

Πολυμεταβλητή Ανάλυση

Πολυμεταβλητή Κανονική Κατανομή

Λουκία Μελιγκοτσίδου

Τμήμα Μαθηματικών, ΕΚΠΑ

1 Πολυμεταβλητή Κανονική Κατανομή

2 Παραγωγή πινάκων - ΕΜΠ για Πολυμεταβλητή Κανονική Κατανομή

Μονοδιάστατη Κανονική κατανομή

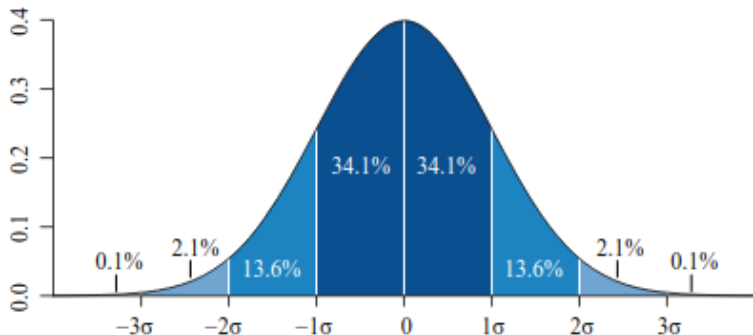
$$X \sim N(\mu, \sigma^2)$$

- $\mu \rightarrow$ μέσος, $\sigma^2 \rightarrow$ διασπορά
- Συνάρτηση πυκνότητας

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\}, \quad x \in \mathbb{R}.$$

- Πολλά φαινόμενα περιγράφονται ικανοποιητικά από την Κανονική κατανομή - υπόθεση σε πολλά στατιστικά μοντέλα (π.χ. **απλό γραμμικό μοντέλο με κανονικά σφάλματα**)
- Ακόμα και εάν η εξαρτημένη μεταβλητή **δεν ακολουθεί την Κανονική κατανομή**, λόγω Κεντρικού Οριακού Θεωρήματος, οι εκτιμητές πολλών στατιστικών μοντέλων ακολουθούν (προσεγγιστικά) **την Κανονική κατανομή για μεγάλο δείγμα**.

Μονοδιάστατη Κανονική κατανομή



Πολυμεταβλητή Κανονική Κατανομή

- Οι περισσότερες μέθοδοι πολυμεταβλητής στατιστικής αναπτύχθηκαν με βάση την κατανομή αυτή.
- Ακρογωνιαίος λίθος για τις περισσότερες μεθόδους

Από κοινού συνάρτηση πυκνότητας πιθανότητας

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\mathbf{\Sigma}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

όπου $\mathbf{x} = (x_1, x_2, \dots, x_p)$, $\boldsymbol{\mu} = E(\mathbf{x})$ διάνυσμα $p \times 1$, και $\mathbf{\Sigma} = \text{Var}(\mathbf{x})$ συμμετρικός και θετικά ορισμένος $p \times p$ πίνακας.

- p άγνωστες παράμετροι για τον μέσο
- p^2 στοιχεία στον $\mathbf{\Sigma}$, όμως $\mathbf{\Sigma}$ συμμετρικός άρα τα k στοιχεία κάτω από την κύρια διαγώνιο ίσα με τα k στοιχεία πάνω από την κύρια διαγώνιο. Άρα $p + 2k = p^2 \Rightarrow k = \frac{p(p-1)}{2} \Rightarrow$ τα άγνωστα στοιχεία του $\mathbf{\Sigma}$ είναι $p + k = \frac{p(p+1)}{2}$.

$$\begin{aligned} F(x_1, x_2, \dots, x_p) &= P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_p \leq x_p) \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_p} f(y_1, y_2, \dots, y_p) dy_p \dots dy_2 dy_1. \end{aligned}$$

Παρατήρηση

Δύσχρηστη στην πολυμεταβλητή ανάλυση, με μικρή χρησιμότητα, καθώς για να υπολογιστούν τα πολύπλοκα πολλαπλά ολοκληρώματα χρειάζονται ειδικές αριθμητικές μέθοδοι.

- **Πολυμεταβλητό αντίστοιχο κατανομής:** Πολυμεταβλητή κατανομή με περιθώριες κατανομές στην αντίστοιχη οικογένεια κατανομών
- Όμως, υπάρχουν περισσότερες από μία πολυμεταβλητές κατανομές για τις οποίες καταχρηστικά έχουμε χρησιμοποιήσει το ίδιο όνομα, επειδή οι περιθώριες κατανομές τις ανήκουν σε κάποια συγκεκριμένη οικογένεια κατανομών → Προσοχή στην ονοματολογία και χρήση πολυμεταβλητών κατανομών

Ειδική περίπτωση: Σ διαγώνιος

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_p^2 \end{pmatrix} \Rightarrow |\Sigma| = \prod_i^p \sigma_i^2 \text{ και}$$
$$\Sigma^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\sigma_p^2 \end{pmatrix}$$
$$f(\mathbf{x}) = \prod_{i=1}^p \frac{1}{(2\pi)^{1/2} \sigma_i} \exp \left\{ -\frac{1}{2\sigma_i^2} (x_i - \mu_i)^2 \right\}$$

το οποίο είναι ένα γινόμενο p μονοδιάστατων συναρτήσεων πυκνότητας από μονοδιάστατες κανονικές κατανομές.

Ειδική περίπτωση: Σ διαγώνιος

$$f(\mathbf{x}) = \prod_{i=1}^p \frac{1}{(2\pi)^{1/2}\sigma_i} \exp \left\{ -\frac{1}{2\sigma_i^2} (x_i - \mu_i)^2 \right\}$$

το οποίο είναι ένα γινόμενο p μονοδιάστατων συναρτήσεων πυκνότητας από μονοδιάστατες κανονικές κατανομές.

