第2章　期待値の基礎事項

この章では, 確率変数の期待値と関連するものを導入し, その基本的な性質を述べる. 節 2.1 では, 確率変数に対する期待値を定義する. 節 2.2 では, 期待値を確率ベクトルに対して定義する. 節 2.3 では, 分布のばらつきを測る量である分散と共分散を導入し, 基本的な性質を説明する. 節 2.4 では, 条件付き分布に対する期待値を導入する. 節 2.5 では, 分布を特定する量である積率母関数を導入し, その基本的な性質を説明する.

2.1　期待値

X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数とする. X が離散型のとき, その p.m.f. を p とし, X の取り得る値を $x_1, x_2, \ldots$ とする. 連続型のとき, その p.d.f. を p とする.

定義 2.1. $g : \mathbb{R} \longrightarrow \mathbb{R}$ を可測[1]関数とする. 確率変数 $g(X)$ の期待値 $\mathsf{E}[g(X)]$ を次のように定義する.

(1) $g \geq 0$ のとき

$$\mathsf{E}[g(X)] = \begin{cases} \displaystyle\sum_{n=1}^{\infty} g(x_n)\mathsf{p}(x_n) & (\text{離散型の場合}) \\ \displaystyle\int_{-\infty}^{\infty} g(x)\mathsf{p}(x)\,dx & (\text{連続型の場合}) \end{cases}$$

と定義する. 右辺は ∞ を許せば, 必ず存在する.

(2) 一般の g に対して

$$g^+(x) = \max\{g(x), 0\}, \quad g^-(x) = \max\{-g(x), 0\}$$

と定義すれば, $g^+ \geq 0$, $g^- \geq 0$ となる. $\mathsf{E}[g^+(X)]$ または $\mathsf{E}[g^-(X)]$ のいずれかが有限ならば,

$$\mathsf{E}[g(X)] := \mathsf{E}[g^+(X)] - \mathsf{E}[g^-(X)]$$

[1]可測関数の定義は定義 C.39 を参照.

と定義する．$\mathsf{E}[g^+(X)] = \mathsf{E}[g^-(X)] = \infty$ のときは，$g(X)$ の期待値は定義されない．$\mathsf{E}[g^+(X)] < \infty$ かつ $\mathsf{E}[g^-(X)] < \infty$ のとき，$\mathsf{E}[g(X)]$ は有限となる．

補題 2.2. X を確率変数とする．可測関数 $g, h : \mathbb{R} \longrightarrow \mathbb{R}$ に対して，$\mathsf{E}[|g(X)|] < \infty$, $\mathsf{E}[|h(X)|] < \infty$ を仮定する．

(1) $a, b \in \mathbb{R}$ に対して

$$\mathsf{E}[ag(X) + bh(X)] = a\,\mathsf{E}[g(X)] + b\,\mathsf{E}[h(X)].$$

(2) $g(x) \le h(x)$ ($\forall x \in \mathbb{R}$) ならば

$$\mathsf{E}[g(X)] \le \mathsf{E}[h(X)].$$

Proof. 積分の性質に帰着される．□

補題 2.3. $0 < q < r$ に対して

$$\mathsf{E}[|X|^r] < \infty \Longrightarrow \mathsf{E}[|X|^q] < \infty.$$

Proof. 積分の性質に帰着される．□

定義 2.4. (1) $k = 1, 2, \ldots$ に対して，$\mathsf{E}[|X|^k] < \infty$ のとき，$\mathsf{E}[X^k]$ を X の k **次モーメント** (または**積率**) という．

(2) $\mathsf{E}[|X|] < \infty$ のとき $\mathsf{E}[X]$ を X **の平均値**[2]という．

(3) $\mathsf{E}[X^2] < \infty$ のとき X **の分散**を

$$\mathsf{Var}[X] := \mathsf{E}[\{X - \mathsf{E}[X]\}^2]$$

で定義する．

(4) 部分集合 $A \subset \mathbb{R}$ に対して

$$\mathbb{1}_A(x) = \begin{cases} 1 & (x \in A) \\ 0 & (x \notin A) \end{cases}$$

を A の**指示関数**[3]という．$A \in \mathcal{B}(\mathbb{R})$ のとき

$$\mathsf{E}[\mathbb{1}_A(X)] = \mathsf{Pr}(X \in A)$$

となる．

[2] 簡単に「平均」ともいう．

[3] 指示関数は $\mathbb{R}$ の任意の部分集合に定義ができることに注意をせよ．

注意 2.5. (1) X は確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の離散型確率変数とする．p を X の p.m.f. とし

$$S_X := \{x \in \mathbb{R};\ \mathsf{p}(x) > 0\}$$

とする．関数 $g : \mathbb{R} \longrightarrow \mathbb{R}$ に対して

$$Y = g(X)$$

とおく．$y \in \mathbb{R}$ に対して

$$A_y := \{x \in S_X;\ g(x) = y\}$$

とおき

$$S_Y := \{y \in \mathbb{R};\ A_y \neq \varnothing\}$$

とする．$y \in S_Y$ に対して

$$\mathsf{q}(y) := \mathsf{Pr}(Y = q) = \mathsf{Pr}(X \in A_y) = \sum_{x \in A_y} \mathsf{Pr}(X = x) = \sum_{x \in A_y} \mathsf{p}(x)$$

と書ける．正項級数は項の順番を並び替えてもその値は変わらないので

$$\sum_{y \in S_Y} |y|\mathsf{q}(y) = \sum_{y \in S_Y} |y| \sum_{x \in S_y} \mathsf{p}(x) = \sum_{y \in S_Y} \sum_{x \in A_y} |g(x)|\mathsf{p}(x) = \sum_{x \in S_X} |g(x)|\mathsf{p}(x)$$

