Η πολυδιάστατη κανονική κατανομή

Ορίζουμε την τυπική πολυδιάστατη κανονική, σαν την κατανομή του τυχαίου διανύσματος $X = \left(X_1, \ldots, X_n\right)^T$, όπου $X_i \sim N\left(0,1\right)$ και όλα τα X_i μεταξύ τους ανεξάρτητα. Τότε

$$
\begin{aligned}
& f_X\left(x\right) = f_{X_1,\ldots,X_n}\left(x_1,\ldots,x_n\right) = N\left(x_1 \mid 0,1\right)\cdots N\left(x_n \mid 0,1\right) \\
& = \frac{1}{\left(2\pi\right)^{n/2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) = \frac{1}{\left(2\pi\right)^{n/2}}\exp\left(-\frac{1}{2}x^T x\right).
\end{aligned}
$$

Εμφανώς οι περιθώριες είναι όλες τυπικές κανονικές εφόσον

$$
\begin{aligned}
& f_{X_i}\left(x_i\right) = \int_{\mathbb{R}^{n-1}} f_X\left(x\right)\prod_{j\neq i} dx_j = N\left(x_i \mid 0,1\right)\int_{\mathbb{R}^{n-1}}\prod_{j\neq i} N\left(x_j \mid 0,1\right)dx_j \\
& = N\left(x_i \mid 0,1\right)\prod_{j\neq i}\int_{\mathbb{R}} N\left(x_j \mid 0,1\right)d\ {}_j x = N\left(x_i \mid 0,1\right).
\end{aligned}
$$

Η μέση τιμή του X είναι το μηδενικό διάνυσμα $\mathcal{O}^T = \left(0,\ldots,0\right)$

$$
\mu_X = \mathbb{E}\left(X\right) = \mathbb{E}\left(\left(X_1,\ldots,X_n\right)^T\right) = \left(\mathbb{E}\left(X_1\right),\ldots,\mathbb{E}\left(X_n\right)\right)^T = \left(0,\ldots,0\right)^T = \mathcal{O}.
$$

Ορίζουμε τον πίνακα συνδιασποράς του τυχαίου διανύσματος X, σαν τον πίνακα $\mathbb{V}\left(X\right)$ (εναλλακτικός συμβολισμός Σ_X είτε Σ) με $ij-$στοιχείο

$$
\left(\mathbb{V}\left(X\right)\right)_{ij} = Cov\left(X_i, X_j\right) = \sigma_{ij}.
$$

Επειδή $\left(X-\mu\right)\in\mathbb{R}^{n\times 1}$ και $\left(X-\mu\right)^T \in \mathbb{R}^{1\times n}$, το γινόμενο (outer product) του διανύσματος $\left(X-\mu\right)$ με το διάνυσμα $\left(X-\mu\right)^T$, θα έχει διάσταση $n\times n$

$$
\left(X-\mu\right)\left(X-\mu\right)^T = \left(\left(X_i-\mu_i\right)\left(X_j-\mu_j\right)\right)\in\mathbb{R}^{n\times n}.
$$

Παίρνοντας μέση τιμή έχουμε:

$$
\begin{aligned}
& \mathbb{V}\left(X\right) = \mathbb{E}\left[\left(\left(X_i-\mu_{X_i}\right)\left(X_j-\mu_{X_j}\right)\right)\right] \\
& = \left(\mathbb{E}\left[\left(X_i-\mu_i\right)\left(X_j-\mu_j\right)\right]\right) = \left(Cov\left(X_i, X_j\right)\right) = \left(\sigma_{ij}\right).
\end{aligned}
$$

Παρατηρούμε ότι ο πίνακας $\mathbb{V}(X)$ είναι συμμετρικός, με διαγώνια στοιχεία τις περιθώριες διασπορές $\sigma_{ii} = \sigma_i^2$.

Πιο συγκεκριμένα, για την τυπική πολυδιάστατη κανονική έχουμε

$$\mathbb{V}(X) = \mathbb{E}\left[(X - \mathcal{O})(X - \mathcal{O})^T\right] = \mathbb{E}\left[X X^T\right] = \mathbb{E}\left[\left(X_i X_j\right)\right] = \left(\mathbb{E}\left[X_i X_j\right]\right) = \left(\delta_{ij}\right) = \mathbb{I}_n$$

Συμβολίζουμε την τυπική πολυδιάστατη κανονική ως

$$N_n\left(x \mid \mathcal{O}_n, \mathbb{I}_n\right) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} x^T x\right) \Leftrightarrow X \sim N_n\left(\mathcal{O}_n, \mathbb{I}_n\right).$$

Θέτοντας για $A = \left(a_{ij}\right)$ με $\det(A) \neq 0$ και $A^{-1} = \left(b_{ij}\right)$

$$Y = T(X) = \mu + AX \Leftrightarrow X = T^{-1}(Y) = A^{-1}(Y - \mu),$$

ή ισοδύναμα σε συντεταγμένες, για $i = 1, \ldots, n$

$$Y_i = T\left(X_1, \ldots, X_n\right) = \mu_i + \sum_{k=1}^{n} a_{ik} X_k \Leftrightarrow X_i = T^{-1}\left(Y_1, \ldots, Y_n\right) = \sum_{k=1}^{n} b_{ik}\left(Y_k - \mu_k\right).$$

Η ιακωβιανή ορίζουσα του αντίστροφου μετασχηματισμού

$$Jac\left(T^{-1}\right) = \frac{\partial\left(x_1, \ldots, x_n\right)}{\partial\left(y_1, \ldots, y_n\right)} = \frac{\partial x}{\partial y},$$

είναι η ορίζουσα των μερικών παραγώγων

$$\frac{\partial x_i}{\partial y_j} = \frac{\partial}{\partial y_j}\left[\sum_{k=1}^{n} b_{ik}\left(y_k - \mu_k\right)\right] = b_{ij} \Rightarrow Jac\left(T^{-1}\right) = \det\left(A^{-1}\right) = \det(A)^{-1}.$$

