数理解析研究所講究録
1032巻 1998年 157-164

Sequential Test と Random Walk の境界決定問題

東邦大 理学部 情報科学専攻 D.1　竹並紀幸

1　序論と問題設定

Sequential Testとは数理統計学における2つの仮説の判定法の一種である．未知の分布を持つ確率変数 $\mathbf{X} = (X_1, X_2, \ldots)$ の確率分布が $\mathbf{P}_0$ (これを仮説 H_0 と表す) あるいは $\mathbf{P}_1$(これを仮説 H_1 と表す) かを観測値によって判断する方法の1つである．この論文では X_j が離散かつ有限個の値のみとると仮定する．

$a < 0 < b$ を選ぶ.次々と観測値 $(x_1, \ldots x_n, \ldots)$ が得られたとする．このとき,観測値を得るたびに

$$S_n = \sum_{i=1}^{n} \log \frac{\mathbf{P}_1\{\omega | X_j(\omega) = x_i\}}{\mathbf{P}_0\{\omega | X_j(\omega) = x_i\}} \tag{1.1}$$

を計算する.そして, $S_n \notin (a, b)$ となるまで観測を続ける．このときの観測回数を

$$T_{ab} = \inf\{n | S_n \notin (a, b)\} \tag{1.2}$$

とする.このとき, $S_{T_{ab}} \leq a$ ならば H_0 を採択, $S_{T_{ab}} \geq b$ ならば H_1 を採択する．観測の結果 H_1 を採択したが実際は H_0 が真である確率

$$\alpha = P_0\{S_{T_{ab}} \geq b\} \tag{1.3}$$

を第1種の誤り確率という．逆に H_0 を採択したが実際は H_1 が真である確率

$$\beta = P_1\{S_{T_{ab}} \leq a\} \tag{1.4}$$

を第2種の誤り確率という．したがって, a, b は α,β に依存している．Sequential Test を用いる場合 α,β を定めて, それに適した a, b を選ばなければならない．

本論文は確率モデルを設定して誤り確率 α, β を数値的に与えた場合に,境界を構成するための計算方法を示すことを目的としている．

以下,我々が取り組んだのは次のものである．

問題　X_j の値は有限個, ε を相対誤差とする．誤り確率 α, β, を指定したとき

$$\left|\frac{\mathbf{P}_0\{S_{T_{ab}} \geq b\} - \alpha}{\alpha}\right| < \varepsilon, \tag{1.5}$$

$$\left|\frac{\mathbf{P}_1\{S_{T_{ab}} \leq a\} - \beta}{\beta}\right| < \varepsilon \tag{1.6}$$

を満足する a, b を決定したい．その計算方法を示す．

注意．一般に $\mathbf{P}_0\{S_{T_{ab}} \geq b\}$, $\mathbf{P}_1\{S_{T_{ab}} \leq a\}$ は極めて小さい確率であり,この小論ではその解を得るためのシミュレーションにImportance Sampling 法を適用した.Importance Sampling 法については Leghtonen[3], Siegmund[4] を参照.

2 準備

この節では問題の解法において手助けとなる近似不等式について説明する.

Sinai[6] より

$$\mathbf{P}_0\{T_{ab} < \infty\} = 1, \quad \mathbf{P}_1\{T_{ab} < \infty\} = 1 \tag{2.1}$$

か成り立っている. したがって, 有限時間内で採択が決定される.

簡略化のため

$$p_0(x) = \mathbf{P}_0\{X_j = x\}, \quad p_1(x) = \mathbf{P}_1\{X_j = x\}, \quad h(x) = \log\frac{p_1(x)}{p_0(x)} \tag{2.2}$$

とおく. $E_{\mathbf{P}_i}$ $(i = 0, 1)$ は分布 $\mathbf{P}_i$に関する期待値とする.

