はじめに

このノートは統計解析と情報統計学の講義録である.

目 次

第 4 章　標本分布論と漸近分布論

4.1　標本分布論の枠組み

4.1.1　ランダム標本

定義 4.1　$X_1, X_2, \ldots, X_n$ が母集団分布 $f(x)$ からの標本の大きさが n の/**ランダム標本**であるとは，$X_1, X_2, \ldots, X_n$ は互いに独立な確率変数列であって，各 X_i, $i = 1, 2, \ldots, n$ は確率関数または確率密度関数 $f(x)$ に従うときをいう．また，$X_1, X_2, \ldots, X_n$ は独立同一に確率関数または確率密度関数 $f(x)$ に従うともいう．

注意 4.1　多くの場合は $n > 1$ である．また，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時確率関数または同時確率密度関数は

$$f_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n) = \prod_{k=1}^{n} f(x_i)$$

である．

4.1.2　統計量と標本分布

定義 4.2　$X_1, X_2, \ldots, X_n$ をある母集団分布からの標本の大きさが n のランダム標本とし，$T(x_1, x_2, \ldots, x_n)$ をランダム標本 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の値域上で定義された実数値または実ベクトル値関数とする．このとき，確率変数または確率ベクトル $Y = T(X_1, X_2, \ldots, X_n)$ を**統計量**という．さらに，統計量 Y の確率分布を Y の**標本分布**とよぶ．

例 4.1　ランダム標本の算術平均は統計量であり，**標本平均**という．通常

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

と記す．

　また，

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

で定義される統計量を**標本分散**という．

補題 4.1　$x_1, x_2, \ldots, x_n$ を実数列とし，$\bar{x}_n = (1/n)(x_1 + x_2 + \cdots + x_n)$ と $s_n^2 = (1/(n-1))\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$ とおく．このとき，

(1)　$\min_a \sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}(x_i - \bar{x}_n)^2$.

(2)　$(n-1)s_n^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2$.

となる．

証明 (1) を証明するために，$\bar{x}_n$ を加えて引けば，

$$\begin{aligned}\sum_{i=1}^n (x_i - a)^2 &= \sum_{i=1}^n (x_i - \bar{x}_n + \bar{x}_n - a)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x}_n)^2 + 2\sum_{i=1}^n (x_i - \bar{x}_n)(\bar{x}_n - a) + \sum_{i=1}^n (\bar{x}_n - a)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x}_n)^2 + n(\bar{x}_n - a)^2 \qquad (4.1)\end{aligned}$$

となる．最後の等式は

$$\sum_{i=1}^n (x_i - \bar{x}_n)(\bar{x}_n - a) = (\bar{x}_n - a)\sum_{i=1}^n (x_i - \bar{x}_n) = 0$$

よりわかる．(4.1) は $a = \bar{x}_n$ の時に最小になることがわかる．

(2) を示すためには，(4.1) において，$a = 0$ とすればよい． □

命題 4.1 $X_1, X_2, \ldots, X_n$ をある母集団分布からの標本の大きさが n のランダム標本とし，$g_i\ (i = 1, 2, \ldots, n)$ を X_i の値域上で定義された実数値関数とする． $g_i^2(X_i)$ の期待値が存在するとき，

$$\begin{aligned}\mathbb{E}[\sum_{i=1}^n g_i(X_i)] &= \sum_{i=1}^n \mathbb{E}[g_i(X_i)], \qquad (4.2) \\ \mathbb{VAR}[\sum_{i=1}^n g(X_i)] &= \sum_{i=1}^n \mathbb{VAR}[g_i(X_i)] \qquad (4.3)\end{aligned}$$

が成立する．

証明 (4.2) は期待値の線形性と

$$\mathbb{E}[\sum_{i=1}^n g_i(X_i)] = \sum_{i=1}^n \mathbb{E}[g_i(X_i)]$$

からわかる(4-1)．

(4.3) を示すために，分散の定義と期待値の線形性から

$$\begin{aligned}\mathbb{VAR}[\sum_{i=1}^n g_i(X_i)] &= \mathbb{E}[\{\sum_{i=1}^n g_i(X_i) - \mathbb{E}[\sum_{i=1}^n g_i(X_i)]\}^2] \\ &= \mathbb{E}[\{\sum_{i=1}^n (g_i(X_i) - \mathbb{E}[g_i(X_i)])\}^2] \\ &= \mathbb{E}\left[\sum_{i=1}^n (g_i(X_i) - \mathbb{E}[g_i(X_i)])^2 + \sum_{i\neq j} g_i(X_i) - \mathbb{E}[g_i(X_i)])g_j(X_j) - \mathbb{E}[g_j(X_j)])\right] \\ &= \sum_{i=1}^n \mathbb{E}[(g_i(X_i) - \mathbb{E}[g_i(X_i)])^2] \\ &\quad + \sum_{i\neq j} \mathbb{E}[\{g_i(X_i) - \mathbb{E}[g_i(X_i)]\}\}g_j(X_j) - \mathbb{E}[g_j(X_j)]\}] \qquad (4.4)\end{aligned}$$

となる．しかし，

$$\sum_{i=1}^n \mathbb{E}[(g_i(X_i) - \mathbb{E}[g_i(X_i)])^2] = \sum_{i=1}^n \mathbb{VAR}[g_i(X_i)]$$

と $i \neq j$ に対して，

$$\mathbb{E}[g_i(X_i) - \mathbb{E}[g_i(X_i)])(g_j(X_j) - \mathbb{E}[g_j(X_j)])] = \mathbb{E}\{g_i(X_i) - \mathbb{E}[g_i(X_i)]\}\mathbb{E}\{g_j(X_j) - \mathbb{E}[g(X_j)]\} = 0$$

となる．ただし，最後の等号は定理 3.1 からわかる．上のふたつの式を (4.4) に代入すれば，(2) は示される．□

系 4.1 $X_1, X_2, \ldots, X_n$ をある母集団分布からの標本の大きさが n のランダム標本とする． X_1^2 の期待値が存在するとき，

$$\begin{aligned}\mathbb{E}[\sum_{i=1}^n a_i X_i)] &= \sum_{i=1}^n a_i \mathbb{E}[X_i)],\\ \mathbb{VAR}[\sum_{i=1}^n a_i X_i] &= \sum_{i=1}^n a_i^2 \mathbb{VAR}[X_i]\end{aligned}$$

が成立する．ただし，$a_1, a_2, \ldots, a_n$ は定数である．

系 4.2 $X_1, X_2, \ldots, X_n$ を平均 μ と分散 $\sigma^2 < \infty$ の母集団分布からの標本の大きさが n のランダム標本とする．このとき，

(1) $\mathbb{E}[\bar{X}_n] = \mu$,

(2) $\mathbb{VAR}[\bar{X}_n] = \dfrac{\sigma^2}{n}$,

(3) $\mathbb{E}[S_n^2] = \sigma^2$.

となる．

証明 (1) を示すために，命題 4.1.2 において，$g_i(x) = x/n \, (i = 1, 2, \ldots, n)$ とすれば，

$$\mathbb{E}\left[\bar{X}_n\right] = \frac{1}{n}\mathbb{E}[\sum_{i=1}^n X_i] = \frac{1}{n} n\mathbb{E}[X_1] = \mathbb{E}[X_1] = \mu$$

となることがわかる．

(2) は分散の性質と定理 を同様に利用すれば，

$$\mathbb{VAR}[\bar{X}_n] = \frac{1}{n^2}\mathbb{VAR}[\sum_{i=1}^n X_i] = \frac{1}{n^2} n\mathbb{VAR}[X_1] = \frac{1}{n}\mathbb{VAR}[X_1] = \frac{\sigma^2}{n}$$

となることがわかる．

(3) を示すために補題 4.1 を使えば，

$$\begin{aligned}\mathbb{E}[S_n^2] &= \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2\right] = \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^n X_i^2 - n\bar{X}_n^2\right]\\ &= \frac{1}{n-1}\left(n\mathbb{E}[X_1^2] - n\mathbb{E}[\bar{X}_n^2]\right) \qquad (4.5)\end{aligned}$$

となる．しかし，

$$\begin{aligned}\mathbb{E}[X_1^2] &= \mathbb{VAR}[X_1] + (\mathbb{E}[X_1])^2 = \sigma^2 + \mu^2,\\ \mathbb{E}[\bar{X}_n^2] &= \mathbb{VAR}[\bar{X}_n] + (\mathbb{E}[\bar{X}_n])^2 = \frac{\sigma^2}{n} + \mu^2\end{aligned}$$

となる．これらと (4.5) をあわせれば，

$$\mathbb{E}[S_n^2] = \frac{1}{n-1}\left(n(\sigma^2 + \mu^2 - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right) = \sigma^2$$

となることがわかる． □

命題 4.2 $X_1, X_2, \ldots, X_n$ を積率母関数 $M_X(t)$ を持つ母集団分布からの標本の大きさが n のランダム標本とする．このとき，

$$M_{\bar{X}_n}(t) = (M_X(t/n))^n$$

が成立する．

証明

$$M_{\bar{X}_n}(t) = \mathbb{E}[e^{t\bar{X}_n}] = \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i/n}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i/n}\right] = [M_X(t/n)]^n$$

からわかる． □

例 4.2 $X_1, X_2, \ldots, X_n$ を正規分布 $N(\mu, \sigma^2)$ からの標本の大きさが n のランダム標本とする．このとき，標本平均 $\bar{X}_n$ は正規分布 $N(\mu, \sigma^2/n)$ に従うことがわかる．なぜならば，

$$M_{\bar{X}_n}(t) \quad = \quad \exp\left[n\left(\frac{t}{n}\mu + \frac{1}{2}\left(\frac{t}{n}\right)^2\sigma^2\right)\right] = \exp\left[t\mu + \frac{(\sigma^2/n)t^2}{2}\right]$$

からわかる．

例 4.3 $X_1, X_2, \ldots, X_n$ を母数 p のベルヌーイ分布からの標本の大きさが n のランダム標本とする．ただし，$0 < p < 1$ である．このとき，標本平均 $n\bar{X}_n$ は母数 n と p の二項分布に従うことがわかる．なぜならば，

$$M_{n\bar{X}_n}(t) \quad = \quad \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}] = \left[pe^t + (1-p)\right]^n$$

からわかる．

4.2 正規分布からのランダム標本

命題 4.3 $X_1, X_2, \ldots, X_n$ を正規分布 $N(\mu, \sigma^2)$ からの標本の大きさが n のランダム標本とし，$\bar{X}_n = (1/n)\sum_{i=1}^n X_i$ と $S_n^2 = (1/(n-1))\sum_{i=1}^n (X_i - \bar{X}_n)^2$ とおく．このとき，以下が成立する：

(1) $\bar{X}_n$ と S_n^2 は独立である．

(2) $\bar{X}_n$ は正規分布 $N(\mu, \sigma^2/n)$ に従う．

(3) $(n-1)S_n^2/\sigma^2$ は自由度 $n-1$ のカイ自乗分布に従う．

証明 (2) は例 4.2 からわかる．次に，(1) を示す．各 i に対して，$Y_i = (X_i - \mu)/\sigma$ とすれば，Y_i は 正規分布 $N(0, 1)$ に従い，$(\bar{X}_n - \mu)/\sigma = \bar{Y}_n/\sigma = (1/n)\sum_{i=1}^n Y_i$ と $S_n^2/\sigma^2 = (1/(n-1))\sum_{i=1}^n (Y_i - \bar{Y}_n)^2$ となるので，一般性を失わず，X_i は正規分布 $N(0, 1)$ に従うとして，$\bar{X}_n$ と S_n^2 の独立性を示せばよいことがわかる．

$X_i - \bar{X}_n$ は正規分布 $N(0, (1-1/n))$ に従うことがわかる．なぜならば，

$$
\begin{aligned}
M_{X_i - \bar{X}_n}(t) &= \mathbb{E}\left[\exp\left(t(1-1/n)X_i - \sum_{j \neq i}(t/n)X_j\right)\right] \\
&= \mathbb{E}[e^{t(1-1/n)X_i}]\prod_{j \neq i}\mathbb{E}[e^{(t/n)X_j}] \\
&= \exp\left(\frac{t^2(1-(1/n))^2}{2}\right)\prod_{j \neq i}\exp\left(\frac{((t/n)^2}{2}\right) \\
&= \exp\left(\frac{(1-1/n)t^2}{2}\right)
\end{aligned}
$$

からわかる．$\bar{X}_n$ と $X_i - \bar{X}_n$ はともに正規分布に従うので，$\bar{X}_n$ と $X_i - \bar{X}_n$ が独立であることをいうためには，$\mathbb{COV}[\bar{X}_n, X_i - \bar{X}_n] = 0$ を示せばよい．しかし，

$$
\begin{aligned}
\mathbb{COV}[\bar{X}_n, X_i - \bar{X}_n] &= \mathbb{E}[\bar{X}_n(X_i - \bar{X}_n)] \\
&= \mathbb{E}[(1/n)\sum_{j=1}^n X_j X_i] - \mathbb{E}[\bar{X}_n^2] = \frac{1}{n}\sum_{j \neq i}\mathbb{E}[X_i X_j] + \frac{1}{n}\mathbb{E}[X_i^2] - \mathbb{E}[\bar{X}_n^2] \\
&= \frac{1}{n}\mathbb{VAR}[X_i] - \mathbb{VAR}[\bar{X}_n] = \frac{1}{n} - \frac{1}{n} = 0
\end{aligned}
$$

となる．よって，$\bar{X}_n$ と $X_i - \bar{X}_n$ が独立である．これから $\bar{X}_n$ と $(X_1 - \bar{X}_n, X_2 - \bar{X}_n, \ldots, X_n - \bar{X}_n)$ は独立(4-2) となり，$\bar{X}_n$ と S_n^2 は独立であることがわかる．

(3) を示すために，

$$
\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 + n(\bar{X}_n - \mu)^2
$$

に注意する．$Y_i = (X_i - \mu)/\sigma$, $i = 1, 2, \ldots, n$, とおけば，$\bar{Y}_n = (1/n)\sum_{i=1}^n Y_i = (\bar{X}_n - \mu)/\sigma$ となり，Y_i と $\bar{Y}_n$ は $N(0, 1)$ と $N(0, 1/n)$ に従う．しがたって，$U = \sum_{i=1}^n (Y_i - \bar{Y}_n)^2$ が自由度 $n-1$ のカイ自乗分布に従うことを示せばよい．いま，$W = \sum_{i=1}^n Y_i^2$, $V = n\bar{Y}_n^2$ とおけば，$t < 1/2$ に対して，W, V の積率母関数は

$$
\begin{aligned}
M_W(t) &= \mathbb{E}[\exp(tW)] = \left(\frac{1}{1-2t}\right)^n, \\
M_V(t) &= \mathbb{E}[\exp(tV)] = \left(\frac{1}{1-2t}\right)
\end{aligned}
$$

となる．さらに，$\bar{Y}_n$ と $(Y_1-\bar{Y}_n,\, Y_2-\bar{Y}_n,\, \ldots,\, Y_n-\bar{Y}_n)$ は独立であることに注意して，$U=\sum_{i=1}^n (Y_i-\bar{Y}_n)^2$ の積率母関数を求める：

$$\begin{aligned} M_W(t) &= \mathbb{E}[\exp(tW)] = \mathbb{E}[\exp\{t\sum_{i=1}^n (Y_i-\bar{Y}_n)^2\} + tn\bar{Y}_n^2] \\ &= \mathbb{E}[\exp\{tU+tV\}] = \mathbb{E}[\exp\{tU\}]\mathbb{E}[\exp(tV)] = M_U(t)M_V(t) \end{aligned}$$

より

$$M_U(t) = \frac{M_W(t)}{M_V(t)} = \left(\frac{1}{1-2t}\right)^{n-1}$$

がわかる． □

4.2.1 t 分布と F 分布

$X_1, X_2, \ldots, X_n$ が正規分布 $N(\mu, \sigma^2)$ からの標本の大きさが n のランダム標本としたとき，

$$\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \tag{4.6}$$

は標準正規分布 $N(0,1)$ に従うことが定理 4.3 (2) からわかる． σ が既知であれば，$\bar{X}_n$ を観測したときに，(4.6) は μ の推測に利用できる．しかし，σ が未知のときは，(4.6) の代わりに

$$\frac{\bar{X}_n-\mu}{S_n/\sqrt{n}} \tag{4.7}$$

を μ の推測に用いる．ただし，$S_n^2=(n-1)^{-1}\sum_{i=1}^n (X_i-\bar{X}_n)^2$ で S_n は S_n^2 の正の平方根である．(4.7) の標本分布を求めるために

$$\frac{\bar{X}_n-\mu}{S_n/\sqrt{n}} = \frac{(\bar{X}_n-\mu)/(\sigma/\sqrt{n})}{\sqrt{S_n^2/\sigma^2}} \tag{4.8}$$

と書き換える．(4.8) の分子は標準正規分布 $N(0,1)$ に従い，分母は $\sqrt{\chi^2_{n-1}/(n-1)}$ と同じ分布で，分母と分子は独立である．ただし，χ^2_{n-1} は自由度 $(n-1)$ のカイ自乗分布である．したがって，(4.8) の分布は $V/\sqrt{U/(n-1)}$ と同じ分布である．ただし，U と V は独立に自由度 $(n-1)$ のカイ自乗分布と標準正規分布に従うものとする．