Συμπέρασμα

Εάν $(X_1, X_2, \dots, X_p) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ και είναι ανά δύο ασυσχετίστες $\Rightarrow$ θα είναι από κοινού ανεξάρτητες.

Έστω τμ $(X, Y) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, όπου $\boldsymbol{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$ και $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}$. Τότε,

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \left\{ \left(\frac{x-\mu_x}{\sigma_x} \right)^2 + \left(\frac{y-\mu_y}{\sigma_y} \right)^2 + 2\rho \left(\frac{x-\mu_x}{\sigma_x} \right) \left(\frac{y-\mu_y}{\sigma_y} \right) \right\} \right]$$

όπου $\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$ η συσχέτιση των X και Y .

- Αν $\rho = 0$, $\rightarrow$ η από κοινού σπι είναι το γινόμενο δύο σπι μονοδιάστατων κανονικών κατανομών.
- $\Rightarrow$ οι δύο τμ είναι ανεξάρτητες.

Μορφή διδιάστατης κανονικής:

- Συμμετρική
- μοιάζει με καμπάνα της οποίας τα χαρακτηριστικά καθορίζονται από τις παραμέτρους
- Η κορυφή της καμπάνας βρίσκεται πάντα στο σημείο μ

Ενδιαφέρον αποτέλεσμα:

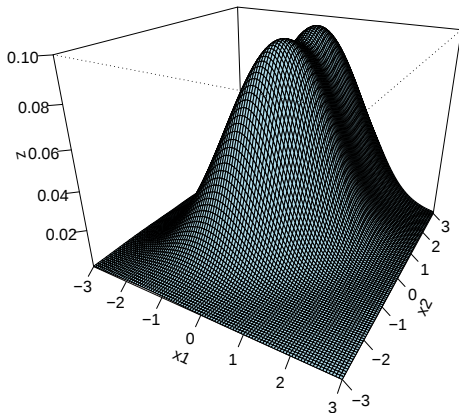
- Για τη διμεταβλητή κανονική κατανομή ισχύει πως αν οι τυχαίες μεταβλητές X και Y είναι ασυσχέτιστες τότε είναι και ανεξάρτητες, ισχύει και το αντίστροφο.

Γενικά το αντίστροφο δεν ισχύει

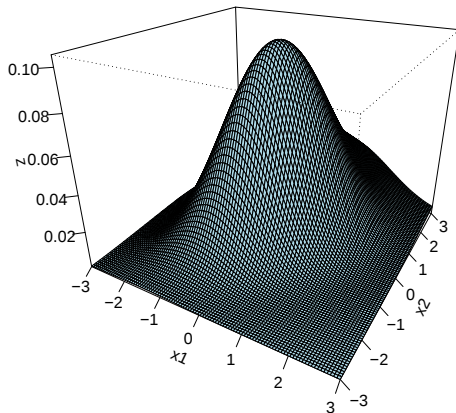
Διδιάστατη κανονική κατανομή με μηδενικό διάνυσμα μέσων και ίσες διασπορές

Δι διάστατη κανονική κατανομή με μηδενικούς μέσους,
ίσες διασπορές και υψηλή συσχέτιση

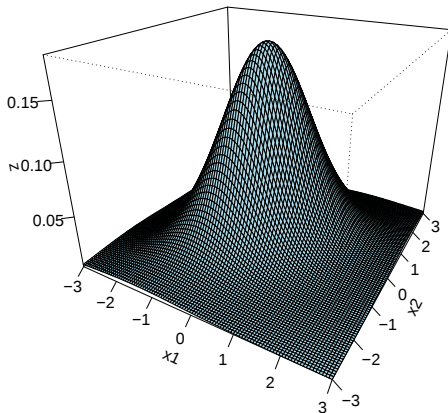
Άλλες συνεχείς κατανομές



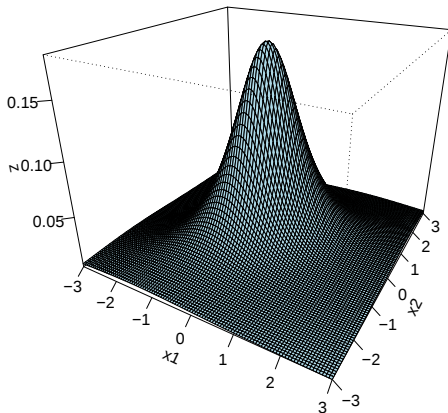
Άλλες συνεχείς κατανομές



Άλλες συνεχείς κατανομές



Άλλες συνεχείς κατανομές



Θεώρημα: Έστω $\mathbf{x}_{p \times 1} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, τότε

$$(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim \chi_p^2$$

Αποτελεί γενίκευση του

$$X \sim N(\mu, \sigma^2) \Rightarrow \frac{X - \mu}{\sigma} \sim N(0, 1) \Rightarrow \frac{(X - \mu)^2}{\sigma^2} \sim \chi_1^2$$

- Η εξίσωση $(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c^2$ περιγράφει μια έλλειψη
- Για την πολυμεταβλητή κανονική κατανομή αυτό αντιστοιχεί σε υψομετρικές καμπύλες σταθερής πυκνότητας
- Με άλλα λόγια, όλα τα σημεία με την ίδια πυκνότητα περιγράφονται από μια έλλειψη
- Περιοχές εμπιστοσύνης (confidence regions) για τα τυχαία διανύσματα