Έτσι η πυκνότητα του Y τυχαίου διανύσματος είναι

$$f_Y(y) = f_X\left(T^{-1}(y)\right)\left|Jac\left(T^{-1}\right)\right| = \frac{1}{\left|\det(A)\right|} f_X\left(A^{-1}(y - \mu)\right)$$
$$= \frac{1}{(2\pi)^{n/2}\left|\det(A)\right|} \exp\left(-\frac{1}{2}\left(A^{-1}(y - \mu)\right)^T\left(A^{-1}(y - \mu)\right)\right).$$

Επειδή

$$\left|\det(A)\right| = \sqrt{\det(A)^2} = \sqrt{\det(A)\det\left(A^T\right)} = \sqrt{\det\left(AA^T\right)},$$

και

$$\left(A^{-1}(y-\mu)\right)^T \left(A^{-1}(y-\mu)\right) = (y-\mu)^T A^{-T} A^{-1}(y-\mu)$$
$$= (y-\mu)^T \left(AA^T\right)^{-1}(y-\mu),$$

παίρνουμε

$$f_Y(y) = \frac{1}{(2\pi)^{n/2}\sqrt{\det\left(AA^T\right)}} \exp\left(-\frac{1}{2}(y-\mu)^T \left(AA^T\right)^{-1}(y-\mu)\right).$$

Συμβολίζοντας με $\Sigma = \Sigma_Y = AA^T$ τον πίνακα συνδιασποράς του τυχαίου διανύσματος Y, η πυκνότητα γίνεται

$$f_Y(y) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}(y-\mu)^T \Sigma^{-1}(y-\mu)\right) = N_n(y \mid \mu, \Sigma)$$

ή ότι $Y \sim N_n(\mu, \Sigma)$. Ο πίνακας συνδιασποράς Σ είναι συμμετρικός και θετικά ορισμένος, το τελευταίο σημαίνει ότι εάν $Q(y \mid \mu, \Sigma) = (y-\mu)^T \Sigma^{-1}(y-\mu)$ τότε $\min\limits_{y \in \mathbb{R}^n} (y-\mu)^T \Sigma^{-1}(y-\mu) = Q(\mu \mid \mu, \Sigma) = 0$.

<u>Εφαρμογή</u>

Εάν $Y \sim N_n(\mu, \Sigma)$, και τα Y_i είναι μεταξύ τους ορθογώνια (γραμμικά ασυσχέτιστα), δηλαδή $Cov\left(Y_i, Y_j\right) = \left\{\begin{matrix} \sigma_i^2, & j = i \\ 0, & j \neq i \end{matrix}\right\}$ τότε τα Y_i είναι μεταξύ τους ανεξάρτητα

Ο πίνακας συνδιασποράς είναι:

$$\Sigma = \left(\sigma_i^2 \delta_{ij}\right) = diag\left(\sigma_1^2, \ldots, \sigma_n^2\right) \Leftrightarrow \Sigma^{-1} = diag\left(\frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_n^2}\right)$$

$$Q(y) = (y-\mu)^T \Sigma^{-1}(y-\mu) = (y-\mu)^T diag\left(\frac{1}{\sigma_1^2}, \ldots, \frac{1}{\sigma_n^2}\right)(y-\mu)$$

$$= \left(\frac{y_1 - \mu_1}{\sigma_1^2}, \ldots, \frac{y_n - \mu_n}{\sigma_n^2} \right) (y - \mu) = \sum_{i=1}^{n} \left(\frac{y_i - \mu_i}{\sigma_i} \right)^2$$

Έτσι έχουμε

$$f_Y(y) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \left(\frac{y_1 - \mu_1}{\sigma_1} \right)^2 \right)$$
$$= \frac{1}{\left(\sigma_1 \sqrt{2\pi}\right) \cdots \left(\sigma_n \sqrt{2\pi}\right)} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \left(\frac{y_i - \mu_i}{\sigma_i} \right)^2 \right)$$
$$= \prod_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y_i - \mu_i}{\sigma_i} \right)^2 \right) = \prod_{i=1}^{n} N\left(y_i \mid \mu_i, \sigma_i^2\right) = \prod_{i=1}^{n} f_{Y_i}(y_i),$$

που σημαίνει ότι τα Y_i είναι μεταξύ τους ανεξάρτητα.

Ορισμός: Ένας πίνακας $\Sigma \in \mathbb{R}^{n\times n}$ λέμε ότι είναι **θετικά ορισμένος** εάν είναι συμμετρικός και η τετραγωνική μορφή[1] $Q(x \mid \Sigma) = x^T \Sigma x$, όπου $x = (x_1, \ldots, x_n)^T$, είναι θετική για κάθε $x \in \mathbb{R}_*^{n\times 1} = \mathbb{R}^{n\times 1} \setminus \{\mathcal{O}_n\}$, δηλαδή:

1. $\Sigma = \Sigma^T$.
2. $Q(x \mid \Sigma) = x^T \Sigma x = \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j > 0,\ \forall x \in \mathbb{R}_*^{n\times 1}$.

Στην περίπτωση θετικά ημιορισμένου πίνακα ζητάμε $Q(x \mid \Sigma) \geq 0,\ \forall x \in \mathbb{R}^{n\times 1}$.

Στον πιο πάνω ορισμό το 2 είναι ισοδύναμο με το να ζητήσουμε η συνάρτηση $Q(x_1, \ldots, x_n \mid \Sigma)$ να έχει ολικό ελάχιστο στο $\mathcal{O}_n$, ή ότι,
$0 = Q(\mathcal{O}_n \mid \Sigma) < Q(x_1, \ldots, x_n \mid \Sigma),\ \forall x \in \mathbb{R}_*^{n\times 1}$ (να είναι δηλαδή γνησίως κυρτή).

Οι θετικά ορισμένοι και ημιορισμένοι πίνακες είναι πολύ σημαντικοί στην στατιστική εφόσον αντιστοιχούν σε **πίνακες συνδιασποράς**.