となる．さらにいずれかの和が有限ならば

$$\sum_{y} y\mathsf{q}(y) = \sum_{y} y \sum_{x \in S_y} \mathsf{p}(x) = \sum_{y} \sum_{x \in S_y} g(x)\mathsf{p}(x) = \sum_{x} g(x)\mathsf{p}(x)$$

となる．よって

$$\mathsf{E}[Y] = \mathsf{E}[g(X)]$$

となる．

(2) X は連続型とする．p_X を X の p.d.f. とし

$$S_X := \{x \in \mathbb{R};\ \mathsf{p}_X(x) > 0\}$$

とする．$g : S_X \longrightarrow \mathbb{R}$ を狭義単調増加かつ C^1 級とする．

$$Y = g(X), \quad S_Y := \{y \in \mathbb{R};\ y = g(x)\,(\exists x \in S_X)\}$$

とする．$g : S_X \longrightarrow S_Y$ と制限[4]すれば g の逆関数 $g^{-1} : S_Y \longrightarrow S_X$ が存在し

$$\mathsf{F}_Y(y) := \mathsf{Pr}(Y \le y) = \mathsf{Pr}\big(g^{-1}(Y) \le g^{-1}(y)\big) = \mathsf{Pr}(X \le g^{-1}(y)) = \mathsf{F}_X(g^{-1}(y)).$$

[4]制限したものを同じ g を用いて表現するのは, 記号の乱用である．記号が煩雑になるので, 記号を乱用した．

さらに $g(g^{-1}(y)) = y$ より

$$\frac{d}{dy}g^{-1}(y) = \frac{1}{\dot{g}(g^{-1}(y))}$$

となる．ただし $\dot{g}(y) = \dfrac{dg}{dy}(y)$ である．これらより, $y \in S_Y$ に対して

$$\begin{aligned}\mathsf{p}_Y(y) &= \frac{d}{dy}\mathsf{F}_Y(y) = \frac{d}{dy}\mathsf{F}_X(g^{-1}(y)) = \dot{\mathsf{F}}_X(g^{-1}(y))\frac{d}{dy}g^{-1}(y)\\ &= \mathsf{p}_X(g^{-1}(y))\frac{1}{\dot{g}\big(g^{-1}(y)\big)}\end{aligned}$$

となる．よって

$$\mathsf{p}_Y(y) = \mathsf{p}_X\big(g^{-1}(y)\big)\frac{1}{\dot{g}(g^{-1}(y))}$$

となる．g が狭義単調減少の場合もふくめると

$$\mathsf{p}_Y(y) = \mathsf{p}_X\big(g^{-1}(y)\big)\frac{1}{\big|\dot{g}\big(g^{-1}(y)\big)\big|}$$

となる． □

2.2　確率ベクトルの期待値

X, Y を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上で定義された確率変数とする．(X, Y) を確率ベクトルという．

(X, Y) が離散型のときその同時 p.m.f. を $\mathsf{p}(x, y)$ とし, 連続型のときその同時 p.d.f. を $\mathsf{p}(x, y)$ とする．

定義 2.6. 可測関数 $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ に対して $g(X, Y)$ の期待値を次のように定義する．

(1) $g \geq 0$ のとき

$$\mathsf{E}[g(X, Y)] = \begin{cases} \displaystyle\sum_{x,y} g(x, y)\mathsf{p}(x, y) & (\text{離散型の場合})\\ \displaystyle\int\!\!\int g(x, y)\mathsf{p}(x, y)\,dxdy & (\text{連続型の場合}). \end{cases}$$

(2) 一般の g に対して

$$g^+(x, y) := \max\{g(x, y), 0\}, \quad g^-(x, y) := \max\{-g(x, y), 0\}$$

と定義すれば $g^+, g^- \geq 0$ となる．$\mathsf{E}[g^+(X, Y)]$ または $\mathsf{E}[g^-(X, Y)]$ のいずれかが有限ならば

$$\mathsf{E}[g(X, Y)] := \mathsf{E}[g^+(X, Y)] - \mathsf{E}[g^-(X, Y)]$$

で定義する．$\mathsf{E}[g^+(X, Y)] = \mathsf{E}[g^-(X, Y)] = \infty$ のときは，$g(X, Y)$ の期待値は定義されない．$\mathsf{E}[g^+(X, Y)] < \infty$ かつ $\mathsf{E}[g^-(X, Y)] < \infty$ のとき，$\mathsf{E}[g(X, Y)]$ は有限の値を取る．

注意 2.7. 3 つ以上の確率変数 $X_1, X_2, \ldots, X_n$ に対しても期待値を定義 2.6 と同様に定義する． □

定理 2.8. $X_1, X_2, \ldots, X_n$ を確率変数とし，各 X_j $(j = 1, 2, \ldots, n)$ の期待値は有限とする．$a_1, a_2, \ldots, a_n$ を定数としたとき

$$\mathsf{E}\left[\sum_{j=1}^{n} a_j X_j\right] = \sum_{j=1}^{n} a_j \mathsf{E}[X_j].$$

Proof. 期待値の線型性 (補題 2.2(1)) よりわかる． □

例 2.9. $0 < p < 1$ と $j = 1, 2, \ldots, n$ に対して，$X_j \sim \mathsf{Bino}(p)$ とする．このとき

$$\mathsf{E}[X_j] = \sum_{x=0}^{1} x \mathsf{Pr}(X = x) = 0 \times (1 - p) + 1 \times p = p.$$

$S = \sum_{j=1}^{n} X_j$ としたとき

$$\mathsf{E}[S] = \sum_{j=1}^{n} \mathsf{E}[X_j] = np.$$

□

2.3 分散と共分散

定義 2.10. X を確率変数とし，$\mathsf{E}[X^2] < \infty$ とする．X **の分散**を

$$\mathsf{Var}[X] := \mathsf{E}[(X - \mu)^2]$$

で定義する．ただし $\mu = \mathsf{E}[X]$ と書いた．さらに $\sqrt{\mathsf{Var}[X]}$ を X **の標準偏差**という．

注意 2.11. 分散 $\mathsf{Var}[X]$ は X の分布の平均 μ まわりの散らばりを測る量である．分散がおおきいほど分布は広がっていることになる． □

定理 2.12. 以下の確率変数は有限の 2 次の積率を持つとする．このとき，次が成立する．

(1) $\mathsf{Var}[X] = \mathsf{E}[X^2] - \{\mathsf{E}[X]\}^2$.