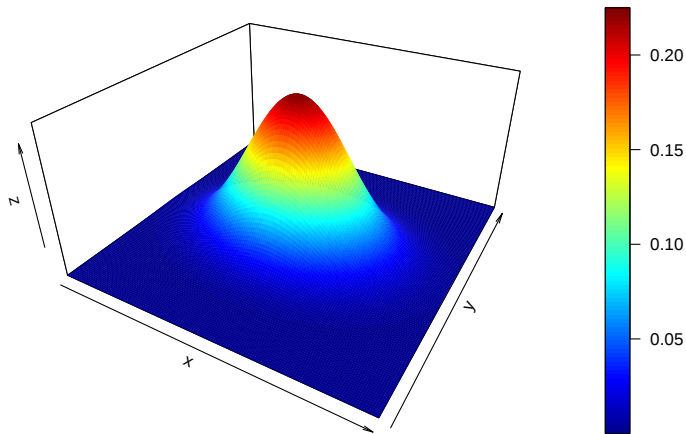
$$P \left\{ (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \leq \chi_{p,\alpha}^2 \right\} = \alpha$$

δηλαδή το $\alpha \times 100\%$ των παρατηρήσεων από την $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ βρίσκεται εντός της έλλειψης

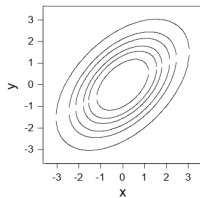
$$(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \chi_{p,\alpha}^2$$

- Υποθέτει πως το διάνυσμα των μέσων και ο πίνακας διακύμανσης είναι γνωστά.

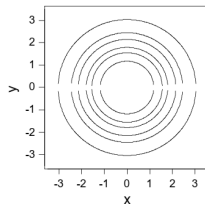
Υψομετρικές καμπύλες σταθερής πυκνότητας



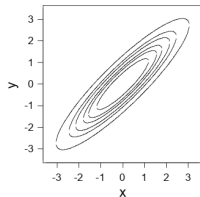
Υψομετρικές καμπύλες σταθερής πυκνότητας



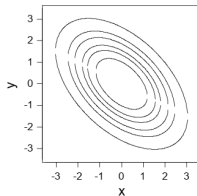
$\rho=0.5$



$\rho=0$

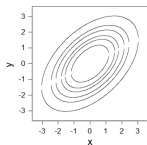


$\rho=0.9$

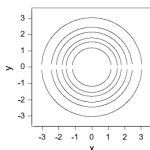


$\rho=-0.5$

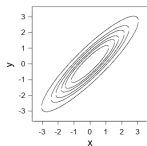
Υψομετρικές καμπύλες σταθερής πυκνότητας



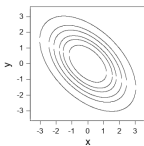
$\rho=0.5$



$\rho=0$



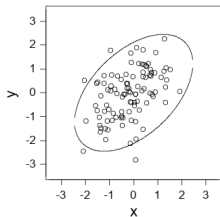
$\rho=0.9$



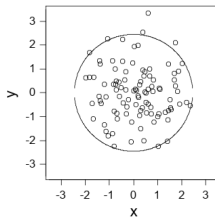
$\rho=0.5$

- Αν οι τυχαίες μεταβλητές X, Y ασυσχέτιστες με ίσες διασπορές $\rightarrow$ η έλλειψη εκφυλίζεται σε κύκλο. Στις περισσότερες διαστάσεις το υπερελλειψοειδές γίνεται
 - υπερσφαίρα
- Η απόλυτη τιμή του ρ καθορίζει τα χαρακτηριστικά της έλλειψης. (όσο τείνει στο 1 η έλλειψη τείνει να εκφυλιστεί σε γραμμή). Οι ελλείψεις αντιστοιχούν σε πιθανότητα 0.5, 0.7, 0.8, 0.9, 0.95, 0.99.

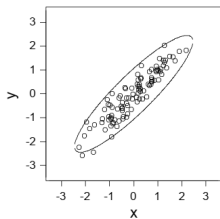
Περιοχές εμπιστοσύνης



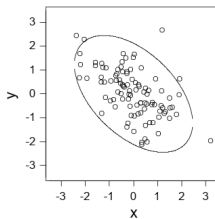
$$\rho=0.5$$



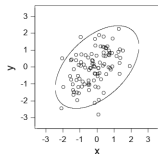
$$\rho=0$$



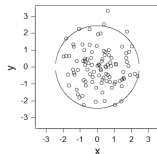
$$\rho=0.9$$



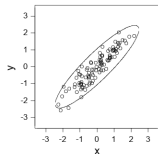
$$\rho=-0.5$$



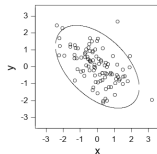
$\rho=0.5$



$\rho=0$



$\rho=0.9$



$\rho=-0.5$

- Προσομοίωση από $N(\mu = \mathbf{0}, \Sigma)$
- Από τρόπο κατασκευής περιμένουμε $\simeq 95\%$ των παρατηρήσεων να είναι μέσα στην έλλειψη.
- Η επιλογή των παραμέτρων δεν επηρεάζει τα συμπεράσματα:
 - το μ μετακινεί το κέντρο της έλλειψης,
 - ο Σ καθορίζει το μήκος των αξόνων

Θεώρημα

Έστω $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ και $\mathbf{a}$ διάνυσμα διαστάσεων $(p \times 1)$ από σταθερούς αριθμούς (όχι τ.μ.). Τότε για την τ.μ. $y = \mathbf{a}^\top \mathbf{x}$ ισχύει πως : $y \sim N(\mathbf{a}^\top \boldsymbol{\mu}, \mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a})$.

