ゼミ 12月17日分

平山 奈々海

December 17, 2024

教科書の説明：黒色

補足：青色

わからなかったところ、気になったところ：赤色

1 最小２乗推定量の性質

さて、最小 2 乗推定量 $\hat{\boldsymbol{\beta}}$ の平均と共分散行列を求めよう。（命題 12.2）これは確率変数のベクトルについての平均と共分散行列を扱うことになるのでその定義から始める。$\boldsymbol{Z} = (Z_1, \ldots, Z_n)^\top$ を n 個の確率変数を縦に並べたベクトルとし、これを確率ベクトルと呼ぶ。

その平均は $\boldsymbol{\mu} = E[\boldsymbol{Z}]$、共分散行列は $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{Z}) = E[(\boldsymbol{Z} - \boldsymbol{\mu})(\boldsymbol{Z} - \boldsymbol{\mu})^\top]$ で定義される。平均（$\boldsymbol{\mu} = E[\boldsymbol{Z}]$）を成分で表示すると

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} E[Z_1] \\ \vdots \\ E[Z_n] \end{pmatrix}$$

共分散行列は $\sigma_{ij} = E[(Z_i - \mu_i)(Z_j - \mu_j)]$ とおくと、

$$\boldsymbol{\Sigma} = (\sigma_{ij}) = \begin{pmatrix} E[(Z_1 - \mu_1)^2] & \cdots & E[(Z_1 - \mu_1)(Z_n - \mu_n)] \\ \vdots & \ddots & \vdots \\ E[(Z_n - \mu_n)(Z_1 - \mu_1)] & \cdots & E[(Z_n - \mu_n)^2] \end{pmatrix}$$

と書けることになる。次の補題は期待値や共分散行列の計算に役立つ。

補題 12.1 $\boldsymbol{Z}$ を n 次元の確率ベクトルで、その平均と共分散行列を $E[\boldsymbol{Z}]=\boldsymbol{\mu}$, $\mathrm{Cov}(\boldsymbol{Z})=\boldsymbol{\Sigma}$ とする。$\boldsymbol{A}$ を $r\times n$ 行列, $\boldsymbol{b}$ を r 次元ベクトル, $\boldsymbol{C}$ を $n\times n$ 行列とすると、次の性質が成り立つ。

1. $\boldsymbol{AZ}+\boldsymbol{b}$ の平均は $E[\boldsymbol{AZ}+\boldsymbol{b}]=\boldsymbol{A\mu}+\boldsymbol{b}$ である。
2. $\boldsymbol{AZ}+\boldsymbol{b}$ の共分散行列は $\mathrm{Cov}(\boldsymbol{AZ}+\boldsymbol{b})=\boldsymbol{A\Sigma A}^\top$ である。
3. $E[\boldsymbol{Z}^\top\boldsymbol{CZ}]=\mathrm{tr}(\boldsymbol{C\Sigma})+\boldsymbol{\mu}^\top\boldsymbol{C\mu}$ である。ただし $n\times n$ 行列 $\boldsymbol{D}=(d_{ij})$ に対して $\mathrm{tr}(\boldsymbol{D})=\sum_{i=1}^n d_{ii}$ で定義される。

証明

まず $\boldsymbol{AZ}+\boldsymbol{b}$ の成分表示は

$$\boldsymbol{AZ}+\boldsymbol{b}=\begin{pmatrix}a_{11}&\cdots&a_{1n}\\\vdots&\ddots&\vdots\\a_{r1}&\cdots&a_{rn}\end{pmatrix}\begin{pmatrix}Z_1\\\vdots\\Z_n\end{pmatrix}+\begin{pmatrix}b_1\\\vdots\\b_r\end{pmatrix}=\begin{pmatrix}a_{11}Z_1+\cdots+a_{1n}Z_n\\\vdots\\a_{r1}Z_1+\cdots+a_{rn}Z_n\end{pmatrix}+\begin{pmatrix}b_1\\\vdots\\b_r\end{pmatrix}$$

