2000年度日本オペレーションズ・リサーチ学会
秋季研究発表会

ランダムウォークに関連した最適停止問題

01303783　愛知大学　　玉置 光司　TAMAKI Mitsushi

1　はじめに

最初に確率論でよく知られているギャンブラーの破滅の結果を述べておこう。ギャンブラーは毎回のゲームにおいて、確率pで1ドル獲得するか、あるいは確率$q=1-p$で1ドル失う。また、毎回のゲームは独立で、ギャンブラーは所持金が目標値Nドルに達するか、あるいは破産して所持金が0ドルとなるまでゲームを続けるものとする。X_nを時刻nでのギャンブラーの所持金とすると、$\{X_n, n=0,1,2,\cdots\}$は、推移確率

$$p_{0,0}=p_{N,N}=1$$

$$p_{j,j+1}=p=1-p_{j,j-1}, \quad j=1,2,\cdots,N-1$$

を持つマルコフ連鎖となる。状態$\{0\},\{N\}$は、吸収状態である。ギャンブラーが所持金iドルからスタートする時、彼がNドル獲得する確率をP_iと記すと、これは次式で与えられる。

$$P_i=\begin{cases}\dfrac{1-r^i}{1-r^N}, & p\neq\frac{1}{2}\text{の時}\\[2mm] \dfrac{i}{N}, & p=\frac{1}{2}\text{の時}\end{cases} \tag{1}$$

ただし、$r=q/p$.

2　Stopping at the best

1節のギャンブラーの破滅問題に関連して、所持金が最大となった時にゲームをストップする確率を最大にする問題を考える。すなわち、

$$P_r\{X_{\tau^*}=\max_{n\geq0}X_n\}=\max_{\tau\in C}P_r\{X_\tau=\max_{n\geq0}X_n\}$$

となる*stopping time* τ^*を求める問題を考える（Cは*stopping time*の集合を表す）。簡単のため、$\max_{n\geq0}X_n$で停止することを成功と呼ぶと、この問題は成功確率を最大にする最適停止問題となる。

2.1　最適方程式

ギャンブラーの現在の所持金がa $(0<a<N)$で、今までの最大である時、この状態を単にaで表す。状態aの最適値関数を$v(a)$とすると、これは次式を満足する。

$$\begin{aligned}v(a)&=\max\{s(a),c(a)\}, \quad 0<a<N\\ s(a)&=\frac{r^a-r^{a+1}}{1-r^{a+1}},\\ c(a)&=pv(a+1)+q(1-s(a-1))v(a),\end{aligned}$$

ただし、$v(N)=v(0)=1$.

$s(a)$は、状態aでストップして成功する確率であるが、これは所持金が$a+1$に到達する前に0に到達する確率であり、（1）より、

$$s(a)=1-\frac{1-r^a}{1-r^{a+1}}$$

として計算される。$c(a)$は、状態aでストップしない場合の成功確率であるが、次のゲームで1ドル獲得するか1ドル失うかによって条件付けることにより上述の結果を得る。

2.2　最適政策

Theorem 2.1

（2）の$v(a)$は、次式で与えられる。

$$v(a)=\begin{cases}\dfrac{r^a-r^{a+1}}{1-r^{a+1}}, & \text{if} \quad 0\leq a\leq a^*-1\\[2mm] \dfrac{1-r^a}{1-r^N}, & \text{if} \quad a^*\leq a\leq N\end{cases}$$

ここで、a^*は、

$$a^*=\min\left\{a:\frac{1-r^a}{1-r^N}\geq\frac{r^a-r^{a+1}}{1-r^{a+1}}\right\}$$

で与えられる。したがって、状態aでの最適政策は、

$$\left\{\begin{array}{c} stop \\ continue \end{array}\right\} \Longleftrightarrow a \left\{\begin{array}{c} < \\ \geq \end{array}\right\} a^*$$

となる。

Theorem2. 1より、最初の所持金がa^*より小さい時はそこでゲームをストップするが、所持金がa^*以上の時は目標値Nになるまでゲームを継続することが最適となる。

3 Stopping at one of m best

mを与えられた正整数として、

$$\begin{aligned} & P_r\{X_{\sigma^*} \geq \max_{n\geq 0} X_n - (m-1)\} \\ & = \max_{\sigma\in C} P_r\{X_\sigma \geq \max_{n\geq 0} X_n - (m-1)\} \end{aligned}$$

となる*stopping time* σ^*を求める問題を考える。これは、2節の問題の一般化になっている($m=1$とおいた場合が2節の問題である)。

3.1 最適方程式

今時刻nで所持金がaであるとする。すなわち、$X_n = a$とする。この時ストップするか否かが問題となるのは、X_nの$X_0, X_1, \cdots, X_n$の中での順位が$1, 2, \cdots, m$の場合である。そこで、$X_n = a$で順位がi $(i = 1, 2, \cdots, m)$である状態を(a, i)と記す。

状態(a,i)での最適値関数を$v_i(a)$とし、この状態でストップした場合、あるいはストップしない場合の値をそれぞれ$s_i(a), c_i(a)$と書くと以下の方程式が成立する。

$$\begin{aligned} v_i(a) &= \max\{s_i(a), c_i(a)\}, \quad 1 \leq i \leq m \\ s_i(a) &= s(a) = \frac{r^a(1-r^m)}{1-r^{a+m}}, \quad 1 \leq i \leq m \\ c_i(a) &= pv_{i-1}(a+1) + qv_{i+1}(a-1), \ 1 \leq i < m \\ c_m(a) &= pv_{m-1}(a+1) + q(1-d(a-1))v_m(a). \end{aligned}$$

ただし、$d(a) = \dfrac{r^a - r^{a+1}}{1-r^{a+1}}$.

3.2 最適政策

次の補題は、状態$(a,i), 1 \leq i < m$においては、最適最策はストップしないことを述べている。

Lemma 3.1

$$v_i(a) = c_i(a) \geq s_i(a), \qquad 1 \leq i < m.$$

したがって、状態(a, m)について最適政策を調べればよい。

Theorem 3.2

$v_m(a)$は、次式で与えられる。

$$v_m(a) = \begin{cases} \dfrac{r^a(1-r^m)}{1-r^{a+m}}, & \text{if} \quad 0 \leq a \leq a_m^* - 1 \\ \dfrac{1-r^a}{1-r^{N-m+1}}, & \text{if} \quad a_m^* \leq a \leq N \end{cases}$$

ここで、a_m^*は、

$$a_m^* = \min\left\{a : \frac{1-r^a}{1-r^{N-m+1}} \geq \frac{r^a(1-r)^m}{1-r^{a+m}}\right\}$$

で与えられる。したがって、状態(a, m)での最適政策は、

$$\left\{\begin{array}{c} stop \\ continue \end{array}\right\} \Longleftrightarrow a \left\{\begin{array}{c} < \\ \geq \end{array}\right\} a_m^*$$

となる。

成功確率は、次式で与えられる。

Lemma 3.3

所持金aでゲームを開始した時、成功する確率$v(a)$は、次式で与えられる。

$$v(a) = \sum_{i=a-m+1}^{N-2m+1} (1-A_m)A_m^{i+m-a-1} v_m(i) + A_m^{N-m-a+1}, \quad m-1 \leq a \leq N-m$$

ただし、$A_m = \dfrac{1-r^{m-1}}{1-r^m}$.

参考文献

[1] Ross, S. M., "Dynamic programming and gambling models", *adv. Appl. Probab.* 6, 593-606, 1974.

[2] Ross, S. M., *Introduction to Stochastic Dynamic Programming*. Academic Press, New York, 1983.