数理解析研究所講究録 1273 巻 2002 年 1-15

Unbiased Estimation in Sequential Binomial Sampling Experiments

筑波大・数理物質科学 谷尾 高志 (Takashi Tanio)
筑波大・数学 赤平 昌文 (Masafumi Akahira)

1 はじめに

統計的逐次実験の基礎となる逐次２項標本抽出計画 (sequential binomial sampling plan) において, その母数 p の関数 $g(p)$ の推定問題が考察されている (Sinha and Sinha [SS75], Ghosh and Sen [GS91], Ghosh, Mukhopadhyay and Sen [GMS97], Akahira and Koike[AK98], Hubert and Pyke[HP00]). たとえば, [SS75] において, ある非ランダム停止則の下で, $g(p) = 1/p$ が不偏推定可能であるための必要条件, 十分条件について論じられ, また, 抽出計画が閉であるための必要十分条件等についても考察された. 一方, 適当な停止則の下で, $g(p)$ が不偏推定可能であるための必要条件として, Bhandari and Bose [BB90] は「$g(p)$ が p の連続関数である」と主張したが, 後に Akahira, Takeuchi and Koike [ATK92] によってその命題の証明の誤りが指摘されるとともに, $g(p)$ が不連続関数であるが, 不偏推定可能であるという反例も示され, [BB93] においてその命題の訂正が行われたという経緯がある.

本論では, [ATK92] に従って, この抽出計画において適当なランダム停止則の下で, 一般の関数 $g(p)$ が不偏推定可能であるための十分条件について論じる. そして, [BB90] においては, 連続であるが微分不可能な関数 $g(p) = \min(p, 1-p)$ $(0 < p < 1)$ の不偏推定量がアドホックな方法で得られたが, 本論では [ATK92] の系統的な方法で構成されることを示し, 他の関数についてもそのことを確かめる. また, あるランダム停止則を伴う, $g(p)$ の不偏推定量の分散に関する不等式を求めて, その例についても述べる.

2 逐次２項標本抽出計画における母数 p の関数 $g(p)$ の推定

この節では, [ATK92] に従って, 逐次２項標本抽出計画における母数 p の関数 $g(p)$ の不偏推定可能性について述べる. まず, $X_1, X_2, \ldots, X_n, \ldots$ を互いに独立に, いずれもベルヌーイ分布 $Ber(p)$ に従う確率変数列とする. この列の任意の実現値は xy 平面上の第一象限における経路として見なせる. すなわち, 原点から出発し, i 回目の試行で $X_i = 0$ ならば y 方向へ一歩, $X_i = 1$ ならば x 方向へ一歩, というように動くとする. また, 経路の集合を

$$R_n := \{(x_1, \ldots, x_n) | x_i = 0, 1; i = 1, \ldots, n\}$$

とし, $R := \cup_{n=1}^{\infty} R_n$ と定義する. そして, S_n を R_n の部分集合とし, 大きさ n の観測列が S_n に属するとき, 標本抽出を停止する. さらに, $S := \cup_{n=1}^{\infty} S_n$ とし, ある正整数 M について $(i_1, \ldots, i_M) \in S_M$ とする. また任意の $k < M$ について $(i_1, \ldots, i_k) \notin S_k$ とする. 各経路 $(X_1, \ldots, X_n) \in S_n$ に対して, 最終点 (X, Y) を $X = \sum_{i=1}^{n} X_i, Y = \sum_{i=1}^{n}(1 - X_i)$ と定義する. 標本抽出を止める最終点を T で表し, 停止則を上記のように定義する.

このとき, $T = (x, y)$ で停止する確率は

$$P\{T = (x, y)\} = N(x, y) p^x (1-p)^y$$

になる. 但し, $N(x,y)$ は, S における, 原点から点 (x,y) までの経路数であり, p には無関係になる. いま, 停止則は $\sum_x\sum_y P\{T=(x,y)\}=1$ という意味で閉であると仮定する. このとき,

$$\sum_{x'\leq x, y'\leq y} N(x',y')_{x+y-x'-y'}C_{x-x'} \leq_{x+y} C_x$$

となる.

次に, 停止則を伴う $g(p)$ の任意の推定量 $\pi(X,Y)$ に対して,

$$E_p[\pi(X,Y)] = \sum_n \sum_{x+y=n} \pi(x,y)N(x,y)p^x(1-p)^y$$

になる. そして, $g(p)$ が不偏推定可能である, すなわち

$$E_p[|\pi(X,Y)|] < \infty$$

を満たすような $g(p)$ の不偏推定量 $\pi(X,Y)$ が存在するための十分条件について考える. いま, $g(p)$ の漸近不偏推定量 $\hat{g}_n(X,Y)$, つまり, 任意の p $(0<p<1)$ について

$$\lim_{n\to\infty} \sum_{x+y=n} \hat{g}_n(x,y)_n C_x p^x(1-p)^y = g(p)$$

をみたす $\hat{g}_n(X,Y)$ が存在すると仮定する. ただし, $\hat{g}_n(x,y)$ は $x+y=n, 0\leq x\leq n$ に対して定義されるとする. このとき

$$g_n(p) := \sum_{x+y=n} \hat{g}_n(x,y)_n C_x p^x(1-p)^y$$

とすれば, 任意の p について $\lim_{n\to\infty} g_n(p) = g(p)$ となる. また,

$$\begin{aligned}
&h_n(p) := g_n(p) - g_{n-1}(p) \quad (n=1,2,\ldots),\\
&h_0(p) = 0,\\
&\hat{g}_0(0,0) = \hat{g}_{n-1}(-1,n) = \hat{g}_{n-1}(n,-1) = 0 \quad (n=1,2,\ldots)
\end{aligned}$$

と定義すれば, $\sum_{n=1}^{\infty} h_n(p) = g(p)$ となる. さらに

$$\begin{aligned}
&G_n(x,y) := \hat{g}_n(x,y) - \frac{x}{n}\hat{g}_{n-1}(x-1,y) - \frac{y}{n}\hat{g}_{n-1}(x,y-1),\\
&G_0(0,0) := 0
\end{aligned}$$

と定義すると, 各 $n=1,2,\ldots$ について, 次のことが成り立つ.