Ερμηνεία

- Η κατανομή γραμμικού συνδυασμού μεταβλητών με από κοινού κατανομή την πολυδιάστατη κανονική ακολουθεί την μονοδιάστατη κανονική κατανομή.
- Είναι γνωστό πως το αποτέλεσμα ισχύει για την περίπτωση ανεξάρτητων κανονικών τυχαίων μεταβλητών, τώρα παρατηρούμε πως το ίδιο ισχύει και για την περίπτωση που οι τ.μ. έχουν συσχέπιση μεταξύ τους.

Θεώρημα

Εστω $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{A}_{q \times p}$ πίνακας και $\mathbf{b}_{q \times 1}$ διάνυσμα από σταθερούς αριθμούς. Τότε για το τῶ ($q \times 1$) $\mathbf{y} = \mathbf{Ax} + \mathbf{b}$ ισχύει πως $\mathbf{y} \sim N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$.

Περιθώριες Από Κοινού Κατανομές

Έστω $\mathbf{x}^T = (X_1, X_2, \dots, X_p) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ και $\mathbf{y}^T = (X_1, X_2, \dots, X_q)$ ($q < p$).

Από κοινού περιθώρια κατανομή της $\mathbf{y}$:

$$\begin{aligned}\mathbf{y} &= \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_q \end{pmatrix} = \begin{pmatrix} 1X_1 + 0X_2 + \dots 0X_q + 0X_{q+1} + \dots 0X_p \\ 0X_1 + 1X_2 + \dots 0X_q + 0X_{q+1} + \dots 0X_p \\ \vdots \\ 0X_1 + 0X_2 + \dots 1X_q + 0X_{q+1} + \dots 0X_p \end{pmatrix} \\ &= \underbrace{\begin{pmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \end{pmatrix}}_{q \times q} \underbrace{\begin{pmatrix} 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix}}_{q \times (p-q)} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}_{p \times 1}\end{aligned}$$

Περιθώριες Από Κοινού Κατανομές (2)

Επομένως, $\mathbf{y} = \mathbf{A}_{q \times p} \mathbf{x}_{p \times 1}$, όπου $\mathbf{A} = (\mathbf{I}_{q \times q} \quad \mathbf{0}_{q \times (p-q)})$. Κάνοντας χρήση του προηγούμενου θεωρήματος

$$\mathbf{y} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)$$

Δηλαδή δημιουργούμε τους μερικούς πίνακες παίρνοντας συγκεκριμένα κομμάτια των πινάκων

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

- ο υποπίνακας $\boldsymbol{\Sigma}_{11}$ περιέχει τις διακυμάνσεις και τις συνδιακυμάνσεις όλων των μεταβλητών που περιέχονται στο διάνυσμα $\mathbf{y}^\top = (X_1, X_2, \dots, X_q)$ και αντίστοιχη ερμηνεία μπορεί να δοθεί στον πίνακα $\boldsymbol{\Sigma}_{22}$.
- οι πίνακες $\boldsymbol{\Sigma}_{12}$ και $\boldsymbol{\Sigma}_{21}$ δεν είναι συμμετρικοί και απλά περιέχουν τις συνδιακυμάνσεις των μεταβλητών του ενός διανύσματος με τις μεταβλητές του άλλου διανύσματος.

Περιθώριες Από Κοινού Κατανομές (3)

Έπεται ότι

$$\mathbf{A}\boldsymbol{\mu} = \underbrace{\begin{pmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \end{pmatrix}}_{q \times q} \underbrace{\begin{pmatrix} 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix}}_{q \times (p-q)} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}_{p \times 1} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{pmatrix}_{q \times 1} = \boldsymbol{\mu}_1$$

Επίσης

$$\begin{aligned} \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T &= \begin{pmatrix} \mathbf{I}_{q \times q} & \mathbf{0}_{q \times (p-q)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{q \times q} \\ \mathbf{0}_{q \times (p-q)} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{q \times q} \\ \mathbf{0}_{q \times (p-q)} \end{pmatrix} = \boldsymbol{\Sigma}_{11}. \end{aligned}$$

Άρα

$$\mathbf{y} \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}).$$

Συμπέρασμα

Δηλαδή κάθε περιθώρια (είτε απλή περιθώρια, δηλαδή για καθεμία μεταβλητή είτε από κοινού περιθώρια, δηλαδή για ζεύγη, τριάδες κλπ μεταβλητών) είναι πολυμεταβλητή κανονική κατανομή. Επομένως και κάθε απλή περιθώρια κατανομή είναι μια απλή κανονική κατανομή

- Στην περίπτωση που οι μεταβλητές δεν είναι στη σειρά πρέπει να προσέξουμε πως θα κατασκευάσουμε τους υποπίνακες που θέλουμε ώστε να είμαστε σίγουροι ότι αντιστοιχούν στην διαμέριση που ζητάμε

Έστω

$$\mathbf{x} : \mathbf{x} = \begin{pmatrix} \underbrace{\mathbf{x}_1}_{q \times 1} \\ \underbrace{\mathbf{x}_2}_{m \times 1} \end{pmatrix}$$

όπου $q + m = p$,

$$\boldsymbol{\mu} = \begin{pmatrix} \underbrace{\boldsymbol{\mu}_1}_{q \times 1} \\ \underbrace{\boldsymbol{\mu}_2}_{m \times 1} \end{pmatrix} \text{ και } \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}^T,$$

όπου $\boldsymbol{\Sigma}_{11}$ και $\boldsymbol{\Sigma}_{22}$ οι πίνακες συνδιακύμανσης των $\mathbf{x}_1$ και $\mathbf{x}_2$, αντίστοιχα (εάν $\boldsymbol{\Sigma}_{21} = \mathbf{0}$ τότε τα $\mathbf{x}_1$ και $\mathbf{x}_2$ είναι ανεξάρτητα).

