Ε.Μ.Π., Σχολή Η.Μ. & Μ.Υ

ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ
1° Σύνολο Αναλυτ. Ασκήσεων

Ακαδ. Έτος 2022-23

Δημήτριος Κοκκίνης
03118896
dimkok00@gmail.com

Άσκηση 1.1

α)

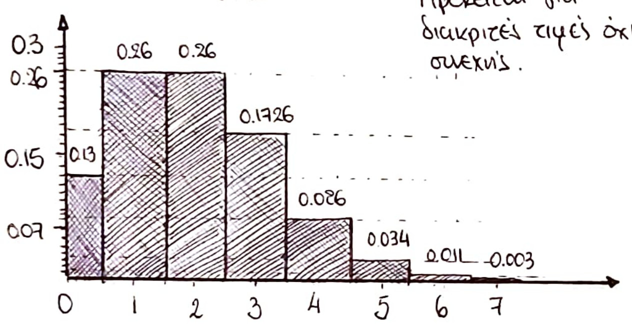
Η συνάρτηση μάζας πιθανότητας για την κατανομή Poisson είναι:

$$p(x|\theta) = \begin{cases} \frac{\theta^x}{x!} e^{-\theta} &, \quad x = 0,1,2,\dots \\ 0 &, \quad \text{αλλού} \end{cases}$$

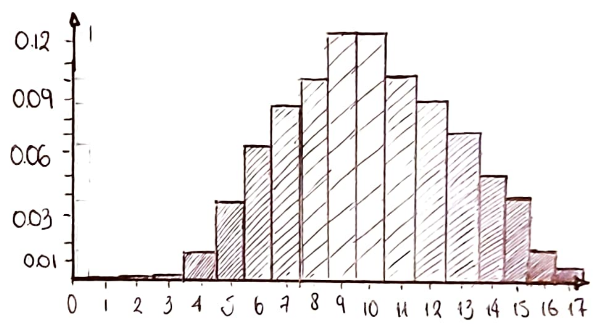
Για τιμές της παραμέτρου $\theta = \begin{cases} 2 \\ 10 \end{cases}$ λαμβάνουμε τα εξής διαγράμματα

θ=2

Πρόκειται για διακριτές τιμές όχι συνεχής.



θ=10



β) Η συνάρτηση πιθανοφάνειας είναι το γινόμενο των Σ.Μ.Π των παρατηρούμενων δεδομένων

$$L(\theta; x_1,\dots,x_N) = \prod_{j=1}^{N} \frac{\theta^{x_j}}{x_j!} e^{-\theta}$$

Για να απλοποιήσουμε τις πράξεις, γράφουμε τον φυσικό λογάριθμο της συνάρτησης πιθανοφάνειας: $\ell(\theta; x_1,\dots,x_N) = \ln\left(\prod_{j=1}^{N} \frac{\theta^{x_j}}{x_j!} e^{-\theta}\right) \Rightarrow \ell(\theta; x_1,\dots,x_N) = \sum_{j=1}^{N} \ln\left(\frac{\theta^{x_j}}{x_j!} e^{-\theta}\right)$

$$\Rightarrow \ell(\theta; x_1,\dots,x_N) = \sum_{j=1}^{N} \left[\ln(\theta^{x_j}) + \ln(e^{-\theta}) - \ln(x_j!)\right] \Rightarrow \ell(\theta; x_1,\dots,x_N) = \sum_{j=1}^{N} \left(x_j \ln(\theta) - \theta - \ln(x_j!)\right)$$

$$\Rightarrow \ell(\theta; x_1,\dots,x_N) = -N\theta - \ln(\theta)\sum_{j=1}^{N} x_j - \sum_{j=1}^{N} \ln(x_j!) \quad (1)$$

Στη συνέχεια υπολογίζουμε τη παράγωγο της σχέσης (1) ως προς θ:

$$\frac{d}{d\theta} \ell(\theta; x_1,\dots,x_N) = \frac{d}{d\theta}\left(-N\theta + \ln(\theta)\sum_{j=1}^{N} x_j - \sum_{j=1}^{N} \ln(x_j!)\right) = -N + \frac{1}{\theta}\sum_{j=1}^{N} x_j$$

Για το μέγιστο, μηδενίζουμε τη παράγωγο και λύνουμε ως προς θ: $-N + \frac{1}{\theta}\sum_{j=1}^{N} x_j = 0 \Rightarrow \boxed{\theta = \frac{1}{N}\sum_{j=1}^{N} x_j}$

Άσκηση 1.2

Αρχικά, βρίσκουμε τη παράγωγο του κόστους ως προς a:

$$\nabla J_s = \nabla \|Ya-b\|^2 = 2Y^T(Ya-b)$$

Άρα με βάση τη μέθοδο gradient descent και για $\eta_k = \frac{1}{k}$ έχουμε:

$$a_{k+1} = a_k - \frac{1}{k} Y^T(Ya-b)$$

Ομοίως, παίρνουμε $a_k = a_{k-1} - \frac{1}{k-1} Y^T(Ya-b)$

Αντικαθιστώντας, προκύπτει:

$$a_{k+1} = a_{k-1} + \left[\frac{1}{k-1}Y^TY + \frac{1}{k}Y^TY + \frac{1}{k(k-1)}(Y^TY)^2\right]a_{k-1} - \left[\frac{1}{k-1}Y^T + \frac{1}{k(k-1)}Y^TYY^T\right]b$$