補題 2.1　Random Walk $\{S_n\}$ は H_0 が成り立っている時は負のドリフトを持ち, またH_1が成り立っているときは正のドリフトを持つ. 即ち

$$E_{\mathbf{P}_0}[h(X_1)] < 0, \quad E_{\mathbf{P}_1}[h(X_1)] > 0. \tag{2.3}$$

証明　logは上に凸な関数なのでイェンゼンの不等式より

$$E_{\mathbf{P}_0}[h(X_1)] = E_{\mathbf{P}_0}[\log\frac{p_1(X_1)}{p_0(X_1)}] \tag{2.4}$$

$$< \log E_{\mathbf{P}_0}[\frac{p_1(X_1)}{p_0(X_1)}] = 0. \tag{2.5}$$

真の不等号は $p_0(x) \neq p_1(x)$ であることによる. 同様に

$$E_{\mathbf{P}_1}[h(X_1)] = E_{\mathbf{P}_1}[\log\frac{p_1(X_1)}{p_0(X_1)}] \tag{2.6}$$

$$= -E_{\mathbf{P}_1}[\log\frac{p_0(X_1)}{p_1(X_1)}] > 0. \tag{2.7}$$

不等式は示された.

補題 2.2　$\Gamma = \{a_1, \ldots, a_K\}$ は X_j の値域とし,

$$h_{max} = \max_{x\in\Gamma} h(x), \quad h_{min} = \min_{x\in\Gamma} h(x)$$

とおく. α, β, εは, $0 < \alpha < 1$, $0 < \beta < 1$, $0 < \varepsilon < 1$　かつ不等式

$$|\frac{\mathbf{P}_0\{S_{T_{ab}} \geq b\} - \alpha}{\alpha}| < \varepsilon, \tag{2.8}$$

$$|\frac{\mathbf{P}_1\{S_{T_{ab}} \leq a\} - \beta}{\beta}| < \varepsilon \tag{2.9}$$

を満たすものとする. このとき次の不等式が成立する.

$$\log(\frac{\beta(1-\varepsilon)}{1-\alpha(1-\varepsilon)}) \leq a \leq \log(\frac{\beta(1+\varepsilon)}{1-\alpha(1+\varepsilon)}\frac{1}{\mathrm{e}^{h_{min}}}), \tag{2.10}$$

$$\log(\frac{1-\beta(1+\varepsilon)}{\alpha(1+\varepsilon)}\frac{1}{\mathrm{e}^{h_{max}}}) \leq b \leq \log(\frac{1-\beta(1-\varepsilon)}{\alpha(1-\varepsilon)}). \tag{2.11}$$

注意. 不等式(2.10) の下からの評価と (2.11) の上からの評価は Siegmund[5] にある.

証明

$$\mathbf{A}_n = \{(x_1,\dots,x_n) \in \Gamma^n | a < \sum_{i=1}^{l} h(x_i) < b \ (1 \le l < n), \quad \sum_{i=1}^{n} h(x_i) \le a\}, \tag{2.12}$$

$$\mathbf{B}_n = \{(x_1,\dots,x_n) \in \Gamma^n | a < \sum_{i=1}^{l} h(x_i) < b \ (1 \le l < n), \quad \sum_{i=1}^{n} h(x_i) \ge b\}, \tag{2.13}$$

と置く.

$$\mathbf{P}_0\{S_{T_{ab}} \ge b\} = \sum_{n=1}^{\infty} \sum_{\mathbf{x} \in \mathbf{B}_n} p_0(x_1) \cdots p_0(x_n) \tag{2.14}$$

$$= \sum_{n=1}^{\infty} \sum_{\mathbf{x} \in \mathbf{B}_n} \frac{p_0(x_1)\cdots p_0(x_n)}{p_1(x_1)\cdots p_1(x_n)} p_1(x_1)\cdots p_1(x_n). \tag{2.15}$$

ここで $x = (x_1,\dots,x_n) \in B_n$のとき

$$\frac{1}{\mathrm{e}^b \mathrm{e}^{h_{max}}} \le \frac{p_0(x_1)\cdots p_0(x_n)}{p_1(x_1)\cdots p_1(x_n)} \le \frac{1}{\mathrm{e}^b}, \tag{2.16}$$

が成り立っているので

$$\frac{1}{\mathrm{e}^b \mathrm{e}^{h_{max}}}(1 - \mathbf{P}_1\{S_{T_{ab}} \le a\}) \le \mathbf{P}_0\{S_{T_{ab}} \ge b\} \le \frac{1}{\mathrm{e}^b}(1 - \mathbf{P}_1\{S_{T_{ab}} \le a\}). \tag{2.17}$$

