2021.07.26(月４)

ZOOM

C.0. 母集団と確率分布（復習）

C.0.1. 確率分布（１次元の場合）

C.0.1.1. 確率分布の表し方

1) 離散分布

確率変数 X がとる値 x_i とその 確率関数 $P(x_i)$ により記述される。

$$P(X=x_i)=P(x_i),\quad \sum_i P(x_i)=1,\quad P(x_i)\geqq 0$$

2) 連続分布

ヒストグラムを、全体の面積が 1 であるように目盛り、その極限を「確率密度関数 $f(x)$」と言う。

$$P(a<X\leqq b)=\int_a^b f(x)dx,\quad \int_{-\infty}^{\infty}f(x)dx=1,\quad f(x)\geqq 0$$

C.0.1.2. 期待値と分散

母集団の中心やバラツキを表す指標を「母数」という。確率分布の中心を表す指標として「母平均 $E(X)$」がある。

$$E(X)=\sum_i x_iP(x_i)\text{（離散分布の場合）}$$

$$E(X)=\int_{-\infty}^{\infty}xf(x)dx\text{（連続分布の場合）}$$

確率分布のバラツキを表す指標として「母分散 $var(X)$」がある。

$$var(X)=E\{X-E(X)\}^2=E(X^2)-\{E(X)\}^2$$

$$D(X)=\sqrt{var(X)}$$

を「(母) 標準偏差」という。

C.0.2. ２次元分布と独立性

２つの確率変数 X,Y の組についての確率分布を 同時確率分布 という。

$P(x_i,y_j)$ と $f(x,y)$ で表現

$$P(X=x_i,\ Y=y_j)=P(x_i,y_j),\ P(a<X<b,\ c<Y<d)=\int_a^b\int_c^d f(x,y)dxdy$$

$$\sum_i\sum_j P(x_i,y_j)=1,\quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy=1$$

周辺確率関数、周辺密度関数

$$P_X(x_i)=\sum_j P(x_i,y_j),\quad P_Y(y_j)=\sum_i P(x_i,y_j)$$

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy,\quad f_Y(y)=\int_{-\infty}^{\infty}f(x,y)dx$$

2021.07.26(月４)

ZOOM

X と Y が 独立（X の分布が Y の取る値によらない。）

$$P(x_i, y_j) = P_X(x_i)P_Y(y_j), \quad f(x,y) = f_X(x)f_Y(y)$$

X と Y の 共分散（X と Y の関連性の尺度）

$$cov(X,Y) = E[\{X - E(X)\}\{Y - E(Y)\}] = E(XY) - E(X)E(Y)$$

母相関係数

$$\rho = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}$$

$\rho = 0$ のことを 無相関 という。独立であれば無相関である。（逆は成り立たない）

C.0.3. n 変数の連続分布と独立性

n 変数は、$f(x_1, \cdots, x_n)$ で表現

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dx_1 \cdots dx_n = 1$$

周辺密度関数

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dx_2 \cdots dx_n$$

$$\cdots$$

$$f_i(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n$$

$$\cdots$$

$$f_n(x_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dx_1 \cdots dx_{n-1}$$

定義：$(X_1, \cdots, X_n)$ が 互いに独立 とは、以下が成り立つこと。

$$f(x_1, \cdots, x_n) = f_1(x_1) \cdots f_n(x_n)$$

定義：$(X_1, \cdots, X_n)$ が 独立同一分布 **(i.i.d.)** とは、以下が成り立つこと。

$$f(x_1, \cdots, x_n) = f(x_1) \cdots f(x_n)$$

C.0.4. 確率変数の独立性について

岩波書店「数学辞典」第３版によると、任意のボレル集合（確率を定義する集合族の元）に対して、同時確率が各確率の積になることが「独立性」の定義である。

これを多次元の場合に拡張すれば、$W_1 \perp\!\!\!\perp W_2$ ならば $X_1 \perp\!\!\!\perp X_2$ が理解できると思われる。

統計推測特論（稲葉） No.C3

2021.07.26(月４)