Από την γραμμική άλγεβρα γνωρίζουμε ότι **εάν ο πίνακας M είναι συμμετρικός, τότε όλες οι ιδιοτιμές του είναι πραγματικές**, δηλαδή[2] $spec(M) \subset \mathbb{R}$, με $spec(M) = \{\lambda \in C : \det(M - \lambda\mathbb{I}) = 0\}$. Έστω $\lambda \in spec(M)$ με

[1] The purely quadratic form associated with M
[2] Με spec συμβολίζουμε το spectrum – φάσμα, δηλαδή το σύνολο ιδιοτιμών ενός πίνακα.

αντίστοιχο ιδιοδιάνυσμα v, τότε $M\,v=\lambda\,v$ που σημαίνει ότι εάν ο M είναι θετικά ορισμένος $Q\left(v\,|\,M\right)>0$ θα έχουμε και

$$Q\left(v\,|\,M\right)=v^T M\,v=\lambda\,v^T v=\lambda\left\|v\right\|^2>0\;\Rightarrow\;\lambda>0\,.$$

Έτσι όλες οι ιδιοτιμές ενός συμμετρικού και θετικά ορισμένου πίνακα M είναι θετικές, δηλαδή $spec\left(M\right)\subset\mathbb{R}^+$, με αποτέλεσμα ο M να είναι και αντιστρέψιμος εφόσον $\det\left(M\right)=\prod_{\lambda\in spec(M)}\lambda\;>\;0$.

Έστω ότι ο M είναι συμμετρικός. Εάν υπάρχουν n διαφορετικά ζεύγη $\left(\lambda_i,v_i\right)$ τέτοια ώστε $M\,v_i=\lambda_i\,v_i$, τότε $M\left[v_1,\ldots,v_n\right]=\left[v_1,\ldots,v_n\right]D$ όπου $D=\left(\lambda_i\delta_{ij}\right)$. Ο πίνακας $P=\left[v_1,\ldots,v_n\right]$ είναι **ορθογώνιος**, δηλαδή $P^{-1}=P^T\Leftrightarrow P=P^{-T}$ και

$$D=P^T M\,P\,.$$

Επειδή $PP^T=\mathbb{I}_n$ θα έχουμε και $\det\left(P\right)^2=1\Leftrightarrow\left|\det\left(P\right)\right|=1$. Εάν επιπροσθέτως ο M είναι και θετικά ορισμένος, θα έχουμε $\det\left(P\right)=1$.

<u>Άσκηση</u>

Να βρεθούν οι ελάχιστες συνθήκες κάτω από τις οποίες ο 2×2 συμμετρικός πίνακας $M=\left(m_{ij}\right)$ είναι θετικά ορισμένος.

$$Q\left(x,y\,|\,M\right)=\left(x,y\right)\begin{pmatrix} m_{11} & m_{12}\\ m_{12} & m_{22}\end{pmatrix}\begin{pmatrix} x\\ y\end{pmatrix}=m_{11}x^2+2m_{12}xy+m_{22}y^2$$

$$=m_{11}\left\{\left(x+\frac{m_{12}}{m_{11}}y\right)^2+\frac{m_{22}}{m_{11}}y^2-\left(\frac{m_{12}}{m_{11}}y\right)^2\right\}=m_{11}\left(x+\frac{m_{12}}{m_{11}}y\right)^2+\left(m_{22}-\frac{m_{12}^2}{m_{11}}\right)y^2$$

$$Q\left(x,y\,|\,M\right)>0,\;\forall\left(x,y\right)\in\mathbb{R}_*^2\Leftrightarrow\left\{\begin{matrix} m_{11}>0\\ m_{22}-\dfrac{m_{12}^2}{m_{11}}>0\end{matrix}\right\}\Leftrightarrow\left\{\begin{matrix} m_{11}>0\\ \det\left(M\right)>0\end{matrix}\right\}.$$

<u>Άσκηση</u>

Δίνεται δ.τ.μ. $X=\left(X_1,\ldots,X_n\right)^T\in\mathbb{R}^{n\times 1}$. Εάν $M=\left(m_{ij}\right)\in\mathbb{R}^{n\times n}$ ορίζουμε την δ.τ.μ. $Y=M\;X$. Δείξτε ότι $\mathbb{E}\left[Y\right]=M\,\mathbb{E}\left[X\right]$

$$\mathbb{E}[Y]=\mathbb{E}[M\,X]=\mathbb{E}\left[\left(\sum_j m_{ij}X_j\right)\right]=\left(\sum_j m_{ij}\mathbb{E}\left[X_j\right]\right)=M\,\mathbb{E}[X].$$

Εάν ο M είναι θετικά ορισμένος τότε μπορεί να παραγοντοποιηθεί σαν $M=L\,L^T$ όπου L είναι πίνακας κάτω τριγωνικός, με θετικά διαγώνια στοιχεία. **Ο πίνακας L καλείται Cholesky παράγοντας του M** είτε τετραγωνική ρίζα του M. Τότε ο M^{-1} έχει Cholesky παράγοντα L^{-1}, εφόσον $M^{-1}=\left(L^{-1}\right)^T L^{-1}$, και για αυτό το λόγο είναι και αυτός θετικά ορισμένος.

Είναι προφανές ότι εάν $M=L\,L^T$ τότε ο M είναι συμμετρικός και θετικά ορισμένος, πράγματι:

$$M^T=\left(L\,L^T\right)^T=L\,L^T=M\,,$$

$$Q(x\,|\,M)=x^T L\,L^T x=\left(L^T x\right)^T L^T x=\left\|L^T x\right\|^2>0,\ \forall x\in\mathbb{R}^{n\times1*}.$$

Επίσης $M^{-1}=\left(LL^T\right)^{-1}=L^{-T}L^{-1}$ διότι $MM^{-1}=\left(L\,L^T\right)\left(L^{-T}L^{-1}\right)=L\left(L^T L^{-T}\right)L^{-1}=L\,L^{-1}=\mathbb{I}$. Παρόμοια και $M^{-1}M=\mathbb{I}$.