(2) 定数 $a, b \in \mathbb{R}\,(a \neq 0)$ に対して

$$\mathsf{Var}[aX + b] = a^2\mathsf{Var}[X].$$

(3) X と Y は独立で $\mathsf{E}[|XY|] < \infty$ とする．このとき

$$\mathsf{E}[XY] = \mathsf{E}[X]\mathsf{E}[Y].$$

(4) $X_1, X_2, \ldots, X_n$ は独立とし, $\mathsf{E}[X_j^2] < \infty\,(j = 1, 2, \ldots, n)$ とする．$a_1, a_2, \ldots, a_n$ は定数としたとき

$$\mathsf{Var}[a_1X_1 + a_2X_2 + \cdots + a_nX_n] = a_1^2\mathsf{Var}[X_1] + a_2^2\mathsf{Var}[X_2] + \cdots + a_n^2\mathsf{Var}[X_n].$$

Proof. (1), (2) は分散を期待値の記号を用いて表現し, 期待値の線型性を用いて計算すればよい．(3) については, 連続型の場合を示す．(X, Y) の同時 p.d.f. は, X と Y の周辺 p.d.f. p_X と p_Y を用いて $\mathsf{p}_X(x)\mathsf{p}_Y(y)$ という形でかけるので

$$\mathsf{E}[XY] = \int\int xy\mathsf{p}_X(x)\mathsf{p}_Y(y)\,dxdy = \int x\mathsf{p}_X(x)\,dy \int y\mathsf{p}_Y(y)\,dy = \mathsf{E}[X]\mathsf{E}[Y].$$

(4) は, 分散を期待値の記号を用いて表現し, (3) に注意する．期待値の線型性を用いて計算すればよい． □

例 2.13. (例 2.9 の続き) 例 2.9 の設定に加えて $X_1, X_2, \ldots, X_n$ は独立とする．定理 2.12(1) に注意すれば

$$\begin{aligned}\mathsf{Var}[X_j] &= \mathsf{E}[X_j^2] - \{\mathsf{E}[X_j]\}^2 = \sum_{x=0}^{1} x^2\mathsf{Pr}(X = x) - p^2\\ &= 0 \times (1-p) + 1 \times p - p^2 = p(1-p).\end{aligned}$$

これと定理 2.12(4) から

$$\mathsf{Var}[S] = \sum_{j=1}^{n}\mathsf{Var}[X_j] = np(1-p).$$

□

定理 2.14. $n \geq 2$ とし, $X_1, X_2, \ldots, X_n$ は i.i.d. 確率変数列とし

$$\mathsf{E}[X_1] = \mu, \quad \mathsf{Var}[X_1] = \sigma^2, \quad 0 < \sigma < \infty$$

とする．$X_1, X_2, \ldots, X_n$ に基づく**標本平均**を

$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$$

で定義し, **標本** (不偏) **分散**を

$$S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}\bigl(X_j - \overline{X}_n\bigr)^2$$

で定義する. このとき

$$(1)\mathsf{E}\bigl[\overline{X}_n\bigr] = \mu, \quad (2)\mathsf{Var}\bigl[\overline{X}_n\bigr] = \frac{\sigma^2}{n}, \quad (3)\mathsf{E}\bigl[S_n^2\bigr] = \sigma^2.$$

Proof. (1), (2) は定理 2.8, 2.12(4) よりわかる. (3) を証明するために

$$\sum_{j=1}^{n}\bigl(X_j - \overline{X}_n\bigr)^2 = \sum_{j=1}^{n}(X_j - \mu)^2 - n\bigl(\overline{X}_n - \mu\bigr)^2$$

に注意する. 定理 2.12(1) より

$$\mathsf{E}\Biggl[\sum_{j=1}^{n}(X_j - \mu)^2\Biggr] = n\sigma^2.$$

(1) と (2) より

$$\mathsf{E}\bigl[\bigl(\overline{X}_n - \mu\bigr)^2\bigr] = \mathsf{Var}\bigl[\overline{X}_n\bigr] = \frac{\sigma^2}{n}.$$

したがって

$$\begin{aligned}\mathsf{E}[S_n^2] &= \frac{1}{n-1}\Biggl\{\sum_{j=1}^{n}\mathsf{E}\bigl[(X_j - \mu)^2\bigr] - n\mathsf{E}\bigl[\bigl(\overline{X}_n - \mu\bigr)^2\bigr]\Biggr\} \\ &= \frac{1}{n-1}\bigl\{n\sigma^2 - \sigma^2\bigr\} = \sigma^2.\end{aligned}$$

□

定義 2.15. X と Y は確率変数とし

$$\mathsf{E}[X] = \mu_X, \quad \mathsf{Var}[X] = \sigma_X^2, \quad \mathsf{E}[Y] = \mu_Y, \quad \mathsf{Var}[Y] = \sigma_Y^2$$

とする. ただし $\mu_X, \mu_Y \in \mathbb{R}$, $0 < \sigma_X, \sigma_Y < \infty$ とする. このとき X **と** Y **の共分散**を

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

で定義し, X **と** Y **の** (Pearson) **の相関係数**を

$$\rho := \rho[X, Y] := \frac{\mathsf{Cov}[X, Y]}{\sigma_X\,\sigma_Y}$$

で定義する.