$\mathbf{P}_1\{S_{T_{ab}} \le a\}$ を (2.9) で評価すれば

$$\frac{1}{\mathrm{e}^b \mathrm{e}^{h_{max}}}(1 - \beta - \varepsilon\beta) < \mathbf{P}_0\{S_{T_{ab}} \ge b\} < \frac{1}{\mathrm{e}^b}(1 - \beta + \varepsilon\beta) \tag{2.18}$$

が得られる. さらに $\mathbf{P}_0\{S_{T_{ab}} \le a\}$ を (2.8) で評価し整理すれば

$$\log(\frac{1-\beta(1+\varepsilon)}{\alpha(1+\varepsilon)} \frac{1}{\mathrm{e}^{h_{max}}}) \le b \le \log(\frac{1-\beta(1-\varepsilon)}{\alpha(1-\varepsilon)}) \tag{2.19}$$

が得られる. 同様に (2.10) も得られる. 証明を終える.

補題 2.2 より境界 a, b の取り得る範囲が明らかになった. そこで, a, b それぞれの最小値, 最大値を

$$a_{min} = \log(\frac{\beta(1-\varepsilon)}{1-\alpha(1-\varepsilon)}), \quad a_{max} = \log(\frac{\beta(1+\varepsilon)}{1-\alpha(1+\varepsilon)} \frac{1}{\mathrm{e}^{h_{min}}}), \tag{2.20}$$

$$b_{min} = \log(\frac{1-\beta(1+\varepsilon)}{\alpha(1+\varepsilon)} \frac{1}{\mathrm{e}^{h_{max}}}), \quad b_{max} = \log(\frac{1-\beta(1-\varepsilon)}{\alpha(1-\varepsilon)}), \tag{2.21}$$

と表す.

$\mathbf{P}_0\{S_{T_{ab}} \ge b\}$ はa, bに関して非増大な関数であり, また$\mathbf{P}_1\{S_{T_{ab}} \le a\}$ はa, bに関して非減少な関数である.

したがって $a \in [a_{min}, a_{max}]$, $b \in [b_{min}, b_{max}]$ のとき

$$\mathbf{P}_0\{S_{T_{a_{max}b_{max}}} \ge b_{max}\} \le \mathbf{P}_0\{S_{T_{ab}} \ge b\} \le \mathbf{P}_0\{S_{T_{a_{min}b_{min}}} \ge b_{min}\}, \tag{2.22}$$

$$\mathbf{P}_1\{S_{T_{a_{min}b_{min}}} \le a_{min}\} \le \mathbf{P}_1\{S_{T_{ab}} \le a\} \le \mathbf{P}_1\{S_{T_{a_{max}b_{max}}} \le a_{max}\}, \tag{2.23}$$

が成立する.

3 境界値 a, b の決定法

1で述べた我々の問題の解法を段取りに従って示す.

はじめに

$$a_0 = a_{min}, \tag{3.1}$$

とおく.

Step 1–a.
a_0を固定して(1.5)を満たすbを2分法に類似した方法により決定する.
$b' = b_{min}$, $b'' = b_{max}$ とおき b の初期値を$b = \frac{b'+b''}{2}$ とおく.

start Monte Carlo Sim.で $\mathbf{P}_0\{S_{T_{a_0 b}} \geq b\}$ の推定値 $\hat{P}_0$ を求める.
if $|(\hat{P}_0 - \alpha)/\alpha| < \varepsilon$
$\Longrightarrow$ **Step 1–a** 終了

else
if $(\hat{P}_0 - \alpha)/\alpha > \varepsilon$
$\Longrightarrow$ $b' = b$と置き直し,改めて$b = \frac{b'+b''}{2}$と置いて**start** に戻る.

else
if $(\hat{P}_0 - \alpha)/\alpha < -\varepsilon$
$\Longrightarrow$ $b'' = b$と置き直し,改めて$b = \frac{b'+b''}{2}$と置いて**start** に戻る.

上の手順によって最終的に決定されたbを b_1 と置く.