Θεώρημα

Έστω $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, χρησιμοποιώντας την προηγούμενη διαμέριση $\mathbf{x}^\top = (\mathbf{x}_1^\top, \mathbf{x}_2^\top)$, για τη δεσμευμένη κατανομή ισχύει ότι

$$\mathbf{x}_1 | \mathbf{x}_2 \sim N(\boldsymbol{\mu}', \boldsymbol{\Sigma}')$$

όπου $\boldsymbol{\mu}' = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$ και $\boldsymbol{\Sigma}' = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$

Παρατηρήσεις:

- Το θεώρημα ισχύει και για την απλή δεσμευμένη κατανομή μιας τ.μ. ή τη δεσμευμένη κατανομή αγνοώντας κάποιες άλλες μεταβλητές.
- Ο πίνακας $\boldsymbol{\Sigma}'$ δεν εξαρτάται από τις τιμές των μεταβλητών ως προς τις οποίες έχουμε δεσμεύσει.

Δεσμευμένες Κατανομές (3)

Έστω 2 τμ $(X, Y) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, με $\boldsymbol{\mu}^\top = (\mu_x, \mu_y)$ και $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}$. Βάσει του προηγούμενου θεωρήματος, $Y|X = x \sim N(\cdot, \cdot)$ με

$$\begin{aligned} E(Y|X = x) &= \mu_y + \frac{\sigma_{xy}}{\sigma_x^2}(x - \mu_x) \\ \text{Var}(Y|X = x) &= \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2} \\ &= \sigma_y^2(1 - \rho^2) \end{aligned}$$

δεδομένου ότι $\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$.

Δεσμευμένες Κατανομές: Χρήσιμα συμπεράσματα

- Η δεσμευμένη διακύμανση δεν εξαρτάται από την τιμή x της τμ X , δηλαδή την πληροφορία. Γενικά αυτό δεν συμβαίνει συχνά!
- $Var(Y|X = x) \leq Var(Y)$ (η ισότητα ισχύει αν $\rho = 0$). Δηλαδή αν υπάρχει έστω και μικρή συσχέτιση ανάμεσα σε δύο μεταβλητές μπορούμε να μειώσουμε τη διακύμανση της μιας έχοντας κάποια πληροφορία για την άλλη.
- Το πρόσημο της συσχέτισης ανάμεσα σε δύο μεταβλητές δεν παίζει ρόλο.
- Η δεσμευμένη αναμενόμενη τιμή είναι γραμμική ως προς την τιμή της τμ X . Στη στατιστική η δεσμευμένη αναμενόμενη τιμή ονομάζεται παλινδρόμηση (regression)

- Υποθέσεις γραμμικού μοντέλου
 - Η αναμενόμενη τιμή του Y δοθέντος του $X = x$ είναι γραμμική ως προς το x
 - Η διακύμανση είναι σταθερή και ανεξάρτητη από το x .
 - Η κατανομή του Y είναι η κανονική.
- Τα α και β της γραμμικής παλινδρόμησης εκτιμούνται με τα δειγματικά αντίστοιχα των τύπων στην έκφραση της δεσμευμένης αναμενόμενης τιμής.

Έστω $\beta_{k \times 1}$ k -διάστατο διάνυσμα και $f(\beta) = f(\beta_1, \dots, \beta_k)$ μια πραγματική συνάρτηση στο $\mathbb{R}$, δηλαδή $f : \mathbb{R}^k \rightarrow \mathbb{R}$. Η παράγωγος της f ως προς β ορίζεται ως

$$\frac{\partial f(\beta)}{\partial \beta} = \begin{Bmatrix} \frac{\partial f(\beta)}{\partial \beta_1} \\ \vdots \\ \frac{\partial f(\beta)}{\partial \beta_k} \end{Bmatrix}_{k \times 1}$$

δηλαδή το k -διάστατο διάνυσμα των μερικών παραγώγων. Σαν σύμβαση, θεωρούμε το $1 \times k$ διάνυσμα (πίνακας γραμμή)

$$\frac{\partial f(\beta)}{\partial \beta^\top} = \left\{ \frac{\partial f(\beta)}{\partial \beta_1}, \dots, \frac{\partial f(\beta)}{\partial \beta_k} \right\}_{1 \times k}.$$

Παραγωγή διανυσματικών συναρτήσεων

Έστω

$$\mathbf{g}(\beta) = \begin{Bmatrix} \mathbf{g}_1(\beta) \\ \vdots \\ \mathbf{g}_n(\beta) \end{Bmatrix}_{n \times 1}$$

όπου $\mathbf{g}$ είναι μια διανυσματική συνάρτηση της μεταβλητής β , $\mathbf{g} : \mathbb{R}^k \rightarrow \mathbb{R}^n$.
Επειδή η $\mathbf{g}$ είναι ουσιαστικά πίνακας στήλη, ορίζουμε την παράγωγο της $\mathbf{g}$ ως προς το $\beta^\top$ (πίνακας γραμμή)

$$\frac{\partial \mathbf{g}(\beta)}{\partial \beta^\top} = \begin{Bmatrix} \frac{\partial \mathbf{g}_1(\beta)}{\partial \beta_1} & \cdots & \frac{\partial \mathbf{g}_1(\beta)}{\partial \beta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathbf{g}_n(\beta)}{\partial \beta_1} & \cdots & \frac{\partial \mathbf{g}_n(\beta)}{\partial \beta_k} \end{Bmatrix}_{n \times k}$$