定義 4.3 p を自然数としたとき，確率変数 T が自由度 p の t 分布に従うとは，T が確率密度関数

$$f_T(t) = \frac{\Gamma((p+1)/2)}{\Gamma(p/2)}\frac{1}{\sqrt{p\pi}}\frac{1}{(1+t^2/p)^{(p+1)/2}}, \qquad -\infty < t < \infty$$

を持つときをいう．

命題 4.4 (t 分布の確率密度関数の導出について) $X_1, X_2, \ldots, X_n, n\geq 2,$ は正規分布 $N(\mu,\sigma^2)$ からの標本の大きさが n のランダム標本とするとき，

$$\frac{\bar{X}_n-\mu}{S/\sqrt{n}}$$

は自由度 $n-1$ の t 分布に従う．

証明

$$U=(n-1)\frac{S_n^2}{\sigma^2},\qquad V=\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}},\qquad p=n-1$$

とおくと U と V は独立で，それぞれは自由度 p のカイ自乗分布 χ_p^2 と標準正規分布 $N(0,1)$ に従い，

$$T=\frac{\bar{X}_n-\mu}{S_n/\sqrt{n}}=\frac{V}{\sqrt{U/p}}$$

となる．したがって，U と V から出発して，$\sqrt{p}V/U$ の確率密度関数を求める．まず，

$$f_{U,V}(u,v)=\frac{1}{\sqrt{2\pi}}e^{-v^2/2}\frac{1}{\Gamma(p/2)2^{p/2}}u^{(p/2)-1}e^{-u/2},\qquad -\infty<v<\infty,\quad 0<u<\infty$$

に注意する．いま

$$t=\frac{v}{\sqrt{u/p}},\qquad w=u$$

とおくと

$$J=\begin{vmatrix}(\partial v/\partial t) & (\partial v/\partial w)\\ (\partial u/\partial t) & (\partial u/\partial w)\end{vmatrix}=\begin{vmatrix}\sqrt{\frac{w}{p}} & *\\ 0 & 1\end{vmatrix}=\sqrt{\frac{w}{p}}$$

から

$$\begin{aligned}f_T(t) &= \int_0^\infty f_{U,V}\left(w,\,t\sqrt{\frac{w}{p}}\right)\sqrt{\frac{w}{p}}\,dw\\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\Gamma(p/2)2^{p/2}}\int_0^\infty e^{-(1/2)t^2w/p}w^{(p/2)-1}e\ -w/2\sqrt{\frac{w}{p}}\,dw\\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\Gamma(p/2)2^{p/2}p^{1/2}}\int_0^\infty e^{-(1/2)(1+t^2/p)w}w^{(p+1)/2-1}\,dw\end{aligned}$$

となる．さらに，

$$z=\left(1+\frac{t^2}{p}\right)w$$

とおけば，

$$\begin{aligned}f_T(t) &= \frac{1}{\Gamma(p/2)\sqrt{p\pi}}\frac{1}{(1+t^2/p)^{(p+1)/2}}\frac{1}{2^{(p+1)/2}}\int_0^\infty z^{(p+1)/2-1}e^{-w/2}\,dz\\ &= \frac{\Gamma((p+1)/2)}{\Gamma(p/2)}\frac{1}{\sqrt{p\pi}}\frac{1}{(1+t^2/p)^{(p+1)/2}}\end{aligned}$$

を得る． □

定義 4.4 p,q を自然数としたとき，確率変数 F が自由度 p と q の F 分布に従うとは，F が確率密度関数

$$f_F(x)=\frac{\Gamma((p+q)/2)}{\Gamma(p/2)\Gamma(q/2)}\left(\frac{p}{q}\right)^{p/2}\frac{x^{(p/2)-1}}{(1+(p/q)x)^{(p+q)/2}},\qquad 0<x<\infty$$

を持つときをいう．

命題 4.5 $n\geq 2$ と $m\geq 2$ とする．$X_1,X_2,\ldots,X_n$ は正規分布 $N(\mu_X,\sigma_X^2)$ からの標本の大きさが n のランダム標本とし，$Y_1,Y_2,\ldots,Y_m$ は $X_i,\ i=1,2,\ldots,n,$ とは独立な正規分布 $N(\mu_Y,\sigma_Y^2)$ からの標本の大きさが m のランダム標本とし，

$$S_X^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X}_n),\qquad \bar{X}_n=\frac{1}{n}\sum_{i=1}^n X_i,\qquad S_Y^2=\frac{1}{m-1}\sum_{i=1}^m(Y_i-\bar{Y}_m),\qquad \bar{Y}_m=\frac{1}{m}\sum_{i=1}^m Y_i,$$

とする．このとき，

$$F = \frac{(S_X^2/\sigma_X^2)}{(S_Y^2/\sigma_Y^2)}$$

は自由度 $n-1$ と $m-1$ の F 分布に従う．

証明　証明は略． □

4.3 順序統計量

定義 4.5 $X_1, X_2, \ldots, X_n$ をランダム標本としたとき，これを小さい順に並べかえたものを

$$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$$

を記し，これらを**順序統計量**という．すなわち，

$$\begin{aligned} X_{(1)} &= \min\{X_1, X_2, \ldots, X_n\}, \\ X_{(2)} &= X_1, X_2, \ldots, X_n \text{ の中で 2 番目に小さいもの}, \\ &\vdots \\ X_{(n)} &= \max\{X_1, X_2, \ldots, X_n\}, \end{aligned}$$

である．

命題 4.6 $X_1, X_2, \ldots, X_n$ を離散型分布 $f_X(x_i) = p_i, i = 1, 2, \ldots,$ からの標本の大きさが n のランダム標本とする．ただし，$x_1 < x_2 < \cdots$ は X の台(4-3) とする．さらに，$P_i = \sum_{j=1}^{i} p_j, i = 1, 2, \ldots, P_0 = 0$ とおく．$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ を標本の順序統計量としたとき，

$$P(X_{(j)} \leq x_i) = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1 - P_i)^{n-k}, \qquad j = 1, 2, \ldots, n, i = 1, 2, \ldots, \tag{4.9}$$

と

$$P(X_{(j)} = x_i) = \sum_{k=j}^{n} \binom{n}{k} [P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k}], \tag{4.10}$$

となる．

証明 $i \in \{1, 2, \ldots\}$ を固定し，Y を確率変数とし，

$$Y = \#\{X_j, j = 1, 2, \ldots, n | X_j \leq x_i\}$$

とする．ここで，確率変数 Z_j を $\{X_j \leq x_i\}$ が起こったとき，$Z_j = 1$，さもなければ，$Z_j = 0$ と定める．$X_1, X_2, \ldots, X_n$ は同一分布に従うので，各 j について，

$$\mathbb{P}(Z_j = 1) = P_i = \mathbb{P}(X_j \leq x_i)$$

である．また，$Z_1, Z_2, \ldots, Z_n$ は独立である．さらに，$Y = \sum_{j=1}^{n} Z_j$ に注意すれば，Y は母数 n, P_i の二項分布 $Bin(n, P_i)$ に従うことがわかる．

事象 $\{X_{(j)} \leq x_i\}$ は事象 $\{Y \geq j\}$ と同じなので，

$$\mathbb{P}(X_{(j)} \leq x_i) = \mathbb{P}(Y \geq j)$$

となり，(4.9) は示された．(4.10) を示すためには，

$$\mathbb{P}(X_{(j)} = x_i) = \mathbb{P}(X_{(j)} \leq x_i) - \mathbb{P}(X_{(j)} \leq x_{i-1})$$

を考えればよい．$i = 1$ の場合は $\mathbb{P}(X_{(j)} = x_i) = X_{(j)} \leq x_i)$ よりわかる．以上から (4.10) を示された． □

命題 4.7 $X_1, X_2, \ldots, X_n$ を分布関数 $F_X(x)$ と確率密度関数 $f_X(x)$ をもつ連続型分布からの標本の大きさが n のランダム標本とする．$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ を標本の順序統計量としたとき，$X_{(j)}, j = 1, 2, \ldots, n,$ の確率密度関数は

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x)[F_X(x)]^{j-1}[1 - F_X(x)]^{n-j}, \tag{4.11}$$

となる．

証明 $X_{(j)}$ の分布関数を求め，その導関数を計算することにより，$X_{(j)}$ の確率密度関数を求めることにする．

実数 x を固定する．Y を確率変数とし，

$$Y = \#\{X_j,\, j = 1, 2, \ldots, n |\, X_j \le x\}$$

とする．ここで，確率変数 Z_j を $\{X_j \le x\}$ が起こったとき，$Z_j = 1$，さもなければ，$Z_j = 0$ と定める．$X_1, X_2, \ldots, X_n$ は同一分布に従うので，各 j について，$\mathbb{P}(Z_j = 1) = F_X(x)$ である．また，$Z_1, Z_2, \ldots, Z_n$ は独立である．さらに，$Y = \sum_{j=1}^{n} Z_j$ に注意すれば，Y は母数 n, P_i の二項分布 $Bin(n, F_X(x))$ に従うことがわかる．これらより

$$F_{X_{(j)}}(x) = \mathbb{P}(Y \ge j) = \sum_{k=j}^{n} \binom{n}{k} [F_X(x)]^k [1 - F_X(x)]^{n-k}$$

となる．したがって，$X_{(j)}$ の確率密度関数は

$$\begin{aligned}
f_{X_{(j)}}(x) &= \frac{d}{dx} F_{X_{(j)}}(x) \\
&= \sum_{k=j}^{n} \binom{n}{k} \Big(k[F_X(x)]^{k-1}[1 - F_X(x)]^{n-k} f_X(x) \\
&\quad - (n-k)[F_X(x)]^k [1 - F_X(x)]^{n-k-1} f_X(x) \Big) \\
&= \binom{n}{j} j f_X(x)[F_X(x)]^{j-1}[1 - F_X(x)]^{n-j} \\
&\quad + \sum_{k=j+1}^{n} \binom{n}{k} k[F_X(x)]^{k-1}[1 - F_X(x)]^{n-k} f_X(x) \\
&\quad - \sum_{k=j}^{n-1} \binom{n}{k} (n-k)[F_X(x)]^k [1 - F_X(x)]^{n-k-1} f_X(x) \\
&= \frac{n!}{(j-1)!(n-j)!} f_X(x)[F_X(x)]^{j-1}[1 - F_X(x)]^{n-j} \\
&\quad + \sum_{k=j}^{n-1} \binom{n}{k+1} (k+1)[F_X(x)]^k [1 - F_X(x)]^{n-k-1} f_X(x) \\
&\quad - \sum_{k=j}^{n-1} \binom{n}{k} (n-k)[F_X(x)]^k [1 - F_X(x)]^{n-k-1} f_X(x)
\end{aligned}$$

となる．最後に，

$$\binom{n}{k+1}(k+1) = \frac{n!}{k!(n-k-1)!} = \binom{n}{k}(n-k)$$

から (4.11) は示される． □

例 4.4 $X_1, X_2, \ldots, X_n$ を $(0, 1)$ 上の一様分布からの標本の大きさが n のランダム標本とする．すなわち，

$$F_X(x) = \begin{cases} x, & 0 < x < 1, \\ 0, & x \leq 0, \\ 1, & x \geq 1 \end{cases},$$

$$f_X(x) = \begin{cases} 1, & 0 < x < 1, \\ 0, & \text{その他}, \end{cases}$$

である．このとき，(4.11) から

$$f_{X_{(j)}} = \begin{cases} \frac{n!}{(j-1)!(n-j)!} x^{j-1}(1-x)^{n-j}, & 0 < x < 1, \\ 0, & \text{その他} \end{cases}, \qquad j = 1, 2, \ldots, n$$

となる．また，補遺の代表的な広義積分 (iii) から

$$\begin{aligned} \mathbb{E}[X_{(j)}] &= \frac{n!}{(j-1)!(n-j)!} \int_0^1 x^j (1-x)^{n-j}, dx \\ &= \frac{n!}{(j-1)!(n-j)!} \frac{\Gamma(j+1)\Gamma(n-j+1)}{\Gamma(n+2)} \\ &= \frac{n!}{(j-1)!(n-j)!} \frac{j!(n-j)!}{(n+1)!} \\ &= \frac{j}{n+1} \end{aligned}$$

となる．

命題 4.8 $X_1, X_2, \ldots, X_n$ を分布関数 $F_X(x)$ と確率密度関数 $f_X(x)$ をもつ連続型分布からの標本の大きさが n のランダム標本とする．$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ を標本の順序統計量としたとき，$X_{(i)}$ と $X_{(j)}, 1 \leq i < j \leq n,$ の同時確率密度関数は

$$\begin{aligned} & f_{X_{(i)}, X_{(j)}}(u, v) \\ &= \begin{cases} \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(u) f_X(v) [F_X(u)]^{i-1} & \\ \qquad \times [F_X(v) - F_X(v)]^{j-i-1} [1 - F_X(x)]^{n-j}, & -\infty < u < v < \infty, \\ 0, & \text{その他} \end{cases} \end{aligned} \tag{4.12}$$

となる．

証明 証明は省略する． □

系 4.3 $X_{(1)}$ と $X_{(n)}$ の同時確率密度関数は

$$f_{X_{(1)}, X_{(n)}}(x_1, x_n) = \begin{cases} n(n-1)[F_X(x_n) - F_X(x_1)]^{n-2} f_X(x_1) f(x_n) & x_1 < x_n, \\ 0 & x_1 \geq x_n \end{cases}$$

で与えられる．

例 4.5 $X_1, X_2, \ldots, X_n, n \geq 2$ は独立同一に $(0, 1)$ 上の一様分布に従うとする．すなわち，各 $X_i, i = 1, 2, \ldots, n$ は確率密度関数と分布関数

$$f_X(x) = \begin{cases} 1 & 0 < x < 1, \\ 0 & \text{（その他）} \end{cases}, \qquad F_X(x) = \begin{cases} 0 & x \leq 0, \\ x & 0 < x < 1, \\ 1 & x \geq 1 \end{cases}$$

を持つ．定理 4.7 から

$$f_{X_{(j)}}(x) = \begin{cases} \frac{n!}{(j-1)!(n-j)!} x^{j-1}(1-x)^{n-1} & 0 < x < 1, \\ 0 & \text{その他）} \end{cases}$$

となる．

4.4 確率変数の列の収束について

以下では，特に断りがない限り $\{X_n\}_{n=1}^{\infty}$ を確率変数列とし，X を確率変数とし，これらは同一の確率空間上で定義されているとする．

定義 4.6 (確率収束)　$n \to \infty$ のとき，$\{X_n\}$ が X に**確率収束**するとは，任意の正数 ϵ に対して

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$$

をみたすときをいい，$X_n \xrightarrow{P} X$ と記す．

定義 4.7 (分布収束)　$\{X_n\}_{n=1}^{\infty}$, X は確率変数とし，$F_X(x)$ を X の分布関数とする．$n \to \infty$ のとき，$\{X_n\}$ が X に**分布収束**するとは，$F_X(x)$ の任意の連続点において，

$$\lim_{n\to\infty} P(X_n \le x) = F_X(x)$$

をみたすときをいい，$X_n \xrightarrow{d} X$ と記す．

注意 4.2　$X_n \xrightarrow{d} F_X(x)$ のように記すこともある．また，X が正規分布 $N(0, \sigma^2)$ に従うとき，$X_n \xrightarrow{d} N(0, ,\sigma^2)$ と記すこともある．また，$F_X(x)$ のことを X_n の極限分布という．

例 4.6　$X_1, X_2, \ldots, X_n$ は独立同一に $[0, 1]$ 上の一様分布に従うとし，

$$M_n = \max\{X_1, X_2, \ldots, X_n\}$$

とする．直感的には，$n \to \infty$ のとき，M_n は 1 に近づくことがわかるであろう．これはつぎのことから保障される．まず，M_n の分布関数は

$$F_{M_n}(x) = \begin{cases} 0 & (x < 0), \\ x^n & (0 \le x \le 1), \\ 1 & (x > 1), \end{cases}$$

と(4-4)なる．したがって，任意の正数 ϵ に対して，

$$\begin{aligned} P(|M_n - 1| > \epsilon) &= P(\{M_n < 1-\epsilon\} \cup \{M_n > 1+\epsilon\}) \\ &= P(M_n < 1-\epsilon) + P(M_n > 1+\epsilon) \\ &= P(M_n < 1-\epsilon) = (1-\epsilon)^n \to 0 \qquad (n \to \infty), \end{aligned}$$

がわかる(4-5)．

次に，$n(1 - M_n)$ の極限分布を求めよう：$x \ge 0$ に対して，

$$\begin{aligned} P(n(1-M_n) \le x) &= P(M_n \ge 1 - x/n) = 1 - P(M_n < 1 - x/n) \\ &= 1 - \left(1 - \frac{x}{n}\right)^n \\ &\to 1 - e^{-x} \qquad (n \to \infty), \end{aligned}$$