Step 1–b.
b_1を固定して(1.6)を満たすaを**Step 1–a**と同様の方法により決定する.
$a' = a_{min}$, $a'' = a_{max}$ とおき a の初期値を$a = \frac{a'+a''}{2}$ とおく.

start Monte Carlo Sim.で $\mathbf{P}_1\{S_{T_{ab_1}} \leq a\}$ の推定値 $\hat{P}_1$ を求める.
if $|(\hat{P}_1 - \beta)/\beta| < \varepsilon$
$\Longrightarrow$ **Step 1–b** 終了

else
if $(\hat{P}_1 - \beta)/\beta > \varepsilon$
$\Longrightarrow$ $a' = a$と置き直し,改めて$a = \frac{a'+a''}{2}$と置いて**start** に戻る.

else
if $(\hat{P}_1 - \beta)/\beta < -\varepsilon$
$\Longrightarrow$ $a'' = a$と置き直し,改めて$a = \frac{a'+a''}{2}$と置いて**start** に戻る.

上の手順によって最終的に決定された a を a_1 と置く.

Step 2–a.
a_1を固定して(1.5)を満たすbを**Step 1–a**と全く同じ手順で決定し b_2 とおく.

Step 2–b.
b_2を固定して(1.6)を満たすaを**Step 1–b**と全く同じ手順で求め a_2 とする.

注意. この時点までに得られた数列$\{a_0, a_1, a_2\}, \{b_1, b_2\}$は

$$a_0 \le a_1 \le a_2, \quad b_2 \le b_1 \tag{3.2}$$

を満たす. なぜなら, まず a_0 の定義より

$$a_0 \le a_1. \tag{3.3}$$

$\mathbf{P}_0\{S_{T_{ab}} \ge b\}$ はa, bに関して非増大な関数であるので **Step (1–a)** において求めた(a_0, b_1) と **Step (1–b)** において求めた(a_1, b_1) の間に

$$\mathbf{P}_0\{S_{T_{a_1b_1}} \ge b_1\} \le \mathbf{P}_0\{S_{T_{a_0b_1}} \ge b_1\} \tag{3.4}$$

が成りたつ. したがって $\mathbf{P}_0\{S_{T_{a_1b_1}} \ge b_1\}$ は (1.5) を満たしているとは限らない. **Step (2–a)** において再び (1.5) を満たすように b を更新すると $\mathbf{P}_0\{S_{T_{ab}} \ge b\}$ はa, bに関して非増大な関数であるので

$$b_2 \le b_1. \tag{3.5}$$

$\mathbf{P}_1\{S_{T_{ab}} \le a\}$ はa, bに関して非減少な関数なので,**step (1–b)** で求めた(a_1, b_1) と**step (2–a)** で求めた(a_1, b_2) との間に

$$\mathbf{P}_1\{S_{T_{a_1b_2}} \le a_1\} \le \mathbf{P}_1\{S_{T_{a_1b_1}} \le a_1\} \tag{3.6}$$

が成り立つ.**Step (2–b)** において再び(1.6)を満たすようにa を更新すると $\mathbf{P}_1\{S_{T_{ab}} \le a\}$ は非減少なので

$$a_0 \le a_1 \le a_2. \tag{3.7}$$

Step を繰り返すと

$$|(\mathbf{P}_0\{S_{T_{a_ib_{i+1}}} \ge b_{i+1}\} - \alpha)/\alpha| < \varepsilon \tag{3.8}$$

に対して, 下に有界な減少列

$$0 \le b_{min} \le b_{i+1} \le b_i \le \cdots \le b_2 \le b_1 \tag{3.9}$$

が得られる.

同様に

$$|(\mathbf{P}_1\{S_{T_{a_{i+1}b_{i+1}}} \le a_{i+1}\} - \beta)/\beta| < \varepsilon \tag{3.10}$$

に対して,上に有界な増大列

$$a_1 \le a_2 \le \cdots \le a_i \le a_{i+1} \le a_{max} \le 0 \tag{3.11}$$

が得られる.

(1.5),(1.6) を同時に満たすa, bが見つかるまで続ける.

4 適用例

$X_1, X_2, \ldots, X_i, \ldots$ を独立同分布の2次元確率ベクトル列とする.その分布は2通りのみ考えられるとして, 以下のような仮説を立てる.