定理 2.16. X, Y, Z は 2 次の積率が有限な確率変数とする.
(1) 共分散は
$$\mathsf{Cov}[X,\,Y]=\mathsf{E}[XY]-\mathsf{E}[X]\mathsf{E}[Y]$$
(2) $\mathsf{Cov}[X,\,Y]=\mathsf{Cov}[Y,\,X]$.
(3) 定数 $a,\,b\neq 0$ に対して
$$\mathsf{Cov}[aX+bY,\,Z]=a\mathsf{Cov}[X,\,Z]+b\mathsf{Cov}[Y,\,Z].$$
(4) 相関係数は
$$-1\leq\rho[X,\,Y]\leq 1$$
をみたす.
(5) ある定数 $a,\,b$ が存在して $Y=aX+b$ となったとき
$$\begin{aligned}a>0&\Longrightarrow\rho[X,\,Y]=1,\\ a<0&\Longrightarrow\rho[X,\,Y]=-1.\end{aligned}$$
(6) X と Y が独立のとき
$$\mathsf{Cov}[X,\,Y]=0.$$

Proof. (1) は期待値の線型性からわかる. (2)(3) は共分散の定義と期待値の線型性よりわかる. (4) は $\mathsf{Var}[X]=\mathsf{Var}[Y]=0$ のときは明らかである. よって $\mathsf{Var}[X]\neq 0$ として証明する. 補題 2.2(2) より
$$\mathsf{Var}[X]t^2-2\mathsf{Cov}[X,\,Y]t+\mathsf{Var}[Y]=\mathsf{E}[\{t(X-\mathsf{E}[X])+Y-\mathsf{E}[Y]\}^2]\geq 0$$
となる. したがって, 判別式をとれば
$$\{\mathsf{Cov}[X,\,Y]\}^2-\mathsf{Var}[X]\mathsf{Var}[Y]\leq 0$$
がわかる. (5) は (2) からわかる. (6) は定理 2.12(2) からわかる. □

注意 2.17. (6) の逆は一般に正しくない. 反例は下の問いをみよ. □

問 2.1. Z を $[0,\,2\pi]$ 上の一様分布とし
$$X=\cos Z,\quad Y=\sin Z$$
とおく.

(1) $\mathsf{E}[X]=0$, $\mathsf{E}[Y]=0$, $\mathsf{E}[XY]=0$ を確かめよ.

(2) $\mathsf{Pr}(X\leq 1/2,\,Y\leq 1/2)=\mathsf{Pr}((5/4)\pi\leq Z\leq(6/4)\pi)$ を確かめよ.

(3) $\mathsf{Pr}(X \le 1/2) = \mathsf{Pr}((1/4)\pi \le Z \le (7/4)\pi)$ を確かめよ[5].

定理 2.18. (1) X, Y は有限な 2 次の積率を持つ確率変数とする．このとき $\mathsf{Var}[X+Y] = \mathsf{Var}[X] + \mathsf{Var}[Y] + 2\mathsf{Cov}[X, Y]$.
(2) $d \ge 2$ とする．$X_1, X_2, , \dots, X_d$ は有限な 2 次の積率を持つ確率変数とする．このとき

$$\mathsf{Var}\left[\sum_{j=1}^{d} a_j X_j\right] = \sum_{j=1}^{d} a_j \mathsf{Var}[X_j] + 2\sum_{j=1}^{d}\sum_{\ell=j+1}^{d} a_j a_\ell \mathsf{Cov}[X_j, X_\ell].$$

Proof. 共分散を期待値で表現し, 展開して期待値の線型性を用いればよい. □

問 2.2. 定理 2.18(1) の証明を具体的に書け.

定義 2.19. $X_1, X_2, \dots, X_d$ を有限な 2 次の積率を持つ確率変数とし

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_d \end{pmatrix}$$

と書く．このとき確率ベクトル $\boldsymbol{X}$ の期待値を

$$\mathsf{E}[\boldsymbol{X}] := \begin{pmatrix} \mathsf{E}[X_1] \\ \mathsf{E}[X_2] \\ \vdots \\ \mathsf{E}[X_d] \end{pmatrix}$$

で定義する．$\boldsymbol{X}$ の共分散を

$$\mathsf{Var}[\boldsymbol{X}] := \mathsf{E}\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^\top\right]$$

で定義する[6]．ただし $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_d)^\top := \mathsf{E}[\boldsymbol{X}]$ である．これは

$$\mathsf{Var}[\boldsymbol{X}] = \begin{pmatrix} \mathsf{Var}[X_1] & \mathsf{Cov}[X_1, X_2] & \cdots & \mathsf{Cov}[X_1, X_d] \\ \mathsf{Cov}[X_2, X_1] & \mathsf{Var}[X_2] & \cdots & \mathsf{Cov}[X_2, X_d] \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{Cov}[X_d, X_1] & \mathsf{Cov}[X_d, X_2] & \cdots & \mathsf{Var}[X_d] \end{pmatrix}$$

である.

[5] これらより，$\mathsf{Pr}(X \le 1/2, Y \le 1/2) \ne \mathsf{Pr}(X \le 1/2)\mathsf{Pr}(Y \le 1/2)$ となり，$\mathsf{Cov}[X, Y] = 0$ だが, X と Y は従属であることがわかる.

