確率・統計 A

若木宏文

wakaki@math.sci.hiroshima-u.ac.jp
http://home.hiroshima-u.ac.jp/ wakaki/lecture/index.shtml

2015.6.16

目 次

3.3. 特性量

定義 3.3 (モーメント)

(1) 原点のまわりの k 次モーメント: $\mu_k = \mathsf{E}(X^k)$.

(2) 原点のまわりの k 次絶対モーメント: $\nu_k = \mathsf{E}(|X|^k)$.

(3) 平均値のまわりの k 次モーメント:
$\alpha_k = \mathsf{E}\{(X - \mathsf{E}(X))^k\} = \mathsf{E}\{(X - \mu)^k\}$.

定理 3.6

確率変数 X は有限な n 次モーメントをもつとする. このとき, X の n 次以下のモーメントも存在する. すなわち

$$\mathsf{E}(|X|^n) < \infty \Rightarrow \mathsf{E}(|X|^m) < \infty,\ m = 0, 1, \ldots, n.$$

(証明は板書で)

定義 3.4

(分散, 歪度, 尖度)

(1) 分散 $\sigma^2 = \mathsf{Var}(X) = \mathsf{E}\{(X-\mu)^2\} = \alpha_2.$

(2) 歪度 $\beta_1 = \mathsf{E}\left\{\left(\dfrac{X-\mu}{\sigma}\right)^3\right\} = \dfrac{\alpha_3}{\sigma^3}.$

(3) 尖度 $\beta_2 = \mathsf{E}\left\{\left(\dfrac{X-\mu}{\sigma}\right)^4\right\} = \dfrac{\alpha_4}{\sigma^4}.$

例 3.5 (2 項分布, 正規分布)

(1) $X \sim N(\mu, \sigma^2)$ の場合

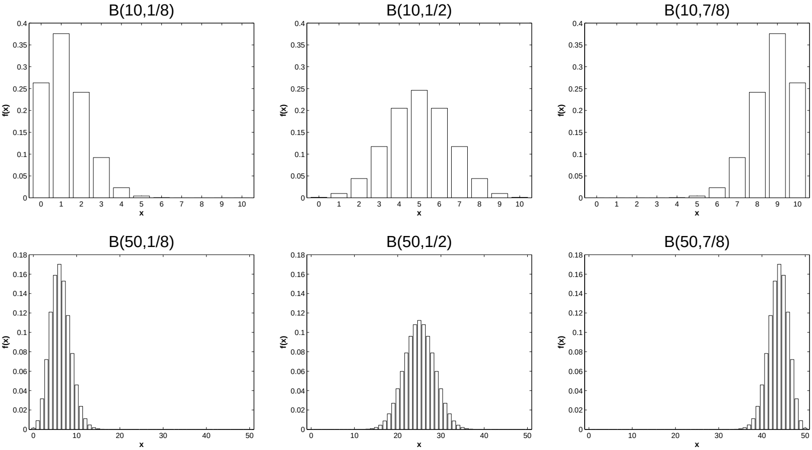
$$\mathsf{E}(X) = \mu, \quad \mathsf{Var}(X) = \sigma^2, \quad \beta_1 = 0, \quad \beta_2 = 3.$$

正規分布の尖度が 3 となるので, 尖度を $\beta_2 - 3$ として定義する場合がある.

(2) $X \sim B(n, p)$ の場合

$$\mathsf{E}(X) = np, \quad \mathsf{Var}(X) = npq, \ (q = 1 - p)$$
$$\beta_1 = \frac{p - q}{\sqrt{npq}}, \quad \beta_2 = \frac{1 - 6pq + 3npq}{npq}.$$

2 項分布の確率関数



定理 3.7 (分散の性質)

確率変数 X の分散に対して, 次が成り立つ. ただし, a は定数である.

(1) $\mathrm{Var}(X+a) = \mathrm{Var}(X)$.

(2) $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$.

(3) $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$.