ZOOM

C.1. 確率分布の性質

Def.C1. 確率変数 $(r.v.)$ を要素とする行列 $X = (X_{ij})$ に対し、$E(X_{ij})$ を要素とする行列を $E(X)$ で表わし、X の期待値という。

r. ベクトル $\boldsymbol{X} = (X_1, \cdots, X_p)'$ に対し、$E(\boldsymbol{X})$ を $\boldsymbol{X}$ の平均ベクトルと呼び、しばしば記号 $\boldsymbol{\mu}$ で表わす。また、

$$\Sigma := var(\boldsymbol{X}) := E(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'$$

を $\boldsymbol{X}$ の分散行列と呼び、しばしば Σ で表わす。

$$\Sigma \text{ の } (i,j) \text{ 要素} = cov(X_i, X_j) = E(X_i - \mu_i)(X_j - \mu_j)$$

Prop.C1. C：定数行列 $\Longrightarrow$ $E(C) = C$

また、E は linear である。即ち、

$$E(X+Y) = E(X) + E(Y)$$
$$E(AXB) = A\ E(X)\ B \qquad \text{ここに } A,\ B\text{：定数行列}$$

Prop.C2. $var(\boldsymbol{X})$ は非負

Def.C2. $cov(\boldsymbol{X}, \boldsymbol{Y}) := E(\boldsymbol{X} - E\boldsymbol{X})(\boldsymbol{Y} - E\boldsymbol{Y})'$

を r. ベクトル $\boldsymbol{X},\ \boldsymbol{Y}$ の共分散行列とよぶ。

特に、$cov(\boldsymbol{X}, \boldsymbol{X}) = var(\boldsymbol{X})$

Prop.C3. $cov(\boldsymbol{a} + B\boldsymbol{X}, \boldsymbol{c} + D\boldsymbol{Y}) = B\ cov(\boldsymbol{X}, \boldsymbol{Y})\ D'$

系 $var(\boldsymbol{a} + B\boldsymbol{X}) = B\ var(\boldsymbol{X})\ B'$

Prop.C4. $var(\boldsymbol{X}) = O \iff \boldsymbol{X} = E(\boldsymbol{X})$ w.p.1(with probabilty 1 の意味)

Prop.C5. (i) $E(\boldsymbol{X}\boldsymbol{X}') = var(\boldsymbol{X}) + (E\boldsymbol{X})(E\boldsymbol{X})'$

(ii) $E||\boldsymbol{X}||^2 = tr\{var(\boldsymbol{X})\} + ||E\boldsymbol{X}||^2$

(iii) $E(\boldsymbol{X} - \theta)(\boldsymbol{X} - \theta)' = var(\boldsymbol{X}) + (E\boldsymbol{X} - \theta)(E\boldsymbol{X} - \theta)'$

(iv) $E||\boldsymbol{X} - \theta||^2 = tr\{var(\boldsymbol{X})\} + ||E\boldsymbol{X} - \theta||^2$

[注意] $\boldsymbol{X}$ を未知母数 $\boldsymbol{\theta}$ の推定量と考えるとき、(iii), (iv) の右辺における第 1 項を分散部分、第２項を偏り部分という。

(iii) を平均２乗誤差 (Mean Square Error) 行列という。

(iv) を総平均２乗誤差 (Total Mean Square Error) という。

$\boldsymbol{X} - \boldsymbol{\theta}$：推定誤差、$E\boldsymbol{X} - \boldsymbol{\theta}$：偏り

Prop.C6. $\boldsymbol{X}: p \times 1$ ベクトル

$\boldsymbol{\mu} = E(\boldsymbol{X}), \quad \Sigma = var(\boldsymbol{X}) \quad \mathrm{rank}(\Sigma) = k$

$\Longrightarrow$

(i) $\exists \boldsymbol{Z}: k \times 1$ r. ベクトル, $\exists A: p \times k$ 定数行列

$$\text{s.t.} \quad \boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Z},\ AA' = \Sigma \qquad \cdots\cdots \tag{1}$$