<u>Παράδειγμα</u>

Να αποδειχθεί ότι ο πίνακας

$$M=\begin{bmatrix}25 & 15 & -5\\ 15 & 18 & 0\\ -5 & 0 & 11\end{bmatrix},$$

είναι θετικά ορισμένος.

Αρκεί να δείξουμε ότι $M=L\,L^T$ με $L=\begin{bmatrix}l_{11} & 0 & 0\\ l_{21} & l_{22} & 0\\ l_{31} & l_{32} & l_{33}\end{bmatrix}$ και $l_{ii}>0$ για $i=1,2,3$.

$$L\,L^T=\begin{bmatrix}\left(l_{11}^2\right) & \left(l_{11}l_{21}\right) & \left(l_{11}l_{31}\right)\\ l_{21}l_{11} & \left(l_{21}^2+l_{22}^2\right) & \left(l_{21}l_{31}+l_{22}l_{32}\right)\\ l_{31}l_{11} & l_{31}l_{21}+l_{32}l_{22} & \left(l_{31}^2+l_{32}^2+l_{33}^2\right)\end{bmatrix}=\begin{bmatrix}25 & 15 & -5\\ 15 & 18 & 0\\ -5 & 0 & 11\end{bmatrix}\Rightarrow$$

Λύνοντας τις εξισώσεις στις παρενθέσεις παίρνουμε: $L=\begin{bmatrix}5 & 0 & 0\\ 3 & 3 & 0\\ -1 & 1 & 3\end{bmatrix}$.

<u>Άσκηση</u>

Να βρεθεί ο αντίστροφος του πίνακα $L=\begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix}$ με $l_{ii}>0$ για $i=1,2,3$. Στη συνέχεια να αντιστραφεί ο πίνακας L του προηγούμενου παραδείγματος.

Ο L είναι αντιστρέψιμος εφόσον $\det(L)=l_{11}l_{22}l_{33}>0$ [3]. Έστω $L^{-1}=(r_{ij})$, τότε από την εξίσωση $(l_{ij})(r_{ij})=\mathbb{I}$ έχουμε

$$L^{-1}=(r_{ij})=\begin{bmatrix} \dfrac{1}{l_{11}} & 0 & 0 \\ -\dfrac{l_{21}}{l_{11}l_{22}} & \dfrac{1}{l_{22}} & 0 \\ \dfrac{l_{32}l_{21}}{l_{11}l_{22}l_{33}}-\dfrac{l_{31}}{l_{11}l_{33}} & -\dfrac{l_{32}}{l_{22}l_{33}} & \dfrac{1}{l_{33}} \end{bmatrix}$$

Αντικαθιστώντας $L=\begin{bmatrix} 5 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & 1 & 3 \end{bmatrix}$ παίρνουμε $L^{-1}=\begin{bmatrix} \frac{1}{5} & 0 & 0 \\ -\frac{1}{5} & \frac{1}{3} & 0 \\ \frac{2}{15} & -\frac{1}{9} & \frac{1}{3} \end{bmatrix}$.

Η πολυδιάστατες student και normal – gamma πυκνότητες

Δίνεται τυχαίο διάνυσμα $X=(X_1,\ldots,X_n)^T$, και τ.μ. Z έτσι ώστε

$$[X \mid Z=z] \sim N_n\left(\mathcal{O}_n, z^{-1}\mathbb{I}_n\right).$$

δηλαδή ο πίνακας συνδιασποράς είναι $\mathbb{V}(X \mid Z=z)=z^{-1}\mathbb{I}_n=\lambda_n^{-1}$, όπου $\lambda_n=z\,\mathbb{I}_n$ ο πίνακας των precisions.

Τότε

$$\begin{aligned} f_{X|Z}(x\mid z) &= N_n\left(x\mid \mathcal{O}_n, z^{-1}\mathbb{I}_n\right)=N_n\left(x\mid\mathcal{O}_n,\lambda_n^{-1}\right) \\ &= \sqrt{\frac{\det(\lambda_n)}{(2\pi)^n}}\exp\left(-\frac{1}{2}x^T\lambda_n x\right)=\left(\frac{z}{2\pi}\right)^{n/2}\exp\left(-\frac{z}{2}x^Tx\right), \end{aligned}$$

[3] Είναι πολύ εύκολο να δείξουμε ότι οι ιδιοτιμές του L είναι τα διαγώνια στοιχεία του.

και θέτουμε

$$f_{X|Z}\left(x\,|\,z\right)=C_1 z^{n/2}\exp\left(-\frac{z}{2}x^T x\right),\ C_1=\left(\frac{1}{2\pi}\right)^{n/2}.$$

Εάν $Z\sim Ga\left(\frac{\nu}{2},\frac{\nu}{2}\right)$ τότε

$$f_Z\left(z\right)=Ga\left(z\,|\,\frac{\nu}{2},\frac{\nu}{2}\right)=C_2 z^{\nu/2-1}\exp\left(-\frac{\nu}{2}z\right),\ C_2=\frac{\left(\nu/2\right)^{\nu/2}}{\Gamma\left(\nu/2\right)},$$

και η από κοινού των X και Z γίνεται

$$f_{X,Z}\left(x,z\right)=f_{X|Z}\left(x\,|\,z\right)f_Z\left(z\right)=C_1C_2 z^{n/2}\exp\left(-\frac{z}{2}x^T x\right)z^{\nu/2-1}\exp\left(-\frac{\nu}{2}z\right)$$

$$=C_1C_2\,z^{\frac{\nu+n}{2}-1}\exp\left(-\frac{z}{2}\left(\nu+x^T x\right)\right)\propto Ga\left(z\,|\,\frac{\nu+n}{2},\frac{\nu+x^T x}{2}\right),$$

