

Πολυδιάστατη Κανονική Κατανομή

Περιεχόμενα

Πολυδιάστατη κανονική κατανομή

Διμεταβλητή κανονική κατανομή

Ορισμός:

Έστω ένα τυχαίο διάνυσμα (στήλης) $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ με ανεξάρτητες και ισόνομες τυχαίες μεταβλητές όπου η κάθε μια ακολουθεί την κανονική κατανομή. Τότε θα λέμε ότι οι τυχαίες μεταβλητές $X_1, X_2, \dots, X_n$ ακολουθούν την πολυμεταβλητή κανονική κατανομή (ή αλλιώς ότι κατανέμονται από κοινού κανονικά) με από κοινού συνάρτηση πυκνότητας πιθανότητας

$$f(\mathbf{x}; \mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \cdot \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right], \quad \mathbf{x} \in \mathbb{R}^n$$

όπου $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_n]'$ είναι ένα $(n \times 1)$ διάνυσμα στήλης του οποίου τα στοιχεία είναι πραγματικοί αριθμοί. Επίσης $\Sigma = [\sigma_{ij}]$, $i, j = 1, 2, \dots, n$ είναι ένας $(n \times n)$ θετικά ορισμένος και συμμετρικός πίνακας (δηλαδή $\Sigma = \Sigma'$), όπου σ_{ij} πραγματικοί αριθμοί.

Θα συμβολίζουμε ότι το τυχαίο διάνυσμα $\mathbf{X}$ ακολουθεί την πολυμεταβλητή κατανομή με μέση τιμή $\boldsymbol{\mu}$ και διακύμανση Σ , ως

$$\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma).$$

Ροπογεννήτρια συνάρτηση:

$$M_{\mathbf{X}}(\mathbf{t}) = M(\mathbf{t}) = E \left[e^{\mathbf{t}' \cdot \mathbf{X}} \right] = \exp \left(\mathbf{t}' \cdot \boldsymbol{\mu} + \frac{1}{2} \mathbf{t}' \cdot \boldsymbol{\Sigma} \cdot \mathbf{t} \right), \quad \mathbf{t} \in \mathbb{R}^n$$

Χαρακτηριστική συνάρτηση:

$$\varphi_{\mathbf{X}}(\mathbf{t}) = E \left[e^{i \mathbf{t}' \cdot \mathbf{X}} \right] = \exp \left(i \mathbf{t}' \cdot \boldsymbol{\mu} + \frac{1}{2} \mathbf{t}' \cdot \boldsymbol{\Sigma} \cdot \mathbf{t} \right), \quad \mathbf{t} \in \mathbb{R}^n$$

Πρόταση:

Αν οι τυχαίες μεταβλητές $X_1, X_2, \dots, X_n$ κατανέμονται από κοινού κανονικά και είναι ανά δύο γραμμικά ανεξάρτητες, δηλαδή $\rho_{ij} = 0$, $i, j = 1, 2, \dots, n$, τότε θα είναι και ανεξάρτητες.

Περιθώριες κατανομές

Πρόταση:

Αν καθεμία από τυχαίες μεταβλητές $X_1, X_2, \dots, X_n$ ακολουθεί τη μονοδιάστατη κανονική κατανομή τότε δεν συνεπάγεται ότι υποχρεωτικά οι $X_1, X_2, \dots, X_n$ ακολουθούν την από κοινού κανονική κατανομή.

Δεσμευμένες κατανομές

Έστω το τυχαίο διάνυσμα $\mathbf{X} = [X_1, X_2, \dots, X_n]'$ το οποίο ακολουθεί την πολυμεταβλητή κατανομή με μέση τιμή $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_n]'$ και διακύμανση $\Sigma = [\sigma_{ij}]$, $i, j = 1, 2, \dots, n$.
Έστω ότι διαμερίζουμε το $(n \times 1)$ διάνυσμα $\mathbf{X}$ σε ένα k -διάστατο διάνυσμα $\mathbf{X}_1$ και σε ένα $(n - k)$ -διάστατο διάνυσμα $\mathbf{X}_2$ δηλαδή

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$$

με $\mathbf{X}_1 = [X_1, X_2, \dots, X_k]'$ και $\mathbf{X}_2 = [X_{k+1}, X_{k+2}, \dots, X_n]'$.

Η μέση τιμή και η διακύμανση τότε για το τυχαίο διάνυσμα $\mathbf{X}$ θα είναι

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

με $\boldsymbol{\mu}_1 = [\mu_1, \mu_2, \dots, \mu_k]'$, $\boldsymbol{\mu}_2 = [\mu_{k+1}, \mu_{k+2}, \dots, \mu_n]'$ και

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

αντίστοιχα.

