2. Multinormaalijakauma

2.1 Alustavaa johdattelua

Monimuuttujamenetelmissä multinormaalijakaumalla on ehkä vielä keskeisempi asema kuin normaalijakaumalla yhden muuttujan tilastollisissa tarkasteluissa. Multinormaalijakauma on suora normaalijakauman yleistys.

Se voidaan johtaa usealla tavalla. Havainnollisinta on määritellä se toisistaan riippumattomien, normaalijakaumaa noudattavien muuttujien erilaisten painotettujen summien yhteisjakaumana esim. seuraavasti:

Olkoot $Z_1,Z_2,...,Z_p$ riippumattomia, standardoitua normaalijakaumaa $N(0,1)$ noudattavia muuttujia. Muodostetaan uudet muuttujat $X_1,X_2,...,X_p$ Z-muuttujien lineaarisina yhdistelminä

$$
\begin{aligned}
X_1 &= c_{11}Z_1 + c_{12}Z_2 + ... + c_{1p}Z_p + \mu_1\\
X_2 &= c_{21}Z_1 + c_{22}Z_2 + ... + c_{2p}Z_p + \mu_2\\
&...\\
X_p &= c_{p1}Z_1 + c_{p2}Z_2 + ... + c_{pp}Z_p + \mu_p
\end{aligned}
$$

eli matriisimuodossa

$$\mathbf{X} = \mathbf{CZ} + \mu$$

missä $\mathbf{X}=(X_1,X_2,...,X_p)$ on X-muuttujien muodostama pystyvektori ja vastaavasti $\mathbf{Z}=(Z_1,Z_2,...,Z_p)$, $\mu=(\mu_1,\mu_2,...,\mu_p)$ sekä $\mathbf{C}$ $p\times p$-kerroinmatriisi.

Muuttujien $X_1,X_2,...,X_p$ yhteisjakaumaa sanotaan multinormaalijakaumaksi ja sen määrittelevät täydellisesti parametrit μ ja $\mathbf{C}$. Itse asiassa tulemme näkemään, että jakauman määrittelemiseksi riittää tuntea odotusarvovektorin μ ohella kovarianssimatriisi $\Sigma = \mathbf{CC}'$.

Multinormaalijakauman syntytapa tulee vielä havainnollisemmaksi käyttämällä hyväksi kerroinmatriisin $\mathbf{C}$ singulaariarvohajotelmaa $\mathbf{C} = \mathbf{UDV}'$, missä $\mathbf{U}$ ja $\mathbf{V}$ ovat $p\times p$-ortogonaalisia matriiseja ja $\mathbf{D}$ (ei-negatiivisten) singulaariarvojen muodostama lävistäjämatriisi.

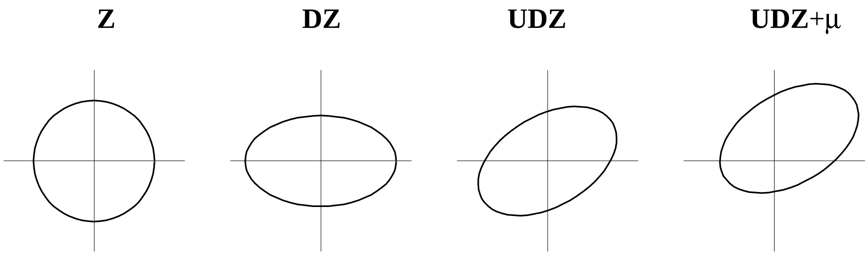
Tällöin

$$\mathbf{X} = \mathbf{CZ} + \mu = \mathbf{UDV}'\mathbf{Z} + \mu .$$

Tulemme osoittamaan, että Z-muuttujien ortogonaalinen muunnos (tässä $\mathbf{V}'\mathbf{Z}$) säilyttää muuttujat riippumattomina (0,1)-normaalisina. Näin ollen X-muuttujat voitaisiin määritellä suoraan muodossa

$$\mathbf{X} = \mathbf{UDZ} + \mu .$$

Tämä merkitsee, että multinormaalijakauman voi aina ajatella syntyvän (0,1)-normaalisista muuttujista kolmessa vaiheessa. Ensin tehdään muuttujittain venytyksiä ja kutistuksia (**DZ**), sitten kierretään koordinaatistoa (**UDZ**) ja lopuksi siirretään jakauman keskipiste pois origosta (lisäämällä μ).