となる．また，$x < 0$ のときは $P(n|1 - M_n| \le x) = 0$ となる．したがって，

$$n|1 - M_n| \xrightarrow{d} F_X(x)$$

を得る．ただし，

$$F_X(x) = \begin{cases} 1 - e^{-x} & (x \ge 0), \\ 0 & (x < 0), \end{cases}$$

である．すなわち，母数 1 の指数分布に分布収束することがわかる．

以下では確率変数列の収束に関する重要な命題を証明するための補題である．

補題 4.2 $\{A_n\}_{n=1}^{\infty}$ と $\{B_n\}_{n=1}^{\infty}$ を事象の列とする．$n \uparrow \infty$ のとき，

$$P(A_n) \to 1, \qquad P(B_n) \to 1$$

ならば，

$$P(A_n \cap B_n) \to 1$$

が成立する．

証明 $P(A_n^c) = 1 - P(A_n) \to 0 (n \to \infty)$ に注意すれば，

$$P\{(A_n \cap B_n)^c\} = P(A_n^c \cup B_n^c) \le P(A_n^c) + P(B_n^c) \to 0, \quad (n \to \infty)$$

よりわかる． □

命題 4.9 $\{X_n\}_{n=1}^{\infty}$ と $\{Y_n\}_{n=1}^{\infty}$ を確率変数列で

$$X_n \xrightarrow{P} c, \qquad Y_n \xrightarrow{P} d, \qquad (n \to \infty)$$

を満足するとする．ただし，c と d は定数とする．このとき，

(i) $X_n \pm Y_n \xrightarrow{P} c \pm d$

(ii) $X_n Y_n \xrightarrow{P} cd$

(iii) $d \neq 0$ ならば，$X_n / Y_n \xrightarrow{P} c/d$

が成立する．

証明 (i) の証明．$|(X_n + Y_n) - (c + d)| \le |X_n - c| + |Y_n - d|$ から，どんな正の数 $\epsilon > 0$ に対しても

$$|X_n - c| < \frac{\epsilon}{2} \quad \text{かつ} \quad |Y_n - d| < \frac{\epsilon}{2} \tag{4.13}$$

ならば，

$$|(X_n + Y_n) - (c + d)| < \epsilon$$

であるので

$$\{|X_n - c| < \frac{\epsilon}{2}\} \cap \{|Y_n - d| < \frac{\epsilon}{2}\} \subset \{|(X_n + Y_n) - (c + d)| < \epsilon\}$$

より，$n \uparrow \infty$ のとき，

$$P\{|(X_n + Y_n) - (c + d)| < \epsilon\} \ge P[\{|X_n - c| < \frac{\epsilon}{2}\} \cap \{|Y_n - d| < \frac{\epsilon}{2}\}] \to 1$$

となる(4-6)．なぜならば，$P\{|X_n - a| < \frac{\epsilon}{2}\} \to 1$ と $P\{|Y_n - b| < \frac{\epsilon}{2}\} \to 1$ から補題 4.2 を用いればわかる．
(ii) の証明．$X_n Y_n - cd = (X_n - c)(Y_n - d) + d(X_n - c) + c(Y_n - d)$ に注意する．どんな正の数 $\epsilon > 0$ に対しても

$$\begin{aligned}\mathbb{P}\{|X_n Y_n - cd| \ge \epsilon\} \quad &\le \quad \mathbb{P}\{|(X_n - c)(Y_n - d)| \ge \frac{\epsilon}{3}\} + \mathbb{P}\{|(X_n - c)| \ge \frac{\epsilon}{3|d|}\} \\ &\quad + \mathbb{P}\{|(Y_n - d)| \ge \frac{\epsilon}{3|c|}\}\end{aligned}$$

となる．どんな正の数 $\delta > 0$ に対しても

$$\begin{aligned}\mathbb{P}\{|(X_n - c)(Y_n - d)| \geq \frac{\epsilon}{3}\} \quad &= \quad \mathbb{P}\{|(X_n - c)(Y_n - d)| \geq \frac{\epsilon}{3} \text{ かつ } |Y_n - d| \geq \delta\} \\ &\quad + \mathbb{P}\{|(X_n - c)(Y_n - d)| \geq \frac{\epsilon}{3} \text{ かつ } |Y_n - d| < \delta\} \\ &\leq \mathbb{P}\{|Y_n - d| \geq \delta\} + \mathbb{P}\{|(X_n - c)(Y_n - d)| \geq \frac{\epsilon}{3} \text{ かつ } |Y_n - d| < \delta\} \\ &\leq \mathbb{P}\{|Y_n - d| \geq \delta\} + \mathbb{P}\{|X_n - c| \geq \frac{\epsilon}{3\delta}\} \to 0, \qquad (n \to \infty)\end{aligned}$$

となることからわかる．

(iii) の証明．$1/Y_n \xrightarrow{P} 1/d$ を示せば (ii) よりわかる．十分小さな正の数 $\delta > 0$ に対して，$|Y_n - d| \leq \delta$ ならば，$|Y_n| \geq (1/2)|d|$ より

$$\mathbb{P}\{|Y_n| \geq \frac{1}{2}|d|\} \geq \mathbb{P}\{|Y_n - d| \leq \delta\} \to 1, \qquad (n \to \infty)$$

となる．また，$|Y_n| \geq (1/2)|d|$ のとき，

$$\left|\frac{1}{Y_n} - \frac{1}{d}\right| \leq \frac{|Y_n - d|}{|Y_n||d|} \leq \frac{2}{|d|^2}|Y_n - d|$$

が成立する．これらを用いれば，

$$\begin{aligned}\mathbb{P}\left\{\left|\frac{1}{Y_n} - \frac{1}{d}\right| \geq \epsilon\right\} \quad &= \quad \mathbb{P}\left\{\left|\frac{1}{Y_n} - \frac{1}{d}\right| \geq \epsilon, |Y_n| \geq \frac{1}{2}|d|\right\} + \mathbb{P}\left\{\left|\frac{1}{Y_n} - \frac{1}{d}\right| \geq \epsilon, |Y_n| < \frac{1}{2}|d|\right\} \\ &\leq \quad \mathbb{P}\left\{|Y_n - d| \geq \frac{|d|^2}{2}\epsilon\right\} + \mathbb{P}\left\{|Y_n| < \frac{1}{2}|d|\right\} \\ &\to \quad 0, \qquad (n \to \infty)\end{aligned}$$

より示せた． □

命題 4.10 （連続写像定理）: g を実数値連続関数とする．このとき，$Y_n \xrightarrow{P} b$（b は定数）ならば，$n \uparrow \infty$ のとき，

$$g(Y_n) \xrightarrow{P} g(b)$$

が成立する．

証明 どんな正の数 $\epsilon > 0$ に対してもある正の数 $\delta > 0$ が存在して，

$$|Y_n - b| \leq \delta \quad \text{ならば} \quad |g(Y_n) - g(b)| \leq \epsilon$$

を満足するので，

$$P\{|Y_n - b| \leq \delta\} \leq P\{|g(Y_n) - g(b)| \leq \epsilon\}$$

より，$n \uparrow \infty$ のとき，

$$P\{|g(Y_n) - g(b)| > \epsilon\} \leq P\{|Y_n - b| > \delta\} \to 0$$

より命題は示せた． □

命題 4.11 （Slutsky の定理）$\{X_n\}_{n=1}^{\infty}$，$\{A_n\}_{n=1}^{\infty}$，$\{B_n\}_{n=1}^{\infty}$ を確率変数列とし $X_n \xrightarrow{d} X$ ，$A_n \xrightarrow{P} a$，$B_n \xrightarrow{P} b$ を満足するとする．ただし，X は確率変数，a と b は定数とする．このとき，

$$A_n + B_n X_n \xrightarrow{d} a + bX$$

が成立する．

証明 まず,

$$A_n + X_n \xrightarrow{L} a + X$$

を示す．どんな正の数 $\epsilon > 0$ に対しても

$$\begin{aligned} F_{(A_n+X_n)}(x) &= \mathbb{P}\{A_n + X_n \le x\} \\ &= \mathbb{P}\{A_n + X_n \le x, A_n \ge a-\epsilon\} + \mathbb{P}\{A_n + X_n \le x, A_n < a-\epsilon\} \end{aligned} \tag{4.14}$$

となることに注意する．十分小さな ϵ ととり，$x \pm \epsilon$ が $F_{(a+X)}(\cdot) = \mathbb{P}\{a + X \le \cdot)$ の連続点になるようにとる[4-7]．(4.14) から

$$F_{(A_n+X_n)}(x) \le \mathbb{P}\{X_n \le x - a + \epsilon\} + \mathbb{P}\{|A_n - a| > \epsilon\}$$

となる．また，$F_{X+a}(\cdot)$ の連続点 x に対して，

$$\begin{aligned} F_{(X_n+a)}(x) &= \mathbb{P}\{X_n + a \le x) = F_{X_n}(x-a) \\ &\to F_X(x-a) = P(X \le x-a) = F_{X+a}(x) \end{aligned}$$

となる ことより，$X_n + a \xrightarrow{d} X + a$ が成り立つ．したがって，

$$\begin{aligned} \limsup_{n\to\infty} F_{(X_n+A_n)}(x) &\le \lim_{n\to\infty} \{\mathbb{P}(X_n \le x - a + \epsilon) + \mathbb{P}(|A_n - a| > \epsilon)\} \\ &= \lim_{n\to\infty} F_{(X_n+a)}(x+\epsilon) + \lim_{n\to\infty} \mathbb{P}(|A_n - a| > \epsilon) \\ &= F_{(X+a)}(x+\epsilon) \end{aligned} \tag{4.15}$$

となる．また，

$$\begin{aligned} 1 - F_{(X_n+A_n)}(x) &= \mathbb{P}\{X_n + A_n > x\} \\ &\le \mathbb{P}\{X_n > x - a - \epsilon\} + \mathbb{P}\{|A_n - a| \ge \epsilon\} \end{aligned}$$

から

$$\liminf_{n\to\infty} F_{(X_n+A_n)}(x) \ge \lim_{n\to\infty} \{F_{(X_n+a)}(x-\epsilon) + \mathbb{P}(|A_n - a| > \epsilon)\} = F_{X+a}(x-\epsilon) \tag{4.16}$$

となる．(4.15) と (4.16) から

$$F_{(X+a)}(x-\epsilon) \le \liminf_{n\to\infty} F_{(X_n+A_n)}(x) \le \limsup_{n\to\infty} F_{(X_n+A_n)}(x) \le F_{(X+a)}(x+\epsilon)$$

を得る．したがって，

$$\lim_{n\to\infty} F_{(X_n+A_n)}(x) = F_{(X+a)}(x)$$

が成り立つ．

つぎに，一般性を失わずに $b = 1$ として，$B_n X_n \xrightarrow{d} X$ を示せば，命題は示される．どんな正の数 $\epsilon > 0$ に対しても

$$\begin{aligned} F_{B_nX_n}(x) &= \mathbb{P}\{B_nX_n \le x\} \\ &= P\left\{B_nX_n \le x, \left|\frac{1}{B_n} - 1\right| \le \frac{\epsilon}{|x|}\right\} + \mathbb{P}\left\{B_nX_n \le x, \left|\frac{1}{B_n} - 1\right| > \frac{\epsilon}{|x|}\right\} \\ &\le \mathbb{P}\{X_n \le x + \epsilon\} + \mathbb{P}\left\{\left|\frac{1}{B_n} - 1\right| > \frac{\epsilon}{|x|}\right\} \end{aligned}$$

より，$n \uparrow \infty$ のとき，

$$\limsup_{n\to\infty} F_{B_nX_n}(x) \le \lim_{n\to\infty}\left[\mathbb{P}\{X_n \le x+\epsilon\} + \mathbb{P}\left\{\left|\frac{1}{B_n}-1\right| > \frac{\epsilon}{|x|}\right\}\right] \to F_X(x+\epsilon)$$

を得る．同様な議論により，

$$\liminf_{n\to\infty} F_{B_nX_n}(x) \ge F_X(x-\epsilon)$$

を得る．したがって，

$$F_X(x-\epsilon) \le \liminf_{n\to\infty} F_{B_nX_n}(x) \le \limsup_{n\to\infty} F_{B_nX_n}(x) \le F_X(x+\epsilon)$$

が成り立つので，

$$F_{B_nX_n}(x) \to F_X(x)$$

を得る．よって，命題は示された． □

命題 4.12 $X_n \xrightarrow{P} X$ ならば，$X_n \xrightarrow{d} X$ が成立する．

証明：$A_n = X_n - X$ とおく．条件より，$A_n \xrightarrow{P} 0$ となり，$X_n = X + A_n$ に対して，Slutsky の定理を用いれば，この命題は示される． □

注意 4.3 上の命題の逆は一般には成立しないが，c をある定数とすると，

$$X_n \xrightarrow{d} c \quad \Leftrightarrow \quad X_n \xrightarrow{P} c$$

である．実際，X_n と X の分布関数を H_n と H かけば，すべての $\epsilon > 0$ に対して，

$$\lim_{n\to\infty} H_n(c-\epsilon-) = H(c-\epsilon-) = 0$$

と

$$\lim_{n\to\infty} H_n(c+\epsilon) = H(c+\epsilon) = 1$$

から

$$\begin{aligned} P(|X_n - c| > \epsilon) &= P(X_n > c+\epsilon) + P(X_n < c-\epsilon) \\ &= 1 - H_n(c+\epsilon-) + H_n(c-\epsilon) \to 0 \end{aligned}$$

となる．

命題 4.13 $X_n \xrightarrow{d} X$ となるための必要十分条件は，すべての有界連続な関数 f に対し，

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$$

が成立することである．

証明 どんな x に対しても，ある $A > 0$ が存在して $|f(x)| \le A$ とできる．また，どんな $\epsilon > 0$ に対しても，ある $B > 0$ とある正の整数 n_0 が存在して，どんな $n \ge n_0$ に対しても $P(|X_n| \ge B) \le \epsilon/(3A)$ ともできる．h を実数値関数とし，$0 \le h(x) \le 1$ でかつ

$$h(x) = \begin{cases} 0 & (\ |x| > B+1\ ) \\ 1 & (\ |x| \le B+1\ ) \end{cases}$$

を満足するものとする．このとき，

$$
\begin{aligned}
|\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &\le |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X_n)h(X_n)]| \\
&\quad + |\mathbb{E}[f(X_n)h(X_n)] - \mathbb{E}[f(X)h(X)]| + |\mathbb{E}[f(X)h(X)] - \mathbb{E}[f(X)]| \\
&\le \frac{\epsilon}{3} + |\mathbb{E}[f(X_n)h(X_n)] - \mathbb{E}[f(X)h(X)]|
\end{aligned}
$$

となる．したがって，コンパクトな台 I をもつ連続関数 g に対して，

$$
|\mathbb{E}[g(X_n)] - \mathbb{E}[g(X)]| < \frac{\epsilon}{3}
$$

が成立することを示せばよい．I はコンパクトで，g は一様連続なので，有限個の矩形領域 I_j でその上での g の変動が $\epsilon/12$ 以下になり，$I \subset \cup_j I_j$ とできる．それぞれの I_j から一点 x_j を取り出し，$g_\epsilon(x) = \sum_j g(x_j)\mathbb{I}_{I_j}(x)$ とする．すると，

$$
\begin{aligned}
|\mathbb{E}[g(X_n)] - \mathbb{E}[g_\epsilon(X_n)]| \le \frac{\epsilon}{12} + \mathbb{P}(X_n \notin I) \le \frac{\epsilon}{6} \\
|\mathbb{E}[g(X)] - \mathbb{E}[g_\epsilon(X)]| \le \frac{\epsilon}{12} + \mathbb{P}(X \notin I) \le \frac{\epsilon}{6}
\end{aligned}
$$

となる[(4-8)]．さらに，n を十分おおきくとれば，

$$
|\mathbb{E}[g_\epsilon(X_n)] - \mathbb{E}[g_\epsilon(X)]| \le \sum_j |\mathbb{P}(X_n \in I_j) - \mathbb{P}(X \in I_j)||g(x_j)| \le \frac{\epsilon}{6}
$$

とできることからわかる． □

命題 4.14 $\{X_n\}_{n=1}^{\infty}$, X は確率変数とし，$g(x)$ を実数値連続関数とする．このとき，つぎが成立する．

$$
X_n \xrightarrow{d} X \qquad \text{ならば，} \qquad g(X_n) \xrightarrow{d} g(X)
$$

.

証明 命題 4.13 から任意の有界連続関数 f に対して，$n \to \infty$ のとき，

$$
\mathbb{E}[f(g(X_n))] \to \mathbb{E}[f(g(X))]
$$

を示せばよい．$f \circ g$ も有界連続関数であることと仮定より上の式は明らか． □

系 4.4 $\{X_n\}_{n=1}^{\infty}$, $\{Y_n\}_{n=1}^{\infty}$ は確率変数とし，$X_n \xrightarrow{d} c$ かつ $Y_n \xrightarrow{d} d$ とする．ただし，c, d は定数である．このとき，つぎが成立する．

(1) $X_n + Y_n \xrightarrow{d} c + d$.

(2) $X_n Y_n \xrightarrow{d} cd$.

