Introduction to mathematical statistics ゼミ (第？回)

担当: 伊藤真道

未定

3 Chap.3 Some Special Distributions

3.5 The Multivariate Normal Distribution

(D くんがやってくれたところの続きの残ってる部分だけです．まずは軽い復習から)

$Z_1, \ldots, Z_n \overset{iid}{\sim} N(0,1)$ である確率変数からなる確率変数ベクトル $\boldsymbol{Z} := (Z_1, \ldots, Z_n)'$ の密度関数は，$\boldsymbol{z} \in \mathbb{R}^n$ について

$$f_{\boldsymbol{Z}}(\boldsymbol{z}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}z_i^2\right\} = \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} z_i^2\right\}$$
$$= \left(\frac{1}{2\pi}\right)^{n/2} \exp\left\{-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\right\} \tag{3.5.1}$$

と表されるのだった．$Z_i \sim N(0,1)$ であり，これらは，無作為標本 (=iid) であることから，$\boldsymbol{Z}$ の平均と分散は，

$$E[\boldsymbol{Z}] = \boldsymbol{0}_n,\ V[\boldsymbol{Z}] = I_n \tag{3.5.2}$$

である．また，$N(0,1)$ の mgf は $E[e^{t_i Z_i}] = \exp(0 \cdot t_i + (1 \cdot t_i)^2/2) = \exp(t_i^2/2)$ であったことを思い出すと，Z_i が iid であることから，$\boldsymbol{Z}$ の mgf は，

$$M_{\boldsymbol{Z}}(\boldsymbol{t}) = E[\exp(\boldsymbol{t}'\boldsymbol{Z})] = \exp\left\{\frac{1}{2}\boldsymbol{t}'\boldsymbol{t}\right\} \tag{3.5.3}$$

となる．特に二つ目の等号は間を飛ばしているので，気になる方は，私に聞くか，自分で調べてください．また，ここで，$\boldsymbol{t} \in \mathbb{R}^n$ である．以上のような性質を持つ確率変数ベクトル $\boldsymbol{Z}$ は n 次元多変量標準正規分布 $N(\boldsymbol{0}_n, I_n)$ に従うといい，$\boldsymbol{Z} \sim N(\boldsymbol{0}_n, I_n)$ と表す．

1 変数の場合と同様に，多変数の場合にも標準でない正規分布も存在する．それを紹介するために，$n \times n$ 半正定値対称行列 (=グラム行列, Gramian)Σ を用意する．線形代数の初歩的な知見か

ら，Σ は常に，

$$\Sigma = \Gamma'\Lambda\Gamma \tag{3.5.4}$$

のように，固有値を対角成分として持つ対角行列 $\Lambda = diag(\lambda_1, \dots, \lambda_n)$ と，それに対応する固有ベクトルを列として持つ直交行列 Γ' を用いて固有値分解を行うことができることがわかっている．但し，$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ である．Σ の半正定値性から，λ_i は非負であり，$\Lambda^{1/2} := diag(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$ を用いて，

$$\Sigma = \Gamma'\Lambda^{1/2}\Gamma\Gamma'\Lambda^{1/2}\Gamma = \Sigma^{1/2}\Sigma^{1/2}$$

のように表すことができる．ここで，グラム行列の平方根を

$$\Sigma^{1/2} := \Gamma'\Lambda^{1/2}\Gamma$$

と定義した．さらに，Σ が正定値行列 (=全ての固有値 > 0) と仮定すると，

$$\left(\Sigma^{1/2}\right)^{-1} = \Gamma'\Lambda^{-1/2}\Gamma$$

と逆行列の計算が可能である．詳しくは，伊藤レクチャー第一回を参照.

　さて，一般の多変量正規分布についての話に戻ろう．先ほどと同様の半正定値対称行列 Σ と，$n \times 1$ の定数ベクトル $\boldsymbol{\mu}$ を用いて，$\boldsymbol{Z} \sim N(\mathbf{0}_n, I_n)$ を変換することを考える．

$$\boldsymbol{X} := \Sigma^{1/2}\boldsymbol{Z} + \boldsymbol{\mu} \tag{3.5.8}$$

と定義された $\boldsymbol{Z}$ の線形変換 $\boldsymbol{X}$ の平均，分散，mgf は，

$$E[\boldsymbol{X}] = \boldsymbol{\mu},\ V[\boldsymbol{X}] = \Sigma \tag{3.5.9}$$

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \exp(\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}) \tag{3.5.10}$$