Tulemme näkemään, että multinormaalijakauman kaikki eriulotteiset reunajakaumatkin ovat (multi)normaalisia. Tämä merkitsee mm. sitä, että multinormaalijakauman em. syntyhistoriassa *Z*-muuttujia voisi olla enemmän kuin lopullisia *X*-muuttujia. Vaikka *Z*-muuttujat eivät olisikaan normaalisia, mutta niitä on "paljon", on osoitettavissa keskeisen raja-arvolauseen tapaan, että *X*-muuttujien yhteisjakauma melko väljin ehdoin lähestyy multinormaalijakaumaa.

Samoin tarkasteltaessa osaa *X*-muuttujista, näiden ehdollinen yhteisjakauma, kun muut *X*-muuttujat asetetaan vakioiksi, on multinormaalinen ja regressiofunktiot (ehdolliset odotusarvot) ovat vakioksi asetettujen *X*-muuttujien lineaarisia funktioita. Tämä viimeinen ominaisuus on myös po. jakauman määritelmän veroinen.

Multinormaalijakauman syntyessä riippumattomien muuttujien lineaaristen yhdistelmien kautta on ilmeistä, että *X*-muuttujien välillä voi vallita vain lineaarisia riippuvuuksia eli korrelaatiokertoimet paljastavat kaiken, mikä koskee muuttujien välisiä riippuvuuksia. Tässä tapauksessa siis korreloimattomuus takaa myös muuttujien riippumattomuuden; seikka, mikä ei välttämättä päde yleisesti moniulotteisissa jakaumissa.

Tämän pohjalta tulee ilmeiseksi, että kaikki multinormaalisuutta edellyttävät tarkastelut saatetaan tehdä muuttujien odotusarvojen, keskihajontojen ja korrelaatiokertoimien avulla. Näiden tunnuslukujen tavanomaiset empiiriset vastineet satunnaisotoksesta laskettuina ovat tyhjentäviä otossuureita eikä esim. korkeamman asteen momentteja tarvita muuta kuin eräissä multinormaalisuutta tutkivissa testeissä.

2.2 Multinormaalijakauman määritelmä ja perusominaisuudet

Tarkennamme äskeistä kuvausta seuraavasti. Olkoot $U_1,U_2,...,U_k$ riippumattomia ja (0,1)-normaalisia satunnaismuuttujia ja $\mathbf{U}=(U_1,U_2,...,U_k)$ niiden muodostama satunnaisvektori. Tällöin odotusarvovektori $\mathrm{E}(\mathbf{U})=\mathbf{0}$ ja kovarianssimatriisi $\mathrm{cov}(\mathbf{U})=\mathbf{I}$.

Jokaisen U_i tiheysfunktio on muotoa

$$\phi(u_i) = (2\pi)^{-1/2}\exp(-1/2u_i^2) .$$

Tällöin U-muuttujien riippumattomuuden perusteella satunnaisvektorin $\mathbf{U}$ tiheysfunktio voidaan kirjoittaa näiden komponenttimuuttujien tiheysfunktioiden tulona

$$\begin{aligned} f(\mathbf{u}) = f(u_1,u_2,\dots,u_k) &= \phi(u_1)\phi(u_2)\dots\phi(u_k) \\ &= (2\pi)^{-k/2}\exp(-1/2(u_1^2 + u_2^2 + \dots + u_k^2)) \\ &= (2\pi)^{-k/2}\exp(-1/2\mathbf{u}'\mathbf{u}) . \end{aligned}$$

Määritellään uusi muuttujavektori $\mathbf{X}=(X_1,X_2,\dots,X_p)$ lineaarikuvauksella

$$\begin{aligned} X_1 &= c_{11}U_1 + c_{12}U_2 + \dots + c_{1k}U_k + \mu_1 \\ X_2 &= c_{21}U_1 + c_{22}U_2 + \dots + c_{2k}U_k + \mu_2 \\ &\dots \\ X_p &= c_{p1}U_1 + c_{p2}U_2 + \dots + c_{pk}U_k + \mu_p \end{aligned}$$

eli

(1) $\mathbf{X} = \mathbf{CU} + \mu$.