証明 Slutzky の定理より明らか． □

命題 4.15 (デルタ法) $\{X_n\}_{n=1}^{\infty}$, Z は確率変数，θ を定数とする．$n \to \infty$ のとき，

$$
\sqrt{n}(X_n - \theta) \xrightarrow{d} Z
$$

が成立すると仮定する．実数値関数 $g(x)$ は $x = \theta$ で微分可能で微係数 $\dot{g}(\theta)$ を持ち，$\dot{g}(\theta) \neq 0$ ならば，

$$
\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{d} \dot{g}(\theta) Z
$$

が成立する．

証明　まず，仮定と命題 4.11 から

$$X_n - \theta = \frac{1}{\sqrt{n}}\sqrt{n}(X_n - \theta) \xrightarrow{d} 0$$

となる．さらに，注意 4.3 から $X_n - \theta \xrightarrow{P} 0$ となる．ここで $g(X_n)$ を $X_n = \theta$ のまわりでテーラー展開する：

$$\sqrt{n}(g(X_n) - g(\theta)) = \dot{g}(\theta)\sqrt{n}(X_n - \theta) + \sqrt{n}\mathrm{Rem} \tag{4.17}$$

ここで

$$\lim_{X_n \to \theta} \frac{\mathrm{Rem}}{|X_n - \theta|} = 0$$

である．これと仮定から

$$\frac{\mathrm{Rem}}{|X_n - \theta|} \xrightarrow{P} 0$$

となる．また，$\sqrt{n}(X_n - \theta)$ は分布収束することと上のことに注意して，再度命題 4.13 を用いると

$$\sqrt{n}\mathrm{Rem} = \sqrt{n}(X_n - \theta)\frac{\mathrm{Rem}}{|X_n - \theta|} \xrightarrow{d} 0$$

となり，注意 4.3 から $\sqrt{n}\mathrm{Rem} \xrightarrow{P} 0$ を得る．(4.17) に命題 4.13 を適用すれば命題は証明される．　□

例 4.7　$\sqrt{n}(X_n - \mu) \xrightarrow{d} X$ とし，X は正規分布 $N(0, \sigma^2)$ に従うとする．ここで，$\mu \neq 0, \sigma^2 > 0$ を仮定する．このとき，

$$\sqrt{n}\left(\frac{1}{X_n} - \frac{1}{\mu}\right) \xrightarrow{d} -\frac{1}{\mu^2}X$$

となる．さらに，正規分布の性質から $-(1/\mu^2)X$ は正規分布 $N(0, \sigma^2/\mu^4)$ に従うことがわかる．

4.5 大数の法則と中心極限定理

定理 4.1 (大数の (弱) 法則) $X_1, X_2, \ldots$ は独立同一分布に従う確率変数列で $\mathbb{E}[|X_1|] < \infty$ する．$n \to \infty$ のとき，

$$\bar{X}_n = \frac{1}{n}\sum_{n=1}^{n} X_i \xrightarrow{P} \mu \tag{4.18}$$

が成立する．ただし，$\mathbb{E}[X_1] = \mu$ である．

証明 証明は略. □

注意 4.4 $X_1, X_2, \ldots$ は独立同一分布に従う確率変数列で $\mathbb{E}[X_1^2] < \infty$ する．いま，$\mathbb{E}[X_1] \leq \mu$, $\mathbb{VAR}[X_1] = \sigma^2 < \infty$ と書くことにする．このとき，任意の正数 ϵ に対して，

$$\mathbb{P}(|\bar{X}_n - \mu| > \epsilon) = \frac{1}{\epsilon^2}\mathbb{E}[(\bar{X}_n - \mu)^2] = \frac{1}{\epsilon^2}\mathbb{VAR}[\bar{X}_n] = \frac{1}{\epsilon}\frac{1}{n}\mathbb{VAR}[X_1] \to 0 \qquad (n \to \infty)$$

を得る．ただし，上の不等号はチェビシェフの不等式を利用した．$\mathbb{E}[X_1^2] < \infty$ を仮定すれば，チェビチェフの不等式より (4.18) はわかるが，大数の弱法則は $\mathbb{E}[|X_1|] < \infty$ ならば，(4.18) が成立することを主張している．

例 4.8 $X_1, X_2, \ldots$ は独立同一分布に従う確率変数列で $\mathbb{E}[X_1] = \mu$, $\mathbb{VAR}[X_1] = \sigma^2 < \infty$ とする．このとき，

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \xrightarrow{P} \sigma^2$$

を示そう．まず，

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2\right)$$

に注意する．$\mathbb{E}[|X_1|] < \infty$ から大数の弱法則を用いれば，$\bar{X}_n \xrightarrow{P} \mu$ になる．さらに，定理 4.10 から $\bar{X}_n^2 \xrightarrow{P} \mu^2$ を得る．また，$\mathbb{E}[X_1^2] = \mathbb{VAR}[X_1] + \{\mathbb{E}[X_1]\}^2 < \infty$ から大数の弱法則を用いれば，

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \mu\mathbb{E}[X_1^2]$$

となる．最後に，系 4.4 から $n \to \infty$ のとき，

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \xrightarrow{P} \mathbb{E}[X_1^2] - \mu^2 = \mathbb{VAR}[X_1]$$

を得る．

定理 4.2 (中心極限定理) $X_1, X_2, \ldots$ は独立同一分布に従う確率変数列で $\mathbb{E}[X_1] = \mu$, $\mathbb{VAR}[X_1] = \sigma^2 < \infty$ とし，

$$Z_n = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu) = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

とする．$n \to \infty$ のとき，

$$Z_n \xrightarrow{d} N(0, 1)$$

が成立する．すなわち，任意の実数 x に対し，

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

が成立する．

証明 証明は略. □

例 4.9 $X_1, X_2, \ldots, X_n$ は独立同一に母数 p, $0 < p < 1$ のベルヌーイ分布に従うとする．すると $T_n = \sum_{i=1}^{n} X_i$ は母数 n と p の二項分布に従う．したがって，$t_1 < t_2$ に対して，

$$\mathbb{P}(t_1 \le T_n \le t_2) = \sum_{x=t_1}^{t_2} f_{T_n}(x)$$

となる．ただし，

$$f_{T_n}(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0,\, 1,\, 2, \ldots, n$$

である．この確率の近似値を中心極限定理を利用して求めよう．$\mathbb{E}[X_1] = p$, $\mathbb{VAR}[X_1] = p(1-p)$ と中心極限定理から

$$\mathbb{P}\left(\frac{\sqrt{n}((1/n)T_n - p)}{\sqrt{p(1-p)}} \le x\right) = \mathbb{P}\left(\frac{T_n - np}{np(1-p)} \le x\right) \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

となる．これより

$$\begin{aligned}
\mathbb{P}(t_1 \le T_n \le t_2) &= \mathbb{P}(t_1 - (1/2) < T_n \le t_2 + (1/2)) \\
&= \mathbb{P}(T_n \le t_2 + (1/2)) - \mathbb{P}(T_n \le t_1 - (1/2)) \\
&= \mathbb{P}\left(\frac{T_n - np}{\sqrt{np(1-p)}} \le \frac{t_2 - np + (1/2)}{\sqrt{np(1-p)}}\right) - \mathbb{P}\left(\frac{T_n - np)}{\sqrt{p(1-p)}} \le \frac{t_2 - np - (1/2)}{\sqrt{np(1-p)}}\right) \\
&\approx \Phi\left(\frac{t_2 - np + (1/2)}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{t_1 - np - (1/2)}{\sqrt{np(1-p)}}\right)
\end{aligned}$$

となる．ただし，

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

とした．

いま，$n = 25$ と $p = 0.2$ と $t_1 = 3, t_2 = 5$ とし，$\mathbb{P}(3 \le T_{25} \le 5)$ の近似値を求めよう：

$$\begin{aligned}
\mathbb{P}(3 \le T_{25} \le 5) &= \mathbb{P}(2.5 < T_{25} \le 5.5) \\
&\approx \Phi\left(\frac{5.5 - 5}{2}\right) - \Phi\left(\frac{2.5 - 5}{2}\right) \\
&= \Phi(0.25) - \Phi(-1.25) \fallingdotseq 0.599 - 0.106 \\
&= 0.493
\end{aligned}$$

となる．

第 5 章　データの縮約

5.1　十分統計量

$X_1, X_2, \ldots, X_n$ を確率 (密度) 関数 $f(x|\theta)$ を持つ分布からのランダム標本とする．ただし，θ は母数空間 $\Theta \subset \mathbb{R}^k$ (k は自然数) の元とする．また，$T = T(X_1, X_2, \ldots, X_n)$ を統計量とする．統計量は n 個の確率変数の「情報」をひとつの確率変数に縮約している．縮約をするときに，どのような「情報」が失われているかを知りたい．別のいい方をすれば，推測に必要な「情報」が失われていない保障があるかどうかを知りたい．このようなことをどのように定式化するかをここでは考えていく．

$\mathcal{X}$ を標本空間とし，T を統計量とすれば，T は標本空間を分割していると考えることができる．このことを簡単な例でみてみよう．

例 5.1　X_1, X_2, X_3 はベルヌーイ試行 $Bi(1, p), 0 < p < 1$ からのランダム標本とする．したがって，標本空間は

$$\mathcal{X} = \{(0,0,0), (0,0,1), (0,1,0), (1,0,0), (0,1,1), (1,0,1), (1,1,0), (1,1,1)\}$$

となる．$S(x_1, x_2, x_3) = x_1 + x_2 + x_3$ とすれば，$S(\cdot)$ の値域は $0, 1, 2, 3$ となり，$\mathcal{X}$ を分割している：

$S(\cdot)$ の値	標本点
0	(0, 0, 0)
1	(0, 0, 1), (0, 1, 0), (1, 0, 0)
2	(0, 1, 1), (1, 0, 1), (1, 1, 0)
3	(1, 1, 1)

たとえば，$S = 1$ とすれば,「1」 が 1 回出現したが，何回目に出現したかという「情報」は失われる．

つぎに，$T = T(X_1, X_2, X_3) = X_1X_2 + X_3$ という統計量を考えて，S と T をそれぞれ与えたときの (X_1, X_2, X_3) の条件付き確率分布をしらべてみよう．

標本点	T の値	$f_{X_1, X_2, X_3 \mid T}(\cdot)$	S 値	$f_{X_1, X_2, X_3 \mid S}(\cdot)$
(0, 0, 0)	0	$\frac{1-p}{1+p}$	0	1
(0, 0, 1)	1	$\frac{1-p}{1+2p}$	1	$\frac{1}{3}$
(0, 1, 0)	0	$\frac{p}{1+p}$	1	$\frac{1}{3}$
(1, 0, 0)	0	$\frac{p}{1+p}$	1	$\frac{1}{3}$
(0, 1, 1)	1	$\frac{p}{1+2p}$	2	$\frac{1}{3}$
(1, 0, 1)	1	$\frac{p}{1+2p}$	2	$\frac{1}{3}$
(1, 1, 0)	1	$\frac{p}{1+2p}$	2	$\frac{1}{3}$
(1, 1, 1)	2	1	3	1

表から S を与えたときの (X_1, X_2, X_3) の条件付き分布は p には依存しないが，T を与えたときの (X_1, X_2, X_3) の条件付き分布は p には依存ことがわかる．

5.1.1 十分統計量の定義

定義 5.1 $X_1, X_2, \ldots, X_n$ は確率 (密度) 関数 $f(x|\theta)$ を持つ分布からの標本の大きさ n のランダム標本とする．ただし，$\theta \in \Theta$ とする．統計量 $S = S(X_1, X_2, \ldots, X_n)$ が θ の十分統計量であるとは，$S = s$ を与えたときの $(X_1, X_2, \ldots, X_n)$ の条件付き分布がどんな s の値に対しても θ に依存しないときをいう．

例 5.2 X_1, X_2 は正規分布 $N(\theta, 1)$ からの標本の大きさ 2 のランダム標本とする．$S = S(X_1, X_2) = X_1 + X_2$ は十分統計量であることを示そう．そのために，変換

$$\begin{cases} S = X_1 + X_2 \\ R = X_1 - X_2 \end{cases}$$

を考える．(S, R) と (X_1, X_2) は一対一対応である．また，正規分布の性質から S と R は独立となるので，

$$f_{R|S}(r|s) = \frac{f_{R,S}(r, s)}{f_S(s)} = \frac{f_R(r)f_S(s)}{f_S(s)} = f_R(r)$$

となる．しかし，

$$f_R(r) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2}}e^{-r^2/4}$$

となるので，S を与えたときの R の条件付き分布は θ に依存しない．さらに，(S, R) と (X_1, X_2) は一対一対応なので，S を与えたときの (X_1, X_2) の条件付き分布は θ に依存しないことがわかる．

統計量が複数あるときも十分性を拡張して定義できる．

定義 5.2 $X_1, X_2, \ldots, X_n$ は確率（密度）関数 $f(x|\theta), \theta \in \Theta$ を持つ分布からの標本の大きさ n のランダム標本とする．統計量 $S_1 = S_1(X_1, X_2, \ldots, X_n), S_2 = S_2(X_1, X_2, \ldots, X_n), \ldots, S_\ell = S_\ell(X_1, X_2, \ldots, X_n)$ が十分統計量であるとは，$S_1 = s_1, S_2 = s_2, \ldots, S_\ell = s_\ell$ を与えたときの $(X_1, X_2, \ldots, X_n)$ の条件付き分布がどんな $s_1, s_2, \ldots, s_\ell$ の値に対しても θ に依存しないときをいう．

例 5.3 ：$X_1, X_2, \ldots, X_n$ をベルヌーイ試行

$$f_\theta(z) = \theta^z(1-\theta)^{1-z}\,\mathbb{I}_{\{0,1\}}(z), \qquad z \in \mathbb{R}, \qquad \theta \in (0, 1)$$

からの標本の大きさ n のランダム標本とし，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ とする．$S(\boldsymbol{X}) = \sum_{i=1}^n X_i$ なる統計量は，直観から θ に対する情報をすべて含んでいると予想される．したがって，十分統計量となることが予想される．これを示すために，$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ とし，

$$\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}|S = s) = \frac{\mathbb{P}(\boldsymbol{X} = \boldsymbol{x}, S = s)}{\mathbb{P}(S = s)}$$

が $\theta \in (0, 1)$ に依存しないことを示せばよい．

$$\mathbb{P}(S = s) = \begin{pmatrix} n \\ s \end{pmatrix}\theta^s(1-\theta)^{n-s}\,\mathbb{I}_{\{0,1,\ldots,n\}}(s)$$

と

$$\begin{aligned} \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}, S = s) &= \prod_{i=1}^n \mathbb{P}(X_i = x_i) \\ &= \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}\,\mathbb{I}_{\{0,1\}}(x_i) \\ &= \theta^s(1-\theta)^{n-s}\prod_{i=1}^n \mathbb{I}_{\{0,1\}}(x_i) \end{aligned}$$

となることに注意する．$B_s = \{(x_1, x_2, \ldots, x_n) : x_i = 0, 1, \sum_{i=1}^n x_i = s\}$ とおけば，

$$\mathbb{P}(\boldsymbol{X} = \boldsymbol{x} | S = s) = \frac{1}{\begin{pmatrix} n \\ s \end{pmatrix}} \mathbb{I}_{B_s}(\boldsymbol{x})$$

となり，θ に依存しないので，$S(\boldsymbol{X})$ は $\theta \in (0, 1)$ に対する十分統計量である．

5.1.2 分解定理

定義に従って十分統計量をみつけるにはあらかじめ十分統計量と思われるものが事前にわかっていなければならない．つぎの定理を用いると比較的容易に十分統計量を見つけることができる．

命題 5.1 σ-有限な測度 ν によって優越される $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の確率分布族を $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ とし，$\boldsymbol{X}$ を $\mathbb{P}_\theta$ からのランダム標本とする．このとき，$S(\boldsymbol{X})$ が $\theta \in \Theta$ に対して十分であるための必要十分条件は $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の非負のボレロ可測関数 h と S の値域上の関数 g_θ （P に依存）が存在し，

$$\frac{dP_\theta}{d\nu}(\boldsymbol{x}) = g_\theta(S(\boldsymbol{x}))h(\boldsymbol{x}) \tag{5.1}$$

と書けることである．

証明 まず，証明は離散型分布の場合は比較的簡単であるので，その場合にについて証明を与える．σ-有限な測度 ν によって優越される $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ 上の確率分布族に対する証明は後で示す．

はじめに S が十分であると仮定する．このとき，

$$\begin{aligned} \mathbb{P}_\theta(\boldsymbol{X} = \boldsymbol{x}) &= \sum_t \mathbb{P}_\theta(\boldsymbol{X} = \boldsymbol{x}, S = s) \\ &= \mathbb{P}_\theta(\boldsymbol{X} = \boldsymbol{x}, S = S(\boldsymbol{x})) \\ &= \mathbb{P}_\theta(S = S(\boldsymbol{x}))\mathbb{P}_\theta(\boldsymbol{X} = \boldsymbol{x} | S = S(\boldsymbol{x})) \end{aligned}$$