(証明は板書で)

定理 3.8 (不等式)

(1) (Markov) X を非負値確率変数とする. 任意の正数 a に対して

$$\mathrm{P}(X \geq a) \leq \mathrm{E}(X)/a.$$

(2) (Chevyshev) 確率変数 X の平均, 分散を $\mathrm{E}(X) = \mu$, $\mathrm{Var}(X) = \sigma^2 < \infty$ とする. 任意の正数 k に対して

$$\mathrm{P}(|X - \mu| \geq k) \leq \sigma^2/k^2.$$

(証明は板書で)

共分散と相関係数

定義 3.5

確率変数 X, Y に対して

$$\mathsf{Cov}(X,Y) = \mathsf{E}[\{X - \mathsf{E}(X)\}\{Y - \mathsf{E}(Y)\}]$$

を X と Y との共分散とよぶ. また

$$\rho(X,Y) = \frac{\mathsf{Cov}(X,Y)}{\sqrt{\mathsf{Var}(X)}\sqrt{\mathsf{Var}(Y)}}$$

を X と Y との相関係数 とよぶ.

$\rho(X,Y) = 0$ であるとき無相関という.
$\rho(X,Y)$ が正であるとき, X と Y の間には正の相関があるといい, 負であるときは負の相関があるという.

分散・共分散の基本性質

定理 3.9

X, Y, Z を確率変数, a を定数とする. このとき, 次が成り立つ.

(1) $\mathsf{Cov}(X, X) = \mathsf{Var}(X)$.

(2) $\mathsf{Cov}(X, Y) = \mathsf{Cov}(Y, X)$.

(3) $\mathsf{Cov}(X + Y, Z) = \mathsf{Cov}(X, Z) + \mathsf{Cov}(Y, Z)$.

(4) $\mathsf{Cov}(aX, Y) = a\mathsf{Cov}(X, Y)$.

(5) $\mathsf{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y)$.

(6) X と Y が独立 $\Rightarrow \mathsf{Cov}(X, Y) = 0$.

(7) X と Y が有限な分散をもつとする. このとき

$$\mathsf{Var}(X + Y) = \mathsf{Var}(X) + \mathsf{Var}(Y) + 2\mathsf{Cov}(X, Y).$$

証明：分散の定義と期待値の基本性質から容易

シュワルツの不等式

定理 3.10

確率変数 X, Y は有限な分散をもつとする．このとき

$$\{\mathsf{Cov}(X,Y)\}^2 \leq \mathsf{Var}(X)\mathsf{Var}(Y)$$

が成り立つ．したがって

$$|\rho(X,Y)| \leq 1.$$

等号が成り立つのは, 次の (1), (2), (3) のいずれかの場合である:

(1) $\mathsf{P}(X = \mathsf{E}(X)) = 1.$

(2) $\mathsf{P}(Y = \mathsf{E}(Y)) = 1.$

(3) 定数 a, b が存在して, $\mathsf{P}(Y = a + bX) = 1.$

(証明は板書で)

定義 3.6 (平均ベクトル, 共分散行列)

p 次元確率ベクトル $\boldsymbol{X} = (X_1, \ldots, X_p)'$ について

$$\mathsf{E}(X_i) = \mu_i, \quad \mathsf{E}\{(X_i - \mu_i)(X_j - \mu_j)\} = \sigma_{ij}, \ i, j = 1, \ldots, p$$

とする. このとき

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & \ddots & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix}$$

をそれぞれ $\boldsymbol{X}$ の平均ベクトル, 共分散行列とよぶ.

共分散行列の性質

注 3.8

確率行列 $X = (X_{ij})$ の平均を $\mathsf{E}(X) = (\mathsf{E}(X_{ij}))$ と定義する．このとき

$$\mathsf{E}(\boldsymbol{X}) = \boldsymbol{\mu}, \quad \mathsf{Cov}(\boldsymbol{X}) = \Sigma = \mathsf{E}\{(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'\}$$

と表せる．また，任意の実ベクトル $\boldsymbol{a} = (a_1, \dots, a_p)'$ に対して

$$\sum_{i,j=1}^{p} \sigma_{ij} a_i a_j = \mathsf{Var}(a_1 X_1 + \cdots + a_p X_p) \geq 0$$

となる．したがって，共分散行列 Σ は半正定値行列である．

確率変数の線形近似

定理 3.11

2 次元確率ベクトル $\boldsymbol{X} = (X_1, X_2)'$ の平均ベクトル, 共分散行列を

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

とする. ただし, $\sigma_{11} > 0$ とする. このとき

$$\min_{\alpha,\beta} \mathsf{E}\{(X_2 - \alpha - \beta X_1)^2\} = \mathsf{E}\{(X_2 - \hat{\alpha} - \hat{\beta} X_1)^2\} = \sigma_2^2(1 - \rho^2)$$