補題 2.1 各 $n=1,2,\ldots$ について

$$h_n(p) = \sum_{x+y=n} G_n(x,y)_n C_x p^x(1-p)^y$$

証明については $g_n(p)$ の定義から導かれるが, 省略する（詳しくは [ATK92] 参照）. 補題 2.1 より, $g(p)$ の推定量を

$$e(x,y) := G_n(x,y)_nC_x/N(x,y) \quad (x+y=n=1,2,\ldots), \tag{2.1}$$
$$e(0,0) := 0$$

とすると,

$$\sum_n \sum_{x+y=n} e(x,y)N(x,y)p^x(1-p)^y = g(p)$$

になる. よって, 次のことを得る.

補題 2.2 $g(p)$ の漸近不偏推定量が存在するとき

$$\sum_n \sum_{x+y=n} |e(x,y)|N(x,y)p^x(1-p)^y < \infty \tag{2.2}$$

ならば, $g(p)$ は不偏推定可能である.

(2.2) に対する十分条件は, 任意の p について

$$\begin{aligned}&\sum_{x+y=n} |e(x,y)|N(x,y)p^x(1-p)^y\\ &\quad = \sum_{x+y=n} |G_n(x,y)|_nC_xp^x(1-p)^y \le c_n\end{aligned}$$

となり,

$$\sum_n c_n < \infty$$

を満たすような正数列 $\{c_n\}$ が存在することである.

3 関数 $g(p)$ の不偏推定可能性

Bhandari and Bose [BB90] は逐次２項標本計画において, 連続であるが, 微分可能でない関数 $g(p)=\min(p,1-p)$ $(0<p<1)$ の不偏推定量をアドホックに求めたが, ここでは, 第２節の系統的な方法を用いて, 実際に $g(p)$ の不偏推定量を構成する. なお, 論文 [BB90] には主要定理に誤りがあることに注意 ([BB93] 参照).

例 3.1 まず,

$$\begin{aligned}g(p) &= \begin{cases} 1-p & (0<p<1/2),\\ 1/2 & (p=1/2),\\ p & (1/2<p<1)\end{cases}\\ &= \frac{1}{2} - \left|p-\frac{1}{2}\right|\end{aligned}$$

より, $h(p) := |p-1/2|$ の不偏推定を考える. いま,

$$\begin{aligned}\hat{g}_n(x,y) &:= \left|\frac{x}{n}-\frac{1}{2}\right|\\ &= \Big(\frac{x}{x+y}-\frac{1}{2}\Big)\mathrm{sgn}\Big(\frac{x}{x+y}-\frac{1}{2}\Big)\\ &= \Big(\frac{x}{x+y}-\frac{1}{2}\Big)\mathrm{sgn}(x-y)\end{aligned}$$

(Ⅰ) $n=x+y\geq 2$ の場合, まず,

$$\begin{aligned}G_n(x,y)&=\hat{g}_n(x,y)-\frac{x}{n}\hat{g}_{n-1}(x-1,y)-\frac{y}{n}\hat{g}_{n-1}(x,y-1)\\&=\Big(\frac{x}{x+y}-\frac{1}{2}\Big)\Big\{\mathrm{sgn}(x-y)\Big\}\\&\quad-\frac{x}{n}\Big(\frac{x-1}{x+y-1}-\frac{1}{2}\Big)\Big\{\mathrm{sgn}(x-y-1)\Big\}\\&\quad-\frac{y}{n}\Big(\frac{x}{x+y-1}-\frac{1}{2}\Big)\Big\{\mathrm{sgn}(x-y+1)\Big\}\end{aligned}$$

になるから, 次の (i)〜(v) の場合について考える.

(i) $y>x-1$ のとき,

$$\begin{aligned}G_n(x,y)&=\Big(\frac{x}{x+y}-\frac{1}{2}\Big)-\frac{x}{n}\Big(\frac{x-1}{x+y-1}-\frac{1}{2}\Big)-\frac{y}{n}\Big(\frac{x}{x+y-1}-\frac{1}{2}\Big)\\&=0.\end{aligned}$$

(ii) $y=x-1$ のとき, $G_n=0$.

(iii) $x=y$ のとき,

$$G_n=\frac{1}{2(n-1)}.$$

(iv) $y=x+1$ のとき, $G_n=0$.

(v) $y>x+1$ のとき, $G_n=0$.

(Ⅱ) $n=x+y=1$ の場合.

$$\begin{aligned}G_1(x,y)&=\hat{g}_1(x,y)-x\hat{g}_0(x-1,y)-y\hat{g}_0(x,y-1)\\&=\begin{cases}\hat{g}_1(1,0)-\hat{g}_0(0,0) & (x=1,y=0\text{ のとき}),\\ \hat{g}_1(0,1)-\hat{g}_0(0,0) & (x=0,y=1\text{ のとき})\end{cases}\\&=0.\end{aligned}$$

よって, (Ⅰ), (Ⅱ) から

$$G_n(x,y)=\begin{cases}\frac{1}{2} & (x+y=1\text{ のとき}),\\ -\frac{1}{2n-2} & (x=y,x+y=n=2,4,6,\ldots\text{ のとき}),\\ 0 & (\text{その他のとき})\end{cases}$$

となる.