となる[(5-1)]．S が十分統計量であることから $\mathbb{P}_\theta(\boldsymbol{X} = \boldsymbol{x} | S = S(\boldsymbol{x}))$ は θ に依存しないので，

$$\begin{aligned} g_\theta(S(\boldsymbol{x})) &= P_\theta(S = S(\boldsymbol{x})) \\ h(\boldsymbol{x}) &= \mathbb{P}(\boldsymbol{X} = \boldsymbol{x} | S = S(\boldsymbol{x})) \end{aligned}$$

とおけばよい．

つぎに，(5.1) が成立すると仮定する．$s = S(\boldsymbol{x})$ とおく．このとき，

$$\begin{aligned} \mathbb{P}_\theta(S = s) &= \sum_{\boldsymbol{y} : S(\boldsymbol{y}) = s} \mathbb{P}_\theta(\boldsymbol{X} = \boldsymbol{y}) \\ &= \sum_{\boldsymbol{y} : S(\boldsymbol{y}) = s} g_\theta(S(\boldsymbol{y}))h(\boldsymbol{y}) \\ &= g_\theta(s) \sum_{\boldsymbol{y} : S(\boldsymbol{y}) = s} h(\boldsymbol{y}) \end{aligned}$$

となる．従って，

$$\begin{aligned} \mathbb{P}_\theta(\boldsymbol{X} = \boldsymbol{x} | S = s) &= \frac{\mathbb{P}_\theta(\boldsymbol{X} = \boldsymbol{x}, S = s)}{\mathbb{P}_\theta(S = s)} \\ &= \frac{g_\theta(s)h(\boldsymbol{x})}{g_\theta(s)\sum_{\boldsymbol{y} : S(\boldsymbol{y}) = s} h(\boldsymbol{y})} \\ &= \frac{h(\boldsymbol{x})}{\sum_{\boldsymbol{y} : S(\boldsymbol{y}) = s} h(\boldsymbol{y})} \end{aligned}$$

となり，θ に依存しないことがわかる． □

例 5.4 ：$X_1, X_2, \ldots, X_n$ を確率密度関数

$$f_X(x|\theta) = \frac{1}{\theta}\mathbb{I}_{(0,\theta)}(x)$$

からのランダム標本とする．ただし，$\theta > 0$ とする．$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の同時確率密度関数は

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= \frac{1}{\theta^n}\mathbb{I}_{(0,\theta)^n}(\boldsymbol{x}) \\ &= \frac{1}{\theta^n}\mathbb{I}\{\max_{1\le i\le n} x_i \le \theta\}\mathbb{I}\{\min_{1\le i\le n} x_i \ge 0\} \\ &= g_\theta(\max(x_1, x_2, \ldots, x_n))h(\boldsymbol{x}) \end{aligned}$$

となる．ただし，$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ である．従って，$X_{(n)} = \max(X_1, X_2, \ldots, X_n)$ は $\theta \in (0, \infty)$ の十分統計量である．

例 5.5 ：$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の分布が k 母数指数分布族に属するとする．すなわち，その同時確率関数または確率密度関数が

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \exp\left[\sum_{i=1}^k c_i(\theta)T_i(\boldsymbol{x}) - d(\boldsymbol{\theta}) + S(\boldsymbol{x})\right]\mathbb{I}\{\boldsymbol{x} \in A\}$$

で与えられる．このとき，$h(\boldsymbol{x}) = \exp[S(\boldsymbol{x})]\mathbb{I}\{\boldsymbol{x} \in A\}$ とみれば，因数分解定理から

$$T = (T_1(\boldsymbol{X}), T_2(\boldsymbol{X}), \ldots, T_k(\boldsymbol{X}))$$

は $\boldsymbol{\theta}$ の十分統計量となる．

例 5.6 $X_1, X_2, \ldots, X_n$ を正規分布 $N(\theta, 1)$ からの大きさ n のランダム標本とする．このとき，

$$\begin{aligned} \Pi_{i=1}^n f(x_i|\theta) &= \Pi_{i=1}^n \frac{1}{\sqrt{2\imath}}\exp\big[-\frac{1}{2}(x_i-\theta)\big] = \frac{1}{(2\pi)^{n/2}}\exp\big[-\frac{1}{2}\sum_{i=1}^n(x_i-\theta)\big] \\ &= \frac{1}{(2\pi)^{n/2}}\exp\big[\theta\sum_{i=1}^n x_i - \frac{n}{2}\theta\big]\exp\big[-\frac{1}{2}\sum_{i=1}^n x_i^2\big] \end{aligned}$$

となる．したがって，$S(X_1, X_2, \ldots, X_n) = \sum_{i=1}^n X_i$ として，

$$g(s|\theta) = \exp\big[\theta s - \frac{n}{2}\theta\big], \qquad h(x_1, x_2, \ldots, x_n) = \exp\big[-\frac{1}{2}\sum_{i=1}^n x_i^2\big]$$

とすれば，分解定理から $S(X_1, X_2, \ldots, X_n) = \sum_{i=1}^n X_i$ は十分統計量となることがわかる．

例 5.7 $X_1, X_2, \ldots, X_n$ を正規分布 $N(\mu, \sigma^2)$ からの大きさ n のランダム標本とする．ただし，$\theta = (\mu, \sigma^2)$ である．

$$\begin{aligned} \Pi_{i=1}^n f(x_i|\theta) &= \Pi_{i=1}^n \frac{1}{\sqrt{2\imath}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left[-\frac{1}{2}\sum_{i=1}^n\left(\frac{x_i-\mu}{\sigma}\right)^2\right] \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\big(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + n\mu^2\big)\right] \end{aligned}$$

となる．したがって，

$$S_1(X_1, X_2, \dots, X_n) = \sum_{i=1}^n X_i, \qquad S_2(X_1, X_2, \dots, X_n) = \sum_{i=1}^n X_i^2$$

とし，

$$g(s_1, s_2 | \theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2}(s_2 - 2\mu s_1 + n\mu^2)\right], \qquad h(x_1, x_2, \dots, x_n) = 1$$

とおけば，$\sum_{i=1}^n X_i$ と $\sum_{i=1}^n X_i^2$ は十分統計量になる．さらに，

$$(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2) \longleftrightarrow (\bar{X}_n, \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2)$$

は一対一対応なので，$(\bar{X}_n, \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2)$ も十分統計量であることがわかる．十分統計量が表現のしかたは一通りでないことに注意する．

例 5.8 $X_1, X_2, \dots, X_n$ を一様分布

$$f(x|\theta) = \frac{1}{\theta_2 - \theta_2}\mathbb{I}_{(\theta_1, \theta_2)}(x)$$

からの大きさ n のランダム標本とする．ただし，$\theta = (\theta_1, \theta_2), \theta_2 > \theta_2$ である．このとき，

$$\begin{aligned}
\Pi_{i=1}^n f(x_i|\theta) &= \Pi_{i=1}^n \frac{1}{\theta_2 - \theta_2}\mathbb{I}_{(\theta_1, \theta_2)}(x_i) \\
&= \frac{1}{(\theta_2 - \theta_2)^2}\Pi_{i=1}^n \mathbb{I}_{(\theta_1, \theta_2)}(x_i) \\
&= \frac{1}{(\theta_2 - \theta_2)^2}\Pi_{i=1}^n \mathbb{I}_{(\theta_1, y_n)}(y_1)\,\mathbb{I}_{(y_1, \theta_2)}(y_n)
\end{aligned}$$

となる[(5-2)]．ただし，$y_1 = \min[x_1, x_2, \dots, x_n]$, $y_n = \max[x_1, x_2, \dots, x_n]$ である．したがって，$(\min[X_1, X_2, \dots, X_n], y_n = \max[X_1, X_2, \dots, X_n])$ が十分統計量となる．

定理 5.1 最尤推定量は十分統計量を通してのみ標本に依存する．

証明 $S_1 = S_1(X_1, X_2, \dots, X_n)$, $S_2 = S_2(X_1, X_2, \dots, X_n)$, $\dots$, $S_\ell = S_\ell(X_1, X_2, \dots, X_n)$ を十分統計量とする．尤度関数は

$$L(\theta| x_1, x_2, \dots, x_n) = \Pi_{i=1}^n f(x_i|\theta) = g(s_1, s_2, \dots, s_n|\theta)h(x_1, x_2, \dots, x_n)$$

と書ける．ただし，$s_i = S_i(x_1, x_2, \dots, x_n)$, $i = 1, 2, \dots, \ell$ とした．$L(\theta| x_1, x_2, \dots, x_n)$ の θ に関する最大化は $g(s_1, s_2, \dots, s_n|\theta)$ θ に関する最大化と同値である．よって，定理は証明された． □

第6章　点推定法

6.1　点推定量の性質

6.1.1　近隣度 (closeness)

$X_1, X_2, \ldots, X_n$ を確率（密度）関数 $f_X(x|\theta)$ からの標本の大きさが n のランダム標本とする．$\Theta \subset \mathbb{R}^k$ を θ の母数空間とし，確率（密度）関数 $f_X(x|\theta)$ は未知母数 θ のみに依存するものとする．$\tau : \Theta \to \mathbb{R}$ とし，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ とおく．

定義 6.1 $t(\boldsymbol{X})$ と $t'(\boldsymbol{X})$ を $\tau(\theta)$ の推定量とする．$t'(\boldsymbol{X})$ が $t(\boldsymbol{X})$ より集中した $\tau(\theta)$ の推定量であるとは，すべての $\lambda > 0$ に対して

$$\mathbb{P}_\theta[|t'(\boldsymbol{X}) - \tau(\theta)| \leq \lambda] \geq \mathbb{P}_\theta[|t(\boldsymbol{X}) - \tau(\theta)| \leq \lambda], \qquad \theta \in \Theta$$

を満足することをいう．$t^* = t^*(\boldsymbol{X}) - \tau(\theta)$ が最も集中しているとは，他のどんな推定量より集中していることをいう．

注意 6.1 t^* は最も望ましい推定量であるが，ほとんどの場合は存在しない．

定義 6.2 (Pitman's closeness) $t'(\boldsymbol{X})$ が $t(\boldsymbol{X})$ より Pitman の意味で $\tau(\theta)$ に近い推定量であるとは

$$\mathbb{P}_\theta\left[|t'(\boldsymbol{X}) - \tau(\theta)| < |t(\boldsymbol{X}) - \tau(\theta)|\right] > \frac{1}{2}, \qquad \theta \in \Theta$$

を満足することである．

$t^*(\boldsymbol{X})$ が Pitman の意味で $\tau(\theta)$ の最近隣推定量であるとは，どんな推定量 $t(\boldsymbol{X})$ よりも Pitman の意味で $\tau(\theta)$ に近いものをいう．

注意 6.2 これもなかなか存在しない．

6.1.2　平均 2 乗誤差

定義 6.3 $t(\boldsymbol{X})$ を $\tau(\theta)$ の推定量とする．推定量 $t(\boldsymbol{X})$ の平均 *2* 乗誤差を

$$\mathbb{E}_\theta[(t(\boldsymbol{X}) - \tau(\theta))^2]$$

これを $MSE(\theta, t)$ と記すことにする．で定義する．

注意 6.3 $\mathbb{E}_\theta$ の θ は考えている標本が分布族の中のどこから来ているかを明示するために用いている．すなわち

$$\begin{aligned}&\mathbb{E}_\theta[(t(\boldsymbol{X}) - \tau(\theta))^2]\\ &\quad = \int \cdots \int [t(x_1, x_2, \ldots, x_n) - \tau(\theta)]^2 f(x_1|\theta) f(x_2|\theta) \times \cdots \times f(x_n|\theta)\, dx_1\, dx_2 \cdots dx_n\end{aligned}$$

である．

注意 6.4 $t_1(\boldsymbol{X})$ と $t_2(\boldsymbol{X})$ の平均 2 乗誤差はどもに θ の関数となる．

例 6.1 $\tau(\theta)=\theta,\,\theta\in\Theta\subset\mathbb{R}$ とする．$\theta_0\in\Theta$ を固定し，

$$t_{\theta_0}(x_1,\,x_2,\,\dots,\,x_n)=\theta_0$$

なる推定量 $t_{\theta_0}(\boldsymbol{X})$ を考える．$t_{\theta_0}(\boldsymbol{X})$ の平均 2 乗誤差は

$$MSE\ (\theta,\,t_{\theta_0})=\mathbb{E}_\theta[(t_{\theta_0}(\boldsymbol{X})-\theta)^2]=\mathbb{E}_\theta[(\theta_0-\theta)^2]=(\theta_0-\theta)^2$$

となる．すなわち，

$$MSE\ (\theta_0,\,t_{\theta_0})=0$$

なる．

いま，すべての推定量 $t(\boldsymbol{X})$ に対して

$$MSE\ (\theta,\,t^*)\le\ MSE\ (\theta,\,t),\qquad\forall\theta\in\Theta$$

を満足するような推定量 $t^*(\boldsymbol{X})$ が存在するならば，

$$MSE\ (\theta,\,t^*)=0$$

とならなければいけない(6-1)．これは常に θ を正しく推定できることある．したがって，θ がわかっていることと同じである．

θ に関して一様に MSE を最小にする推定量を見つけることができない理由はすべての推定量の集合は大きすぎるからである．したがって，対象とする推定量の集合を小さくしよう．

定義 6.4 推定量 $t(\boldsymbol{X})$ が $\tau(\theta)$ の不偏推定量であるとは，すべての $\theta\in\Theta$ に対して

$$\mathbb{E}_\theta[t(\boldsymbol{X})]=\tau(\theta)$$

を満足することである．

例 6.2 $t_{\theta_0}(\boldsymbol{X})$ は θ の不偏推定量ではない．なぜならば，$\theta\ne\theta_0$ のとき

$$\mathbb{E}_\theta[t_{\theta_0}(\boldsymbol{X})]=\theta_0\ne\theta$$

からわかる．

定理 6.1 $t(\boldsymbol{X})$ は $\tau(\theta)$ の推定量とし，$\mathbb{E}_\theta|t(\boldsymbol{X})|^2<\infty$ とする．このとき，

$$MSE\ (\theta,\,t)=\mathbb{VAR}(t(\boldsymbol{X}))+\{\mathbb{E}_\theta[t(\boldsymbol{X})]-\tau(\theta)\}^2$$

が成立する．さらに，$t(\boldsymbol{X})$ が $\tau(\theta)$ の不偏推定量ならば，

$$MSE\ (\theta,\,t)=\mathbb{VAR}(t(\boldsymbol{X}))$$

が成立する．

証明

$$\begin{aligned}\mathrm{MSE}\ (\theta,\, t) &= \mathbb{E}_\theta[\{t(\boldsymbol{X}) - \mathbb{E}_\theta[t(\boldsymbol{X})] + \mathbb{E}_\theta[t(\boldsymbol{X})] - \tau(\theta)\}^2] \\ &= \mathbb{E}_\theta[\{t(\boldsymbol{X}) - \mathbb{E}_\theta[t(\boldsymbol{X})]\}^2] \\ &\quad +2(\mathbb{E}_\theta[t(\boldsymbol{X})] - \tau(\theta))\mathbb{E}_\theta[t(\boldsymbol{X}) - \mathbb{E}_\theta[t(\boldsymbol{X})]] \\ &\quad +\{\mathbb{E}_\theta[t(\boldsymbol{X})] - \tau(\theta)\}^2 \\ &= \mathbb{VAR}(t(\boldsymbol{X})) + \{\mathbb{E}_\theta[t(\boldsymbol{X})] - \tau(\theta)\}^2\end{aligned}$$

□

定義 6.5 $BIAS = \mathbb{E}_\theta[t(\boldsymbol{X}) - \tau(\theta)]$ を推定量 $t(\boldsymbol{X})$ のバイアスという.

例 6.3 $X_1, X_2, \ldots, X_n (n \geq 2)$ を $N(\mu, \sigma^2)$ からのランダム標本とする.

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad U^2 = \frac{1}{n-1}\sum_{i=}^{n}(X_i - \bar{X}_n)^2 \qquad S^2 = \frac{1}{n}\sum_{i=}^{n}(X_i - \bar{X}_n)^2$$

とおく.

$\mathbb{E}_{\mu,\,\sigma^2}[\bar{X}_n] = \mu$ より $\bar{X}_n$ は μ の不偏推定量である．また，$\bar{X}_n$ の MSE（$\tau(\mu,\,\sigma^2) = \mu$）は

$$MSE\ (\mu,\, \bar{X}_n) = \mathbb{E}_{\mu,\,\sigma^2}[(\bar{X}_n - \mu)^2] = \frac{\sigma^2}{n}$$

である.

一方，

$$\mathbb{E}_{\mu,\,\sigma^2}[S^2] = \mathbb{E}_{\mu,\,\sigma^2}[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - (\bar{X}_n - \mu)^2] = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

となり，σ^2 の最尤推定量は σ^2 の不偏推定量ではない．しかし，

$$\mathbb{E}_{\mu,\,\sigma^2}[U^2] = \mathbb{E}_{\mu,\,\sigma^2}[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu)^2 - \frac{n}{n-1}(\bar{X}_n - \mu)^2] = \frac{n}{n-1}\sigma^2 - \frac{1}{n-1}\sigma^2 = \sigma^2$$

となり，U^2 は σ^2 の不偏推定量である.