Ισχύει πάντα ότι

$$\frac{\partial \mathbf{g}^\top(\beta)}{\partial \beta} = \left\{ \frac{\partial \mathbf{g}(\beta)}{\partial \beta^\top} \right\}^\top.$$

Έστω $f(\beta) \in \mathbb{R}$, ο Εσσιανός πίνακας ορίζεται ως

$$\frac{\partial^2 f(\beta)}{\partial \beta \partial \beta^\top} = \frac{\partial^2 f(\beta)}{\partial \beta^\top \partial \beta} = \left\{ \begin{array}{ccc} \frac{\partial^2 f(\beta)}{\partial \beta_1 \partial \beta_1} & \cdots & \frac{\partial^2 f(\beta)}{\partial \beta_1 \partial \beta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\beta)}{\partial \beta_k \partial \beta_1} & \cdots & \frac{\partial^2 f(\beta)}{\partial \beta_k \partial \beta_k} \end{array} \right\}_{k \times k}$$

ο οποίος είναι ένας $k \times k$ **συμμετρικός** πίνακας (η σειρά της παραγωγίσης δεν παίζει ρόλο).

Έστω $f(\beta) = \mathbf{c}^\top \beta = c_1\beta_1 + \dots + c_k\beta_k$, όπου $\beta_{k \times 1}$ και $\mathbf{c}_{k \times 1}$ (σταθερά). Άρα,

$$\frac{\partial \mathbf{c}^\top \beta}{\partial \beta} = \left\{ \begin{array}{c} \frac{\partial(c_1\beta_1 + \dots + c_k\beta_k)}{\partial \beta_1} \\ \vdots \\ \frac{\partial(c_1\beta_1 + \dots + c_k\beta_k)}{\partial \beta_k} \end{array} \right\} = \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix} = \mathbf{c}.$$

Επειδή $\mathbf{c}^\top \beta = \beta^\top \mathbf{c} \Rightarrow \frac{\partial \beta^\top \mathbf{c}}{\partial \beta} = \mathbf{c}$.

Παραδείγμα

Έστω $g(\beta) = \mathbf{A}\beta$, όπου $\mathbf{A}_{n \times k}$ σταθερός πίνακας.

$$g(\beta) = \mathbf{A}\beta = \begin{pmatrix} a_{11}\beta_1 + \cdots + a_{1k}\beta_k \\ \vdots \\ a_{n1}\beta_1 + \cdots + a_{nk}\beta_k \end{pmatrix}_{n \times 1}$$

Επομένως,

$$\frac{\partial \mathbf{A}\beta}{\partial \beta^\top} = \left\{ \begin{array}{ccc} \frac{\partial(a_{11}\beta_1 + \cdots + a_{1k}\beta_k)}{\partial \beta_1} & \cdots & \frac{\partial(a_{11}\beta_1 + \cdots + a_{1k}\beta_k)}{\partial \beta_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial(a_{n1}\beta_1 + \cdots + a_{nk}\beta_k)}{\partial \beta_1} & \cdots & \frac{\partial(a_{n1}\beta_1 + \cdots + a_{nk}\beta_k)}{\partial \beta_k} \end{array} \right\} = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nk} \end{pmatrix} = \mathbf{A}.$$

Ανάλογα, $\frac{\partial \beta^\top \mathbf{A}}{\partial \beta} = \mathbf{A}^\top$.

Έστω $f(\beta) = \beta^\top \mathbf{A} \beta$ για κάποιον πίνακα $\mathbf{A}$. Τότε αποδεικνύεται ότι

$$\begin{aligned}\frac{\partial \beta^\top \mathbf{A} \beta}{\partial \beta} &= (\mathbf{A} + \mathbf{A}^\top) \beta \\ \frac{\partial \beta^\top \mathbf{A} \beta}{\partial \beta^\top} &= \beta^\top (\mathbf{A} + \mathbf{A}^\top)\end{aligned}$$

Εάν $\mathbf{A}$ συμμετρικός, $\frac{\partial \beta^\top \mathbf{A} \beta}{\partial \beta} = 2\mathbf{A} \beta$. Επιπλέον,

$$\frac{\partial^2 \beta^\top \mathbf{A} \beta}{\partial \beta \partial \beta^\top} = \mathbf{A} + \mathbf{A}^\top.$$

Παράδειγμα: το κανονικό γραμμικό μοντέλο

Έστω το μοντέλο πολλαπλής γραμμικής παλινδρόμησης

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i, i = 1, 2, \dots, n$$

$$\epsilon_i \sim N(0, \sigma^2)$$

$$\epsilon_i \perp\!\!\!\perp \epsilon_j \ (i \neq j) \text{ ανεξάρτητες}$$

Το μοντέλο μπορεί να γραφεί στην μορφή πινάκων ως

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}_{n \times 1}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}_{n \times (p+1)}$$

Παράδειγμα: το κανονικό γραμμικό μοντέλο (2)

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}_{(p+1) \times 1}, \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}_{n \times 1},$$

Επομένως,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Παρατηρήσεις

- Στην θεωρία της γραμμικής παλινδρόμησης, ο πίνακας σχεδιασμού (design matrix) θεωρείται σταθερός - εξετάζουμε την κατανομή $\mathbf{Y}|\mathbf{X}$, άρα $\Rightarrow$
- $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ και $Var(\mathbf{Y}) = Var(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_n$ αφού τα σφάλματα είναι ανεξάρτητα μεταξύ τους.
- Έπεται από τα θεωρήματα γραμμικού συνδυασμού κανονικών