そこで, (2.1) より $h(p)$ の推定量として, 次の $e(\cdot,\cdot)$ を考える.

$$\begin{aligned}&e(x,y)=G_n(x,y)_nC_x/N(x,y) \qquad (x+y=n=1,2,3,\ldots),\\&e(0,0)=0.\end{aligned}$$

次に, この推定量 e が $E[|e(x,y)|]<\infty$ を満たすことを示す. このことを示すには

$$E[|e(x,y)|]=\sum_{x+y=n}|e(x,y)|N(x,y)p^x(1-p)^y$$

$$= \begin{cases} \sum_{x+y=n} |G_n(x,y)|_n C_x p^x (1-p)^y & (n \geq 2), \\ \sum_{x+y=1} \frac{1}{2} {}_1C_x p^x (1-p)^y (= \frac{1}{2}) & (n = 1) \end{cases}$$

となるので, $c_n := \sum_{x+y=n} = |G_n(x,y)|_n C_x p^x (1-p)^y$ とおいて, $n \geq 2$ について $c_n \leq c'_n, \sum_{n=2}^{\infty} c'_n < \infty$ をみたす $\{c'_n\}$ が存在することを示せばよい. いま, $n = 3, 5, 7. \ldots$ のときは $c_n = 0$ であるので, $n = 2, 4, 6, \ldots$ に対して

$$\begin{aligned} & \sum_{x+y=n} |G_n(x,y)|_n C_x p^x (1-p)^y \\ &= \sum_{x+y=\frac{n}{2}} \frac{1}{2(n-1)} {}_n C_x p^x (1-p)^y \quad (n = 2, 4, 6, \ldots) \\ &= \frac{1}{2(n-1)} {}_n C_{n/2} p^{n/2} (1-p)^{n/2} \\ &= \frac{1}{2(2m-1)} {}_{2m} C_m p^m (1-p)^m \quad (n = 2m = 2, 4, 6, \ldots) \\ &=: d_m \quad (m = 1, 2, 3, \ldots) \end{aligned}$$

となる. ここで, スターリング (Stirling) の公式 $m! = \sqrt{2\pi} e^{-m} m^{m+(1/2)} \{1 + O(1/m)\}$ を用いて

$$\frac{(2m)!}{(m!)^2} = \frac{2^{2m+(1/2)}}{\sqrt{2\pi} m^{(1/2)}} \left\{1 + O(\tfrac{1}{m})\right\}$$
$$p^m (1-p)^m \leq \frac{1}{2^{2m}}$$

となるから

$$\begin{aligned} \sum_{m=1}^{\infty} d_m &\leq \sum_{m=1}^{\infty} \frac{1}{2(m-1)} \frac{2^{2m+(1/2)}}{\sqrt{2\pi} m^{(1/2)}} \left\{1 + O(\frac{1}{m})\right\} \frac{1}{2^{2m}} \\ &= \frac{1}{2\sqrt{\pi}} \sum_{m=1}^{\infty} O\Big(\frac{1}{m\sqrt{m}}\Big) < \infty \end{aligned}$$

になる. よって, 補題 2.2 より, $h(p)$ は不偏推定可能である, つまり, $g(p)$ は不偏推定可能である.

次に, 実際に $g(p)$ の不偏推定量を求める. まず

$$h(p) = E[e(X,Y)] = \frac{1}{2} + \sum_{n=2}^{\infty} \sum_{x+y=n} G_n(x,y)_n C_x p^x (1-p)^x$$

より,

$$g(p) = \frac{1}{2} - h(p) = \sum_{n=2}^{\infty} \sum_{x+y=n} \left\{ - \frac{G_n(x,y)_n C_x}{N(x,y)} \right\} N(x,y) p^x (1-p)^y$$

となるから, $e^*(x,y) := -G_n(x,y)_n C_x / N(x,y)$ とすれば, $e^*(x,y)$ は $g(p)$ の不偏推定量となる. そこで, 実際に求めてみると

$$
\begin{aligned}
e^*(x,y) &= \begin{cases} 0 & (x+y=1\text{ のとき}), \\ \frac{{}_nC_x}{2(n-1)N(x,y)} & (x=y, T=(x,x)\text{ のとき}), \\ 0 & (\text{その他のとき}) \end{cases} \\
&= \begin{cases} \frac{1}{2(2x-1)N(x,y)}\frac{(2x)!}{(x!)^2} & (x=y, T=(x,y)\text{ のとき}), \\ 0 & (\text{その他のとき}) \end{cases} \\
&= \begin{cases} \frac{(2x-2)!}{(x-1)!x!N(x,y)} & (x=y, T=(x,y)\text{ のとき}), \\ 0 & (\text{その他のとき}) \end{cases}
\end{aligned}
$$

となり, これは, Bhandari and Bose [BB90] がアドホックに得たものと一致している.
また,

$$
\begin{aligned}
E[e^*(x,y)] &= \sum_{n=2}^{\infty}\sum_{x+y=n} e^*(x,y)N(x,y)p^x(1-p)^y \\
&= \sum_{m=1}^{\infty}\frac{(2m-2)!}{(m-1)!m!N(m,m)}N(m,m)p^m(1-p)^m \\
&= \sum_{m=1}^{\infty}\frac{(2m-1)!}{(m-1)!m!}\left(\frac{\theta}{4}\right)^m \qquad (\theta := 4p(1-p)) \\
&= \frac{1}{2}\sum_{m=1}^{\infty}\frac{(2m)!}{2^{2m}m!(2m-1)}\frac{\theta^m}{m!} \\
&= \frac{1}{2}(1-\sqrt{1-\theta}) \\
&= \frac{1}{2}-\frac{1}{2}|2p-1| \\
&= \min(p, 1-p) \\
&= g(p)
\end{aligned}
$$

より, $e^*(x,y)$ が $g(p)$ 不偏推定量となっていることは確かめられる.