例 6.4 $X_1, X_2, \ldots, X_n$ をベルヌーイ試行(6-2) $\mathrm{Bi}(1,\, p)$ からのランダム標本とする．ただし，$0 < p < 1$ とする．標本平均 $\bar{X}_n$ は p の不偏推定量となる．$\tau(p) = p$ として，$\bar{X}_n$ の MSE を求めよう.

$$MSE(p,\, \bar{X}_n) = \mathbb{E}[(\bar{X}_n - p)^2] = \mathbb{VAR}[\bar{X}_n] = \frac{p(1-p)}{n}$$

となる．いま，$Y = \sum_{i=1}^{n} X_i$ として，

$$\hat{p}_B = \frac{Y + \alpha}{\alpha + \beta + n}$$

なる推定量を考えよう．ただし，α と β は（n に依存する）定数とする．定理 6.1 を利用して，

$$\begin{aligned}MSE(p,\, \hat{p}_B) &= \mathbb{VAR}[\hat{p}_B] + \{\mathbb{E}[\hat{p}_B] - p\}^2 \\ &= \mathbb{VAR}\left[\frac{Y+\alpha}{\alpha+\beta+n}\right] + \left[\frac{np+\alpha}{\alpha+\beta+n} - p\right]^2 \\ &= \frac{1}{(\alpha+\beta+n)^2}\mathbb{VAR}[Y] + + \left[\frac{np+\alpha}{\alpha+\beta+n} - p\right]^2 \\ &= \frac{np(1-p)}{(\alpha+\beta+n)^2} + + \left[\frac{np+\alpha}{\alpha+\beta+n} - p\right]^2\end{aligned}$$

となる．ここで $\alpha = \beta = \sqrt{n/4}$ とおけば，

$$MSE(p, \hat{p}_B) = \frac{n}{4(\sqrt{n} + n^2)}$$

となる[(6-3)]．

6.1.3 推定量の一致性

定義 6.6 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ とする．$\tau(\theta)$ の推定量の列 $\{t_n(\boldsymbol{X})\}_{n=1}^{\infty}$ が平均 2 乗誤差の意味で $\tau(\theta)$ の一致推定量であるとは，すべての $\theta \in \Theta$ に対して，

$$\lim_{n\to\infty} \mathbb{E}_\theta[(t_n(\boldsymbol{X}) - \tau(\theta))^2] = 0$$

を満足することである．

注意 6.5 推定量 $t_n(\boldsymbol{X})$ が平均 2 乗誤差の意味で一致性を持てば，定理 6.1 から

$$\lim_{n\to\infty} \mathbb{E}_\theta[t_n(\boldsymbol{X})] = \tau(\theta), \qquad \lim_{n\to\infty} \mathbb{VAR}_\theta[t_n(\boldsymbol{X})] = 0$$

となる．

例 6.5 $X_1, X_2, \ldots, X_n$ を平均 μ，分散 σ^2 の分布からのランダム標本とする．ただし，$\sigma < \infty$ とする．$\{\bar{X}_n\}_{n=1}^{\infty}$ を μ の推定量の列とすれば，

$$\mathbb{E}_{\mu,\sigma^2}[\{\bar{X}_n - \mu\}^2] = \mathbb{VAR}[\bar{X}_n] = \frac{\sigma^2}{n}$$

よって，$\bar{X}_n$ は平均 2 誤差の意味で一致性を持つ．

定義 6.7 $\tau(\theta)$ の推定量の列 $\{t_n(\boldsymbol{X})\}_{n=1}^{\infty}$ が弱一致性を持つとは，どんな正の数 ϵ に対しても

$$\lim_{n\to\infty} \mathbb{P}_\theta[|t_n(\boldsymbol{X}) - \tau(\theta)| < \epsilon] = 1, \qquad {}^\forall\theta \in \Theta$$

が成立することである．簡単に，このような推定量 $t_n(\boldsymbol{X})$ を弱一致推定量という．

注意 6.6 $t_n(\boldsymbol{X})$ が平均 2 乗誤差の意味で一致推定量であれば，$t_n(\boldsymbol{X})$ は弱一致推定量である．なぜならば，Markov の不等式から

$$\begin{aligned}
\mathbb{P}_\theta[|t_n(\boldsymbol{X}) - \tau(\theta)| < \epsilon] &= 1 - \mathbb{P}_\theta[|t_n(\boldsymbol{X}) - \tau(\theta)| \geq \epsilon] \\
&= 1 - \mathbb{P}_\theta[|t_n(\boldsymbol{X}) - \tau(\theta)|^2 \geq \epsilon^2] \\
&\geq 1 - \frac{1}{\epsilon^2}\mathbb{E}_\theta[(t_n(\boldsymbol{X}) - \tau(\theta))^2] \nearrow 1, \qquad n \to \infty
\end{aligned}$$

からわかる．

6.2 最尤法

確率ベクトル $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ の統計モデルを $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ とする．$\mathcal{P}$ に含まれる P_θ に対応する確率密度関数もしくは確率関数を $p(\boldsymbol{x}|\,\theta)$ と記すことにする．ただし，$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ である．$\boldsymbol{X} = \boldsymbol{x}$ が観測されたときの尤度関数を

$$L_n(\theta|\,\boldsymbol{x}) = p(\boldsymbol{x}|\,\theta), \qquad \theta \in \Theta$$

で定めることにする．$L_n(\cdot\,|\,\boldsymbol{x})$ は標本空間から $\{\theta \mapsto p(\boldsymbol{x}|\,\theta) : \boldsymbol{x} \in S\}$ なる関数族への対応となる．$\boldsymbol{x}$ があたられたとき，$L_n(\theta|\,\boldsymbol{x})$ は θ の関数とみなす．これを簡単に $L_n(\theta)$ と書くことにする．$L_n(\theta)$ は $\boldsymbol{x}$ が与えられたとき，いろいろな θ の「確からしさ」もしくは「尤もらしさ」を表現するものである．

特に，$X_1, X_2, \ldots, X_n$ が独立同一に確率密度関数 $f(x|\,\theta)$ に従うならば，尤度関数は

$$L_n(\theta|\,\boldsymbol{x}) = \Pi_{i=1}^n f(x_i|\,\theta)$$

で与えられる．

最尤法とは，与えられたデータを実現させせるために「尤もらしい」母数の値を母数の推定値として用いる手法である．すなわち，$\boldsymbol{X} = \boldsymbol{x}$ が与えられたとき，尤度関数を最大にする値 $\widehat{\theta}(\boldsymbol{x})$ を見つけることである：

$$L_n(\widehat{\theta}(\boldsymbol{x})|\,\boldsymbol{x}) = p(\boldsymbol{x}|\,\widehat{\theta}(\boldsymbol{x})) = \max\{p(\boldsymbol{x}|\,\theta) : \theta \in \Theta\} = \max\{L_n(\theta|\,\boldsymbol{x}) : \theta \in \Theta\}$$

$\widehat{\theta}(\boldsymbol{x})$ を θ の最尤推定値といい，$\widehat{\theta}(\boldsymbol{X})$ を θ の最尤推定量という．

例 6.6 確率変数 X が正規分布 $N(\theta, \sigma^2)$ に従うとする．ただし，σ^2 は既知とする．このとき，尤度関数は

$$L_1(\theta|\,x) = \frac{1}{\sigma}\psi\left(\frac{x-\theta}{\sigma}\right)$$

となる．ただし，$\psi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ である．このとき，最大は

$$\widehat{\theta}(x) = x$$

のとき唯一達成される．したがって，$\widehat{\theta}(X) = X$ は最尤推定量となる．

つぎに，$X_1, X_2, \ldots, X_n$ は独立同一に正規分布 $N(\theta, \sigma^2)$ に従うとする．ここでも σ^2 は既知とする．このとき，尤度関数は

$$\begin{aligned} L_n(\theta|\,\boldsymbol{x}) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2}\sum_{i=1}^n \frac{x_i-\theta}{\sigma^2}\right] \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2}\frac{(\bar{x}_n-\theta)^2}{\sigma^2/n}\right] \exp\left[-\frac{1}{2}\sum_{i=1}^n \frac{(x_i-\bar{x}_n)^2}{\sigma^2}\right] \end{aligned}$$

となる．よって，最大は

$$\widehat{\theta}(\boldsymbol{x}) = \bar{x}_n, \qquad \bar{x}_n = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

で達成される．したがって，最尤推定量は $\widehat{\theta}(\boldsymbol{X}) = \bar{X}_n$ となる．ただし，$\bar{X}_n = (1/n)(X_1 + X_2 + \cdots + X_n)$ である．

尤度関数に対数をとったものを対数尤度とよび，

$$l_n(\theta) = \log L_n(\theta|\,\boldsymbol{x})$$

と記す(6-4)．特に，$X_i,\, i=1,2,\ldots,n,$ が独立同一に確率密度関数 $f(x|\theta)$ に従う場合には

$$l_n(\theta) = \log \Pi_{i=1}^n f(x_i|\,\theta) = \sum_{i=1}^n \log f(x_i|\,\theta)$$

となる．

もし，Θ が開集合で $l_n(\theta)$ が θ に関して微分可能で $\widehat{\theta}(\boldsymbol{x})$ が存在するならば，$\widehat{\theta}(\boldsymbol{x})$ は方程式

$$\frac{\partial}{\partial\theta_i} l_n(\theta) = 0, \qquad i = 1, 2, \ldots, k$$

をみたす．この方程式を尤度方程式という．

例 6.7 標識 1, 2, 3 のどれかをもつ個体から構成される母集団を考える．それぞれの標識の出現確率は Hardy-Weinberg 比率で与えられるとする：

$$p(1|\theta) = \theta^2, \qquad p(2|\theta) = 2\theta(1-\theta), \qquad p(3|\theta) = (1-\theta)^2, \qquad 0 < \theta < 1$$

たとえば，3 つの個体を観測し，$x_1 = 1, x_2 = 2, x_3 = 1$ を得たとする．このとき

$$L_3(\theta|\,\boldsymbol{x}) = p(1|\,\theta)p(2|\,\theta)p(1|\,\theta) = 2\theta^5(1-\theta)$$

となる．尤度方程式は

$$\frac{\partial}{\partial\theta} l_3(\theta) = \frac{5}{\theta} - \frac{1}{1-\theta} = 0$$

となり，唯一の解 $\widehat{\theta} = 5/6$ を得る．これは

$$\frac{\partial^2}{\partial\theta^2} l_3(\theta) = -\frac{5}{\theta^2} - \frac{1}{(1-\theta)^2} < 0, \qquad 0 < \theta < 1$$

よりわかる．

一般に，n 個の観測 $x_1, x_2, \ldots, x_n$ を得たとする．いま

$$n_j = \#\{x_i = j : i = 1, 2, \ldots, n\}, \qquad j = 1, 2, 3$$

とする．尤度関数は

$$L_n(\theta|\,\boldsymbol{x}) = \theta^{2n_1}\{2\theta(1-\theta)\}^{n_2}(1-\theta)^{2n_3} = 2^{n_2}\theta^{2n_1+n_2}(1-\theta)^{n_2+2n_3}$$

より

$$\frac{\partial}{\partial\theta} l_n(\theta) = \frac{2n_1+n_2}{\theta} - \frac{n_2+2n_3}{1-\theta} = \frac{1}{\theta(1-\theta)}\{(2n_1+n_2) - 2(n_1+n_2+n_3)\theta\}$$

より，$2n_1 + n_2 > 0,\, n_2 + 2n_3 > 0$ のとき，最尤推定値は唯一存在して，

$$\widehat{\theta}(\boldsymbol{x}) = \frac{2n_1+n_2}{2(n_1+n_2+n_3)}$$

となる．もし，$2n_1 + n_2 = 0$ のとき，尤度関数は

$$2^{n_2}(1-\theta)^{(2n_1+n_2)+(n_2+2n_3)} = 2^{n_2}(1-\theta)^{2(n_1+n_2+n_3)}$$

となり，$\theta = 0$ のとき，尤度関数は最大になり，$\Theta = (0, 1)$ なので，最尤推定値は存在しない．また，$n_2 + 2n_3 = 0$ のときは，尤度関数は $2^{n_2}\theta^{2(n_1+n_2+n_3)}$ なり，最尤推定値は存在しない．

例 6.8 $X_1, X_2, \ldots, X_n$ は独立同一分布に従い，各 $X_i, i = 1, 2, \ldots, n,$ は $1, 2, \ldots, k$ の値をとり，その確率は $\theta_j = \mathbb{P}\{X_i = j\}, j = 1, 2, \ldots, k,$ で与えられるとする．ここで，$n \geq k-1$ を仮定する．いま，$N_j = \sum_{i=1}^n I\{X_i = j\}$ おく．$\boldsymbol{X} = \boldsymbol{x}$ が与えられたとする．このとき，$n_j = \sum_{i=1}^n I\{x_i = j\}$ とおけば，対数尤度は

$$l_n(\boldsymbol{\theta}) = \log L_n(\boldsymbol{\theta}|\,\boldsymbol{x}) = \log p(\boldsymbol{x}|\,\boldsymbol{\theta}) = \sum_{j=1}^k n_j \log \theta_j$$

となる．ただし，$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ で

$$\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j \tag{6.1}$$

である．

まず，$n_j > 0, j = 1, 2, \ldots, k,$ を仮定する．このとき，θ_j のどれかがゼロならば，$p(\boldsymbol{x}|\,\boldsymbol{\theta}) = 0$ となる．したがって，最尤推定値は $\theta_j > 0$ となるので，上の仮定のもとでは，最尤推定値は $[0, 1]^k$ の内点である．したがって，最尤推定値は尤度方程式

$$\frac{\partial}{\partial \theta_j} l_n(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_j} \sum_{\ell=1}^k n_\ell \log \theta_\ell = \sum_{\ell=1}^k \frac{n_\ell}{\theta_\ell} \frac{\partial \theta_\ell}{\partial \theta_j} = 0, \qquad j = 1, 2, \ldots, k-1$$

となる．(6.1) から $\partial\theta_k/\partial\theta_j = -1, j = 1, 2, \ldots, k-1,$ となる．よって，尤度方程式は

$$\frac{\widehat{\theta}_k}{\widehat{\theta}_j} = \frac{n_k}{n_j}, \qquad j = 1, 2, \ldots, k-1$$

となる．これを再度 (6.1) に代入すれば，$\widehat{\theta}_k = n_k/n$ となり，

$$\widehat{\theta}_j = \frac{n_j}{n}, \qquad j = 1, 2, \ldots, k$$

となる．ただし，$n = \sum_{\ell=1}^k n_\ell$ とした．つぎに，$\theta_j = n_j/n, j = 1, 2, \ldots, k$ が実際に $l_n(\boldsymbol{\theta})$ を最大にしていることを確認するために，$l_n(\boldsymbol{\theta})$ は $(\theta_1, \theta_2, \ldots, \theta_{k-1})$ に関して concave であることを示す．$1 \leq r \leq k-1, 1 \leq j \leq k-1$ に対して，

$$\frac{\partial}{\partial \theta_r}\frac{\partial}{\partial \theta_j} l_n(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_r}\left(\frac{n_j}{\theta_j} - \frac{n_k}{\theta_k}\right) = \begin{cases} -\left(\frac{n_r}{\theta_r^2} + \frac{n_k}{\theta_k^2}\right) < 0, & r = j \\ -\frac{n_k}{\theta_k^2} < 0, & r \neq j \end{cases}$$

となる．

ある j に対して，$n_j = 0$ のとき，$\widehat{\theta}_j = n_j/n$ が最尤推定値であることも確認することができる．

定理 6.2 $T(\boldsymbol{X})$ を未知母数 θ の十分統計量とする．このとき，θ の最尤推定量が一意に存在するならば，θ の最尤推定量は T の関数である．

証明 $p(\boldsymbol{x}|\,\theta)$ を確率関数または確率密度関数とする．因子分解定理から θ と T を通してのみ $\boldsymbol{x}$ に依存する関数 g と $\boldsymbol{x}$ のみに依存する関数 h が存在して，

$$p(\boldsymbol{x}|\,\theta) = h(\boldsymbol{x})g(T(\boldsymbol{x})|\,\theta)$$

と書ける．これより θ に関して $p(\boldsymbol{x}|\,\theta)$ を最大化することは θ に関して $g(T(\boldsymbol{x})|\,\theta)$ を最大化することと同値になる．また，$T(\boldsymbol{x}) = t$ と与えられたとき，$g(t|\,\theta)$ が 2 つ以上 θ で最大になるとすれば，それに対応する $\boldsymbol{x}$ において $g(T(\boldsymbol{x})|\,\theta)$ も 2 つ以上の θ で最大化されるので，仮定と矛盾する．したがって，$g(t|\,\theta)$ を θ について最大にする点はひとつである． □

定理 6.3 関数 $g: \mathbb{R} \to \mathbb{R}$ が 1 対 1 のとき，$\widehat{\theta}$ が θ の最尤推定量であれば，$g(\widehat{\theta})$ は $g(\theta)$ の最尤推定量である．

証明 g は 1 対 1 だから，逆関数 g^{-1} が存在し，$\tau = g(\theta)$ のとき，$\theta = g^{-1}(\tau)$ となる．これより

$$L_n(\theta|\,\boldsymbol{x}) = p(\boldsymbol{x}|\,\theta) = p(\boldsymbol{x}|\,g^{-1}(\tau)) = \tilde{L}_n(\tau|,\boldsymbol{x})$$