Oletetaan, että $p \le k$ ja matriisin $\mathbf{C}$ aste $r(\mathbf{C})=p$. Muussa tapauksessa muuttujat X olisivat lineaarisesti toisistaan riippuvia eikä jakauma olisi aidosti p-ulotteinen.

Muuttujien X odotusarvovektori on

$$E(\mathbf{X}) = \mathbf{C}E(\mathbf{U}) + \mu = \mathbf{C}\cdot\mathbf{0} + \mu = \mu$$

ja kovarianssimatriisi

$$\Sigma = \mathrm{cov}(\mathbf{X}) = E(\mathbf{X}-\mu)(\mathbf{X}-\mu)' = E(\mathbf{CUU}'\mathbf{C}') = \mathbf{C}(E\mathbf{UU}')\mathbf{C}' = \mathbf{CC}' .$$

Koska $r(\mathbf{C})=p$, on $\Sigma=\mathbf{CC}'>\mathbf{0}$ (eli positiivisesti definiitti).

Määrätään nyt X-muuttujien yhteisjakauman tiheysfunktio.

Todistetaan ensin apulause:
Olkoot $U_1,U_2,\dots,U_k$ riippumattomia ja $N(0,1)$. Tällöin myös muuttujat $\mathbf{V}=(V_1,V_2,\dots,V_k)=\mathbf{QU}$, ovat riippumattomia ja $N(0,1)$, jos matriisi $\mathbf{Q}$ on ortogonaalinen (eli $\mathbf{Q}'\mathbf{Q}=\mathbf{QQ}'=\mathbf{I}$).

Koska kääntäen $\mathbf{U}=\mathbf{Q}'\mathbf{V}$ ja

$$f_{\mathbf{U}}(\mathbf{u}) = c \exp(-1/2\mathbf{u}'\mathbf{u}),$$

tulee muuttujien V tiheysfunktioksi (sijoittamalla tähän tiheysfunktioon $\mathbf{u}=\mathbf{Q}'\mathbf{v}$ ja kertomalla vastaavalla funktionaalideterminantilla, joka kuvausmatriisin $\mathbf{Q}'$ ortogonaalisuudesta johtuen on 1)

$$f_{\mathbf{V}}(\mathbf{v}) = c \exp[-\tfrac{1}{2}(\mathbf{Q'v})'(\mathbf{Q'v})]$$
$$= c \exp(-\tfrac{1}{2}\mathbf{v'v}) = f_{\mathbf{U}}(\mathbf{v}) .$$

Osoitetaan nyt, että jos (1) pätee, on olemassa satunnaisvektori $\mathbf{V}=(V_1,V_2,...,V_p)$, jonka komponentit ovat riippumattomia ja (0,1)-normaalisia siten, että **X** voidaan lausua myös niiden avulla muodossa

(2) $\mathbf{X} = \mathbf{AV} + \mu$,

missä **A** on $p\times p$-matriisi ja $\det(\mathbf{A})\neq 0$.

Tämä todistetaan lähtemällä $p\times k$-matriisin **C** singulaariarvohajotelmasta $\mathbf{C} = \mathbf{SDT'}$, missä **D** on singulaariarvojen $d_1 \geq d_2 \geq ... \geq d_p > 0$ ($r(\mathbf{C})=p$) muodostama lävistäjämatriisi ja **S** $p\times p$-ortogonaalinen sekä **T** $k\times p$-pystyriveittäin ortogonaalinen eli $\mathbf{T'T}=\mathbf{I}$. Valitsemalla nyt $\mathbf{V}=\mathbf{T'U}$ ja $\mathbf{A}=\mathbf{SD}$ saadaan haluttu esitys $\mathbf{X} = \mathbf{CU} + \mu = \mathbf{SDT'U} + \mu = \mathbf{AV} + \mu$. Tässä muuttujat V ovat apulauseen perusteella riippumattomia ja $N(0,1)$, sillä matriisi **T** on aina täydennettävissä $k\times k$-ortogonaaliseksi matriisiksi.