例 3.2 p の関数

$$
g(p) := \begin{cases} 1/(1-p) & (0<p<1/2), \\ 2 & (p=1/2), \\ 1/p & (1/2<p<1) \end{cases}
$$

の不偏推定可能性について考える. この関数 g は区間 $(0,1)$ において連続であるが, 微分可能でないことに注意. いま, 各 $n=1,2,\ldots$ について

$$
\hat{g}_n(x,y) := \begin{cases} \frac{n}{y} & (0 \le \frac{x}{n} < \frac{1}{2}-\epsilon_n), \\ 2 & (\frac{1}{2}-\epsilon_n \le \frac{x}{n} \le \frac{1}{2}+\epsilon_n), \\ \frac{n}{x} & (\frac{1}{2}+\epsilon_n < \frac{x}{n} \le 1) \end{cases}
$$

とおく. 但し, $x+y=n$ で, ϵ_n は $\epsilon_n \to 0$ となる適当な正数とする. まず, $\hat{g}_n(x,y)$ が有界であることを示す.

(i) $0 < x/n < (1/2) - \epsilon_n$ のとき

$$0 < \frac{x}{n} < \frac{1}{2} - \epsilon_n, \quad 0 < \frac{n-y}{n} < \frac{1}{2} - \epsilon_n, \quad 0 < 1 - \frac{y}{n} < \frac{1}{2} - \epsilon_n,$$

$$\frac{1}{2} + \epsilon_n < \frac{y}{n} < 1, \quad 1 < \frac{n}{y} < \frac{2}{1+2\epsilon_n} < 2.$$

(ii) $(1/2) + \epsilon_n < x/n < 1$ のとき

$$\frac{2}{1+2\epsilon_n} < \frac{n}{x} < 1.$$

よって, $\hat{g}_n(x,y)$ は有界である. また, $n \to \infty$ のとき $X/n \xrightarrow{p} p$(確率収束) より, $g_n(p)$ $E[\hat{g}_n(X,Y)]$ とすると

$$\lim_{n\to\infty} g_n(p) = g(p)$$

となる. 次に,

$$G_n(x,y) = \hat{g}_n(x,y) - \frac{x}{n}\hat{g}_{n-1}(x-1,y) - \frac{y}{n}\hat{g}_{n-1}(x,y-1)$$

として,

$$\hat{g}_{n-1}(x-1,y) = \begin{cases} \frac{n-1}{y} & (0 < \frac{x-1}{n-1} < \frac{1}{2} - \epsilon_{n-1}), \\ 2 & (\frac{1}{2} - \epsilon_{n-1} < \frac{x-1}{n-1} < \frac{1}{2} + \epsilon_{n-1}), \\ \frac{n-1}{x-1} & (\frac{1}{2} + \epsilon_{n-1} < \frac{x-1}{n-1} < 1) \end{cases}$$

$$= \begin{cases} \frac{n-1}{y} & (1 < x < \frac{n+1}{2} - (n-1)\epsilon_{n-1}), \\ 2 & (\frac{n+1}{2} - (n-1)\epsilon_{n-1} < x < \frac{n+1}{2} + (n-1)\epsilon_{n-1}), \\ \frac{n-1}{x-1} & (\frac{n+1}{2} + (n-1)\epsilon_{n-1} < x < n) \end{cases}$$

$$\hat{g}_{n-1}(x,y-1) = \begin{cases} \frac{n-1}{y-1} & (0 < \frac{x}{n-1} < \frac{1}{2} - \epsilon_{n-1}), \\ 2 & (\frac{1}{2} - \epsilon_{n-1} < \frac{x}{n-1} < \frac{1}{2} + \epsilon_{n-1}), \\ \frac{n-1}{x} & (\frac{1}{2} + \epsilon_{n-1} < \frac{x}{n-1} < 1) \end{cases}$$

$$= \begin{cases} \frac{n-1}{y-1} & (0 < x < \frac{n-1}{2} - (n-1)\epsilon_{n-1}), \\ 2 & (\frac{n-1}{2} - (n-1)\epsilon_{n-1} < x < \frac{n-1}{2} + (n-1)\epsilon_{n-1}), \\ \frac{n-1}{x} & (\frac{n-1}{2} + (n-1)\epsilon_{n-1} < x < n-1), \end{cases}$$

また,

$$\hat{g}_0(0,0) = \hat{g}_{n-1}(-1,n) = \hat{g}_{n-1}(n,-1) = 0 \quad (n = 1, 2, \cdots)$$

より

$$\begin{aligned} G_0(0,0) &= 0, \\ G_n(0,y) &= \hat{g}_n(0,y) - \frac{y}{n}\hat{g}_{n-1}(0,y-1) = \hat{g}_n(0,y) - \hat{g}_{n-1}(0,y-1) = 0, \\ G_n(x,0) &= \hat{g}_n(x,0) - \frac{x}{n}\hat{g}_{n-1}(x-1,0) = \hat{g}_n(x,0) - \hat{g}_{n-1}(x-1,0) = 0 \end{aligned}$$

となる. ここで, $0 < n\epsilon_n < 1/2$ とすると, $x > 1,\ y > 1$ になる.
(i) $x < y - 1$ のとき, $x < (n-1)/2$ となり,

$$\begin{aligned} G_n(x,y) &= \frac{n}{y} - \frac{x}{n} \cdot \frac{n-1}{y} - \frac{y}{n} \cdot \frac{n-1}{y-1} \\ &= \frac{x+y}{y} - \frac{x}{x+y} \cdot \frac{x+y-1}{y} - \frac{y}{x+y} \cdot \frac{x+y-1}{y-1} \\ &= \frac{(y-1)(x+y)^2 - x(y-1)(x+y-1) - y^2(x+y-1)}{y(y-1)(x+y)} \\ &= -\frac{x}{y(y-1)(x+y)} \end{aligned}$$