και έτσι

$$f_{X,Z}\left(x,z\right)=\frac{C_1C_2}{C_3}Ga\left(z\,|\,\frac{\nu+n}{2},\frac{\nu+x^T x}{2}\right),\ C_3=\frac{\left(\frac{\nu+x^T x}{2}\right)^{\frac{\nu+n}{2}}}{\Gamma\left(\frac{\nu+n}{2}\right)}.$$

Ολοκληρώνοντας στο $\mathbb{R}^+$ την προηγούμενη πυκνότητα, ως προς z, παίρνουμε:

$$f_{X,Z}\left(x,z\right)=\frac{C_1C_2}{C_3}=\frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\left(2\pi\right)^{n/2}}\frac{\left(\nu/2\right)^{\nu/2}}{\Gamma\left(\nu/2\right)}\left(\frac{\nu+x^T x}{2}\right)^{-\frac{\nu+n}{2}}$$

$$=\frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\nu/2\right)}\frac{\left(\nu/2\right)^{\nu/2}}{\left(2\pi\right)^{n/2}}\left(\frac{\nu}{2}\right)^{-\frac{\nu+n}{2}}\left(1+\frac{1}{\nu}x^T x\right)^{-\frac{\nu+n}{2}}$$

$$=\frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\nu/2\right)}\frac{1}{\left(\nu\pi\right)^{n/2}}\left(1+\frac{1}{\nu}x^T x\right)^{-\frac{\nu+n}{2}}=St_n\left(x\,|\,\mathcal{O}_n,\mathbb{I}_n,\nu\right).$$

Η πυκνότητα $St_n\left(x \mid \mathcal{O}_n, \mathbb{I}_n, \nu\right)$ είναι η τυπική $n-$διάστατη student με ν βαθμούς ελευθερίας.

Για τις περιθώριες κατανομές έχουμε για $d\,x_{-i}=\prod_{j\neq i} dx_j$

$$f_{X_i}\left(x_i\right)=\int_{\mathbb{R}^{n-1}} f_X\left(x\right)d\,x_{-i}=\int_{\mathbb{R}^{n-1}}\left\{\int_{z>0} f_{X,Z}\left(x,z\right)d\,z\right\}d\,x_{-i}$$
$$=\int_{\mathbb{R}^{n-1}}\left\{\int_{z>0} f_{X|Z}\left(x\mid z\right)f_Z\left(z\right)dz\right\}d\,x_{-i}=\int_{z>0} f_Z\left(z\right)\left\{\int_{\mathbb{R}^{n-1}} f_{X|Z}\left(x\mid z\right)d\,x_{-i}\right\}dz\,.$$

Επειδή

$$\int_{\mathbb{R}^{n-1}} f_{X|Z}\left(x\mid z\right)d\,x_{-i}=\int_{\mathbb{R}^{n-1}} N_n\left(x\mid \mathcal{O}_n,\, z^{-1}\mathbb{I}_n\right)d\,x_{-i}=N\left(x_i\mid 0, z^{-1}\right),$$

η περιθώρια τ.μ. X_i θα έχει κατανομή

$$f_{X_i}\left(x_i\right)=\int_{z>0} N\left(x_i\mid 0,z^{-1}\right)Ga\left(z\mid \frac{\nu}{2},\frac{\nu}{2}\right)dz=St\left(x_i\mid 0,1,\nu\right)$$

Εμφανώς

$$\mathbb{E}\left(X\right)=\mathbb{E}\left(\left(X_1,\dots,X_n\right)^T\right)=\left(\mathbb{E}\left(X_1\right),\dots,\mathbb{E}\left(X_n\right)\right)^T=\left(0,\dots,0\right)^T=\mathcal{O}$$

ενώ

$$\mathbb{V}\left(X\right)=\mathbb{E}\left[X\,X^T\right]=\mathbb{E}\left[\mathbb{E}\left(X\,X^T\mid Z\right)\right]=\mathbb{E}\left[\mathbb{E}\left(\left(X_i\,X_j\right)\mid Z\right)\right]$$
$$=\mathbb{E}\left[\left(\mathbb{E}\left(X_i\,X_j\mid Z\right)\right)\right]$$

όμως $\left[X\mid Z=z\right]\sim N_n\left(\mathcal{O}_n,\, z^{-1}\mathbb{I}_n\right)$ και έτσι $\mathbb{E}\left(X_i\,X_j\mid Z=z\right)=z^{-1}\delta_{ij}$, που δίνει

$$\mathbb{V}\left(X\right)=\mathbb{E}\left[\left(Z^{-1}\delta_{ij}\right)\right]=\mathbb{E}\left[Z^{-1}\right]\left(\delta_{ij}\right)=\frac{\nu}{\nu-2}\mathbb{I}_n\,.$$

Για την γενική μορφή της πολυδιάστατης student, θέτουμε

$A=\left(a_{ij}\right)$ με $\det\left(A\right)\neq 0$ και $A^{-1}=\left(b_{ij}\right)$

$$Y = T\left(X\right) = \mu + AX \Leftrightarrow X = T^{-1}\left(Y\right) = A^{-1}\left(Y-\mu\right),$$

ή ισοδύναμα σε συντεταγμένες, για $1 = 1,\ldots,n$

$$Y_i = T\left(X_1,\ldots,X_n\right) = \mu_i + \sum_{k=1}^{n} a_{ik} X_k \Leftrightarrow X_i = T^{-1}\left(Y_1,\ldots,Y_n\right) = \sum_{k=1}^{n} b_{ik}\left(Y_k - \mu_k\right).$$

Η ιακωβιανή ορίζουσα του αντίστροφου μετασχηματισμού είναι

$$Jac\left(T^{-1}\right) = \frac{\partial\left(x_1,\ldots,x_n\right)}{\partial\left(y_1,\ldots,y_n\right)} = \frac{\partial x}{\partial y}$$

και

$$\frac{\partial x_i}{\partial y_j} = \frac{\partial}{\partial y_j}\left[\sum_{k=1}^{n} b_{ik}\left(y_k - \mu_k\right)\right] = b_{ij} \Rightarrow Jac\left(T^{-1}\right) = \det\left(A^{-1}\right) = \det\left(A\right)^{-1}.$$