という r 次元ベクトルになる。

「ベクトルの期待値をとること」の定義とは、それぞれの成分に対して期待値をとったもののベクトルということですか？

この線形変換した確率ベクトル $\boldsymbol{AZ}+\boldsymbol{b}$ の平均は

$$\begin{aligned}E[\boldsymbol{AZ}+\boldsymbol{b}]&=\begin{pmatrix}E[a_{11}Z_1+\cdots+a_{1n}Z_n+b_1]\\\vdots\\E[a_{r1}Z_1+\cdots+a_{rn}Z_n+b_r]\end{pmatrix}\\&=\begin{pmatrix}a_{11}E[Z_1]+\cdots+a_{1n}E[Z_n]+b_1\\\vdots\\a_{r1}E[Z_1]+\cdots+a_{rn}E[Z_n]+b_r\end{pmatrix}\\&=\begin{pmatrix}a_{11}&\cdots&a_{1n}\\\vdots&\ddots&\vdots\\a_{r1}&\cdots&a_{rn}\end{pmatrix}\begin{pmatrix}E[Z_1]\\\vdots\\E[Z_n]\end{pmatrix}+\begin{pmatrix}b_1\\\vdots\\b_r\end{pmatrix}\\&=\boldsymbol{A}E[\boldsymbol{Z}]+\boldsymbol{b}=\boldsymbol{A\mu}+\boldsymbol{b}\end{aligned}$$

となる。

(2) の共分散行列は、(1) より $E[\boldsymbol{AZ}+\boldsymbol{b}]=\boldsymbol{A\mu}+\boldsymbol{b}$ を用いて

$$\begin{aligned}\boldsymbol{Cov}(\boldsymbol{AZ}+\boldsymbol{b}) &= E\{(\boldsymbol{AZ}+\boldsymbol{b}-(\boldsymbol{A\mu}+\boldsymbol{b}))(\boldsymbol{AZ}+\boldsymbol{b}-(\boldsymbol{A\mu}+\boldsymbol{b}))^\top\} \\ &= E\{(\boldsymbol{A}(\boldsymbol{Z}-\boldsymbol{\mu}))(\boldsymbol{A}(\boldsymbol{Z}-\boldsymbol{\mu}))^\top\} \\ &= E[\boldsymbol{A}(\boldsymbol{Z}-\boldsymbol{\mu})(\boldsymbol{Z}-\boldsymbol{\mu})^\top\boldsymbol{A}^\top] \\ &= \boldsymbol{A}E[(\boldsymbol{Z}-\boldsymbol{\mu})(\boldsymbol{Z}-\boldsymbol{\mu})^\top]\boldsymbol{A}^\top \\ &= \boldsymbol{A\Sigma A}^\top\end{aligned}$$

となる。

(3) については、

$$\begin{aligned}E[\boldsymbol{Z}^\top\boldsymbol{CZ}] &= E[(\boldsymbol{Z}-\boldsymbol{\mu}+\boldsymbol{\mu})^\top\boldsymbol{C}(\boldsymbol{Z}-\boldsymbol{\mu}+\boldsymbol{\mu})] \\ &= E[(\boldsymbol{Z}-\boldsymbol{\mu})^\top\boldsymbol{C}(\boldsymbol{Z}-\boldsymbol{\mu})+(\boldsymbol{Z}-\boldsymbol{\mu})^\top\boldsymbol{C\mu}+\boldsymbol{\mu}^\top\boldsymbol{C}(\boldsymbol{Z}-\boldsymbol{\mu})+\boldsymbol{\mu}^\top\boldsymbol{C\mu}] \\ &= E[(\boldsymbol{Z}-\boldsymbol{\mu})^\top\boldsymbol{C}(\boldsymbol{Z}-\boldsymbol{\mu})]+\boldsymbol{\mu}^\top\boldsymbol{C\mu}\end{aligned}$$

さらにトレースの性質 $\mathrm{tr}(\boldsymbol{AC})=\mathrm{tr}(\boldsymbol{CA})$ を用いると、

$$\begin{aligned}E[(\boldsymbol{Z}-\boldsymbol{\mu})^\top\boldsymbol{C}(\boldsymbol{Z}-\boldsymbol{\mu})] &= E[\mathrm{tr}\{(\boldsymbol{Z}-\boldsymbol{\mu})^\top\boldsymbol{C}(\boldsymbol{Z}-\boldsymbol{\mu})\}] \quad \text{スカラーだから？} \\ &= E[\mathrm{tr}\{\underline{\boldsymbol{C}(\boldsymbol{Z}-\boldsymbol{\mu})}\ \underline{(\boldsymbol{Z}-\boldsymbol{\mu})^\top}\}] \quad \mathrm{tr}(\boldsymbol{AC})=\mathrm{tr}(\boldsymbol{CA})\ \text{より} \\ &= \mathrm{tr}[\boldsymbol{C}E\{(\boldsymbol{Z}-\boldsymbol{\mu})(\boldsymbol{Z}-\boldsymbol{\mu})^\top\}] \quad E[tr(A)]=tr(E[A])\ \text{やから？} \\ &= \mathrm{tr}(\boldsymbol{C\Sigma})\end{aligned}$$

であるから (3) が成り立つ。