となる．証明は略．気になったら私に聞いてください．

もし，ここで，<u>Σ が正定値行列ならば</u>，$\Sigma^{-1/2}$ が存在し，$\boldsymbol{X}, \boldsymbol{Z}$ は一対一に対応し，$\boldsymbol{X}$ を

$$\boldsymbol{Z} = \Sigma^{-1/2}(\boldsymbol{X} - \mu),\ det(\Sigma^{-1/2}) = det(\Sigma)^{-1/2}$$

と変換することで，$\boldsymbol{Z}$ を求めることができる．以上から，多変量正規分布の密度関数は，

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}det(\Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\},\ for\ \boldsymbol{x} \in \mathbb{R}^n \tag{3.5.12}$$

と表される．以上を一旦まとめる．

多変量正規分布の密度関数，平均，分散，mgf

多変量正規分布 $N(\boldsymbol{\mu}, \Sigma)$ の密度関数は，

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2} det(\Sigma)^{1/2}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \right\}, \ for \ \boldsymbol{x} \in \mathbb{R}^n \tag{3.5.12}$$

であり，平均，分散は，

$$\begin{aligned} E[\boldsymbol{X}] &= \boldsymbol{\mu}, \\ V[\boldsymbol{X}] &= \Sigma \end{aligned}$$

であり，mgf は

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \exp(\boldsymbol{t}'\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t})$$

である．

¡note¿

以下頻繁に利用される定理を紹介する．証明は，前回の D くん担当のゼミでやっているため，省略するが，気になる方は私に聞くか，テキストを読んでください．

Theorem 3.5.1

$\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$ とし，$\boldsymbol{Y} := \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ という変換を考える．この時，$\boldsymbol{Y} \sim N_m(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')$ である．

証明は，mgf を使う．

¡note¿

Corollary 3.5.1

Theorem 3.5.1 で $A = [I_m : O_{mp}]$,

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix},\ \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix},\ \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

であったとする．この時，$\boldsymbol{X}_1 = A\boldsymbol{X}$ であり，$\boldsymbol{X}_1 \sim N_m(\boldsymbol{\mu}_1, \Sigma_{11})$ である．

証明は，Theorem3.5.1 の結果を利用する．

¡note¿

Theorem 3.5.2

(上記のセッティングで，)$\boldsymbol{X}_1$ と $\boldsymbol{X}_2$ が独立であるとは，$\Sigma_{12} = O$ であることを言う．

証明は，$\Sigma_{12}' = \Sigma_{21}$ であることに注意して，$\boldsymbol{X}_1, \boldsymbol{X}_2$ の jmgf をみる．

¡note¿

Theorem 3.5.3

(上記のセッティングで，)Σ が正定値行列であることを仮定する．この時，$\boldsymbol{X}_1|\boldsymbol{X}_2$ の条件付き分布は，

$$N_m(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \tag{3.5.12}$$

となる．

以降面倒なので，ベクトルの太字はやめます．本当に面倒なので．その時の状況で判断してください．わからなかったら，書き方が悪い！と私を詰ってください．

Proof 3.1. まず，確率変数ベクトル $W = X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2$ と X_2 について考える．これらの確率変数ベクトルの分布は，$X = [X_1, X_2]'$ の線形変換，

$$\begin{bmatrix} W \\ X_2 \end{bmatrix} = \begin{bmatrix} I_m & -\Sigma_{12}\Sigma_{22}^{-1} \\ O_m & I_p \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

によって得られる．つまり，*Theorem 3.5.1(多変量正規分布に従う確率変数ベクトルの線形変換)*

が適用できて，$E[W]=\mu_1-\Sigma_{12}\Sigma_{22}^{-1}\mu_2, E[X_2]=\mu_2$ であり，共分散行列は，

$$
\begin{aligned}
V[(W,X_2)'] &= \begin{bmatrix} I_m & -\Sigma_{12}\Sigma_{22}^{-1} \\ O_m & I_p \end{bmatrix} V[X] \begin{bmatrix} I_m & -\Sigma_{12}\Sigma_{22}^{-1} \\ O_m & I_p \end{bmatrix}' \\
&= \begin{bmatrix} I_m & -\Sigma_{12}\Sigma_{22}^{-1} \\ O_m & I_p \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I_m & -\Sigma_{12}\Sigma_{22}^{-1} \\ O_m & I_p \end{bmatrix}' \\
&= \begin{bmatrix} \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & O' \\ O & \Sigma_{22} \end{bmatrix}
\end{aligned}
$$

ここで，$\Sigma_{12}=O$ から *Theorem 3.5.2* が適用できて，W, X_2 は独立であることがわかる．よって，$W|X_2$ の分布と，W の周辺分布は等しい．つまり，

$$
W|X_2 \sim N_m(\mu_1-\Sigma_{12}\Sigma_{22}^{-1}\mu_2, \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})
$$