となる. ここに

$$\hat{\alpha} = \mu_2 - \hat{\beta}\mu_1, \ \ \hat{\beta} = \rho\sigma_2/\sigma_1.$$

(証明は板書で)

相関係数の性質

注 3.9

確率変数 X_2 を X_1 の一次結合 $\alpha + \beta X_1$ で近似するときの誤差を E$\{(X_2 - \alpha - \beta X_1)^2\}$ で測るとき，これを最小とする最良近似

$$\mu_2 + \rho(\sigma_2/\sigma_1)(X_1 - \mu_1)$$

の誤差は $\sigma_2^2(1-\rho^2)$ である．$|\rho|$ が大きくなると，誤差は小さくなる．

注 3.7

$|\rho(X,Y)| \le 1$ (定理 3.10)

$|\rho(X,Y)|$ の大きさは： X と Y の線形関係の強さ (定理 3.11)

$|\rho(X,Y)| \approx 1 \;\Rightarrow\;$ 強い相関

$\rho(X,Y) \approx 1 \;\Rightarrow\;$ 強い正の相関

$\rho(X,Y) \approx -1 \;\Rightarrow\;$ 強い負の相関

3.4. 条件付き分布と平均

事象を与えたときの条件付き分布

X : $(\Omega, \mathcal{B}, \mathsf{P})$ 上の確率変数, $B \in \mathcal{B}$, $\mathsf{P}(B) > 0$
$A \in \mathbb{B}_1$
事象 B を与えたときの, 事象 "$X \in A$" の条件付き確率：

$$\mathsf{P}(X \in A \,|\, B) = \frac{\mathsf{P}(\text{"}X \in A\text{"} \cap B)}{\mathsf{P}(B)}$$

B を固定して, $\mathsf{P}_{X|B}(A) = \mathsf{P}(X \in A \,|\, B)$ を A の関数とみると,
$\mathsf{P}_{X|B}$ は $(\mathbb{R}^1, \mathbb{B}_1)$ 上の確率となる.

$\mathsf{P}_{X|B}$ を B を与えたときの X の条件付き分布とよぶ.

離散型分布の場合

(X, Y)：2 次元離散型確率変数,
$D = \{(x_i, y_j);\ i, j = 1, 2, \ldots\}$：取り得る値の集合
$f_{X,Y}(x, y)$：(X, Y) の確率関数, $f_Y(y)$：Y の周辺確率関数

事象 "$Y = y_j$" を与えたときの, 事象"$X \in A$" の条件付き確率は

$$\begin{aligned} \mathsf{P}_{X|Y=y_j}(A) &= \mathsf{P}(X \in A \,|\, Y = y_j) \\ &= \frac{\mathsf{P}(X \in A, Y = y_j)}{\mathsf{P}(Y = y_j)} = \sum_{x_i \in A} \frac{f_{X,Y}(x_i, y_j)}{f_Y(y_j)} \end{aligned}$$

$\mathsf{P}_{X|Y=y_j}$ は $(\mathbb{R}^1, \mathbb{B}_1)$ 上の離散型分布であり, その確率関数は

$$f_{X|Y}(x|y_j) \equiv \frac{f_{X,Y}(x_i, y_j)}{f_Y(y_j)}$$

によって与えられる.

$\mathsf{P}_{X|Y=y_j}$ を $Y = y_j$ を与えたときの X の条件付き分布とよぶ.
$f_{X|Y}(x|y_j)$ を $Y = y_j$ を与えたときの X の条件付き確率関数とよぶ.

条件付き期待値 (離散型の場合)

(X, Y) は離散型で, $f_{X|Y}(x_i|y_j)$ を $Y = y_j$ を与えたときの X の条件付き確率関数とする.

$$\sum_{i=1}^{\infty} |x_i| f_{X|Y}(x_i|y_j) < \infty$$

であるとき, $Y = y_j$ を与えたときの X の条件付き平均は存在するといい,

$$\mathsf{E}(X \mid Y = y_j) \equiv \sum_{i=1}^{\infty} x_i f_{X|Y}(x_i|y_j)$$

を $Y = y_j$ を与えたときの X の条件付き平均, あるいは条件付き期待値とよぶ.