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n).$$

Παράδειγμα: το κανονικό γραμμικό μοντέλο (3)

Για να βρούμε εκτιμητές των παραμέτρων, αρκεί να ελαχιστοποιήσουμε το άθροισμα των τετραγωνικών αποστάσεων

$$Q(\beta) = (\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta) = \mathbf{Y}^\top \mathbf{Y} + \beta \mathbf{X}^\top \mathbf{X} \beta - 2\beta^\top \mathbf{X}^\top \mathbf{Y}$$

Παραγωγίζουμε ως προς β

$$\begin{aligned}\frac{\partial Q(\beta)}{\partial \beta} &= 2\mathbf{X}^\top \mathbf{X} \beta - 2\mathbf{X}^\top \mathbf{Y} = \mathbf{0} \Rightarrow \\ (\mathbf{X}^\top \mathbf{X}) \hat{\beta} &= \mathbf{X}^\top \mathbf{Y} \Rightarrow \\ \hat{\beta} &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}\end{aligned}$$

- Πρέπει ο πίνακας $\mathbf{X}^\top \mathbf{X}$ να είναι **θετικά ορισμένος**.
- Αυτό μπορεί να επιβεβαιωθεί εφόσον καμία μεταβλητή (στήλη του $\mathbf{X}$) εκ των $X_1, X_2, \dots, X_p$ δεν είναι γραμμικός συνδυασμός των υπολοίπων (θυμηθείτε το ζήτημα της πολυσυγγραμμικότητας).

Παράδειγμα: το κανονικό γραμμικό μοντέλο (4)

Συμπεράσματα

- Η κατανομή του β είναι $(p + 1)$ -διάστατη κανονική ως γραμμικός συνδυασμών κανονικών κατανομών

$$\begin{aligned}E(\hat{\beta}) &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top E(\mathbf{Y}) = (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X}) \beta = \beta \\ \text{Var}(\hat{\beta}) &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\sigma^2 I_n) \mathbf{X} \{(\mathbf{X}^\top \mathbf{X})^{-1}\}^\top \\ &= \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X}) (\mathbf{X}^\top \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}\end{aligned}$$

Επομένως,

$$\hat{\beta} \sim N\{\beta, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}\}$$

Έστω $\Sigma_{p \times p} = (\sigma_{ij})$ ένας $p \times p$ πίνακας ανεξάρτητων μεταβλητών (σ_{ij}) , $i, j = 1, 2, \dots, p$. Επειδή

$$\frac{\partial \sigma_{st}}{\partial \sigma_{ij}} = \begin{cases} 1, \text{ εάν } s = i \text{ και } t = j \\ 0, \text{ διαφορετικά} \end{cases}$$

ορίζουμε ότι

$$\frac{\partial \Sigma}{\partial \sigma_{ij}} = \mathbf{u}_i \mathbf{u}_j^\top \{1 \text{ στο } (i, j) \text{ στοιχείο και } 0 \text{ αλλού}\}$$

όπου $\mathbf{u}_i$ είναι η i στήλη του $\mathbf{I}_p$ και $\mathbf{u}_j$ η j στήλη του $\mathbf{I}_p$.

Παραγωγή ίχνους πίνακα

Έστω $\mathbf{\Sigma}_{p \times p} = (\sigma_{ij})$ ένας $p \times p$ πίνακας ανεξάρτητων μεταβλητών και $\mathbf{A}_{p \times p}$ σταθερός πίνακας (δεν μεταβάλλεται ως προς σ_{ij}). Τότε

$$\begin{aligned}\frac{\partial \text{tr}(\mathbf{A}\mathbf{\Sigma})}{\partial \sigma_{ij}} &= \text{tr} \left(\mathbf{A} \frac{\partial \mathbf{\Sigma}}{\partial \sigma_{ij}} \right) = \text{tr}(\mathbf{A} \mathbf{u}_i \mathbf{u}_j^\top) = \text{tr}(\mathbf{u}_j^\top \mathbf{A} \mathbf{u}_i) \\ &= \mathbf{u}_j^\top \mathbf{A} \mathbf{u}_i = a_{ji} \text{ \{το } (j, i) \text{ στοιχείο του } \mathbf{\Sigma} \text{ \}}\end{aligned}$$

Άρα σε μορφή πινάκων

$$\frac{\partial \text{tr}(\mathbf{A}\mathbf{\Sigma})}{\partial \mathbf{\Sigma}} = \frac{\partial \text{tr}(\mathbf{\Sigma}\mathbf{A})}{\partial \mathbf{\Sigma}} = \mathbf{A}^\top.$$

Παραγωγή ορίζουσας πίνακα

Έστω $\mathbf{\Sigma}_{p \times p} = (\sigma_{ij})$ πίνακας ανεξάρτητων μεταβλητών. Υπενθυμίζεται ότι $\det(\mathbf{\Sigma}) = \sum_{t=1}^m \sigma_{it} \xi_{it}$, όπου $\xi_{it} = (-1)^{t+1} \det(\mathbf{\Sigma}_{it}^*)$ είναι **ανεξάρτητα (δεν μεταβάλλονται ως προς σ_{it})**. $\mathbf{\Sigma}_{it}^* \rightarrow$ πίνακας που προκύπτει διαγράφοντας την i -οστή γραμμή και t -οστή στήλη του $\mathbf{\Sigma}$. Επομένως,

$$\frac{\partial \det(\mathbf{\Sigma})}{\partial \sigma_{ij}} = \xi_{ij}.$$

Στην γραμμική άλγεβρα, ο ανάστροφος πίνακας συμπαραγόντων (cofactor matrix), $\text{adj}(\mathbf{\Sigma})$ ορίζεται ως $\text{adj}(\mathbf{\Sigma}) = (\xi_{ji})$. Επιπλέον, εάν $\mathbf{\Sigma}$ αντιστρέψιμος, ισχύει ότι $\mathbf{\Sigma}^{-1} = \frac{\text{adj}(\mathbf{\Sigma})}{\det(\mathbf{\Sigma})}$, άρα συνεπάγεται ότι

$$\frac{\partial \det(\mathbf{\Sigma})}{\partial \mathbf{\Sigma}} = \text{adj}(\mathbf{\Sigma})^\top = |\mathbf{\Sigma}| (\mathbf{\Sigma}^{-1})^\top.$$