[6] 確率ベクトルと同様に確率変数を成分とする行列を確率行列という．確率行列の期待値はそれぞれの成分の期待値を取ったものを配置した行列と定義している.

補題 2.20. $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top$ は確率ベクトルで各成分は有限な 2 次の積率を持つとし

$$\mathsf{E}[\boldsymbol{X}] = \boldsymbol{\mu}, \quad \mathsf{Var}[\boldsymbol{X}] = \boldsymbol{\Sigma}$$

とする．ただし $\boldsymbol{\mu} \in \mathbb{R}^d$, $\boldsymbol{\Sigma}$ は $d \times d$ の半正定値行列[7]である．
(1) 任意の定数ベクトル $\boldsymbol{a} \in \mathbb{R}^d$ に対して

$$\mathsf{E}[\boldsymbol{a}^\top \boldsymbol{X}] = \boldsymbol{a}^\top \boldsymbol{\mu}, \quad \mathsf{Var}[\boldsymbol{a}^\top \boldsymbol{X}] = \boldsymbol{a}^\top \boldsymbol{\Sigma} \boldsymbol{a}.$$

(2) $k \in \mathbb{N}$ とする． 任意の定数の $k \times d$ 行列 $\mathbf{A}$ に対して

$$\mathsf{E}[\mathbf{A}\boldsymbol{X}] = \mathbf{A}\mathsf{E}[\boldsymbol{X}], \quad \mathsf{Var}[\mathbf{A}\boldsymbol{X}] = \mathbf{A}\mathsf{Var}[\boldsymbol{X}]\mathbf{A}^\top.$$

Proof. 期待値の線型性を用いて計算すればよい． □

問 2.3. $d = 2$ として, 補題 2.20 を確認せよ．

系 2.21. 確率ベクトル $\boldsymbol{X}$ は, 任意の定数ベクトル $\boldsymbol{a}(\neq \mathbf{0})$ に対して,

$$\mathsf{Pr}(\boldsymbol{a}^\top \boldsymbol{X} = 0) = 0$$

をみたす．このとき, $\boldsymbol{\Sigma}$ は正定値である．

注意 2.22. $\boldsymbol{X}$ の共分散行列 $\boldsymbol{X}$ は, 補題 2.20(2) から半正定値であることがわかる．系 2.21 の仮定は, 確率ベクトル $\boldsymbol{X}$ が d 次元空間より次元の低い空間に集中しないことを意味している．したがって, 確率ベクトル $\boldsymbol{X}$ の値を取る空間が退化していなければ, その共分散行列は正定値となることがわかる． □

2.4 条件付き期待値

定義 2.23. (1) X と Y を確率変数とし, 条件付き p.d.f.(条件付き p.m.f.) を $\mathsf{p}_{X|Y}(p_{X|Y})$ とする．$Y = y$ を与えたときの X の条件付き期待値を

$$\mathsf{E}[X|Y = y] = \begin{cases} \displaystyle\sum_x x\mathsf{p}_{X|Y}(x|y) & \text{(離散型の場合)} \\ \displaystyle\int x\mathsf{p}_{X|Y}(x|y)\,dx & \text{(連続型の場合)} \end{cases}$$

で定義する．ただし考えている $Y = y$ で条件付き p.d.f. または p.m.f. は定義され $\mathsf{E}[|X|] < \infty$ とする．

[7] $d \times d$ の対称行列 $\mathbf{A}$ が半正定値であるとは, $\forall \boldsymbol{a} \in \mathbb{R}^d$ に対して $\boldsymbol{a}^\top \mathbf{A} \boldsymbol{a} \geq 0$ が成立するときをいう．また $d \times d$ の対称行列 $\mathbf{A}$ が正定値であるとは, $\forall \boldsymbol{a} \in \mathbb{R}^d$ $(\boldsymbol{a} \neq \mathbf{0})$ に対して $\boldsymbol{a}^\top \mathbf{A} \boldsymbol{a} > 0$ が成立するときをいう．

(2) (Borel 可測) 関数 $g:\mathbb{R}^2\longrightarrow\mathbb{R}$ に対して, $g(X,Y)$ の条件付き期待値を

$$\mathsf{E}[g(X,Y)|\,Y=y]=\begin{cases}\displaystyle\sum_x g(x,y)\mathsf{p}_{X|Y}(x|\,y) & (\text{離団型の場合})\\ \displaystyle\int g(x,y)\mathsf{p}_{X|Y}(x|\,y)\,dx & (\text{連続型の場合})\end{cases}$$

で定義する．ただし, $\mathsf{E}[|g(X,Y)|]<\infty$ のとき, 考えている $Y=y$ での条件付き p.d.f. または p.m.f. は定義されるとする．

注意 2.24. $\mathsf{E}[X]$ は定数であるが, $\mathsf{E}[X|\,Y=y]$ は一般に y の関数である．

$$h(y):=\mathsf{E}[X|\,Y=y]$$

とおいたときに $h(y)$ に Y を代入したものは確率変数になる．これを

$$\mathsf{E}[X|\,Y]:=h(Y)$$

と定義する．したがって $\omega\in\Omega$ に対して, $y=Y(\omega)$ と書けば

$$\mathsf{E}[X|\,Y]:\Omega\ni\omega\mapsto\mathsf{E}[X|\,Y(\omega)]=\mathsf{E}[X|\,Y=y]\in\mathbb{R}$$

は可測となる． □

例 2.25. 連続型確率変数 Y は p.d.f.

$$\mathsf{p}(y)=\begin{cases}1 & (0<y<1),\\ 0 & (\text{その他の場合})\end{cases}$$

を持つとする．$Y=y\,(0<y<1)$ を観測したとき

$$X|\,Y=y\ \sim\ \mathsf{Unif}(y,1)$$

とする．すなわち $0<y<1$ のとき

$$\mathsf{p}_{X|Y}(x|\,y)=\begin{cases}\dfrac{1}{1-y} & (y<x<1),\\ 0 & (\text{その他の場合})\end{cases}$$