と書けるので，

$$\sup_{\tau} \tilde{L}_n(\tau|,\boldsymbol{x}) = \sup_{\tau} L_n(g^{-1}(\tau)|,\boldsymbol{x}) = \sup_{\theta} L_n(\theta|\,\boldsymbol{x})$$

よって，$\widehat{\theta} = g^{-1}(\widehat{\tau})$ のとき最大化される．したがって，$\widehat{\tau} = g(\widehat{\theta})$ のとき最大化される． □

例 6.9 （指数分布族の最尤推定量） $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ は k 母数指数分布族

$$f(x|\,\boldsymbol{\eta}) = S(x)\exp[\sum_{j=1}^{k} \eta_j T_j(x) - A(\boldsymbol{\eta})]$$

からの大きさ n のランダム標本とする．ただし，$\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k)$ とし，自然母数空間 $\eta \in \mathcal{E}$ は $\mathbb{R}^k$ の開集合とする．である．このとき，$\boldsymbol{X}$ の同時確率（密度）関数は

$$\begin{aligned} p(\boldsymbol{x}|\,\boldsymbol{\eta}) &= \Pi_{i=1}^{n} f(x_i|\,\boldsymbol{\eta}) \\ &= \Pi_{i=1}^{n} S(x_i)\exp[\sum_{j=1}^{k} n\eta_j \bar{T}_j(\boldsymbol{x}) - nA(\boldsymbol{\eta})] \end{aligned}$$

となる．ただし，$\bar{T}_j(\boldsymbol{x}) = \sum_{i=1}^{n} T_j(x_i)$, $j = 1, 2, \ldots, k$ である．したがって，対数尤度は

$$l_n(\boldsymbol{\eta}) = n\sum_{j=1}^{k} \eta_j \bar{T}_j(\boldsymbol{x}) - nA(\boldsymbol{\eta}) + \text{（定数項）}$$

となる．系 **??** を用いれば，

$$\frac{\partial}{\partial \eta_j} l_n(\boldsymbol{\eta}) = n\bar{T}_j(\boldsymbol{x}) - n\frac{\partial}{\partial \eta_j} A(\boldsymbol{\eta}) = n\bar{T}_j(\boldsymbol{x}) - n\mathbb{E}_\eta[T_j(X)]$$

と

$$\frac{\partial^2}{\partial \eta_j \eta_m} l_n(\boldsymbol{\eta}) = -n\frac{\partial^2}{\partial \eta_j \eta_m} A(\boldsymbol{\eta}) = -n\mathbb{COV}_\eta(T_j(X), T_m(X))$$

となるので，行列 $((\partial^2/\partial\boldsymbol{\eta}'\boldsymbol{\eta})A(\boldsymbol{\eta}))$ は負の定符号となる．したがって，尤度方程式

$$\frac{1}{n}\sum_{i=1}^{n} T_j(x_i) = \frac{\partial}{\partial \eta_j} A(\boldsymbol{\eta})$$

または

$$\frac{1}{n}\sum_{i=1}^{n} T_j(x_i) = \mathbb{E}_\eta[T_j(X)], \qquad j = 1, 2, \ldots, k$$

は唯一の解を持ち，これは $\boldsymbol{\eta}$ の最尤推定値となる．

6.3 不偏推定と Cramér-Rao の定理

$X_1, X_2, \ldots, X_n$ を確率（密度）関数 $f(x|\theta), \theta \in \Theta$ からのランダム標本とし，$T = T(\boldsymbol{X})$ を標本 $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ に基づく $\tau(\theta)$ の推定量とする．T の MSE は

$$\mathbb{E}_\theta[\{T(\boldsymbol{X}) - \tau(\theta)\}^2] = \mathbb{VAR}\theta[T(\boldsymbol{X})] + \{\mathbb{E}_\theta[T(\boldsymbol{X})] - \tau(\theta)\}^2$$

と書けた．さらに，T が $\tau(\theta)$ の不偏推定量であれば，

$$\mathbb{E}_\theta[\{T((\boldsymbol{X}) - \tau(\theta)\}^2] = \mathbb{VAR}\theta[T(\boldsymbol{X})]$$

となる．

定義 6.8 (一様最小分散不偏推定量　Uniformly Minimum Variance Unbiased Estimator)　$X_1, X_2, \ldots, X_n$ を確率（密度）関数 $f(x|,\theta), \theta \in \Theta$ からのランダム標本とし，$\tau(\theta)$ の推定量 $T^* = T^*(X_1, X_2, \ldots, X_n)$ が $\tau(\theta)$ の一様最小分散不偏推定量（略して UMVUE）であるとはつぎを満足することである．
(i)　$\mathbb{E}_\theta[T^*] = \tau(\theta)$
(ii)　他の $\tau(\theta)$ の不偏推定量 $T = T(X_1, X_2, \ldots, X_n)$ に対し

$$\mathbb{VAR}_\theta[T^*] \leq \mathbb{VAR}_\theta[T].$$

6.3.1 不偏推定量の分散の下限

$X_1, X_2, \ldots, X_n$ を確率密度関数 $f(x|\theta), \theta \in \Theta \subset \mathbb{R}$ からのランダム標本とする．以下の仮定をおく．
(A1)　集合 $\{x : f(x|\theta) > 0\}$ は θ に依存しない．
(A2)　すべての x と θ に対し，$(\partial/\partial\theta)f(x|\theta)$ は存在する．
(A3)　$\mathbb{E}_\theta[T(X_1, X_2, \ldots, X_n)] < \infty$ なる推定量 T に対し

$$\frac{\partial}{\partial\theta}\int\cdots\int T(x_1, x_2, \ldots, x_n)\Pi_{i=1}^n f(x_i|\theta)\,dx_i = \int\cdots\int T(x_1, x_2, \ldots, x_n)\frac{\partial}{\partial\theta}\Pi_{i=1}^n f(x_i|\theta)\,dx_i.$$

(A4)　すべての $\theta \in \Theta$ に対し

$$0 < \mathbb{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f(X_1|\theta)\right\}^2\right] < \infty.$$

(A5)　$\tau(\theta)$ は微分可能．このとき，θ に関して一様に

$$\mathbb{VAR}_\theta[T] \geq \frac{[\tau'(\theta)]^2}{\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\theta)\right)^2\right]}.$$

が成立する．また，等号成立は

$$\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\,\theta) = K(\theta, n)[t(\boldsymbol{X} - \tau(\theta)] \tag{6.2}$$

のときに限る．

証明 $f(\boldsymbol{x}|\,\theta) = \Pi_{i=1}^n f(x_i|\,\theta)$ と $d\boldsymbol{x} = dx_1 dx_2 \cdots dx_n$ 記すことにする．

$$\begin{aligned}
\tau'(\theta) &= \frac{\partial}{\partial\theta}\int\cdots\int T(\boldsymbol{x}) f(\boldsymbol{x}|\,\theta)\,d\boldsymbol{x} \\
&= \int\cdots\int T(\boldsymbol{x})\frac{\partial}{\partial\theta} f(\boldsymbol{x}|\,\theta)\,d\boldsymbol{x} - \tau(\theta)\frac{\partial}{\partial\theta}\int\cdots\int f(\boldsymbol{x}|\,\theta)\,d\boldsymbol{x} \\
&= \int\cdots\int [T(\boldsymbol{x}) - \tau(\theta)]\frac{\partial}{\partial\theta} f(\boldsymbol{x}|\,\theta)\,d\boldsymbol{x} \\
&= \int\cdots\int [T(\boldsymbol{x}) - \tau(\theta)]\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{x}|\,\theta)\right) f(\boldsymbol{x}|\,\theta)\,d\boldsymbol{x} \\
&= \mathbb{E}_\theta\left[\{T(\boldsymbol{X}) - \tau(\theta)\}\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\,\theta)\right)\right]
\end{aligned}$$

となる．Cauchy-Schwarz の不等式から

$$[\tau'(\theta)]^2 \leq \mathbb{E}_\theta[\{T(\boldsymbol{X}) - \tau(\theta)\}^2]\,\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\,\theta)\right)^2\right]$$

よって

$$\mathbb{VAR}_\theta[T] \geq \frac{[\tau'(\theta)]^2}{\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\,\theta)\right)^2\right]}.$$

しかし

$$\begin{aligned}
\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}|\,\theta)\right)^2\right] &= \mathbb{E}_\theta\left[\left\{\sum_{i=1}^n \frac{\partial}{\partial\theta}\log f(X_i|\,\theta)\right\}\right] \\
&= \sum_{i=1}^n\sum_{j=1}^n \mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X_i|\,\theta)\frac{\partial}{\partial\theta}\log f(X_j|\,\theta)\right] \\
&= n\mathbb{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f(X_1|\,\theta)\right\}^2\right].
\end{aligned}$$

なぜならば，$i \neq j$ のとき X_i と X_j は独立であることと

$$\begin{aligned}
\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(X_1|\,\theta)\right] &= \int\left(\frac{\partial}{\partial\theta}\log f(x|\,\theta)\right) f(x|\,\theta)\,dx = \int\frac{\partial}{\partial\theta} f(x|\,\theta)\,dx \\
&= \frac{\partial}{\partial\theta}\int f(x|\,\theta)\,dx = 0
\end{aligned}$$

最後に Cauchy-Schwarz の不等式の等号成立条件から

$$\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\,\theta) = K(\theta,\,n)[t(\boldsymbol{x}) - \tau(\theta)] + b(\theta) \tag{6.3}$$

が等号成立条件となることがわかる．しかし，両辺の期待値をとれば，

$$0 = \mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\log f(\boldsymbol{X}\,\theta)\right] = K(\theta,\,n)\mathbb{E}_\theta\,[t(\boldsymbol{x}) - \tau(\theta)] + b(\theta) = b(\theta)$$

となるので，$b(\theta) = 0$ がわかる．よって，等号成立条件は (6.2) であることが示せた． □

注意 6.7 T が $\tau(\theta)$ の一様最小分散不偏推定量ならば，一意である．これは以下の議論からわかる．T' を別の $\tau(\theta)$ の一様最小分散不偏推定量とし，$T^* = (1/2)(T + T')$ とおく．すると，$\mathbb{E}[T^*] = \tau(\theta)$ となり，T^* も $\tau(\theta)$ の不偏推定量となる．したがって，すべての θ に対して，

$$\mathbb{VAR}_\theta(T^*) \geq \mathbb{VAR}_\theta(T)$$

となる．しかし，

$$\begin{aligned}\mathbb{VAR}_\theta(T^*) &= \mathbb{VAR}_\theta((1/2)T+(1/2)T') \\ &= \frac{1}{4}\mathbb{VAR}_\theta(T)+\frac{1}{4}\mathbb{VAR}_\theta(T')+\frac{1}{2}\mathbb{COV}_\theta(T,T') \\ &\le \frac{1}{4}\mathbb{VAR}_\theta(T)+\frac{1}{4}\mathbb{VAR}_\theta(T')+\frac{1}{2}\sqrt{\mathbb{VAR}_\theta(T)}\sqrt{\mathbb{VAR}_\theta(T')} \\ &= \mathbb{VAR}_\theta(T) \end{aligned} \tag{6.4}$$

となる．不等号は Cauchy-Schwartz の不等式からわかり，最後の等号は $\mathbb{VAR}_\theta(T)=\mathbb{VAR}_\theta(T')$ よりわかる．(6.4) と (6.4) から

$$\mathbb{VAR}_\theta(T)=\mathbb{VAR}_\theta(T^*) \tag{6.5}$$

が成立することがわかる．したがって，(6.4) の Cauchy-Schwartz の不等式は等号が成立している．Cauchy-Schwartz の不等式の等号成立条件から，ある定数 $a(\theta)$ と $b(\theta)$ が存在して

$$T'=a(\theta)T+b(\theta)$$

と書ける．しかし，$\mathbb{VAR}_\theta(T')=a(\theta)\mathbb{VAR}_\theta(T)$ と (6.5) から $a(\theta)=1$ となる．さらに，

$$\tau(\theta)=\mathbb{E}_\theta(T')=\mathbb{E}_\theta(T)+b(\theta)=\tau(\theta)+b(\theta)$$

から $b(\theta)=0$ となる．したがって，$T=T'$ となることがわかる．

例 6.10 $X_1, X_2, \ldots, X_n$ を指数分布 $f(x|\theta)=\theta e^{-\theta x}I_{(0,\infty)}(x)$ からの大きさ n のランダム標本とする．ただし，$\theta>0$ である．

まず，$\tau(\theta)=\theta$ とする．θ の推定量 $T=T(X_1, X_2, \ldots, X_n)$ の Cramér-Rao の下限は

$$\mathbb{VAR}_\theta(T)\ge\frac{1}{n\mathbb{E}_\theta[\{(\partial/\partial\theta)\log f(X_1|\theta)\}^2]}$$

となる．さらに，

$$\frac{\partial}{\partial\theta}\log f(x|\theta)=\frac{1}{\theta}-x$$

より

$$\mathbb{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f(X_1|\theta)\right\}^2\right]=\mathbb{E}_\theta\left[\left\{\frac{1}{\theta}-X_1\right\}^2\right]=\mathbb{VAR}(X)=\frac{1}{\theta^2}$$

となる(6-5)．したがって，

$$\mathbb{VAR}_\theta\ge\frac{1}{n(1/\theta^2)}=\frac{\theta^2}{n}$$

となる．

つぎに，$\tau(\theta)=1/\theta$ とする．$\tau(\theta)=1/\theta)$ の推定量 $S=S(X_1, X_2, \ldots, X_n)$ の Cramér - Rao の下限は

$$\mathbb{VAR}_\theta(S)\ge\frac{\{\tau'(\theta)\}^2}{n(1/\theta^2)}=\frac{1}{n\theta^2}$$

となる．さらに，

$$\sum_{i=1}^n\frac{\partial}{\partial\theta}\log f(X_i|\theta)=\sum_{i=1}^n\frac{\partial}{\partial\theta}(\log\theta-\theta X_i)=\sum_{i=1}^n\left(\frac{1}{\theta}-X_i\right)=-n\left(\bar{X}_n-\frac{1}{\theta}\right)$$

より，(6.2) において $K(\theta,n)=-n$ とおけば，$S=\bar{X}_n$ は $1/\theta$ の UMVUE となる．

例 6.11 $X_1, X_2, \ldots, X_n$ をポアソン分布 $f(x|\theta) = e^{-\lambda}\lambda^x/x!(x = 0, 1, ldots)$ からの大きさ n のランダム標本とする．ただし，$\theta > 0$ である．

$\tau(\theta) = e^{-\theta} = \mathbb{P}[X_1 = 0]$ とする．

$$\frac{\partial}{\partial\theta}\log f(x|\theta) = -1 + \frac{x}{\theta}$$

に注意すれば，

$$\mathbb{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f(X_1|\theta)\right\}^2\right] = \mathbb{E}_\theta\left[\left\{\frac{X_1}{\theta} - 1\right\}^2\right] = \frac{1}{\theta^2}\mathbb{E}_\theta[(X_1 - \theta)^2] = \frac{1}{\theta}$$

となる(6-6)．よって，$\tau(\theta)$ の推定量推定量 $T = T(X_1, X_2, \ldots, X_n)$ の Cramér-Rao) の下限は

$$\mathbb{VAR}_\theta(T) \geq \frac{(-e^{-\theta})^2}{n(1/\theta)} = \frac{\theta e^{-2\theta}}{n}$$

となる．

いま，$\tau(\theta)$ の推定量推定量

$$T_0 = \frac{1}{n}\sum_{i=1}^n I\{X_i = 0\}$$

を考える．

$$\mathbb{E}_\theta[T_0] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\theta[I\{X_i = 0\}] = \frac{1}{n}\sum_{i=1}^n \big[1\cdot\mathbb{P}_\theta(X_i = 0) + 0\cdot\mathbb{P}_\theta(X_i \neq 0)\big] = e^{-\theta}$$

となる．したがって，T_0 は $\tau(\theta) = e^{-\theta}$ の不偏推定量となる．また，

$$\begin{aligned}
\mathbb{VAR}_\theta &= \mathbb{E}_\theta\left[\left\{\frac{1}{n}\sum_{i=1}^n I\{X_i = 0\}\right\}^2\right] - \{\mathbb{E}_\theta(T_0)\}^2 \\
&= \sum_{i=1}^n \mathbb{E}_\theta\left[\frac{1}{n^2}I\{X_i = 0\}\right] + \sum_{i\neq j}\mathbb{E}_\theta\left[\frac{1}{n^2}I\{X_i = 0\}I\{X_j = 0\}\right] - e^{-2\theta} \\
&= \frac{1}{n}e^{-\theta} + \frac{n(n-1)}{n^2}e^{-2\theta} - e^{-2\theta} = \frac{1}{n}e^{-\theta}(1 - e^{-\theta})
\end{aligned}$$