Εἰναι

$$\Sigma_{11} = E[(X_1 - \boldsymbol{\mu}_1)(X_1 - \boldsymbol{\mu}_1)'] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{bmatrix}$$

$$\begin{aligned} \Sigma_{22} &= E[(X_2 - \boldsymbol{\mu}_2)(X_2 - \boldsymbol{\mu}_2)'] \\ &= \begin{bmatrix} \sigma_{k+1,k+1} & \sigma_{k+1,k+2} & \cdots & \sigma_{k+1,n} \\ \sigma_{k+2,k+1} & \sigma_{k+2,k+2} & \cdots & \sigma_{k+2,n} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{n,k+1} & \sigma_{n,k+2} & \cdots & \sigma_{n,n} \end{bmatrix} \end{aligned}$$

και

$$\begin{aligned}\Sigma_{12} &= E[(X_1 - \boldsymbol{\mu}_1)(X_2 - \boldsymbol{\mu}_2)'] \\ &= \begin{bmatrix} \sigma_{1,k+1} & \sigma_{1,k+2} & \cdots & \sigma_{1,n} \\ \sigma_{2,k+1} & \sigma_{2,k+2} & \cdots & \sigma_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{k,k+1} & \sigma_{k,k+2} & \cdots & \sigma_{k,n} \end{bmatrix}\end{aligned}$$

Ο πίνακας Σ_{12} σε αντίθεση με τους πίνακες Σ_{11} και Σ_{22} δεν είναι συμμετρικός.

Ορισμός

Έστω το τυχαίο διάνυσμα $\mathbf{X} = [X_1, X_2, \dots, X_n]'$ το οποίο ακολουθεί την πολυμεταβλητή κανονική κατανομή με παραμέτρους $\boldsymbol{\mu}, \boldsymbol{\Sigma}$, δηλαδή $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, όπως ορίστηκαν προηγουμένως με τις αντίστοιχες διαμερίσεις τους. Τότε η δεσμευμένη (από κοινού) συνάρτηση πυκνότητας πιθανότητας, $f_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1|\mathbf{x}_2)$, του $\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2$ είναι η k -διάσταση κανονική

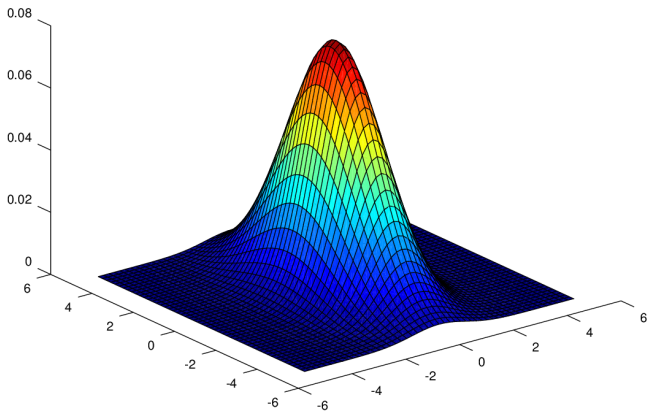
$$f_{\mathbf{X}_1|\mathbf{X}_2}(\mathbf{x}_1|\mathbf{x}_2) = \frac{1}{(2\pi)^{\frac{1}{2}k} \sqrt{|\Sigma_{11.2}|}} \cdot \exp \left\{ -\frac{1}{2} \left[(\mathbf{x}_1 - \boldsymbol{\mu}_1) - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)' \right] \times \right. \\ \left. \times \Sigma_{11.2}^{-1} \left[(\mathbf{x}_1 - \boldsymbol{\mu}_1) - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \right] \right\}$$

Μέση τιμή:

$$E(\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2) = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)$$

Πίνακας διακυμάνσεων-συνδιακυμάνσεων:

$$\text{Cov}(\mathbf{X}_1|\mathbf{X}_2 = \mathbf{x}_2) = \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$



Τι παρατηρείται για τη δεσμευμένη μέση τιμή και τον δεσμευμένο πίνακα διακυμάνσεων-συνδιακυμάνσεων;

Διμεταβλητή κανονική κατανομή

Έστω ένα τυχαίο διάνυσμα $\mathbf{X} = [X_1, X_2]'$ το οποίο ακολουθεί τη διμεταβλητή κανονική κατανομή, με μέση τιμή $\boldsymbol{\mu}$ και διακύμανση Σ

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{και} \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

δηλαδή $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$.

Δεσμευμένη μέση τιμή και διακύμανση:

$$E(X_1 | X_2 = x_2) = \mu_1 + \sigma_{12}\sigma_{22}^{-1}(x_2 - \mu_2)$$
$$Var(X_1 | X_2 = x_2) = \sigma_{11.2} = \sigma_{11} - \sigma_{12}\sigma_{22}^{-1}\sigma_{21}$$