となる. 同様にして,
(ii) $x = y - 1$ のとき, $x = (n-1)/2$ となり,

$$\begin{aligned} G_n(x,y) &= \frac{n}{y} - \frac{x}{n} \cdot \frac{n-1}{y} - \frac{y}{n} \cdot 2 \\ &= -\frac{1}{(x+1)(2x+1)}, \end{aligned}$$

(iii)$x = y$ のとき, $x = n/2$ となり,

$$\begin{aligned} G_n(x,y) &= 2 - \frac{x}{n} \cdot \frac{n-1}{y} - \frac{y}{n} \cdot \frac{n-1}{x} \\ &= \frac{1}{x}, \end{aligned}$$

(iv) $x = y + 1$ のとき, $x = (n+1)/2$ となり,

$$\begin{aligned} G_n(x,y) &= \frac{n}{x} - \frac{x}{n} \cdot 2 - \frac{y}{n} \cdot \frac{n-1}{x} \\ &= -\frac{1}{x(2x-1)}, \end{aligned}$$

(v) $x > y + 1$ のとき, $x > (n+1)/2$ となり,

$$\begin{aligned} G_n(x,y) &= \frac{n}{x} - \frac{x}{n} \cdot \frac{n-1}{x-1} - \frac{y}{n} \cdot \frac{n-1}{x} \\ &= -\frac{y}{x(x-1)(x+y)} \end{aligned}$$

となる. よって,

$$G_n(x,y) = \begin{cases} 0 & (x = 0, n), \\ -\frac{x}{y(y-1)(x+y)} & (0 < x < \frac{n-1}{2}), \\ -\frac{1}{(x+1)(2x+1)} & (x = \frac{n-1}{2}), \\ \frac{1}{x} & (x = \frac{n}{2}), \\ -\frac{1}{x(2x-1)} & (x = \frac{n+1}{2}), \\ -\frac{y}{x(x-1)(x+y)} & (\frac{n+1}{2} < x < n) \end{cases}$$

次に, 推定量 $e(x,y) = G_n(x,y){}_nC_x/N(x,y)$ が $E[|e(x,y)|] < \infty$ を満たすことを示す
まず,

$$
\begin{aligned}
E[|e(x,y)|] &= \sum_{n=1}^{\infty} \sum_{x+y=n} |G_n(x,y)|{}_nC_x p^x (1-p)^y \\
&=: \sum_{n=1}^{\infty} c_n
\end{aligned}
$$

とする. (i) $n = 2k-1$ $(k = 3,4,5,\ldots)$ のとき,

$$
\begin{aligned}
c_n = &\sum_{x=1}^{k-2} \frac{x}{(n-x)(n-x-1)n} \cdot \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&+ \frac{1}{k(2k-1)} \cdot \frac{n!}{(k-1)!(n-k+1)!} p^{k-1} (1-p)^{n-k+1} \\
&+ \frac{1}{k(2k-1)} \cdot \frac{n!}{k!(n-k)!} p^{k} (1-p)^{n-k} \\
&+ \sum_{y=1}^{k-2} \frac{y}{(n-y)(n-y-1)n} \cdot \frac{n!}{y!(n-y)!} p^{n-y} (1-p)^{x} \\
=: &I_1 + I_2 + I_3 + I_4.
\end{aligned}
$$

(ii) $n = 2k$ $(k = 2,3,4,\ldots)$ のとき,

$$
\begin{aligned}
c_n = &\sum_{x=1}^{k-1} \frac{x}{(n-x)(n-x-1)n} \cdot \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&+ \frac{1}{k} \cdot \frac{n!}{k!(n-k)!} p^{k} (1-p)^{k} \\
&+ \sum_{y=1}^{k-1} \frac{y}{(n-y)(n-y-1)n} \cdot \frac{n!}{y!(n-y)!} p^{n-y} (1-p)^{y} \\
=: &I_5 + I_6 + I_7,
\end{aligned}
$$

になる. そこで, $I_1,\ldots,I_7$ の評価をスターリングの公式と $p^x(1-p)^{n-x} \le (x/n)^x(1$ $(x/n))^{n-x}$ を用いて行う. まず,

$$
\begin{aligned}
I_1 &= \sum_{x=1}^{k-2} \frac{x}{(2k-x-1)(2k-x-2)(2k-1)} \cdot \frac{(2k-1)!}{x!(2k-x-1)!} p^x (1-p)^{2k-x-1} \\
&= \sum_{x=1}^{k-2} \frac{x}{(2k-x-1)(2k-x-2)(2k-1)} \\
&\quad \times \frac{\sqrt{2\pi} e^{-(2k-2)} (2k-2)^{2k-2+(1/2)}}{\sqrt{2\pi} e^{-x} x^{x+(1/2)} \cdot \sqrt{2\pi} e^{-(2k-x-1)} (2k-x-1)^{(2k-x-1)+(1/2)}} p^x (1-p)^{2k-x-1} \\
&\le \sum_{x=1}^{k-2} \frac{x}{(2k-x-1)(2k-x-2)(2k-1)} \\
&\quad \times \frac{1}{\sqrt{2\pi}} \cdot \frac{(2k-1)^{2k-(1/2)}}{x^{x+(1/2)} (2k-x-1)^{2k-x-(1/2)}} \cdot \left(\frac{x}{2k-1}\right)^x \left(\frac{2k-x-1}{2k-1}\right)^{2k-x-1}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{2\pi}} \sum_{x=1}^{k-2} \frac{x^{1/2}}{(2k-x-1)^{3/2}(2k-x-2)(2k-1)^{1/2}} \\
&\leq \frac{1}{\sqrt{2\pi}} \sum_{x=1}^{k-2} \frac{(k-2)^{1/2}}{\{2k-(k-2)-1\}^{3/2}\{2k-(k-2)-2\}(2k-1)^{1/2}} \\
&= \frac{1}{\sqrt{2\pi}} \frac{(k-2)^{3/2}}{(k+1)^{3/2}k(2k-1)^{1/2}} \\
&= O(k^{-3/2})
\end{aligned}
$$