Huomattakoon lisäksi, että $\Sigma = \mathrm{cov}(\mathbf{X}) = \mathbf{AA'} = \mathbf{CC'}$ ja $\det(\Sigma)=\det(\mathbf{A})^2$.

Koska matriisi **A** on säännöllinen, saadaan kääntäen $\mathbf{V} = \mathbf{A}^{-1}(\mathbf{X}-\mu)$ ja

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{V}}(\mathbf{v}(\mathbf{x}))\cdot|\partial(v_1,v_2,...,v_p)/\partial(x_1,x_2,...,x_p)|$$
$$= (2\pi)^{-p/2}\exp(-\tfrac{1}{2}\mathbf{v'v})\det(\mathbf{A}^{-1})$$
$$= (2\pi)^{-p/2}\det(\Sigma)^{-1/2}\exp[-\tfrac{1}{2}(\mathbf{x}-\mu)'(\mathbf{A}^{-1})'\mathbf{A}^{-1}(\mathbf{x}-\mu)]$$
$$= (2\pi)^{-p/2}\det(\Sigma)^{-1/2}\exp[-\tfrac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)] .$$

Siis
$E(\mathbf{X})=\mu$ ja $\mathrm{cov}(\mathbf{X})=\Sigma > \mathbf{0}$
määräävät X-muuttujien yhteisjakauman yksikäsitteisesti.

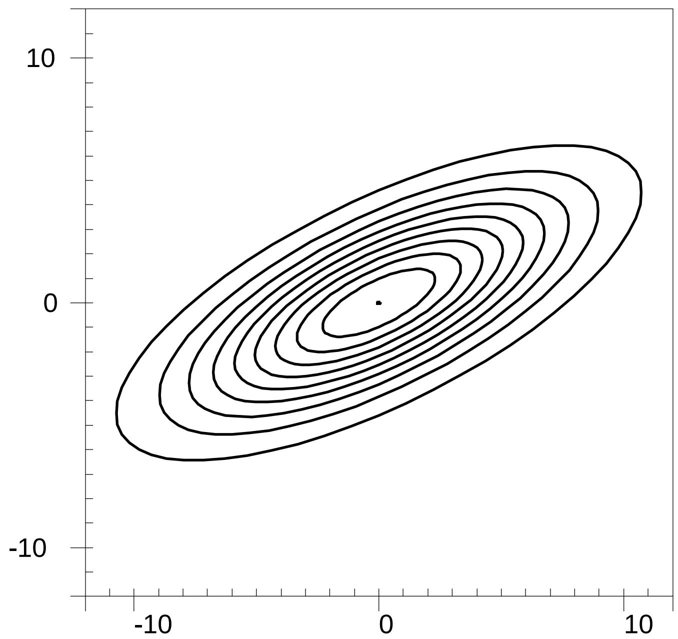
Sanomme, että **X** noudattaa p-ulotteista normaalijakaumaa l. multinormaalijakaumaa $N(\mu,\Sigma)$.
Merkitään $\mathbf{X} \sim N(\mu,\Sigma)$, $f_{\mathbf{X}}(\mathbf{x})=n(\mathbf{x} \mid \mu,\Sigma)$, jolloin esim. $\mathbf{V} \sim N(\mathbf{0},\mathbf{I}_p)$.

Kovarianssimatriisin lävistäjällä ovat muuttujien varianssit $\sigma_{11},\sigma_{22},...,\sigma_{pp}$. Näille käytetään myös merkintöjä
$\sigma_{ii}=\sigma_i^2$, $i=1,2,...,p$,
eli $\sigma_1,\sigma_2,...,\sigma_p$ tarkoittavat muuttujien keskihajontoja. Keskihajontojen muodostamaa lävistäjämatriisia merkitään
$\mathbf{D}_\sigma = \mathrm{diag}(\sigma_1,\sigma_2,...,\sigma_p)$,
jolloin muuttujien **X** korrelaatiomatriisi **P**, iso kreikkalainen ρ (rho), saadaan

kaavasta

$$\mathbf{P} = \mathbf{D}_\sigma^{-1}\Sigma\mathbf{D}_\sigma^{-1} .$$

Multinormaalisen satunnaisvektorin **X** tiheysfunktiota hallitsee positiivisesti definiitti neliömuoto $(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)$. Tiheysfunktio on suurimmillaan, kun $\mathbf{x}=\mu$ ja sen arvot vähenevät tästä pisteestä etäännyttäessä siten, että (hyper)-ellipsit eli hajontaellipsit $(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)$ = vakio toimivat tasa-arvokäyrinä.