Έτσι

$$\begin{aligned}
f_Y\left(y\right) &= St_n\left(T^{-1}\left(y\right) \mid \mathcal{O}_n, \mathbb{I}_n, \nu\right)\left|Jac\left(T^{-1}\right)\right| \\
&= \frac{1}{\left|\det\left(A\right)\right|} St_n\left(A^{-1}\left(y-\mu\right) \mid \mathcal{O}_n, \mathbb{I}_n, \nu\right) \\
&= \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\left(\nu\pi\right)^{n/2}} \frac{1}{\left|\det\left(A\right)\right|}\left(1 + \frac{1}{\nu}\left(A^{-1}\left(y-\mu\right)\right)^T \left(A^{-1}\left(y-\mu\right)\right)\right)^{-\frac{\nu+n}{2}}.
\end{aligned}$$

Παρατηρούμε ότι

$$\left|\det\left(A\right)\right| = \sqrt{\det\left(A\right)^2} = \sqrt{\det\left(A\right)\det\left(A^T\right)} = \sqrt{\det\left(AA^T\right)},$$

και ότι

$$\begin{aligned}
&\left(A^{-1}\left(y-\mu\right)\right)^T\left(A^{-1}\left(y-\mu\right)\right) = \left(y-\mu\right)^T A^{-T} A^{-1}\left(y-\mu\right) \\
&= \left(y-\mu\right)^T \left(AA^T\right)^{-1}\left(y-\mu\right),
\end{aligned}$$

που δίνουν

$$f_Y(y) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \sqrt{\frac{1}{(\nu\pi)^n \det\left(AA^T\right)}} \left(1 + \frac{1}{\nu}(y-\mu)^T \left(AA^T\right)^{-1} (y-\mu)\right)^{-\frac{\nu+n}{2}}$$

Συμβολίζοντας με $\Sigma = \Sigma_Y = AA^T$ τον πίνακα συνδιασποράς, έχουμε

$$f_Y(y) = St_n(y \mid \mu, \Sigma, \nu) = \frac{\Gamma\left(\frac{\nu+n}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\det(\nu\pi\Sigma)}} \left(1 + \frac{1}{\nu}(y-\mu)^T \Sigma^{-1} (y-\mu)\right)^{-\frac{\nu+n}{2}}$$

Για τις περιθώριες Y_i έχουμε

$$\begin{aligned}
f_{Y_i}(x_i) &= \int_{\mathbb{R}^{n-1}} St_n(y \mid \mu, \Sigma, \nu)\, dy_{-i} \\
&= \int_{\mathbb{R}^{n-1}} \left\{ \int_{z>0} N_n\left(x \mid \mu, z^{-1}\Sigma\right) Ga\left(z \mid \frac{\nu}{2}, \frac{\nu}{2}\right) dz \right\} dx_{-i} \\
&= \int_{z>0} \left\{ \int_{\mathbb{R}^{n-1}} N_n\left(x \mid \mu, z^{-1}\Sigma\right) dx_{-i} \right\} Ga\left(z \mid \frac{\nu}{2}, \frac{\nu}{2}\right) dz \\
&= \int_{z>0} N\left(x_i \mid \mu_i, z^{-1}\sigma_i^2\right) Ga\left(z \mid \frac{\nu}{2}, \frac{\nu}{2}\right) dz = St\left(x_i \mid \mu_i, \sigma_i^2, \nu\right)
\end{aligned}$$

Εμφανώς

$$\mathbb{E}(Y) = \mathbb{E}\left((Y_1, \ldots, Y_n)^T\right) = \left(\mathbb{E}(Y_1), \ldots, \mathbb{E}(Y_n)\right)^T = (\mu_1, \ldots, \mu_n)^T = \mu$$

ενώ

$$\mathbb{V}(Y) = \mathbb{V}(\mu + AX) = A\,\mathbb{V}(X)\,A^T = A\frac{\nu}{\nu-2}\left(\delta_{ij}\right)A^T = \frac{\nu}{\nu-2}AA^T = \frac{\nu}{\nu-2}\Sigma\,.$$

Η $(n+1)-$διάστατη κατανομή

$$Ng_n\left(x, z \mid \mu, \Sigma, \frac{\nu}{2}, \frac{\nu}{2}\right) = N_n\left(x \mid \mu, z^{-1}\Sigma\right) Ga\left(z \mid \frac{\nu}{2}, \frac{\nu}{2}\right)$$

είναι ειδική περίπτωση της $(n+1)-$διάστατης normal – gamma κατανομής:

$$Ng_n\left(x,z\mid\mu,\Sigma,a,b\right)=N_n\left(x\mid\mu,\, z^{-1}\Sigma\right)Ga\left(z\mid a,b\right).$$

Για το x – marginal της $Ng_n\left(x,z\mid\mu,\Sigma,a,b\right)$ έχουμε

$$f_X\left(x\right)=\int\limits_{z>0} Ng_n\left(x,z\mid\mu,\Sigma,a,b\right)=\int\limits_{z>0} N_n\left(x\mid\mu,\, z^{-1}\Sigma\right)Ga\left(z\mid a,b\right)d\, z$$

και θέτοντας, $u=sz$ παίρνουμε

$$f_X\left(x\right)=\int\limits_{u>0} N_n\left(x|\mu,\,\left(\frac{u}{s}\right)^{-1}\Sigma\right)Ga\left(u|a,\frac{b}{s}\right)du\,,$$

για $s=\dfrac{b}{a}$ η περιθώρια $f_X\left(x\right)$ τελικά γίνεται

$$f_X\left(x\right)=\int\limits_{u>0} N_n\left(x|\mu,\, u^{-1}\frac{b}{a}\Sigma\right)Ga\left(u|a,a\right)du=St_n\left(x\mid\mu,\frac{b}{a}\Sigma,2a\right).$$

