$$

$$\Leftrightarrow \sum_{j=1}^{n} \left[\frac{x_j}{d_0^2} - \frac{1}{d_0} \right] (1 - \xi_j^{(g)}) = 0$$

$$\Leftrightarrow \sum_{j=1}^{n} x_j (1 - \xi_j^{(g)}) = d_0 \sum_{j=1}^{n} (1 - \xi_j^{(g)})$$

$$\Leftrightarrow \boxed{\hat{d}_0 = \frac{\sum_{j=1}^{n} x_j (1 - \xi_j^{(g)})}{\sum_{j=1}^{n} (1 - \xi_j^{(g)})}}$$