である．よって

$$\mathsf{E}[X|\,Y=y]=\int_y^1 y\mathsf{p}_{X|Y}(x|\,y)\,dx=\frac{1}{1-y}\int_y^1 dx=\frac{1+y}{2}.$$

これより

$$\mathsf{E}[X|\,Y]=\frac{1+Y}{2}$$

となる． □

定理 2.26. (1) 有限な期待値を持つ確率変数 X, Y と確率変数 Z に対して

$$\mathsf{E}[X+Y|\,Z] = \mathsf{E}[X|\,Z] + \mathsf{E}[Y|\,Z].$$

(2) 有限な期待値を持つ確率変数 X, Y に対して

$$\mathsf{E}[\mathsf{E}[Y|\,X]] = \mathsf{E}[Y], \quad \mathsf{E}[\mathsf{E}[X|\,Y]] = \mathsf{E}[X].$$

(3) 一般の (Borel 可測) 関数 $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ を考える．$\mathsf{E}[|g(X, Y)|] < \infty$ のとき

$$\mathsf{E}[\mathsf{E}[g(X, Y)|\,Y]] = \mathsf{E}[g(X, Y)].$$

(4) $\mathsf{E}[XY|\,Y] = Y\mathsf{E}[X|\,Y]$.

Proof. (1) は期待値の線型性よりわかる．連続型の場合について (2) の第 1 番目の等式を示す．他の場合もほとんど同じように証明できる．f を (X, Y) の同時 p.d.f. とする．このとき

$$\mathsf{p}(x, y) = \mathsf{p}_Y(y)\mathsf{p}_{X|\,Y}(x|\,y)$$

となる．ただし p_X は X の周辺 p.d.f. で $\mathsf{p}\quad_{Y|\,X}$ は $X = x$ を与えたときの Y の条件付き p.d.f. である．すると

$$\begin{aligned}
\mathsf{E}[\mathsf{E}[X|\,Y]] &= \int \mathsf{E}[X|\,Y=y]\mathsf{p}_Y(y)\,dy \\
&= \int \left\{ \int y\mathsf{p}_{X|\,Y}(x|\,y)\,dx \right\} \mathsf{p}_Y(y)\,dy \\
&= \int\int y\mathsf{p}_{X|\,Y}(x|\,y)\mathsf{p}_Y(y)\,dxdy \\
&= \int\int x \left\{ \int \mathsf{p}(x, y)\,dy \right\}\,dx \\
&= \int x\mathsf{p}_X(x)\,dx \\
&= \mathsf{E}[X].
\end{aligned}$$

積分の順序交換は $g(X, Y)$ が有限の期待値を持つことから保証される．
□

例 2.27. (例 2.25 の続き)

$$\mathsf{E}[X] = \mathsf{E}\left[\frac{1+Y}{2}\right] = \frac{3}{4}$$

となる．一方 $0 < x < y < 1$ に対して

$$\mathsf{p}(x, y) = \mathsf{p}_{X|\,Y}(x|\,y)p_Y(y) = \frac{1}{1-y}.$$

よって

$$
\begin{aligned}
\mathsf{E}[X] &= \int_0^1 \left\{ \int_y^1 x \frac{1}{1-y}\, dx \right\} dy \\
&= \int \frac{1}{1-y} \left[\frac{x^2}{2} \right]_y^1 dy \\
&= \int_0^1 \frac{1}{1-y} \frac{1-y^2}{2}\, dy = \left[\frac{(1+y)^2}{4} \right]_0^1 = \frac{3}{4}.
\end{aligned}
$$

□

定義 2.28. X は有限の 2 次の積率を持つとする. $Y=y$ を与えたときの X の条件付き p.d.f. $\mathsf{p}_{X|Y}$(p.m.f. $\mathsf{p}_{X|Y}$) が定義できる y を考える. このとき, $Y=y$ を与えたときの条件付き分散を

$$
\mathsf{Var}[X|\,Y=y] = \begin{cases} \displaystyle\sum_x \{x-\mu(y)\}^2 \mathsf{p}_{X|Y}(x|\,y) & (\text{離散型の場合}) \\ \displaystyle\int \{x-\mu(y)\}^2 \mathsf{p}_{X|Y}(x|\,y)\, dx & (\text{連続型の場合}) \end{cases}
$$

で定義する. ただし $\mu(y) = \mathsf{E}[X|\,Y=y]$ である. これは

$$
\mathsf{Var}[X|\,Y] = \mathsf{E}[X^2|\,Y] - \{\mathsf{E}[X|\,Y]\}^2
$$

とも書ける.

定理 2.29. X, Y を確率変数とし $\mathsf{E}[X^2] < \infty$ とする. このとき

$$
\mathsf{Var}[X] = \mathsf{E}[\mathsf{Var}[X|\,Y]] + \mathsf{Var}[\mathsf{E}[X|\,Y]]
$$

が成立する. ただし $h(y) = \mathsf{Var}[X|\,Y=y]$ としたとき $\mathsf{Var}[X|\,Y] := h(Y)$ と定義した.