となる．$\theta > 0$ に対し

$$\frac{1}{n}e^{-\theta}(1 - e^{-\theta}) \geq \frac{1}{n}\theta e^{-2\theta}$$

である(6-7)．

つぎに，$\tau(\theta) = \theta$ とおく．

$$\sum_{i=1}^n \frac{\partial}{\partial\theta}\log f(X_i|\theta) = \sum_{i=1}^n\left(-1 + \frac{X_i}{\theta}\right) = \frac{n}{\theta}(\bar{X}_n - \theta)$$

となり，(6.2) において $K(\theta, n) = n/\theta$ とおけば，$\bar{X}_n$ は $\tau(\theta) = \theta$ の UMVUE となることがわかる．

6.4 十分性と完備性

定理 6.4 (Rao-Blackwell) $X_1, X_2, \ldots, X_n$ を確率（密度）関数 $f(x|\theta), \theta \in \Theta$ からの大きさ $n\,(n \geq 2)$ のランダム標本とし，$S_i = S_i(X_1, X_2, \ldots, X_n), i = 1, 2, \ldots, k$ を十分統計量とする．$T = T(X_1, X_2, \ldots, X_n)$ を $\tau(\theta)$ の不偏推定量で $\mathbb{VAR}(T) < \infty$ なるものとする．推定量 $T' = T'(X_1, X_2, \ldots, X_n)$ を $T' = \mathbb{E}[T|\, S_1, \ldots, S_k]$ で定義する．このとき，
(i) T' は統計量で $S_1, \ldots, S_k$ を通してのみ $(X_1, X_2, \ldots, X_n)$ に依存する．したがって，$T' = T'(S_1, \ldots, S_k)$ を記すことにする．
(ii) $\mathbb{E}_\theta[T'] = \tau(\theta)$ となる．すなわち，T' は $\tau(\theta)$ の不偏推定量である．
(iii) すべての $\theta \in \Theta$ に対し

$$\mathbb{VAR}(T') \leq \mathbb{VAR}(T)$$

が成立する．ある θ に対し，$\mathbb{P}(T = T') \neq 1$ ならば，

$$\mathbb{VAR}(T') < \mathbb{VAR}(T)$$

が成立する．

証明 $S_1, \ldots, S_k$ の十分性から $S_1, \ldots, S_k$ を与えたときの $(X_1, X_2, \ldots, X_n)$ の条件付き分布は θ に依存しない．したがって，$S_1, \ldots, S_k$ の十分性から $S_1, \ldots, S_k$ を与えたときの T の条件付き分布は θ に依存しないので，$\mathbb{E}_\theta[T|\, S_1, \ldots, S_k]$ は θ に依存しないので，統計量となる．また，これは $S_1, \ldots, S_k$ の関数となることは明らである．

$$\mathbb{T}' = \mathbb{E}_\theta[\mathbb{E}[T|\, S_1, \ldots, S_k]] = \mathbb{E}_\theta[T] = \tau(\theta)$$

となる．したがって，T' は $\tau(\theta)$ の不偏推定量となる．

$$\begin{aligned}
\mathbb{VAR}_\theta[T] &= \mathbb{E}_\theta[(T - \mathbb{E}_\theta(T))^2] \\
&= \mathbb{E}_\theta[(T - T' + T' - \mathbb{E}_\theta(T))^2] \\
&= \mathbb{E}_\theta[\{T - T'\}^2] + 2\mathbb{E}_\theta[(T - T')(T' - \mathbb{E}_\theta(T))] + \mathbb{VAR}_\theta(T')
\end{aligned}$$

となる．しかし，

$$\begin{aligned}
\mathbb{E}_\theta[(T - T')(T' - \mathbb{E}_\theta(T))] &= \mathbb{E}_\theta[\mathbb{E}[(T - T')(T' - \mathbb{E}_\theta(T))|\, S_1, \ldots, S_k]] \\
&= \mathbb{E}_\theta[(T' - \mathbb{E}_\theta(T))\mathbb{E}[T - T'|\, S_1, \ldots, S_k]] = 0
\end{aligned}$$

となる．よって，

$$\mathbb{VAR}_\theta[T] = \mathbb{E}_\theta[\{T - T'\}^2] + \mathbb{VAR}_\theta(T') \geq \mathbb{VAR}_\theta(T')$$

を得る．さらに，

$$\mathbb{E}_\theta[\{T - T'\}^2] \Longleftrightarrow \mathbb{P}_\theta(T - T' = 0) = 1$$

に注意すればよい． □

注意 6.8 Rao-Blackwell の定理の証明の最後の部分は以下のように確認できる．W を確率変数とし，$\mathbb{E}[W^2] < \infty$ とする．このとき，

$$\mathbb{E}[W^2] \Longleftrightarrow P(W = 0) = 1$$

を示せばよい．まず，$\Leftarrow$ は自明である．したがって，$\Rightarrow$ を示せばよい．任意の正数 ϵ に対し

$$\mathbb{E}[W^2] \geq \mathbb{E}[W^2 I\{W \geq \epsilon\}] \geq \epsilon^2 \mathbb{E}[I\{W \geq \epsilon\}] = \epsilon^2 \mathbb{P}\{W \geq \epsilon\}$$

となることより，$\mathbb{P}\{W \geq \epsilon\} = 0$ となる．同様にすれば，$\mathbb{P}\{W \leq \epsilon\} = 0$ を得る．したがって，$\mathbb{P}\{-\epsilon \leq W \leq \epsilon\} = 1$ となる．いま，$A_n = \{-1/n \leq W \leq 1/n\}\,(n = 1, 2, \ldots)$ とおけば，$A_1 \supset A_2 \supset \cdots$ となる．また，$\{W = 0\} = \cap_{n=1}^{n} A_n$ であり，$\mathbb{P}_\theta(A_n) = 1$ である．これらより

$$\mathbb{P}_\theta(W = 0) = \mathbb{P}_\theta\{\cap_{n=1}^{n} A_n\} = \lim_{n\to\infty} P(A_n) = 0$$

となることよりわかる．

例 6.12 $X_1, X_2, \ldots, X_n$ をベルヌーイ試行 $f(x|\,\theta) = \theta^x(1-\theta)^{1-x} I_{\{0,1\}}(x)$ からの大きさ $n(n \geq 2)$ のランダム標本とし，$\tau(\theta) = \theta$ の推定を考える．明らかに，$T(X_1, X_2, \ldots, X_n) = X_1$ は θ の不偏推定量になる．しかし，この推定量は効率がわるいので，十分統計量 $S = \sum_{i=1}^{n} X_i$ を用いて，修正することを考える．

$$\tilde{T} = \mathbb{E}_[X_1|\,S]$$

とおく．この推定量を具体的に計算してみよう．

$$\begin{aligned}
\mathbb{P}[X_1 = 0|\sum_{i=1}^{n} X_i = s] &= \frac{\mathbb{P}[X_1 = 0 \quad \text{かつ} \quad \sum_{i=1}^{n} X_i = s]}{\mathbb{P}[\sum_{i=1}^{n} X_i = s]} \\
&= \frac{\mathbb{P}[X_1 = 0 \quad \text{かつ} \quad \sum_{i=2}^{n} X_i = s]}{\mathbb{P}[\sum_{i=1}^{n} X_i = s]} \\
&= \frac{\mathbb{P}[X_1 = 0]\mathbb{P}[\sum_{i=2}^{n} X_i = s]}{\mathbb{P}[\sum_{i=1}^{n} X_i = s]} \\
&= \frac{(1-\theta)\begin{pmatrix} n-1 \\ s \end{pmatrix}\theta^s(1-\theta)^{n-1-s}}{\begin{pmatrix} n \\ s \end{pmatrix}\theta^s(1-\theta)^{n-s}} = \frac{n-s}{n}
\end{aligned}$$

となる．また，

$$\begin{aligned}
\mathbb{P}[X_1 = 1|\sum_{i=1}^{n} X_i = s] &= \frac{\mathbb{P}[X_1 = 1 \quad \text{かつ} \quad \sum_{i=1}^{n} X_i = s]}{\mathbb{P}[\sum_{i=1}^{n} X_i = s]} \\
&= \frac{\mathbb{P}[X_1 = 1 \quad \text{かつ} \quad \sum_{i=2}^{n} X_i = s-1]}{\mathbb{P}[\sum_{i=1}^{n} X_i = s]} \\
&= \frac{\mathbb{P}[X_1 = 1]\mathbb{P}[\sum_{i=2}^{n} X_i = s-1]}{\mathbb{P}[\sum_{i=1}^{n} X_i = s]} \\
&= \frac{\theta\begin{pmatrix} n-1 \\ s-1 \end{pmatrix}\theta^s(1-\theta)^{n-1-(s-1)}}{\begin{pmatrix} n \\ s \end{pmatrix}\theta^s(1-\theta)^{n-s}} = \frac{n}{n}
\end{aligned}$$

となる．したがって

$$\tilde{T} = \mathbb{E}_[X_1|\,S] = 0\cdot\mathbb{P}[X_1 = 0|\,S] + 1\cdot\mathbb{P}[X_1 = 1|\,S] = \frac{S}{n}$$

定義 6.9 (完備性) $X_1, X_2, \ldots, X_n$ を確率 (密度) 関数 $f(x|\,\theta), \theta \in \Theta$ からの大きさ n のランダム標本とし，$T = T(X_1, X_2, \ldots, X_n)$ を統計量とする．統計量 T の標本分布の確率 (密度) 関数の族が完備であるとは，次の条件をみたすことである．統計量 T の値域上で定義された可測関数 g が，すべての $\theta \in \Theta$ に対し，$\mathbb{E}_\theta[g(T)] = 0$ をみたすならば，$\mathbb{P}_\theta(g(T) = 0) = 1$ が成立する．

さらに，統計量 T が完備であるとは，統計量 T の標本分布の確率 (密度) 関数の族が完備であることをいう．

例 6.13 $X_1, X_2, \ldots, X_n$ をベルヌーイ試行 $f(x|\theta) = \theta^x(1-\theta)^{1-x}I_{\{0,1\}}(x), \theta(0,1)$ からの大きさ $n\,(n \geq 2)$ のランダム標本とする．たとえば，統計量 $\tilde{T} = X_1 - X_2$ は完備ではない．なぜならば，$\mathbb{E}_\theta[X_1 - X_2] = 0$ であるが，$\mathbb{P}(X_1 - X_2 = 0) \neq 1$ であることよりわかる．

次に，$T = \sum_{i=1}^n X_i$ を考える．$g(T)$ をすべての $\theta \in (0,1)$ に対し $\mathbb{E}_\theta[g(T)] = 0$ をみたす可測関数とする．T が完備であることを示すために，$t = 0, 1, \ldots, n$ に対して $g(t) = 0$ を示せばよい．いま，$0 < \theta < 1$ に対し

$$0 = \mathbb{E}_\theta[g(T)] = \sum_{t=0}^n \begin{pmatrix} n \\ t \end{pmatrix} \theta^t(1-\theta)^{n-t} = (1-\theta)^t \sum_{t=0}^n g(t) \begin{pmatrix} n \\ t \end{pmatrix} \left(\frac{\theta}{1-\theta}\right)$$

に注意すれば，すべての $0 < \theta < 1$ に対し，

$$\sum_{t=0}^n g(t) \begin{pmatrix} n \\ t \end{pmatrix} \left(\frac{\theta}{1-\theta}\right)^t = 0$$

となることがわかる．または

$$\sum_{t=0}^n g(t) \begin{pmatrix} n \\ t \end{pmatrix} \alpha^t = 0, \qquad \alpha = \frac{\theta}{1-\theta}$$

となる．$\{\alpha^t, t = 0, 1 \ldots, n\}$ は一次独立なので，

$$g(t) \begin{pmatrix} n \\ t \end{pmatrix} = 0, \qquad t = 0, 1 \ldots, n$$

となる．$\begin{pmatrix} n \\ t \end{pmatrix} \neq 0$ なので，$g(t) = 0\,(t = 0, 1 \ldots, n)$ となる．

例 6.14 $X_1, X_2, \ldots, X_n$ を一様分布 $f(x) = I_{(0,1)}(x)$ からの大きさ n のランダム標本とする．ただし，$\Theta = \{\theta, \theta > 0\}$ である．$Y_n = \max(X_1, X_2, \ldots, X_n)$ が完備であることを示そう．そのために，すべての $\theta \in \Theta$ に対し，$\mathbb{E}_\theta[g(Y_n)] = 0$ ならば，$\mathbb{P}_\theta\{g(Y_n) = 0\} = 1$ を示せばよい．

$$\mathbb{E}_\theta[g(Y_n)] = \int g(y) f_{Y_n}(y)\,dy = \int g(y)\theta^{-n} n y^{n-1} I_{(0,\theta)}(y)\,dy$$

となり，すべての $\theta \in \Theta$ に対し

$$\mathbb{E}_\theta[g(Y_n)] = 0 \Longleftrightarrow \frac{n}{\theta^n} \int_0^\theta g(y) y^{n-1}\,dy = 0$$

となる．したがって，θ に関して微分すれば

$$g(y)\theta^{n-1} = 0 \Longleftrightarrow g(y) = 0$$

となることがわかる．

定理 6.5 Lehmann-Sheffée $X_1, X_2, \ldots, X_n$ を $f(x|\theta)$ からの大きさ n のランダム標本とし，$\tau : \Theta \to \mathbb{R}$ とする．$S = S(X_1, X_2, \ldots, X_n)$ が完備十分統計量で $T^* = T^*(S)$ が $\tau(\theta)$ の不偏推定量ならば，T^* は $\tau(\theta)$ の UMVUE である．

証明 $T' = T'(S)$ を $\tau(\theta)$ の不偏推定量とする．すると任意の $\theta \in \Theta$ に対し

$$\mathbb{E}_\theta[T' - T^*] = 0$$

となる．S の完備性から

$$\mathbb{P}_theta[T'(S) = T^*(S)] = 0$$

となる．S の関数である不偏推定量は唯一である．

次に，T を任意の $\tau(\theta)$ の推定量とする．Rao-Blackwell の定理から $\mathbb{E}[T|\,S]$ は $\tau(\theta)$ の不偏推定量で S の関数となるので

$$T^* = \mathbb{E}[T|\,S], \quad a.s.$$

となる．さらに，Rao-Blackwell の定理から

$$\mathbb{VAR}_\theta[T^*] \leq \mathbb{VAR}_\theta[T]$$

となる．よって，T^* は UMVUE である．□

例 6.15 $X_1, X_2, \ldots, X_n$ をポアソン分布 $f(x|\,\theta) = e^{-\theta}\theta^x/x!\,(x = 0, 1, \ldots)$ からの大きさ $n\,(n \geq 2)$ のランダム標本とする．$\sum_{i=1}^n X_i$ は母数 $n\theta$ のポアソン分布に従う．いま，$T = \sum_{i=1}^n X_i$ とする．

$$\begin{aligned}
\mathbb{E}_\theta[g(T)] &= \sum_{t=0}^\infty g(t)\frac{e^{-n\theta}(n\theta)^t}{t!} = e^{-n\theta}\sum_{t=0}^\infty g(t)\frac{n^t}{t!}\theta^t = 0 \\
&\Leftrightarrow \sum_{t=0}^\infty g(t)\frac{n^t}{t!}\theta^t = 0 \qquad \theta \text{ の多項式は一次独立} \\
&\Leftrightarrow g(t) = 0
\end{aligned}$$

となる．したがって，$\sum_{i=1}^n X_i$ は完備統計量である．

$\tau(\theta) = e^{-\theta} = \mathbb{P}_\theta\{X_1 = 0\}$ の推定を考えよう．$I_{\{0\}}(X_1)$ は $e^{-\theta}$ の不偏推定量なので，Lehmann-Sheffée の定理から $\mathrm{T}^* = \mathbb{E}[I_{\{0\}}(X_1)|\,\sum_{i=1}^n X_i]$ は $\tau(\theta)$ の UMVUE となる．T^* がどのような推定量になるかを計算しよう．

$$\begin{aligned}
\mathbb{P}\{X_1 = 0\sum_{i=1}^n X_i = s\} &= \frac{\mathbb{P}\{X_1 = 0 \quad \text{かつ} \quad \sum_{i=1}^n X_i = s\}}{\mathbb{P}\{\sum_{i=1}^n X_i = s\}} \\
&= \frac{\mathbb{P}\{X_1 = 0 \quad \text{かつ} \quad \sum_{i=2}^n X_i = s\}}{\mathbb{P}\{\sum_{i=1}^n X_i = s\}} \\
&= \frac{\mathbb{P}\{X_1 = 0\}\mathbb{P}\{\sum_{i=2}^n X_i = s\}}{\mathbb{P}\{\sum_{i=1}^n X_i = s\}} \\
&= \frac{e^{-\theta}e^{-(n-1)\theta}[(n-1)\theta]^s/s!}{e^{-n\theta}[n\theta]^s/s!} = \left(\frac{n-1}{n}\right)^s
\end{aligned}$$

となる．したがって

$$\mathbb{E}[I_{\{0\}}(X_1)|\sum_{i=1}^n X_i = s]\mathbb{P}\{X_1 = 0|\sum_{i=1}^n X_i = s\} = \left(\frac{n-1}{n}\right)^s$$

となる．故に

$$T^* = \left(\frac{n-1}{n}\right)^{\sum_{i=1}^n X_i}$$

は $\tau(\theta) = e^{-\theta}$ の UMVUE である．