όπου φαίνεται ότι για $k \to \infty$ οι συντελεστές των a_{k-1} και b μηδενίζονται και επομένως η μέθοδος συγκλίνει σε διάνυσμα a για το οποίο ισχύει: $Y^T(Ya-b)=0$

Άσκηση 1.3

Αρχικά παράγουμε τα 15 δείγματα με χρήση της γεννήτριας np.random.normal, τα 10 από τη κατανομή $N(1.5, 0.1)$ και τα 5 από την $N(2.5, 0.2)$ τα οποία είναι:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
1.44	1.75	1.34	0.9	1.32	1.46	1.08	1.56	1.65	1.53	2.35	3.2	1.93	2.57	2.44

Ακολουθούμε τον αλγόριθμο EM.

Αρχικά, θεωρούμε τις εξής αρχικές τιμές:

$$(\mu_1)^0 = 1, \quad (\mu_2)^0 = 3$$
$$(\sigma_1^2)^0 = (\sigma_2^2)^0 = 1$$
$$(P_1)^0 = (P_2)^0 = 0.5$$

οι οποίες δίνουν αρχική log-πιθανοφάνεια:

$$\ln p(X|(\theta)^0) = \sum_{n=1}^{15} \ln\left((P_1)^0 \frac{1}{\sqrt{2\pi(\sigma_1^2)^0}} e^{-\frac{(x_n-(\mu_1)^0)^2}{2(\sigma_1^2)^0}} + (P_2)^0 \frac{1}{\sqrt{2\pi(\sigma_2^2)^0}} e^{-\frac{(x_n-(\mu_2)^0)^2}{2(\sigma_2^2)^0}}\right)$$

$$= \boxed{-21.62}$$

Επανάληψη 1

E step

Υπολογίζουμε τις posterior πιθανότητες να έχει "γεννήσει" το κάθε δείγμα η μία ή η άλλη κατανομή με βάση τις προηγούμενες τιμές παραμέτρων:

$$\gamma(z_{ni}) = \frac{p_i \cdot N(x_n \mid (\mu_i)^0, (\sigma_i^2)^0)}{p_1 \cdot N(x_n \mid (\mu_1)^0, (\sigma_1^2)^0) + p_2 \cdot N(x_n \mid (\mu_2)^0, (\sigma_2^2)^0)}, \quad n = 1, \dots, 15, \quad i = 1,2$$

Έτσι έχουμε:

$$\gamma(z_{11}) = \frac{0.5 \cdot \frac{1}{\sqrt{2\pi \cdot 1}} e^{-\frac{(1.44-1)^2}{2 \cdot 1}}}{0.5 \cdot \frac{1}{\sqrt{2\pi \cdot 1}} e^{-\frac{(1.44-1)^2}{2 \cdot 1}} + 0.5 \cdot \frac{1}{\sqrt{2\pi \cdot 1}} e^{-\frac{(1.44-3)^2}{2 \cdot 1}}} = 0.754$$

$$\gamma(z_{12}) = 1 - \gamma(z_{11}) = 0.246$$

$$\gamma(z_{21}) = \frac{0.5 \cdot \frac{1}{\sqrt{2\pi \cdot 1}} e^{-\frac{(1.75-1)^2}{2 \cdot 1}}}{0.5 \cdot \frac{1}{\sqrt{2\pi \cdot 1}} e^{-\frac{(1.75-1)^2}{2 \cdot 1}} + 0.5 \frac{1}{\sqrt{2\pi \cdot 1}} e^{-\frac{(1.75-3)^2}{2 \cdot 1}}} = 0.6225$$

$$\gamma(z_{22}) = 1 - \gamma(z_{21}) = 0.3775$$

$$\vdots$$

(Συνεχίζουμε για όλα τα δείγματα)

$$\gamma(z_{15\,1}) = \frac{0.5 \cdot \frac{1}{\sqrt{2\pi \cdot 1}} e^{-\frac{(2.44-1)^2}{2 \cdot 1}}}{0.5 \cdot \frac{1}{\sqrt{2\pi \cdot 1}} e^{-\frac{(2.44-1)^2}{2 \cdot 1}} + 0.5 \cdot \frac{1}{\sqrt{2\pi \cdot 1}} e^{-\frac{(2.44-3)^2}{2 \cdot 1}}} = 0.2932$$

$$\gamma(z_{15\,2}) = 1 - \gamma(z_{15\,1}) = 0.7068$$

M step

Ανανεώνουμε τις παραμέτρους με κριτήριο τη μεγιστοποίηση της πιθανοφάνειας:

$$(\mu_1)' = \frac{\sum_{n=1}^{15} \gamma(z_{n1}) x_n}{\sum_{n=1}^{15} \gamma(z_{n1})} = \frac{1.44 \cdot 0.754 + \dots + 2.44 \cdot 0.2932}{0.754 + \dots + 0.2932} = 1.5298$$