になる. 同様にして

$$
I_2 \leq O(k^{-5/2}), I_3 \leq O(k^{-5/2}), I_4 \leq O(k^{-3/2}),
$$

$$
I_5 \leq O(k^{-3/2}), I_6 \leq O(k^{-3/2}), I_7 \leq O(k^{-3/2})
$$

が示される. よって, $n = 2k-1, n = 2k$ のいずれの場合も $c_n \leq O(k^{-3/2})$ $(n \geq 4)$ になるから, $\sum_{n=4}^{\infty} c_n < \infty$. また, $\sum_{n=1}^{3} c_n < \infty$ になるので,

$$
E[|e(x,y)|] = \sum_{n=1}^{\infty} c_n < \infty
$$

が成り立つ. よって, 補題2.2より $g(p)$ は不偏推定可能であることが分かる.

4 あるランダム停止則を伴う不偏推定量の分散に関する不等式

ここでは, 第2, 3節における逐次2項標本抽出計画における停止則とは異なるランダム停止則を伴う不偏推定量の分散に関する不等式について考える. 各 $i = 1, 2, \ldots$ について, i 回目の試行を終えたとき確率 p_i で標本抽出を停止し, 確率 $(1-p_i)$ で $i+1$ 回目の標本抽出を行う, という停止則を考える. ただし, $p_i > 0 (i = 1, 2, \ldots)$ とする. いま, 各 $n = 1, 2, 3, \ldots$ について, $n-1$ 回目まで停止せず n 回目に停止する確率を Q_n とすると,

$$
\begin{aligned}
Q_n &= p_n \prod_{i=1}^{n-1} (1-p_i) \quad (n = 2, 3, \ldots) \\
Q_1 &= p_1
\end{aligned}
$$

になる. ここでは, 閉計画についてのみ考えるので,

$$
\sum_{n=1}^{\infty} Q_n = 1
$$

を仮定する. このとき, $T = (x,y)(x+y = n = 1, 2, \ldots)$ で停止する確率は, $\nu(x,y) := Q_n \cdot {}_nC_x$ とすると,

$$
P\{T = (x,y)\} = \nu(x,y) p^x (1-p)^y
$$

になる. このように停止則を変えても, 第2節の結果は $N(x,y)$ を $\nu(x,y)$ に置き換えても成立する. よって, $g(p)$ の不偏推定量は

$$
e(x,y) = \frac{G_n(x,y)_n C_x}{\nu(x,y)} = \frac{G_n(x,y)}{Q_n}
$$

になる. このとき, この不偏推定量 e の分散は

$$\mathrm{Var}\{e(x,y)\} = E[e(x,y)^2] - \{g(p)\}^2$$

となり, 分散を小さくする. すなわち, $E[e(x,y)^2]$ を小さくするような停止則を考える. そこで,

$$E[e(x,y)^2] = \sum_{n=1}^{\infty} \frac{1}{Q_n} \sum_{x+y=n} G_n(x,y)^2 {}_nC_x p^x (1-p)^y$$

となる. ここで, $c_{n,p} := \sum_{x+y=n} G_n(x,y)^2 {}_nC_x p^x (1-p)^y$ とおくと, 次のことが成立する.

定理 4.1 $\sum_{n=1}^{\infty}(c_{n,p})^{1/2} < \infty$ ならば, $g(p)$ の不偏推定量 $e(x,y)$ について

$$E[e(x,y)^2] \geq \Big\{\sum_{n=1}^{\infty}(c_{n,p})^{1/2}\Big\}^2$$

が成立する. ただし, 等号成立は $Q_n = (1/\lambda)(c_{n,p})^{1/2}$ のときに限る. ただし, λ は定数とする.

証明 Schwarz の不等式から

$$\begin{aligned} E[e(x,y)^2] &= \sum_{n=1}^{\infty} \frac{1}{Q_n} c_{n,p} \sum_{n=1}^{\infty} Q_n \\ &\geq \Big(\sum_{n=1}^{\infty} \frac{1}{Q_n^{1/2}} c_{n,p}^{1/2} Q_n^{1/2}\Big)^2 \\ &= \Big(\sum_{n=1}^{\infty} c_{n,p}^{1/2}\Big)^2 \end{aligned}$$

が成り立つ. ここで, 等号成立は

$$\frac{1}{Q_n^{1/2}} c_{n,p} = \lambda Q_n^{1/2}$$

のときのみ, すなわち,

$$Q_n = \frac{1}{\lambda} c_{n,p}^{1/2}$$

のときに限る. ただし, λ は定数とする. □

系 4.1 定理 4.1 の条件を満たすとき, $g(p)$ の不偏推定量 $e(x,y)$ について

$$\mathrm{Var}\{e(x,y)\} \geq \Big\{\sum_{n=1}^{\infty}(c_{n,p})^{\frac{1}{2}}\Big\}^2 - \{g(p)\}^2$$

が成り立つ. ただし, 等号成立は $Q_n = (1/\lambda)(c_{n,p})^{1/2}$ のときに限る. ただし, λ は定数とする.