Normal πολυδιάστατο μοντέλο με γνωστό πίνακα συνδιασποράς, και άγνωστο διάνυσμα μέσων. Εδώ έχουμε να εκτιμήσουμε τη location παράμετρο του normal μοντέλου για γνωστό πίνακα συνδιασποράς $\Sigma=\lambda^{-1}$ δηλαδή η κατανομή δειγματοληψίας είναι

$$x_j^T=\left(x_{1j},\ldots,x_{dj}\right),\ 1\le j\le n\,,$$
$$x_j|\vartheta \overset{iid}{\sim} N_d\left(\cdot\mid\vartheta,\Sigma\right),\ \vartheta^T=\left(\vartheta_1,\ldots,\vartheta_d\right)\in\Theta=\mathbb{R}^d$$

$$\pi\left(x_i|\vartheta\right)=N_d\left(x_i\mid\vartheta,\lambda^{-1}\right)=\sqrt{\det\left(\frac{\lambda}{2\pi}\right)}\exp\left(-\frac{1}{2}\left(x_i-\vartheta\right)^T\lambda\left(x_i-\vartheta\right)\right),$$

όπου $\sqrt{\det\left(\dfrac{\lambda}{2\pi}\right)}=\dfrac{1}{\sqrt{\det\left(2\pi\lambda^{-1}\right)}}=\dfrac{1}{\sqrt{\det\left(2\pi\Sigma\right)}}$.

Η πιθανοφάνεια θα είναι

$$\pi\left(x_1,\ldots,x_n \mid \vartheta\right)=\prod_{i=1}^{n} N_d\left(x_i \mid \vartheta,\lambda^{-1}\right)=\prod_{i=1}^{n}\sqrt{\det\left(\frac{\lambda}{2\pi}\right)}\exp\left(-\frac{1}{2}\left(x_i-\vartheta\right)^T\lambda\left(x_i-\vartheta\right)\right)$$

$$=\det\left(\frac{\lambda}{2\pi}\right)^{n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(x_i-\vartheta\right)^T\Sigma^{-1}\left(x_i-\vartheta\right)\right)$$

Για την κλασσική εκτίμηση έχουμε για $x=\left(x_1,\ldots,x_n\right)$

$$\log\pi\left(x|\vartheta\right)\overset{\vartheta}{\propto}-\frac{1}{2}\sum_{i=1}^{n}Q\left(x_i \mid \vartheta,\lambda\right),$$

όπου $Q\left(x_i \mid \vartheta,\lambda\right)=\left(x_i-\vartheta\right)^T\lambda\left(x_i-\vartheta\right)$,

$$\sum_{i=1}^{n}Q\left(x_i \mid \vartheta,\lambda\right)=\sum_{i=1}^{n}x_i^T\lambda x_i-\left(\sum_{i=1}^{n}x_i^T\right)\lambda\vartheta-\vartheta^T\lambda\sum_{i=1}^{n}x_i+n\vartheta^T\lambda\vartheta .$$

Τότε για $\left(\frac{\partial}{\partial\vartheta}\right)^T=\left(\frac{\partial}{\partial\vartheta_1},\ldots,\frac{\partial}{\partial\vartheta_d}\right)$ έχουμε

$$\frac{\partial}{\partial\vartheta}\sum_{i=1}^{n}Q\left(x_i \mid \vartheta,\lambda\right)=-\frac{\partial}{\partial\vartheta}\left(\sum_{i=1}^{n}x_i^T\right)\lambda\vartheta-\frac{\partial}{\partial\vartheta}\vartheta^T\lambda\sum_{i=1}^{n}x_i+n\frac{\partial}{\partial\vartheta}\vartheta^T\lambda\vartheta .$$

Θέτοντας

$$n\overline{x}^T=\sum_{i=1}^{n}x_i^T=\sum_{i=1}^{n}\left(x_{1i},\ldots,x_{d\,i}\right)=\left(\sum_{i=1}^{n}x_{1i},\ldots,\sum_{i=1}^{n}x_{d\,i}\right)=\left(n\overline{x}_1,\ldots,n\overline{x}_d\right)$$

Παίρνουμε

$$\frac{\partial}{\partial\vartheta}\sum_{i=1}^{n}Q\left(x_i \mid \vartheta,\lambda\right)=-n\frac{\partial}{\partial\vartheta}\overline{x}^T\lambda\vartheta-n\frac{\partial}{\partial\vartheta}\vartheta^T\lambda\overline{x}+n\frac{\partial}{\partial\vartheta}\vartheta^T\lambda\vartheta$$

Επειδή ο λ είναι συμμετρικός πίνακας, θα έχουμε

1. $\frac{\partial}{\partial\vartheta}\overline{x}^T\lambda\vartheta=\frac{\partial}{\partial\vartheta}\sum_{i=1}^{n}\left(\overline{x}^T\lambda\right)_i\vartheta_i=\left(\sum_{i=1}^{n}\left(\overline{x}^T\lambda\right)_i\right)=\left(\overline{x}^T\lambda\right)^T=\lambda^T\overline{x}=\lambda\overline{x}$

2. $\frac{\partial}{\partial\vartheta}\vartheta^T\lambda\overline{x}=\frac{\partial}{\partial\vartheta}\sum_{i=1}^{n}\vartheta_i\left(\lambda\overline{x}\right)_i=\left(\left(\lambda\overline{x}\right)_i\right)=\lambda\overline{x}$

3. $\dfrac{\partial}{\partial \vartheta}\vartheta^T \lambda \vartheta = \dfrac{\partial}{\partial \vartheta}\left[\sum_{i=1}^{n}\lambda_{ii}\vartheta_i^2 + 2\sum_{i<j}\lambda_{ij}\vartheta_i\vartheta_j\right] = \left(2\sum_{k=1}^{n}\lambda_{ik}\vartheta_k\right) = 2\lambda\vartheta$