$$(\mu_2)' = \frac{\sum_{n=1}^{15} \gamma(z_{n2}) x_n}{\sum_{n=1}^{15} \gamma(z_{n2})} = \frac{1.44 \cdot 0.246 + \dots + 2.44 \cdot 0.7068}{0.246 + \dots + 0.7068} = 2.1317$$

$$(\sigma_1^2)' = \frac{\sum_{n=1}^{15} \gamma(z_{n1}) (x_n - (\mu_1)')^2}{\sum_{n=1}^{15} \gamma(z_{n1})} = \frac{(1.44-1.5298)^2 \cdot 0.754 + \dots + (2.44-1.5298)^2 \cdot 0.2932}{0.754 + \dots + 0.2932} = 0.1866$$

$$(\sigma_2^2)' = \frac{\sum_{n=1}^{15} \gamma(z_{n2})\,(x_n-(\mu_2)')^2}{\sum_{n=1}^{15} \gamma(z_{n2})} = \frac{(1.44-2.1317)^2\cdot 0.246 + \dots + (2.41-2.1317)^2\cdot 0.7068}{0.246+\dots+0.7068} = 0.1099$$

$$(p_1)' = \frac{\sum_{n=1}^{15} \gamma(z_{n1})}{N} = \frac{0.754+\dots+0.2932}{15} = 0.6031$$

$$(p_2)' = \frac{\sum_{n=1}^{15} \gamma(z_{n2})}{N} = 1-(p_1)' = 0.3969$$

Η νέα πιθανοφάνεια θα είναι: $\ln p(x\,|\,(\theta)') = \sum_{n=1}^{15} \ln\Big((p_1)' \cdot \frac{1}{\sqrt{2\pi\cdot(\sigma_1^2)'}}\, e^{-\frac{(x_n-(\mu_1)')^2}{2\cdot(\sigma_1^2)'}}$

$$+ (p_2)' \cdot \frac{1}{\sqrt{2\pi(\sigma_2^2)'}}\, e^{-\frac{(x_n-(\mu_2)')^2}{2\cdot(\sigma_2^2)'}} \Big) = \boxed{-12.65}$$

Στις παραπάνω πράξεις παρέλειψα τους υπολογισμούς με όλα τα δείγματα, οι οποίοι είχαν γίνει σε πρόχειρο, λόγω μήκους.

Ακολουθούμε την ίδια διαδικασία για την επανάληψη 2 και προκύπτουν οι εξής τιμές:

E step

$\gamma(z_{11}) = 0.798$

$\gamma(z_{12}) = 0.202$

$\gamma(z_{21}) = 0.7026$

$\gamma(z_{22}) = 0.2974$

$\vdots$

$\gamma(z_{15\,1}) = 0.2155$

$\gamma(z_{15\,2}) = 0.7845$

M step

Ανανέωση παραμέτρων: $(\mu_1)^2 = 1.5016$, $(\mu_2)^2 = 2.1836$, $(\sigma_1^2)^2 = 0.1354$, $(\sigma_2^2)^2 = 0.131$

$(p_1)^2 = 0.6084$, $(p_2)^2 = 0.3916$

Η πιθανοφάνεια στην επανάληψη 2 θα είναι: $\ln p(x\,|\,(\theta)^2) = \sum_{n=1}^{15} \ln\Big((p_1)^2 \cdot \frac{1}{\sqrt{2\pi(\sigma_1^2)^2}}\, e^{-\frac{(x_n-(\mu_1)^2)^2}{2(\sigma_1^2)^2}} + (p_2)^2 \frac{1}{\sqrt{2\pi(\sigma_2^2)^2}}\, e^{-\frac{(x_n-(\mu_2)^2)^2}{2(\sigma_2^2)^2}}\Big)$

$$= \boxed{-12.19}$$

Παρατηρούμε ότι η τιμή της πιθανοφάνειας αρχίζει και συγκλίνει σε μία σταθερά. Επαναλαμβάνουμε για ακόμα μια φορά τη διαδικασία.

Επανάληψη 3

	$\gamma(z_{n1})$	$\gamma(z_{n2})$
x_1	0.8385	0.1615
x_2	0.733	0.267
x_3	0.8508	0.1492
x_4	0.8312	0.1688
x_5	0.8536	0.1464
x_6	0.8349	0.1651
x_7	0.8552	0.1448
x_8	0.8112	0.1888
x_9	0.7805	0.2195
x_{10}	0.8194	0.1806
x_{11}	0.1671	0.8326
x_{12}	0.0002	0.9998
x_{13}	0.6026	0.3974
x_{14}	0.0464	0.9536
x_{15}	0.1038	0.8962

E step

M step

$(\mu_1)^3 = 1.4671$

$(\mu_2)^3 = 2.2375$

$(\sigma_1^2)^3 = 0.1025$

$(\sigma_2^2)^3 = 0.404$

$(p_1)^3 = 0.6086$

$(p_2)^3 = 0.3914$

Η νέα πιθανοφάνεια θα έχει τιμή $\boxed{-11.77}$

Στο σημείο αυτό σταματάμε αλλά η διαδικασία μπορεί να συνεχιστεί με τον ίδιο τρόπο μέχρι η πιθανοφάνεια να μην αυξάνεται. Το ότι από τη $1^{η}$ επανάληψη και έπειτα η αύξηση της πιθανοφάνειας είναι σχετικά μικρή και δείχνει ότι είμαστε κοντά στο σημείο σύγκλισης.