注意 系 4.1 で与えられた不等式において, $Q_n = (1/\lambda)(c_{n,p})^{1/2}$ を満たす G_n を用いて不偏推定量を得ることができれば, 等号が成立するが, 一般に $c_{n,p}$ を求めるのは容易ではない.

次に, 例 3.1 で得た $g(p)$ の不偏推定量が定理 4.1 の条件を満たすことを示す.

例 3.1(続). まず, 例 3.1 より

$$g(p) = \begin{cases} 1-p & (0 < p < 1/2), \\ 1/2 & (p = 1/2), \\ p & (1/2 < p < 1) \end{cases}$$

の不偏推定量は,

$$e^*(x,y) = \begin{cases} \frac{{}_nC_x}{2(n-1)N(x,y)} & (x = y, T = (x,y) \text{ のとき}), \\ 0 & (\text{その他のとき}) \end{cases}$$

であるので, $G'_n(x,y)$ を,

$$G'_n(x,y) = \begin{cases} \frac{1}{2(n-1)} & (x = y, T = (x,y) \text{ のとき}), \\ 0 & (\text{その他のとき}) \end{cases}$$

とし, $c_{n,p} = \sum_{x+y=n} G'_n(x,y)^2 {}_nC_x p^x (1-p)^y$ について定理 4.1 の条件

$$\sum_{n=1}^{\infty} (c_{n,p})^{1/2} < \infty$$

が成り立つことを示す.

$$c_{n,p} = \begin{cases} \left(\frac{1}{2(n-1)}\right)^2 {}_nC_{n/2} p^{n/2} (1-p)^{n/2} & (n = 2,4,6,\ldots \text{ のとき}), \\ 0 & (n = 1,3,5,\ldots \text{ のとき}) \end{cases}$$

であるので, スターリングの公式と $p^x(1-p)^{n-x} \le (x/n)^x (1-(x/n))^{n-x}$ より,

$$\begin{aligned}
\sum_{n=1}^{\infty} (c_{n,p})^{1/2} &= \sum_{m=1}^{\infty} \left\{ \left(\frac{1}{2(2m-1)}\right)^2 {}_{2m}C_m p^m (1-p)^m \right\}^{1/2} \\
&= \sum_{m=1}^{\infty} \left\{ \frac{1}{4(2m-1)^2} \frac{(2m)!}{(m!)^2} p^m (1-p)^m \right\}^{1/2} \\
&\le \sum_{m=1}^{\infty} \left\{ \frac{1}{4(2m-1)^2} \frac{\sqrt{2\pi} e^{-2m} (2m)^{2m+1/2}}{(\sqrt{2\pi} e^{-m} m^{m+1/2})^2} \frac{1}{2^{2m}} \right\}^{1/2} \\
&= \sum_{m=1}^{\infty} \left\{ \frac{1}{4(2m-1)^2} \frac{1}{\sqrt{2\pi}} \frac{2^{2m+1/2} m^{2m+1/2}}{m^{2m+1}} \frac{1}{2^{2m}} \right\}^{1/2} \\
&= \sum_{m=1}^{\infty} \left\{ \frac{1}{\sqrt{2\pi}} \frac{2^{1/2}}{4(2m-1)^2 m^{1/2}} \right\}^{1/2} \\
&= \sum_{m=1}^{\infty} \left\{ O(m^{-5/2}) \right\}^{1/2} \\
&= \sum_{m=1}^{\infty} O(m^{-5/4}) \\
&< \infty
\end{aligned}$$

よって, この場合では, 定理 4.1, 系 4.1 が適用でき, $g(p)$ の不偏推定量の分散に関する不等式が成り立つ.

例 3.2 についても, $g(p)$ の不偏推定量が定理 4.1 の条件を満たすことを示す.

例 3.2(続). まず, 例 3.2 より

$$g(p)=\begin{cases} 1/(1-p) & (0<p<1/2),\\ 2 & (p=1/2),\\ 1/p & (1/2<p<1)\end{cases}$$

について,

$$G_n(x,y)=\begin{cases} 0 & (x=0,n),\\ -\frac{x}{y(y-1)(x+y)} & (0<x<\frac{n-1}{2}),\\ -\frac{1}{(x+1)(2x+1)} & (x=\frac{n-1}{2}),\\ \frac{1}{x} & (x=\frac{n}{2}),\\ -\frac{1}{x(2x-1)} & (x=\frac{n+1}{2}),\\ -\frac{y}{x(x-1)(x+y)} & (\frac{n+1}{2}<x<n)\end{cases}$$

となるので,

$$c_{n,p}=\sum_{x+y=n} G_n(x,y)^2{}_nC_xp^x(1-p)^y$$

になる. 次に,

(i) $n=2k-1\ \ (k=3,4,5,\dots)$ のとき,

$$\begin{aligned} c_{n,p}=&\sum_{x=1}^{k-2}\Big\{\frac{x}{(n-x)(n-x-1)n}\Big\}^2\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}\\ &+\Big\{\frac{1}{k(2k-1)}\Big\}^2\frac{n!}{(k-1)!(n-k+1)!}p^{k-1}(1-p)^{n-k+1}\\ &+\Big\{\frac{1}{k(2k-1)}\Big\}^2\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}\\ &+\sum_{y=1}^{k-2}\Big\{\frac{y}{(n-y)(n-y-1)n}\Big\}^2\frac{n!}{y!(n-y)!}p^{n-y}(1-p)^x\\ =:&\ J_1+J_2+J_3+J_4,\end{aligned}$$