Kuvassa on sellaisen 2-ulotteisen normaalijakauman tasa-arvokäyriä, jossa muuttujien hajonnat ovat 5 ja 3 sekä korrelaatiokerroin 0.7. Käyrät vastaavat todennäköisyystasoja 0.1,0.2,...,0.9 eli todennäköisyysmassasta 90% on uloimman hajontaellipsin sisällä.

Edelläkäyty tarkastelu osoittaa, että *p*-ulotteinen satunnaisvektori **X** voidaan aina määritellä *p* riippumattoman (0,1)-normaalisen muuttujan avulla.

Annetulla multinormaalisella **X**-vektorilla parametrit μ ja Σ ovat yksikäsitteiset, mutta **V** ja **A** voidaan ajatella valittavaksi useilla tavoilla. Olettaessamme, että satunnaisvektori **X** noudattaa multinormaalijakaumaa $N(\mu,\Sigma)$, emme siis voi tuntea tästä jakaumasta saatujen havaintojen täsmällistä syntytapaa, mutta kaikissa jakauman ominaisuuksia koskevissa tarkasteluissa on lupa käyttää konstruktiota (2), kun vain **A** täyttää ehdon $\Sigma=\mathbf{AA}'$.

Kun siis $\mathbf{X} \sim N(\mu,\Sigma)$, **A**-kuvaus voidaan saada esim. matriisin Σ Cholesky-hajotelmasta $\Sigma=\mathbf{AA}'$, missä **A** on yläkolmiomatriisi tai spektraalihajotelmasta $\Sigma=\mathbf{S}\Lambda\mathbf{S}'$, missä **S** on ortogonaalinen ja Λ ominaisarvojen muodostama lävistäjämatriisi, jolloin $\mathbf{A}=\mathbf{S}\Lambda^{1/2}$.

Edellä on oletettu matriisi **A** täysiasteiseksi, jolloin sillä ja kovarianssimatriisilla Σ on käänteismatriisi. Tällöin jakauma on aidosti p-ulotteinen ja sille voidaan kirjoittaa edellä todettu tiheysfunktion lauseke.

Voimme jo johdannossa mainitulla tavalla vielä yksinkertaistaa määritelmää (2) matriisin **A** singulaariarvohajotelman $\mathbf{A}=\mathbf{SDT}'$ avulla. Tällöin

$$\mathbf{X} = \mathbf{AV} + \mu = \mathbf{SDT'V} + \mu = \mathbf{SDW} + \mu$$

eli

(3) $\mathbf{X} = \mathbf{SDW} + \mu$,

missä $\mathbf{W} \sim N(\mathbf{0},\mathbf{I})$ edellä olevan apulauseen nojalla, **D** on positiivisten singulaariarvojen $d_1 \geq d_2 \geq ... \geq d_p > 0$ muodostama lävistäjämatriisi ja **S** $p{\times}p$-ortogonaalinen matriisi.

Kuten myöhemmin tulemme näkemään, muuttujat $\mathbf{DW}=(d_1W_1,...,d_pW_p)$ ovat muuttujien **X** pääkomponentteja, joiden voimakkuuksia (itse asiassa keskihajontoja ja geometrisesti hajontaellipsoidien pääakseleiden pituuksia) vastaavat singulaariarvot.

Esittämämme konstruktiivinen määritelmä antaisi mahdollisuuden käsitellä vaivatta myös vajaa-asteisia tapauksia, joissa osa singulaariarvoista on nollia, mutta jatkossa tarkastelemme lähes poikkeuksetta vain täysiulotteista multinormaalijakaumaa.