Οι τελικές τιμές παραμέτρων που παίρνουμε είναι:

$$\boxed{\begin{aligned} \mu_1 &= 1.4671 \\ \mu_2 &= 2.2375 \\ \sigma_1^2 &= 0.1025 \\ \sigma_2^2 &= 0.404 \\ p_1 &= 0.6086 \\ p_2 &= 0.3914 \end{aligned}}$$

Η διαδικασία δειγματοληψίας που ακολουθήσαμε αρχικά αντιστοιχεί σε διαφορετικές πραγματικές τιμές παραμέτρων. Η απόκλιση αυτή, οφείλεται στον μικρό αριθμό δειγμάτων που διαθέταμε.

Άσκηση 1.4

α) Έχουμε: $D_1 = \{x_{11}, x_{12}, \dots, x_{1n}\}$

$D_2 = \{x_{21}, x_{22}, \dots, x_{2n}\}$

$\vdots$

$D_d = \{x_{d1}, x_{d2}, \dots, x_{dn}\}$

Επίσης ορίζουμε το διάνυσμα $s = (s_1, \dots, s_d)^T$ ως το άθροισμα των n δειγμάτων, δηλαδή έχουμε:

$$s = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_d \end{bmatrix} = \begin{bmatrix} x_{11} + x_{12} + \dots + x_{1n} \\ x_{21} + x_{22} + \dots + x_{2n} \\ \vdots \\ x_{d1} + x_{d2} + \dots + x_{dn} \end{bmatrix}$$

$$P(D|\theta) = P(x_1, x_2, \dots, x_n|\theta) \underset{\text{ανεξάρτητα}}{=} P(x_1|\theta)\cdot P(x_2|\theta)\cdots P(x_n|\theta) = \prod_{i=1}^{d} \theta_i^{x_{i1}}(1-\theta_i)^{1-x_{i1}} \cdot \prod_{i=1}^{d} \theta_i^{x_{i2}}(1-\theta_i)^{1-x_{i2}} \cdots \prod_{i=1}^{d} \theta_i^{x_{in}}(1-\theta_i)^{1-x_{in}}$$

$$= \theta_1^{x_{11}} (1-\theta_1)^{1-x_{11}} \cdot \theta_2^{x_{21}} \cdot (1-\theta_2)^{1-x_{21}} \cdots \theta_d^{x_{d1}} (1-\theta_d)^{1-x_{d1}}$$

$$\theta_1^{x_{12}} (1-\theta_1)^{1-x_{12}} \cdot \theta_2^{x_{22}} (1-\theta_2)^{1-x_{22}} \cdots \theta_d^{x_{d2}} (1-\theta_d)^{1-x_{d2}}$$

$$\ldots$$

$$\theta_1^{x_{1n}} (1-\theta_1)^{1-x_{1n}} \cdot \theta_2^{x_{2n}} (1-\theta_2)^{1-x_{2n}} \cdots \theta_d^{x_{dn}} (1-\theta_d)^{1-x_{dn}}$$

$$= \theta_1^{x_{11}+x_{12}+\ldots+x_{1n}} \cdot (1-\theta_2)^{n-x_{21}-x_{22}-\ldots-x_{2n}} \cdot$$

$$\ldots$$

$$\theta_d^{x_{d1}+x_{d2}+\ldots+x_{dn}} \cdot (1-\theta_d)^{n-x_{d1}-x_{d2}-\ldots-x_{dn}}$$

$$= \theta_1^{S_1} (1-\theta_1)^{n-S_1} \cdot \theta_2^{S_2} (1-\theta_2)^{n-S_2} \cdots \theta_d^{S_d} (1-\theta_d)^{n-S_d} = \prod_{i=1}^{d} \theta_i^{S_i} (1-\theta_i)^{n-S_i}$$

Δηλαδή $\boxed{P(D|\theta) = \prod_{i=1}^{d} \theta_i^{S_i} (1-\theta_i)^{n-S_i}}$

β) Απο κανόνα Bayes έχουμε: $P(\theta|D) = \dfrac{P(D|\theta)\cdot P(\theta)}{P(D)}$ (1)

- $P(D|\theta) = \prod_{i=1}^{d} \theta_i^{S_i} (1-\theta_i)^{n-S_i}$ απο το α ερώτημα

- $P(\theta) = \begin{cases} 1, & 0 \le \theta \le 1 \\ 0, & \text{αλλού} \end{cases}$ (uniform a priori)

- $P(D) = p(D_1 \cup D_2 \cup \ldots \cup D_d) \underset{\text{ανεξάρτητα } D}{=} P(D_1) + \ldots + P(D_d) = \theta_1^{S_1}(1-\theta_1)^{n-S_1} + \ldots + \theta_d^{S_d}(1-\theta_d)^{n-S_d}$

$$\underset{\substack{\theta_1,\ldots,\theta_d \\ \text{παίρνουν} \\ \text{συνεχείς τιμές} \\ \text{απο 0 έως 1}}}{=} \int_0^1 \theta^{S_i} (1-\theta)^{n-S_i} d\theta \quad , i=1,2,\ldots,d \quad (2)$$