(ii) $n=2k\ \ (k=2,3,4,\dots)$ のとき,

$$\begin{aligned} c_{n,p}=&\sum_{x=1}^{k-1}\Big\{\frac{x}{(n-x)(n-x-1)n}\Big\}^2\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}\\ &+\Big\{\frac{1}{k}\Big\}^2\frac{n!}{k!(n-k)!}p^k(1-p)^k\\ &+\sum_{y=1}^{k-1}\Big\{\frac{y}{(n-y)(n-y-1)n}\Big\}^2\frac{n!}{y!(n-y)!}p^{n-y}(1-p)^y\\ =:&\ J_5+J_6+J_7\end{aligned}$$

となるから, $J_1,\dots,J_7$ の評価をスターリングの公式と $p^x(1-p)^{n-x}\le(x/n)^x(1-(x/n))^{n-x}$ を用いて行うと,

$$
\begin{aligned}
J_1 &= \sum_{x=1}^{k-2} \frac{x^2}{(2k-x-1)^2(2k-x-2)^2(2k-1)^2} \cdot \frac{(2k-1)!}{x!(2k-x-1)!} p^x (1-p)^{2k-x-1} \\
&= \sum_{x=1}^{k-2} \frac{x^2}{(2k-x-1)^2(2k-x-2)^2(2k-1)^2} \\
&\quad \times \frac{\sqrt{2\pi} e^{-(2k-2)}(2k-2)^{2k-2+(1/2)}}{\sqrt{2\pi} e^{-x} x^{x+(1/2)} \cdot \sqrt{2\pi} e^{-(2k-x-1)}(2k-x-1)^{(2k-x-1)+(1/2)}} p^x (1-p)^{2k-x-1} \\
&\leq \sum_{x=1}^{k-2} \frac{x^2}{(2k-x-1)^2(2k-x-2)^2(2k-1)^2} \\
&\quad \times \frac{1}{\sqrt{2\pi}} \cdot \frac{(2k-1)^{2k-(1/2)}}{x^{x+(1/2)}(2k-x-1)^{2k-x-(1/2)}} \cdot \left(\frac{x}{2k-1}\right)^x \left(\frac{2k-x-1}{2k-1}\right)^{2k-x-1} \\
&= \frac{1}{\sqrt{2\pi}} \sum_{x=1}^{k-2} \frac{x^{3/2}}{(2k-x-1)^{5/2}(2k-x-2)^2(2k-1)^{3/2}} \\
&\leq \frac{1}{\sqrt{2\pi}} \sum_{x=1}^{k-2} \frac{(k-2)^{3/2}}{\{2k-(k-2)-1\}^{5/2}\{2k-(k-2)-2\}^2(2k-1)^{3/2}} \\
&= \frac{1}{\sqrt{2\pi}} \frac{(k-2)^{5/2}}{(k+1)^{5/2}k^2(2k-1)^{3/2}} \\
&= O(k^{-7/2})
\end{aligned}
$$

になる. 同様にして,

$$
J_2 \leq O(k^{-9/2}), J_3 \leq O(k^{-9/2}), J_4 \leq O(k^{-7/2}),
$$

$$
I_5 \leq O(k^{-7/2}), I_6 \leq O(k^{-5/2}), I_7 \leq O(k^{-7/2})
$$

になる. よって, $n = 2k-1, 2k$ のいずれの場合も $c_{n,p} \leq O(k^{-5/2})$ $(n \geq 4)$ になるから, $(c_{n,p})^{1/2} \leq O(k^{-5/4})$ となり, $\sum_{n=4}^{\infty}(c_{n,p})^{1/2} < \infty$ になる. また, $\sum_{n=1}^{3}(c_{n,p})^{1/2} < \infty$ になるので,

$$
\sum_{n=1}^{\infty}(c_{n,p})^{1/2} < \infty
$$

が成り立つ. よって, この場合では, 定理 4.1, 系 4.1 が適用でき, $g(p)$ の不偏推定量の分散に関する不等式が成り立つ.

参考文献

[AK98] Akahira, M. and Koike, K. (1998). On the properties of statistical sequential decision procedures. *Sugaku Expositions* **11**, 197-213.

[ATK92] Akahira, M., Takeuchi, K. and Koike, K. (1992). Unbiased estimation in sequential binomial sampling. *Rep. Stat. Appl. Res., JUSE,* **39**, 1-13.

[BB90] Bhandari, S. K. and Bose, A. (1990). Existence of unbiased estimates in sequential binomial experiments. *Sankhyā, Ser. A.* **52**, 127-130.

[BB93] Bhandari, S. K. and Bose, A. (1993). Corrigenda *ibid. Sankhyā, Ser. A.* **55**, 327.

[GMS97] Ghosh, M., Mukhopadyay, N. and Sen, P. K. (1997). *Sequential Estimation.* Wiley, New York.

[GS91] Ghosh, M. and Sen, P. K. (1991). Handbook of *Sequential Analysis.* Marcel Dekker, New York.

[Gu67] Gupta, M. K. (1967). Unbiased estimate for 1/p. *Ann. Inst. Statist. Math.,* **19**, 413-416.

[HP00] Hubert, S. L. and Pyke, R. (2000). Sequential estimation of functions of p for Bernoulli trials. In: *Game Theory, Optimal Stopping, Probability and Statistics,* 263-294, IMS Lecture Notes Monogr. Ser., **35**, Inst. Math. Statist., Beachwood, OH.

[SS75] Sinha, B. K. and Sinha, B. K. (1975). Some problems of unbiased sequential binomial estimation. *Ann. Inst. Statist. Math.,* **27**, 245-258.

[W46] Wolfowitz, J. (1946). On sequential binomial estimation. *Ann. Math. Statist.,* **17**, 489-493.