仮説H_0

$$X_i = \begin{cases} (1,0) & \text{with prob.} \quad 0.35 \\ (-1,0) & \text{with prob.} \quad 0.25 \\ (0,1) & \text{with prob.} \quad 0.25 \\ (0,-1) & \text{with prob.} \quad 0.15 \end{cases} \tag{4.1}$$

仮説H_1

$$X_i = \begin{cases} (1,0) & \text{with prob.} \quad 0.3 \\ (-1,0) & \text{with prob.} \quad 0.2 \\ (0,1) & \text{with prob.} \quad 0.3 \\ (0,-1) & \text{with prob.} \quad 0.2 \end{cases} \tag{4.2}$$

それぞれの場合について期待値を計算すると

$$E_{\mathbf{P}_0}[X_1] = E_{\mathbf{P}_1}[X_1] = (0.1, 0.1). \tag{4.3}$$

したがって,軌跡を観測するだけでは判別が難しい. 前節の手順に従いSequential Test によって判別すべき問題のように思われる.

本節ではこのモデルに対して,誤り確率を定め境界を構成するための具体的な計算結果を示す.

以下,前節において定義した変数に対して上記のモデルに関する数値を設定して,補題2.2 において示した近似式を用いて境界値を数値として求める.

X_iの値域Γは

$$\Gamma = \{(1,0), (-1,0), (0,1), (0,-1)\}, \quad |\Gamma| = 4, \tag{4.4}$$

となる.

したがって,

$$h(X_i) = \log \frac{p_1(X_i)}{p_0(X_i)} \in \{\log \frac{0.3}{0.35}, \log \frac{0.2}{0.25}, \log \frac{0.3}{0.25}, \log \frac{0.2}{0.15}\}, \tag{4.5}$$

$$h_{max} = \log \frac{4}{3}, \quad h_{min} = \log \frac{4}{5}. \tag{4.6}$$

$$S_n = \sum_{i=1}^{n} h(X_i). \tag{4.7}$$

境界の概算.

第一種の誤り確率α, 第二種の誤り確率β, 相対誤差εに対して

$$\alpha = 0.01, \quad \beta = 0.01, \quad \varepsilon = 0.01, \tag{4.8}$$

と指定する. 境界a, bのとりうる範囲を概算する.

補題2.2より

$$|\frac{\mathbf{P}_0\{S_{T_{ab}} \geq b\} - \alpha}{\alpha}| < \varepsilon, \quad |\frac{\mathbf{P}_1\{S_{T_{ab}} \leq a\} - \beta}{\beta}| < \varepsilon$$

$$\Longrightarrow$$

$$-4.60528992 \leq a \leq -4.361924993, \tag{4.9}$$

$$4.297386432 \leq b \leq 4.605271191. \tag{4.10}$$

が成立する.

a, b それぞれの最小値, 最大値を

$$a_{min} = -4.60528992, \quad a_{max} = -4.361924993, \tag{4.11}$$

$$b_{min} = 4.297386432, \quad b_{max} = 4.605271191, \tag{4.12}$$

と表す.

数値Simulationによると, この例の場合

$$a = -4.511406656, \quad b = 4.496229135 \tag{4.13}$$

という結果が得られた.

注意.

今回扱った例では相対誤差$\varepsilon = 0.01$として収束が確認できたが,$\varepsilon = 0.001$としたときプログラムは止まらなかった.

これは分布が離散のため,$\mathbf{P}_0\{S_{T_{ab}} \geq b\}$, $\mathbf{P}_1\{S_{T_{ab}} \leq a\}$ も離散の値となることが原因と考えられる.

停止条件

$$|\frac{\mathbf{P}_0\{S_{T_{ab}} \geq b\} - \alpha}{\alpha}| < \varepsilon, \quad |\frac{\mathbf{P}_1\{S_{T_{ab}} \leq a\} - \beta}{\beta}| < \varepsilon$$

の取り方に限界があることを示している.

参考文献

[1] Hammersley,J.M. and Handscomb,D.C.. Monte Carlo Methods.Methuen(1964).

[2] Kolmogorov,A.N.. 関数解析の基礎. 岩波書店 (1962).

[3] Lehtonen,T. and Nyrhinen,H.. Simulating level-crossing probabilitys by importance sampling. Adv.Appl.Probab.**24**, 858-874(1992).

[4] Siegmund,D.. Importance sampling in the Monte Carlo study of sequential tests. Ann.Statist.**4**,673-684(1976).

[5] Siegmund,D.. Sequential Analysis.Springer(1985).

[6] Sinai,Ya.G.. シナイ確率論入門コース.Springer(1995).