Δίνεται επίσης: $\int_0^1 \theta^m (1-\theta)^n d\theta = \dfrac{m!\,n!}{(m+n+1)!}$ (3)

Για $m=S_i$ και n θα είναι $n-S_i$ η (2) γράφεται λόγω της (3) ως:

$$\int_0^1 \theta^{Si}(1-\theta)^{n-Si}\,d\theta = \frac{Si!\,(n-Si)!}{(Si+n-Si+1)!} = \frac{Si!\,(n-Si)!}{(n+1)!} = P(D)$$

Αντικαθιστώντας στην (1) έχουμε:

$$P(\theta|D) = \prod_{i=1}^{d} \theta_i^{Si}(1-\theta_i)^{n-Si} \cdot \frac{1}{Si!\,(n-Si)!/(n+1)!} = \prod_{i=1}^{d} \theta_i^{Si}(1-\theta_i)^{n-Si} \cdot \frac{(n+1)!}{Si!\,(n-Si)!}$$

άρα αποδείχτηκε το ζητούμενο

γ) Για $d=1$ και $n=1$ έχουμε:

$$P(\theta_1|D) = \frac{(1+1)!}{S_1!\,(1-S_1)!}\,\theta_1^{S_1}(1-\theta_1)^{1-S_1} = \frac{2\theta_1^{S_1}(1-\theta_1)^{1-S_1}}{S_1!\,(1-S_1)!} \propto 2\theta_1^{S_1}(1-\theta_1)^{1-S_1}$$

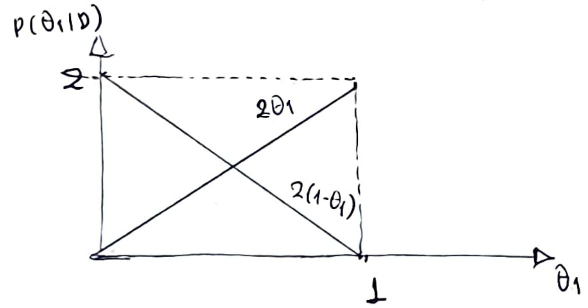
και $S_1 = X_{11}$.

Καθώς η X_{11} ακολουθεί κατανομή Bernoulli $[X_{11} = \{0,1\}]$ θα ισχύει:

- Για $X_{11} = 0 = S_1$ είναι $P(\theta_1|D) \propto 2(1-\theta_1)$
- Για $X_{11} = 1 = S_1$ είναι $P(\theta_1|D) \propto 2\theta_1$

Επίσης θ1 ακολουθεί ~~κατα~~ ομοιόμορφη κατανομή, $p(\theta_1) = \begin{cases} 1, & 0 \le \theta_1 \le 1 \\ 0, & \text{αλλού} \end{cases}$ και παίρνει τιμές στο $[0,1]$

Άρα η γραφική παράσταση της $P(\theta_1|D)$ για τις 2 περιπτώσεις ($S_1=0$ και $S_1=1$) είναι:



ε) Από Bayes έχουμε:

$$p(\theta|x) = \frac{p(x|\theta)\,p(\theta)}{p(x)} \propto p(x|\theta)\,p(\theta) \quad \text{(παραλείπουμε τον παρονομαστή)}$$

Άρα: $p(\theta|x) = p(x|\theta)\,p(\theta) = \prod_{i=1}^{n} p(x_i|\theta)\,p(\theta) = p(\theta)\prod_{i=1}^{n}\theta^{x_i}(1-\theta)^{1-x_i} = p(\theta)\,\theta^{x_1}(1-\theta)^{1-x_1}\cdot\ldots\cdot\theta^{x_n}(1-\theta)^{1-x_n}$

$$= p(\theta)\,\theta^{x_1+x_2+\ldots+x_n}(1-\theta)^{n-(x_1+x_2+\ldots+x_n)} = p(\theta)\,\theta^{S}(1-\theta)^{n-S}$$

$$\underset{\text{uniform}}{=} 1\cdot\theta^{S}(1-\theta)^{n-S}$$

Γενικεύοντας για $i=1,2,\ldots,d$ διαστάσεις θα έχουμε $p(\theta_i|x) = \theta_i^{s_i}(1-\theta_i)^{n-s_i}$

Η συνάρτηση αυτή είναι ο πυρήνας μιας Beta κατανομής με παραμέτρους

$B(\alpha = 1+s_i, \ \beta = 1+n-s_i)$

Άρα για την εκτίμηση Bayes θα έχουμε:

$$\theta_{i\,bayes} = E(\theta_i|x) = \frac{\alpha}{\alpha+\beta} = \frac{s_i+1}{s_i+1+n-s_i+1} \Longrightarrow \theta_{i\,bayes} = \frac{s_i+1}{n+2}$$