例 3.10 (3 項分布)

$(X,Y) \sim M_3(n,(p,q))$ (3 項分布), $\quad p>0,\ q>0,\ 1-p-q>0.$

$$
\begin{aligned}
f_{X|Y}(i|j) &\equiv \mathsf{P}(X=i|Y=j)\\
&= \frac{n!}{i!j!(n-i-j)!}p^i q^j (1-q-p)^{n-j-i}\left\{\frac{n!}{j!(n-j)!}q^j(1-q)^{n-j}\right\}^{-1}\\
&= \frac{(n-j)!}{i!(n-j-i)!}\left(\frac{p}{1-q}\right)^i\left(1-\frac{p}{1-q}\right)^{n-j-i}.
\end{aligned}
$$

$Y=j$ を与えたときの X の条件付き分布は 2 項分布 $B(n-j,p/(1-q))$ であり

$$\mathsf{E}(X|Y=j)=(n-j)\frac{p}{1-q}.$$

連続型の場合

(X, Y) : 2 次元連続型確率変数
$f_{X,Y}(x, y)$: (X, Y) の確率密度関数, $f_Y(y)$: Y の周辺確率密度関数
$f_Y(y) > 0$ である y に対して

$$f_{X|Y}(x|y) \equiv \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

は, 確率密度関数の性質

(1) $f_{X,Y}(x|y) \geq 0$ $({}^\forall x \in \mathbb{R})$, (2) $\displaystyle\int_{-\infty}^{\infty} f_{X,Y}(x|y)dx = 1$

を満たす.
$f_{X|Y}(x|y)$ を $Y = y$ を与えたときの X の条件付き確率密度関数,

$f_{X|Y}(x|y)$ によって定められる連続型分布を
$f_{X|Y}(x|y_j)$ を $Y = y_j$ を与えたときの X の条件付き分布とよぶ.

注 3.8

$f_{X,Y}(x,y)$, $f_Y(y)$ が連続であれば

$$
\begin{aligned}
\lim_{h\to 0+} \mathsf{P}(X \le x | y < Y \le y+h) &= \lim_{h\to 0+} \frac{\int_{-\infty}^{x} \left\{ \int_y^{y+h} f_{X,Y}(u,v)dv \right\} du}{\int_y^{y+h} f_Y(v)dv} \\
&= \lim_{h\to 0+} \frac{\int_{-\infty}^{x} f_{X,Y}(u, y+\theta_1 h) h du}{f_Y(y+\theta_2 h)h} = \int_{-\infty}^{x} f_{X|Y}(u|y)du
\end{aligned}
$$

が成り立つ. ただし, $0 \le \theta_i \le 1,\ i = 1, 2.$

条件付き期待値 (連続型の場合)

(X, Y) は連続型で, $f_{X|Y}(x|y)$ を $Y = y$ を与えたときの X の条件付き確率密度関数とする.

$$\int_{-\infty}^{\infty} |x| f_{X|Y}(x|y) dx < \infty$$

であるとき, $Y = y$ を与えたときの X の条件付き平均は存在するといい,

$$\mathsf{E}(X \mid Y = y) \equiv \int_{-\infty}^{\infty} x f_{X|Y}(x|y) dx$$

を $Y = y$ を与えたときの X の条件付き平均, あるいは条件付き期待値とよぶ.

例 3.11 (2 次元正規分布)

$(X_1, X_2)' \sim N_2(\boldsymbol{\mu}, \Sigma)$ (2 次元正規分布),

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

$X_2 = x_2$ を与えたときの X_1 の条件付き分布は

$$N\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2),\ (1-\rho^2)\sigma_1^2\right)$$

条件付き平均は

$$\mathsf{E}(X_1 | X_2 = x_2) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2).$$

(証明は板書で)

条件付き平均の性質

定理 3.12

$f_{X,Y}(x,y)$: (X,Y) の同時確率 (密度) 関数
$f_{X|Y}(x|y)$: $Y=y$ を与えたときの X の条件付き確率 (密度) 関数

$$\mathrm{E}(g(X)|Y=y) = \begin{cases} \sum_{j=1}^{\infty} g(x_j) f_{X|Y}(x_j|y), & \text{離散型のとき} \\ \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)dx, & \text{連続型のとき} \end{cases}$$

次が成り立つ.

$g(y)=\mathrm{E}(X|Y=y)$ とするとき, $g(Y)$ は確率変数となるが, これを $\mathrm{E}(X|Y)$ と表す.

定理 3.13

$f_{X,Y}(x,y)$: (X,Y) の同時確率密度関数 (確率関数) このとき

$$\mathrm{E}\{\mathrm{E}(X|Y)\} = \mathrm{E}(X).$$

が成り立つ.

(証明は板書で)