Χρησιμοποιώντας τον κανόνα της αλυσίδας,

$$\frac{\partial \log \det(\mathbf{\Sigma})}{\partial \mathbf{\Sigma}} = \frac{1}{|\mathbf{\Sigma}|} \frac{\partial \det(\mathbf{\Sigma})}{\partial \mathbf{\Sigma}} = (\mathbf{\Sigma}^{-1})^T = \mathbf{\Sigma}^{-T}.$$

Παραγωγή αντιστρόφου πίνακα

Εάν $\mathbf{F} = \mathbf{F}(\mathbf{x})_{p \times p}$ πίνακας ο οποίος είναι συνάρτηση ενός διανύσματος $\mathbf{x}$, αποδεικνύεται εύκολα ότι (διότι $\mathbf{F}\mathbf{F}^{-1} = \mathbf{I}$)

$$\frac{\partial \mathbf{F}^{-1}}{\partial x_j} = -\mathbf{F}^{-1} \frac{\partial \mathbf{F}}{\partial x_j} \mathbf{F}^{-1},$$

για παράδειγμα,

$$\begin{aligned} \frac{\partial \operatorname{tr}(\mathbf{\Sigma}^{-1} \mathbf{A})}{\partial \sigma_{jj}} &= \operatorname{tr} \left(\frac{\partial \mathbf{\Sigma}^{-1}}{\partial \sigma_{jj}} \mathbf{A} \right) = -\operatorname{tr} \left(\mathbf{\Sigma}^{-1} \mathbf{u}_j \mathbf{u}_j^T \mathbf{\Sigma}^{-1} \mathbf{A} \right) \\ &= -\mathbf{u}_j^T \mathbf{\Sigma}^{-1} \mathbf{A} \mathbf{\Sigma}^{-1} \mathbf{u}_j = -[\mathbf{\Sigma}^{-1} \mathbf{A} \mathbf{\Sigma}^{-1}]_{jj} \end{aligned}$$

Επομένως,

$$\frac{\partial \operatorname{tr}(\mathbf{\Sigma}^{-1} \mathbf{A})}{\partial \mathbf{\Sigma}} = -(\mathbf{\Sigma}^{-1} \mathbf{A} \mathbf{\Sigma}^{-1})^T.$$

Εκτίμηση μέγιστης πιθανοφάνειας για τις παραμέτρους της πολυδιάστατης κανονικής κατανομής

Έστω $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^\top$, $i = 1, 2, \dots, n$ τυχαίο δείγμα από την $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Η πιθανοφάνεια του μοντέλου είναι ίση με

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^n f(\mathbf{x}_i | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \prod_{i=1}^n \left[\frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right\} \right] \end{aligned}$$

Σε λογαριθμική κλίμακα (αγνοώντας σταθερούς όρους)

$$\ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = -\frac{n}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}).$$

$$\begin{aligned}\frac{\partial \ell(\mu, \Sigma)}{\partial \mu} &= -\frac{1}{2} \frac{\partial}{\partial \mu} \sum_{i=1}^n \left(\mathbf{x}_i^\top \Sigma^{-1} \mathbf{x}_i + \mu^\top \Sigma \mu - 2\mu^\top \Sigma^{-1} \mathbf{x}_i \right) \\ &= -\frac{1}{2} \sum_{i=1}^n (2\Sigma^{-1} \mu - 2\Sigma^{-1} \mathbf{x}_i) = \Sigma^{-1} \sum_{i=1}^n (\mathbf{x}_i - \mu) = \mathbf{0} \\ \Rightarrow \hat{\mu} &= n^{-1} \sum_{i=1}^n \mathbf{x}_i = \bar{\mathbf{x}}\end{aligned}$$

ο (διανυσματικός) δειγματικός μέσος των δεδομένων.

Παραγωγή ως προς Σ

Για την παραγωγή ως προς Σ , αντικαθιστούμε την παράμετρο μ με την εκτίμησή της

$$\ell(\hat{\mu}, \Sigma) = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \text{tr} \left\{ \Sigma^{-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top \right\}$$

Επειδή ο πίνακας Σ είναι συμμετρικός, με βάση τους κανόνες παραγωγής της ορίζουσας, του ίχνους, και του αντιστρόφου πινάκων, θέτοντας $\mathbf{A} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$ (συμμετρικός πίνακας), προκύπτει ότι,

$$\begin{aligned} \frac{\partial \ell(\hat{\mu}, \Sigma)}{\partial \Sigma} &= -\frac{n}{2} \frac{\partial \log |\Sigma|}{\partial \Sigma} - \frac{1}{2} \frac{\partial}{\partial \Sigma} \text{tr} (\Sigma^{-1} \mathbf{A}) \\ &= -\frac{n}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} \mathbf{A} \Sigma^{-1} = \mathbf{0}, \end{aligned}$$

και πολλαπλασιάζοντας από αριστερά και δεξιά με Σ , καταλήγουμε στο $\hat{\Sigma} = n^{-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^\top$.