$\boxed{\text{Άσκηση 1.5}}$

a) Θέλουμε να δείξουμε ότι $a^T A a = trace(A a a^T)$

Ορίζουμε τους πίνακες:

$$a = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \quad m \times n$$

Ο transpose του a θα είναι

$$a^T = m \times n \text{ διαστάσεων} \qquad \text{και} \qquad A = \begin{bmatrix} A_{11} & A_{12} & \ldots & A_{1k} \\ A_{21} & A_{22} & \ldots & A_{2k} \\ \vdots & & & \\ A_{\ell 1} & A_{\ell 2} & \ldots & A_{\ell k} \end{bmatrix} \quad \ell \times k$$

Έχουμε $a^T A a = (n\times m)\cdot(\ell\times k)\cdot(m\times n) = (n\times k)\cdot(m\times n) = n\times n$ (1)

Εφόσον το ίχνος του πίνακα είναι το άθροισμα των στοιχείων της διαγωνίου του, το trace είναι ένας αριθμός, άρα μπορούμε να τον εκφράσουμε ως πίνακα διαστάσεων (1×1) (2)

Επειδή $a^T A a = trace(A a a^T)$ από τη σχέση (1) και (2) προκύπτει ότι $n=1$, δηλαδή ο $a^T A a$ είναι πίνακας (1×1)

Για να ορίζεται το $trace(A a a^T)$ σημαίνει ότι ο $A a a^T$ είναι τετραγωνικός.

Είναι: $A a a^T = (\ell\times k)\cdot(m\times n)\,(n\times m) = (\ell\times n)\,(n\times m) = \ell\times m \xrightarrow[\text{τετραγωνικός}]{} \ell = m$ (3)

Επίσης, για να ορίζεται ο πολλαπλασιασμός Aa $((\ell\times k)(m\times n))$ θα πρέπει $k=m$ (4)

Από (3), (4) $\Rightarrow k=\ell=m$

Επομένως πλέον γνωρίζουμε καλύτερα τις διαστάσεις των πινάκων.

$a : m \times 1 \qquad a^T : 1 \times m \quad$ και $\quad A : m \times m$

Κάνουμε τους πολλαπλασιασμούς : $\boxed{a^T A a}$:

- $a^T A = [\, a_{11} A_{11} + a_{21} A_{21} + \dots + a_{m1} A_{m1} \quad a_{11} A_{12} + a_{21} A_{22} + \dots + a_{m1} A_{m2} \quad \dots \quad a_{11} A_{1m} + a_{21} A_{2m} + \dots + a_{m1} A_{mm}$

- $$\begin{aligned} a^T A a = {} & a_{11} (a_{11} A_{11} + a_{21} A_{21} + \dots + a_{m1} A_{m1}) \\ & + a_{21} (a_{11} A_{12} + a_{21} A_{22} + \dots + a_{m1} A_{m2}) \\ & + \dots + a_{m1} (a_{11} A_{1m} + a_{21} A_{2m} + \dots + a_{m1} A_{mm}) \quad (5) \end{aligned}$$

$\boxed{A a a^T}$

- $$A a = \begin{bmatrix} A_{11} a_{11} + A_{12} a_{21} + \dots + a_{m1} A_{m1} \\ A_{21} a_{11} + A_{22} a_{21} + \dots + A_{2m} a_{m1} \\ \vdots \\ A_{m1} a_{11} + A_{m2} a_{21} + \dots + A_{mm} a_{m1} \end{bmatrix}$$

- $$A a a^T = \begin{bmatrix} a_{11} (A_{11} a_{11} + A_{12} a_{21} + \dots + A_{1m} a_{m1}) & a_{21} (A_{11} a_{11} + A_{12} a_{21} + \dots + A_{1m} a_{m1}) & \dots & a_{m1} (A_{11} a_{11} + A_{12} a_{21} + \dots + A_{1m} a_{m1}) \\ a_{11} (A_{21} a_{11} + A_{22} a_{21} + \dots + A_{2m} a_{m1}) & a_{21} (A_{21} a_{11} + A_{22} a_{21} + \dots + A_{2m} a_{m1}) & \dots & a_{m1} (A_{21} a_{11} + A_{22} a_{21} + \dots + A_{2m} a_{m1}) \\ \vdots & \vdots & & \vdots \\ a_{11} (A_{m1} a_{11} + A_{m2} a_{21} + \dots + A_{mm} a_{m1}) & a_{21} (A_{m1} a_{11} + A_{m2} a_{21} + \dots + A_{mm} a_{m1}) & \dots & a_{m1} (A_{m1} a_{11} + A_{m2} a_{21} + \dots + A_{mm} a_{m1}) \end{bmatrix}$$

$$\begin{aligned} \mathrm{trace}(A a a^T) = {} & a_{11} (A_{11} a_{11} + A_{12} a_{21} + \dots + A_{1m} a_{m1}) \\ & + a_{21} (A_{21} a_{11} + A_{22} a_{21} + \dots + A_{2m} a_{m1}) \\ & + \dots + a_{m1} (A_{m1} a_{11} + A_{m2} a_{21} + \dots + A_{mm} a_{m1}) \quad (6) \end{aligned}$$

Απο (5), (6) $\Rightarrow \boxed{a^T A a = \mathrm{trace}(A a a^T)}$

β) Για πολυμεταβλητή κανονική κατανομή, η συνάρτηση πυκνότητας πιθανότητας δίνεται από το τύπο: $f_x(x) = \frac{1}{(2\pi)^{d/2}\,|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\}$

όπου d είναι οι διαστάσεις και Σ ο πίνακας συνδιασποράς και Σ^{-1} ο αντίστροφος.