Proof.

$$
\begin{aligned}
\mathsf{Var}[X] &= \mathsf{E}\big[\{X - \mathsf{E}[X]\}^2\big] \\
&= \mathsf{E}\Big[\big\{X - \mathsf{E}[X|\,Y] + \mathsf{E}[X|\,Y] - \mathsf{E}[X]\big\}^2\Big] \\
&= \mathsf{E}\big[\{X - \mathsf{E}[X|\,Y]\}^2\big] + \mathsf{E}\big[\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}^2\big] \\
&\quad + 2\mathsf{E}\big[\{X - \mathsf{E}[X|\,Y]\}\{\mathsf{E}[X|\,Y] - \mathsf{E}[X]\}\big].
\end{aligned}
$$

しかし

$$
\begin{aligned}
\mathsf{E}\big[\{X-\mathsf{E}[X|Y]\}^2\big] &= \mathsf{E}\big[X^2-2X\mathsf{E}[X|Y]+\{\mathsf{E}[X|Y]\}\big] \\
&= \mathsf{E}\Big[\mathsf{E}\big[X^2-2X\mathsf{E}[X|Y]+\{\mathsf{E}[X|Y]\}^2\big]\Big|\,Y\Big] \\
&\quad (\text{定理 } 2.26(3)) \\
&= \mathsf{E}\Big[\mathsf{E}\big[X^2\big|\,Y\big]-2\mathsf{E}[X|Y]\,\mathsf{E}[X|Y]+\{\mathsf{E}[X|Y]\}^2\Big] \\
&= \mathsf{E}\Big[\mathsf{E}[X^2|Y]-\{\mathsf{E}[X|Y]\}^2\Big] \\
&= \mathsf{E}\Big[\mathsf{Var}[X|Y]\Big], \\
\mathsf{E}\Big[\{\mathsf{E}[X|Y]-\mathsf{E}[X]\}^2\Big] &= \mathsf{E}\Big[\big\{\mathsf{E}[X|Y]-\underbrace{\mathsf{E}[\mathsf{E}[X|Y]]}_{=\mathsf{E}[X]}\big\}^2\Big] \qquad (\text{定理 } 2.26(3)) \\
&= \mathsf{Var}\Big[\mathsf{E}[X|Y]\Big], \\
\mathsf{E}\big[\{X-\mathsf{E}[X|Y]\}\{\mathsf{E}[X|Y]-\mathsf{E}[X]\}\big] &= \mathsf{E}\Big[\mathsf{E}\big[\{X-\mathsf{E}[X|Y]\}\{\mathsf{E}[X|Y]-\mathsf{E}[X]\}\Big|\,Y\Big] \\
&\quad (\text{定理 } 2.26(3)) \\
&= \mathsf{E}\Big[\{\mathsf{E}[X|Y]-\mathsf{E}[X]\}\,\underbrace{\mathsf{E}\big[X-\mathsf{E}[X|Y]\big|Y\big]}_{=0}\Big] \\
&= 0.
\end{aligned}
$$

最後から 2 番目の等号は定理 2.26(3) を用いた. □

2.5　積率母関数

定義 2.30. X を確率変数とし, ある $t_0>0$ が存在して, $\mathsf{E}[e^{tX}]<\infty$ $(\forall|t|<t_0)$ とする. このとき, X **の積率母関数** (Moment Generating Function (MGF)) を

$$\mathsf{m}_X(t) := \mathsf{E}[e^{tX}] \quad (-t_0<t<t_0)$$

と定義する.

注意 2.31. $\mathsf{m}_X(t)$ が存在するとき, 期待値と微分の記号の入れ替えが保

証され,

$$\begin{aligned}\dot{\mathsf{m}}_X(0) &= \dot{\mathsf{m}}_X(t)\big|_{t=0} = \left[\frac{d}{dt}m_X(t)\right]\bigg|_{t=0} = \left[\frac{d}{dt}\mathsf{E}\big[e^{tX}\big]\right]\bigg|_{t=0} = \mathsf{E}\left[\frac{d}{dt}e^{tX}\right]\bigg|_{t=0} \\ &= \mathsf{E}\big[Xe^{tX}\big]\big|_{t=0} = \mathsf{E}[X]\end{aligned}$$

となる. この議論を繰り返せば $k = 2, 3, \dots$ に対して

$$\mathsf{m}_X^{(k)}(0) = \mathsf{E}\big[X^k\big].$$

□

例 2.32. $X \sim \mathsf{Exp}(1)$ とする. $t < 1$ に対して

$$\mathsf{m}_X(t) = \mathsf{E}\big[e^{tX}\big] = \int_0^\infty e^{tx}e^{-x}\,dx = \int_0^\infty e^{-(1-t)x}\,dx = \frac{1}{1-t}.$$

$t \geq 1$ のときは, e^{tX} の期待値は発散する. したがって

$$\mathsf{m}_X(t) = \frac{1}{1-t} \qquad (t < 1)$$

となる. いま

$$\dot{\mathsf{m}}_X(0) = 1, \quad \ddot{\mathsf{m}}_X(0) = 2$$

なので

$$\mathsf{E}[X] = 1, \quad \mathsf{E}[X^2] = 2, \quad \mathsf{Var}[X] = \mathsf{E}[X^2] - \{\mathsf{E}[X]\}^2 = 1.$$

□

補題 2.33. (1) $a, b \in \mathbb{R}\,(a \neq 0)$ とする. $Y = aX + b$ としたとき

$$\mathsf{m}_Y(t) = e^{tb}\mathsf{m}_X(at).$$

(2) $X_1, X_2, \dots, X_d$ は独立とし $Y = \sum_{j=1}^d Y_j$ とする. このとき

$$\mathsf{m}_Y(t) = \prod_{j=1}^d \mathsf{m}_{X_j}(t).$$

Proof. (1) は指数関数と期待値の性質よりわかる. (2) は指数関数の性質と定理 2.12(3) よりわかる. □

定理 2.34. X と Y を確率変数とする. ある数 $t_0 > 0$ が存在して

$$\mathsf{m}_X(t) = \mathsf{m}_Y(t) \quad (|t| < t_0)$$

ならば

$$X \overset{d}{=} Y.$$

ただし X と Y の c.d.f. を F_X と F_Y としたとき

$$X \overset{d}{=} Y \iff \mathsf{F}_X(x) = \mathsf{F}_Y(x) \qquad (\forall x \in \mathbb{R})$$

である.