(ii) (1) 式が成立すれば

$$\boldsymbol{Z} = (A'A)^{-1}A'(\boldsymbol{X} - \boldsymbol{\mu})$$

$$E(\boldsymbol{Z}) = \boldsymbol{0}, \qquad var(\boldsymbol{Z}) = I_k \qquad \cdots\cdots \qquad \text{規準化 (normalize)}$$

(iii) $(\boldsymbol{Z}, A),\ (\boldsymbol{Z}^*, A^*)$ がともに (1) をみたすならば

$$\exists T \in O(k), \quad s.t. \quad A^* = AT,\ \boldsymbol{Z}^* = T'\boldsymbol{Z}$$

（証明）(i) $\Sigma = V\Lambda V' = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & O \\ O & O \end{bmatrix} \begin{bmatrix} V_1' \\ V_2' \end{bmatrix}$ ：スペクトル分解

とする。

$$\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{bmatrix} = \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} (\boldsymbol{X} - \boldsymbol{\mu}) \qquad \cdots\cdots \tag{2}$$

とおくと、$\boldsymbol{\xi}_2 = \boldsymbol{0} \qquad (w.p.1)$

（なぜならば、$E(\boldsymbol{\xi}_2) = V_2'E(\boldsymbol{X} - \boldsymbol{\mu}) = \boldsymbol{0}, \quad var(\boldsymbol{\xi}_2) = V_2'\Sigma V_2 = O$ ）

(2) より、　$\boldsymbol{X} = \boldsymbol{\mu} + V\boldsymbol{\xi}, \qquad V\boldsymbol{\xi} = V_1\boldsymbol{\xi}_1 = V_1\Lambda_1^{\frac{1}{2}}\Lambda_1^{-\frac{1}{2}}\boldsymbol{\xi}_1$

ここで、$A = V_1\Lambda_1^{\frac{1}{2}},\ \ \boldsymbol{Z} = \Lambda_1^{-\frac{1}{2}}\boldsymbol{\xi}_1$ と置けば、$A\boldsymbol{Z} = V_1\boldsymbol{\xi}_1$ でかつ $AA' = \Sigma$ も成立する。

(ii) (1) 式より、$A\boldsymbol{Z} = \boldsymbol{X} - \boldsymbol{\mu}$ の両辺に左から A' を掛けると、$(A'A)\boldsymbol{Z} = A'(\boldsymbol{X} - \boldsymbol{\mu})$ が得られる。ここで、$A'A = (\Lambda_1^{\frac{1}{2}}V_1')(V_1\Lambda_1^{\frac{1}{2}}) = \Lambda_1$ は逆行列を持つので、結論を得る。

(iii) $A^*(A^*)' = \Sigma = AA'$

問題 A13 より、$\exists T \in O(k), \quad s.t. \quad A^* = AT$

$A^*\boldsymbol{Z}^* = \boldsymbol{X} - \boldsymbol{\mu} = A\boldsymbol{Z}.$　ここで、$A^* = AT$ ゆえ $AT\boldsymbol{Z}^* = A\boldsymbol{Z}.$

左から $(A'A)^{-1}A'$ をかけると、$T\boldsymbol{Z}^* = \boldsymbol{Z}, \qquad \boldsymbol{Z}^* = T'\boldsymbol{Z}. \qquad (Q.E.D.)$

（参考）問題 A13. $\left.\begin{array}{l} A, B: n \times m \\ A'A = B'B \end{array}\right\} \Longrightarrow \quad \exists P \in O(n), B = PA$

【略証】：$A = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_m)$ とおく。これらの中の極大一次独立な列ベクトルを、シュミットの直交化を行って選ぶ。B も、A で残したのと同じ列番号のベクトルが残る。

これらは、n 次元である。n 本に満たない時は、独立なベクトルを任意に追加する。

A, B は正則で $A'A = B'B$ となり、$B = PA$ を満たす $P \in O(n)$ が唯一存在する。

この P で、最初の行列 A, B に対しても、$B = PA$ が成り立つ。**【略証終】**