Θεωρούμε $x_1, x_2, \dots, x_n$ ανεξάρτητα δεδομένα μιας κανονικής κατανομής $f_x(x)$.

Η joint pdf θα είναι: $f(x_1, x_2, \dots, x_n) = f(x_1) f(x_2) \dots f(x_n) = \prod_{i=1}^{n} f(x_i)$

Έτσι, δεδομένου του πίνακα Σ έχουμε: $p(x_1, x_2, \dots, x_n \mid \Sigma) = p(x_1 \mid \Sigma)\, p(x_2 \mid \Sigma) \dots p(x_n \mid \Sigma)$

$$\underset{\substack{x_1, x_2, \dots, x_n \\ \text{ακολουθούν} \\ \text{κανονική κατανομή}}}{=} \left.\begin{aligned} &\frac{1}{(2\pi)^{d/2}\,|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\} \\ &\dots \\ &\cdot \frac{1}{(2\pi)^{d/2}\,|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\} \end{aligned}\right\} n \text{ φορές}$$

$$= \frac{1}{(2\pi)^{nd/2}\,|\Sigma|^{n/2}} \exp\left\{-\frac{1}{2}\sum_{k=1}^{n}(x_k-\mu)^T \Sigma^{-1}(x_k-\mu)\right\}$$

Ισχύει ότι $a^T A a = \mathrm{trace}(A a a^T)$

Θεωρώντας όπου A τον Σ^{-1}, όπου a τον $(x_k-\mu)$ και όπου a^T τον $(x_k-\mu)^T$, $k=1,2,\dots,n$, έχουμε:

$$\mathrm{trace}\left(\Sigma^{-1}(x_k-\mu)(x_k-\mu)^T\right) = (x_k-\mu)^T \cdot \Sigma^{-1} \cdot (x_k-\mu), \quad k=1,2,\dots,n \qquad (2)$$

$$\Sigma^{-1}\sum_{k=1}^{n}(x_k-\mu)(x_k-\mu)^T = \sum_{k=1}^{n}\Sigma^{-1}(x_k-\mu)(x_k-\mu)^T \Rightarrow -\frac{1}{2}\Sigma^{-1}\sum_{k=1}^{n}(x_k-\mu)(x_k-\mu)^T = -\frac{1}{2}\sum_{k=1}^{n}\Sigma^{-1}(x_k-\mu)(x_k-\mu)^T$$

Άρα $-\frac{1}{2}\sum_{k=1}^{n}\mathrm{trace}\left(\Sigma^{-1}(x_k-\mu)(x_k-\mu)^T\right) \overset{(2)}{=} -\frac{1}{2}\sum_{k=1}^{n}(x_k-\mu)^T \Sigma^{-1}(x_k-\mu) \qquad (3)$

Από τη σχέση (3) γράφουμε ισοδύναμα τη $p(x_1, x_2, \dots, x_n \mid \Sigma)$ ως εξής:

$$p(x_1, x_2, \dots, x_n \mid \Sigma) = \frac{1}{(2\pi)^{nd/2}\,|\Sigma|^{n/2}} \exp\left\{-\frac{1}{2}\mathrm{trace}\left(\Sigma^{-1}\sum_{k=1}^{n}(x_k-\mu)(x_k-\mu)^T\right)\right\}$$

οπότε το ζητούμενο αποδείχτηκε

γ) Δείξαμε ότι $p(x_1,x_2,\dots,x_n) = \dfrac{1}{(2\pi)^{nd/2}|\Sigma|^{n/2}} \exp\left\{-\frac{n}{2}\,\mathrm{trace}(\Sigma^{-1}\hat{\Sigma})\right\}$ (1)

Απο ορισμό ο πίνακας συνδιασποράς είναι θετικά ορισμένος και συμμετρικός. Επομένως και η εκτίμηση του $\hat{\Sigma}$ που θεωρήσαμε ίση με $\frac{1}{n}\sum_{k=1}^{n}(x_k-\mu)(x_k-\mu)^T$ θα είναι θετικά ορισμένη.

Απο το φασματικό θεώρημα της γραμμικής άλγεβρας, ο positive-definite συμμετρικός πίνακας $\hat{\Sigma}$ έχει έναν μοναδικό positive-definite συμμετρικό πίνακα $\hat{\Sigma}^{1/2}$.