Proof. これは信じることにする. □

例 2.35. $n_1, n_2 \in \mathbb{N}$, $0 < p < 1$ とし $X_1 \sim \mathsf{Bino}(n_1, p)$, $X_2 \sim \mathsf{Bino}(n_2, p)$ は独立とする. $Y = X_1 + X_2$ としたとき

$$\mathsf{m}_Y(t) = \mathsf{m}_{X_1}(t)\mathsf{m}_{X_2}(t) = \left(pe^t + q\right)^{n_1}\left(pe^t + q\right)^{n_2} = (pe^t + q)^{n_1+n_2}.$$

ただし $q = 1 - p$. よって $Y \sim \mathsf{Bino}(n_1 + n_2, p)$. □

例 2.36. $\lambda_1, \lambda_2 > 0$ とし, $X_1 \sim \mathsf{Poi}(\lambda_1)$, $X_2 \sim \mathsf{Poi}(\lambda_2)$ は独立とする. $Y = X_1 + X_2$ としたとき

$$\mathsf{m}_Y(t) = \mathsf{m}_{X_1}(t)\mathsf{m}_{X_2}(t) = e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)} = e^{(\lambda_1+\lambda_2)(e^t-1)}$$

より $Y \sim \mathsf{Poi}(\lambda_1 + \lambda_2)$ となる. □

2.6 章末注釈と参考文献

2.7 演習問題

演習問題 2.1. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 X は

$$\mathsf{Pr}(X = c) = 1; \qquad c \text{ は定数}$$

をみたすとする. このとき, X の平均 $\mathsf{E}[X]$ と分散 $\mathsf{Var}[X]$ を求めよ.

演習問題 2.2. 確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の確率変数 X は $\mathsf{Var}[X] = \sigma^2$ $(0 < \sigma < \infty)$ とする. c を定数としたとき, $\mathsf{Var}[X + c]$ と $\mathsf{Var}[cX]$ を σ と c を用いて表せ.

演習問題 2.3. (1) $X \sim \mathsf{Ga}(\alpha, \beta)$ とする. X の k 次の積率は以下で与えられることを示せ.

$$\mathsf{E}[X^k] = \frac{(a + k - 1)(a + k - 2)\cdots\alpha}{\beta^k}.$$

(2) 上の問いの結果から

$$\mathsf{E}[X] = \frac{\alpha}{\beta}, \quad \mathsf{E}[X^2] = \frac{(\alpha+1)\alpha}{\beta^2}, \quad \mathbb{V}[X] = \frac{\alpha}{\beta^2}$$

を示せ.
(3) X の積率母関数は以下で与えられることを示せ.

$$\mathsf{m}_X(t) = \left(\frac{1}{1-\beta t}\right)^{\alpha} \quad (t < 1/\beta).$$

演習問題 2.4. X を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の連続型確率変数とし, X の p.d.f. を

$$\mathsf{p}(x) = \begin{cases} 2x & (0 < x < 1) \\ 0 & \text{(その他の場合)} \end{cases}$$

とする. さらに, 離散型確率変数 Y を

$$Y := \mathbb{1}_{(1/2,\,\infty)}(X) := \begin{cases} 1 & \left(X > \frac{1}{2}\right) \\ 0 & \left(X \leq \frac{1}{2}\right) \end{cases}$$

で定める. このとき, 以下の問に答えよ.
(1) 確率変数 X の平均 $\mathsf{E}[X]$ と分散 $\mathsf{Var}[X]$ を求めよ.
(2) 確率変数 Y の平均 $\mathsf{E}[Y]$ と分散 $\mathsf{Var}[Y]$ を求めよ.
(3) 確率変数 X と Y の共分散 $\mathsf{Cov}[X, Y]$ を求めよ.

演習問題 2.5. X_1, X_2 を確率空間 $(\Omega, \mathcal{A}, \mathsf{Pr})$ 上の連続型確率変数とし,

$$X_1 \sim \mathsf{N}(x_0 + \mu, \sigma^2), \qquad X_2 |\, X_1 = x_1 \sim \mathsf{N}(x_1 + \mu, \sigma^2)$$

とする. ただし, $x_0, x_1, \mu \in \mathbb{R}$ である. このとき, 以下の問いに答えよ. ただし, 正規分布 $\mathsf{N}(\mu, \sigma^2)$ に従う確率変数 Z の期待値と分散が $\mathbb{E}[Z] = \mu$, $\mathsf{Var}[Z] = \sigma^2$ となることは証明なしで用いてよい. また, 期待値, 分散, 共分散, 条件付き期待値に係る資料に書いてある性質も証明なしで用いてよい. なお, どの性質を用いたかは明示すること.
(1) 確率変数 $X_2 - X_1$ の期待値 $\mathbb{E}[X_2 - X_1]$ を求めよ.
(2) X_2 の分散 $\mathsf{Var}[X_2]$ を求めよ.
(3) X_1 と X_2 の共分散 $\mathsf{Cov}[X_1, X_2]$ を求めよ.
(4) $X_2 - X_1$ の分散 $\mathsf{Var}[X_2 - X_1]$ を $\mathsf{Var}[X_1]$, $\mathsf{Var}[X_2]$, $\mathsf{Cov}[X_1, X_2]$ で表現せよ.
(5) $X_2 - X_1$ の分散 $\mathsf{Var}[X_2 - X_1]$ を求めよ.