Tutkiessamme multinormaalijakauman ominaisuuksia käytämme usein apuna konstruktiivisia määritelmiä (1), (2) ja (3), jotka yleensä tekevät tarkastelut yksinkertaisemmiksi kuin jos perustaisimme ne multinormaalijakauman tiheysfunktion esitykseen. Useimmat oppikirjat lähtevät liikkeelle suoraan esim. tiheysfunktiosta tai karakteristisesta funktiosta, jolloin helposti kadotetaan jakauman luonnollinen tausta.

2.2.1 Reunajakaumat

Tulemme useasti tarkastelemaan p komponentin satunnaisvektoria **X** kahden osavektorin $\mathbf{X}^{(1)}$ ja $\mathbf{X}^{(2)}$ yhdistelmänä siten, että $\mathbf{X}^{(1)}$ käsittää q ($q<p$) ensimmäistä muuttujaa $\mathbf{X}^{(1)}=(X_1,X_2,...,X_q)$ ja $\mathbf{X}^{(2)}$ loput p-q muuttujaa $\mathbf{X}^{(2)}=(X_{q+1},X_{q+2},...,X_p)$. Mikä tahansa muuttujien osajoukko saadaan näiden tarkastelujen piiriin järjestämällä muuttujavektorin **X** komponentit sopivasti uudelleen.

Ositettujen matriisien merkintätapoja noudattaen on siis

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix},$$

jolloin odotusarvovektorin μ ja kovarianssimatriisin Σ ositetut esitykset ovat

$$\mu = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Osoitamme nyt, että muuttujavektorin $\mathbf{X}^{(1)}$ jakauma on $N(\mu^{(1)},\Sigma_{11})$. Tämä tapahtuu määritelmän (2) avulla eli kirjoittamalla $\mathbf{X} = \mathbf{AV} + \mu$ ositetussa muodossa

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \mathbf{V} + \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix},$$

jolloin

$$\mathbf{X}^{(1)} = \mathbf{A}_1\mathbf{V} + \mu^{(1)} .$$

Tällöin määritelmän (1) mukaan

$$\mathbf{X}^{(1)} \sim N(\mu^{(1)}, \mathbf{A}_1\mathbf{A}_1') = N(\mu^{(1)},\Sigma_{11}) .$$

2.2.2 Muuttujien vaihto

Konstruktiivisen määritelmän mukaan on mitä ilmeisintä, että multinormaalisuus säilyy muuttujien lineaarisissa kuvauksissa. Näytämme täsmällisemmin, että jos $\mathbf{X} \sim N(\mu,\Sigma)$ ja $\mathbf{Y}=\mathbf{BX}$, missä $\mathbf{B}$ on täysiasteinen $m\times p$-matriisi ($r(\mathbf{B})=m$, $m\leq p$), niin $\mathbf{Y} \sim N(\mathbf{B}\mu,\mathbf{B}\Sigma\mathbf{B}')$.

Tämän todistamiseksi käytämme määritelmää (2) eli $\mathbf{X}=\mathbf{AV}+\mu$, jolloin

$$\mathbf{Y} = \mathbf{BX} = \mathbf{BAV} + \mathbf{B}\mu$$

eli $\mathbf{Y}$ syntyy määritelmän (1) mukaan (0,1)-normaalisista V-muuttujista käyttäen kuvausmatriisia $\mathbf{BA}$ ja lisäystä $\mathbf{B}\mu$. Siis $\mathbf{Y} \sim N(\mathbf{B}\mu,\mathbf{BAA}'\mathbf{B}')$ eli $\mathbf{Y} \sim N(\mathbf{B}\mu,\mathbf{B}\Sigma\mathbf{B}')$, sillä $\mathbf{AA}'=\Sigma$.

Erityisesti havaitaan, että jokainen $\mathbf{X}$-muuttujien lineaarinen kombinaatio noudattaa tavallista yksiulotteista normaalijakaumaa seuraavasti. Olkoon $\alpha=(\alpha_1,\alpha_2,...,\alpha_p)$ p komponentin pystyvektori. Tällöin

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + ... + \alpha_p X_p = \alpha'\mathbf{X} \sim N(\alpha'\mu, \alpha'\Sigma\alpha) .$$