Χρησιμοποιώντας τη κυκλική ιδιότητα του trace, γράφουμε την (1) ως εξής:

$$p(x_1,\dots,x_n|\Sigma) = \frac{1}{(2\pi)^{nd/2}|\Sigma|^{n/2}} \exp\left\{-\frac{n}{2}\,\mathrm{trace}(\hat{\Sigma}^{1/2}\Sigma^{-1}\hat{\Sigma}^{1/2})\right\}$$

$$= \frac{1}{(2\pi)^{nd/2}} \cdot \frac{1}{|\hat{\Sigma}|^{n/2}} \det(\hat{\Sigma}^{1/2}\Sigma^{-1}\cdot\hat{\Sigma}^{1/2})^{n/2} \exp\left\{-n/2\,\mathrm{trace}(\hat{\Sigma}^{1/2}\Sigma^{-1}\hat{\Sigma}^{1/2})\right\}$$

$$= \frac{1}{(2\pi)^{nd/2}|\hat{\Sigma}|^{n/2}} (\lambda_1,\dots,\lambda_d)^{nd/2} \exp\left\{-\frac{n}{2}\sum_{i=1}^{d}\lambda_i\right\}$$ άρα αποδείχτηκε το ζητούμενο

δ) Για να βρω πού μεγιστοποιείται η πιθανοφάνεια, παραγωγίζω τη σχέση που αποδείχτηκε στο ερώτημα (γ) ως προς λ και μηδενίζω τη παράγωγο.

$$\frac{\partial p(x_1,x_2,\dots,x_n|\Sigma)}{\partial \lambda_i} = 0 \Rightarrow \frac{\partial \frac{1}{(2\pi)^{nd/2}|\hat{\Sigma}|^{n/2}} (\lambda_1\dots\lambda_d)^{n/2}\exp(-n/2\sum_{i=1}^{d}\lambda_i)}{\partial \lambda_1} = 0$$

$$\Rightarrow \frac{n}{2}\lambda_1^{\frac{n}{2}-1}(\lambda_2\dots\lambda_d)^{n/2}\exp\left(-\frac{n}{2}\sum_{i=1}^{d}\lambda_i\right) + (\lambda_1\dots\lambda_d)^{n/2}\left(-\frac{n}{2}\exp\left(-\frac{n}{2}\sum_{i=1}^{d}\lambda_i\right)\right) = 0$$

$$\Rightarrow \frac{n}{2}\lambda_1^{n/2-1}(\lambda_2\dots\lambda_d)^{n/2}\exp\left(-\frac{n}{2}\sum_{i=1}^{d}\lambda_i\right) = \frac{n}{2}(\lambda_1\dots\lambda_d)^{n/2}\exp\left(-\frac{n}{2}\sum_{i=1}^{d}\lambda_i\right)$$

$$\Rightarrow \lambda_1^{n/2-1}(\lambda_2\dots\lambda_d)^{n/2} = (\lambda_1\dots\lambda_d)^{n/2} \Rightarrow \lambda_1^{n/2-1}(\lambda_2\dots\lambda_d)^{n/2} = \lambda_1^{n/2}(\lambda_2\dots\lambda_d)^{n/2} \Rightarrow \lambda_1^{n/2-1} = \lambda_1^{n/2}$$

Για να ισχύει αυτό πρέπει $\boxed{\lambda_1 = 1}$

Αντίστοιχα για $\dfrac{\partial p(x_1,x_2,\dots,x_n|\Sigma)}{\partial\lambda_2} = 0 \Rightarrow \dots \Rightarrow \lambda_2^{n/2-1} = \lambda_2^{n/2} \Rightarrow \boxed{\lambda_2 = 1}$

Παίρνωντας όλες τις παραγώγους ως προς λ_i , $i=1,2,\ldots,d$ καταλήγουμε στο ότι $\lambda_1=\lambda_2=\ldots=\lambda_d=1$. Για αυτά τα λ_i η πιθανοφάνεια μεγιστοποιείται.

Όπως είπαμε ο Σ είναι θετικά ορισμένος και συμμετρικός.

Συνεπώς και ο Σ^{-1} είναι και αυτός θετικά ορισμένος, όπως και ο $\hat{\Sigma}$.

Άρα, το γινόμενο $\Sigma^{-1}\hat{\Sigma}$ τους είναι ένας θετικά ορισμένος πίνακας.

Αφού ο $\Sigma^{-1}\hat{\Sigma}$ είναι συμμετρικός, θετικά ορισμένος υπάρχει ένας πίνακας P που τον διαγωνοποιεί ως εξής: $\Sigma^{-1}\hat{\Sigma}=PDP^{-1}$ όπου $D=\mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_d)$.

Δείξαμε ότι η πιθανοφάνεια μεγιστοποιείται για $\lambda_1,\lambda_2,\ldots,\lambda_d=1$ Άρα $D=I$.

Η σχέση $\Sigma^{-1}\hat{\Sigma}=PDP^{-1}$ γράφεται: $\Sigma_{MLE}^{-1}\hat{\Sigma}=PIP^{-1}=(PP^{-1})PP^{-1}=I\cdot I=I$

$\Rightarrow \Sigma_{MLE}^{-1}\hat{\Sigma}=I \Rightarrow \Sigma_{MLE}=\hat{\Sigma}$ άρα πράγματι ο εκτιμητής μέγιστης πιθανοφάνειας δίνεται από τη σχέση: $\boxed{\hat{\Sigma}=\frac{1}{n}\sum_{k=1}^{n}(x_k-\mu)(x_k-\mu)^T}$