さらに，$X_1|X_2 = W+\Sigma_{12}\Sigma_{22}^{-1}X_2|X_2 = W+\Sigma_{12}\Sigma_{22}^{-1}X_2$ から，再び，*Theorem3.5.1* を適用して，

$$
\begin{aligned}
X_1|X_2 &\sim N_m(\mu_1-\Sigma_{12}\Sigma_{22}^{-1}\mu_2+\Sigma_{12}\Sigma_{22}^{-1}X_2, \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \\
&= N_m(\mu_1+\Sigma_{12}\Sigma_{22}^{-1}(X_2-\mu_2), \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \quad \square
\end{aligned}
$$

(最初の W と X_2 の導入とか天才すぎて自力で思いつかんやろ...)

¡note¿

Theorem 3.5.4

(Theorem 3.5.3 と同様にセッティングで．) 確率変数 W を $W := (X-\mu)'\Sigma^{-1}(X-\mu)$ と定義すると，

$$
W \sim \chi_n^2
$$

である．

Proof 3.2. Σ の正定値性から，$\Sigma=\Sigma^{1/2}\Sigma^{1/2}$ とかける．これを用いて，$Z=\Sigma^{-1/2}(X-\mu)$ に対して，*Theorem 3.5.1* を適用すると，$Z\sim N_n(0_n, I_n)$ である．また，$W=(X-\mu)'\Sigma^{-1}(X-\mu)=(X-\mu)'\Sigma^{-1/2}\Sigma^{-1/2}(X-\mu)=Z'Z=\sum_i^n Z_i^2$ とかけ，Z_i はそれぞれ独立であることに注意すると，定理の結果を得る．

¡note¿

3.5.1 Applications

$X \sim N_n(\mu, \Sigma)$ で，Σ は正定値行列とする．この時，以前紹介したように，$\Sigma = \Gamma'\Lambda\Gamma$ と固有値分解可能である．$Y := \Gamma(X-\mu)$ と定義し，$\Gamma\Sigma\Gamma' = \Lambda$ であることに注意すると，$Y \sim N_n(0_n, \Lambda)$ である．Λ が対角行列であることから，$Y_1, \ldots, Y_n$ はそれぞれ独立であり，$N(0, \lambda_i)$ という分布に従う．このように定義された確率変数ベクトル Y を<u>**主成分** *(principal components)*</u>と呼ぶ．

また，確率変数ベクトルの<u>**総変動** *(Total Variation)*</u>とは，その成分の分散の総和である．X の総変動は，

$$TV(X) = \sum_{i=1}^{n} \sigma_i^2 = tr\Sigma = tr\Gamma'\Lambda\Gamma = tr\Lambda\Gamma\Gamma' = \sum_{i=1}^{n} \lambda_i = TV(Y)$$

と変形でき，主成分 Y と同一の総変動を持つことがわかる．

次に，Y の最初の成分 $Y_1 = v_1'(X-\mu)$ について考えよう．これは，$(X-\mu)$ の成分の $\|v_1\| = 1$ なる v_1 による線型結合である．さらに，別の $(X-\mu)$ の線型結合を考える．仮にこれを $a'(X-\mu)$ とし，こちらも $\|a\| = 1$ であるとする．$\{v_1, v_2 \ldots, v_n\}$ は $\mathbb{R}^n$ の基底であり，$a \in \mathbb{R}^n$ であることから，何らかのスカラー $(c_1, \ldots, c_n)$ を用いて，$a = \sum_{i=1}^{n} c_i v_i$ と書けるはずである．さらに，$\{v_1, v_2 \ldots, v_n\}$ は直交しているため，

$$a'v_i = \left(\sum_{j=1}^{n} c_j v_j\right)' v_i = c_i v_i' v_i = c_i$$

となる．線形変換 $a'X$ の分散は，

$$\begin{aligned} Var(a'X) &= a'\Sigma a \\ &= \sum_{i=1}^{n} \lambda_i (a'v_i)^2 \\ &= \sum_{i=1}^{n} \lambda_i a_i^2 \leq \lambda_1 \sum_{i=1}^{n} a_i^2 = \lambda \|a\|^2 = \lambda_1 = Var(Y_1) \end{aligned} \tag{3.5.24}$$

よって，Y_1 は直交ベクトル a による線形変換の中で最大の分散を持つ．そのため，Y_1 は X の First principal component と呼ばれる．

Theorem 3.5.5

(上につらつら書いたセッティングで)$j = 2, \ldots, n,\ i = 1, 2, \ldots, j-1$ に対して，$a \perp v_i$ なる直交ベクトルに関して，$Var[a'X] \leq \lambda_j = Var[Y_j]$ となる．

¡note¿