Επίσης παρατηρούμε ότι από τις παραγωγίσεις 1. και 2., η παραγώγιση 3. έχει τον κανόνα:

$$\left(\vartheta^T\lambda\vartheta\right)' = \left(\vartheta^T\right)'\left(\lambda\vartheta\right) + \left(\vartheta^T\lambda\right)\left(\vartheta\right)' = \lambda\vartheta + \left(\vartheta^T\lambda\right)^T = 2\lambda\vartheta$$

$$\frac{\partial}{\partial \vartheta}\sum_{i=1}^{n}Q\left(x_i \mid \vartheta, \lambda\right) = -2n\lambda\overline{x} + 2n\lambda\vartheta = -2n\lambda\left(\overline{x} - \vartheta\right)$$

$$\frac{\partial}{\partial \vartheta}\log\pi\left(x|\vartheta\right) = n\lambda\left(\overline{x} - \vartheta\right) = \mathcal{O}_d \Rightarrow \hat{\vartheta}_{MLE} = \overline{x} = \left(\overline{x}_1, \ldots, \overline{x}_d\right)^T$$

Η posterior για prior

$$\pi\left(\vartheta\right) = N_d\left(\vartheta|\ \mu_0, \lambda_0^{-1}\right) = \sqrt{\det\left(\frac{\lambda_0}{2\pi}\right)}\exp\left(-\frac{1}{2}\left(\vartheta - \mu_0\right)^T\lambda_0\left(\vartheta - \mu_0\right)\right)$$

γίνεται

$$\begin{aligned}
\pi\left(\vartheta \mid x\right) &\propto \exp\left(-\frac{1}{2}\left(\left(\vartheta - \mu_0\right)^T\lambda_0\left(\vartheta - \mu_0\right) + \sum_{i=1}^{n}\left(x_i - \vartheta\right)^T\lambda\left(x_i - \vartheta\right)\right)\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\vartheta^T\lambda_0\vartheta - \vartheta^T\lambda_0\mu_0 - \mu_0^T\lambda_0\vartheta - n\overline{x}^T\lambda\vartheta - n\vartheta^T\lambda\overline{x} + n\vartheta^T\lambda\vartheta\right)\right) \\
&= \exp\left(-\frac{1}{2}\left(\vartheta^T\left(\lambda_0 + n\lambda\right)\vartheta - \vartheta^T\left(\lambda_0\mu_0 + n\lambda\overline{x}\right) - \left(\mu_0^T\lambda_0 + n\overline{x}^T\lambda\right)\vartheta\right)\right)
\end{aligned}$$

Εάν $\pi\left(\vartheta \mid x\right) = N_d\left(\vartheta|\ \mu_n, \lambda_n^{-1}\right)$ θα έχουμε και

$$\pi\left(\vartheta \mid x\right) \propto \exp\left(-\frac{1}{2}\left(\vartheta^T\lambda_n\vartheta - \vartheta^T\lambda_n\mu_n - \mu_n^T\lambda_n\vartheta\right)\right).$$

Έτσι βρίσκουμε ότι:

$$\left\{\begin{array}{l}\lambda_n=\lambda_0+n\lambda\\ \lambda_n\mu_n=\lambda_0\mu_0+n\lambda\overline{x}\\ \mu_n^T\lambda_n=\mu_0^T\lambda_0+n\overline{x}^T\lambda\end{array}\right\}\Rightarrow\left\{\begin{array}{l}\lambda_n=\lambda_0+n\lambda\\ \mu_n=\left(\lambda_0+n\lambda\right)^{-1}\left(\lambda_0\mu_0+n\lambda\overline{x}\right)\end{array}\right\}$$

ή ότι

$$\pi\left(\vartheta\,|\,x\right)=N_d\left(\vartheta|\left(\lambda_0+n\lambda\right)^{-1}\left(\lambda_0\mu_0+n\lambda\overline{x}\right),\left(\lambda_0+n\lambda\right)^{-1}\right)$$

$$\mathbb{E}\left(\vartheta\,|\,x\right)=\left(\lambda_0+n\lambda\right)^{-1}\left(\lambda_0\mu_0+n\lambda\overline{x}\right)=\left(\lambda_0+n\lambda\right)^{-1}\lambda_0\mathbb{E}\left(\vartheta\,|\,x\right)+\left(\lambda_0+n\lambda\right)^{-1}n\lambda\overline{x}$$

Θέτοντας $\gamma_n=\left(\lambda_0+n\lambda\right)^{-1}\lambda_0$ και επειδή

$$\left(\lambda_0+n\lambda\right)^{-1}\lambda_0+\left(\lambda_0+n\lambda\right)^{-1}n\lambda=\left(\lambda_0+n\lambda\right)^{-1}\left(\lambda_0+n\lambda\right)=\mathbb{I}_n\,,$$

η posterior μέση τιμή θα είναι

$$\mathbb{E}\left(\vartheta\,|\,x\right)=\gamma_n\mathbb{E}\left(\vartheta\,|\,x\right)+\left(\mathbb{I}_n-\gamma_n\right)\hat{\vartheta}_{MLE}\,.$$

Για την prior predictive έχουμε

$$\mathbb{E}\left(x\right)=\mathbb{E}\left(\mathbb{E}\left(x\,|\,\vartheta\right)\right)=\mathbb{E}\left(\vartheta\right)=\mu_0$$

$$\mathbb{V}\left(x\right)=\mathbb{V}\left(\mathbb{E}\left(x\,|\,\vartheta\right)\right)+\mathbb{E}\left(\mathbb{V}\left(x\,|\,\vartheta\right)\right)=\mathbb{V}\left(\vartheta\right)+\mathbb{E}\left(\lambda^{-1}\right)=\lambda_0^{-1}+\lambda^{-1}$$

Είναι εύκολο να δούμε ότι για μια παρατήρηση $x\in\mathbb{R}^d$ έχουμε

$$\pi\left(x\right)=\int_{\mathbb{R}^d}N_d\left(x|\,\vartheta,\lambda^{-1}\right)N_d\left(\vartheta|\,\mu_0,\lambda_0^{-1}\right)d\vartheta=N_d\left(x|\,\vartheta,\lambda^{-1}+\lambda_0^{-1}\right).$$