Optimal Stopping Related to the Random Walk

愛知大学経営学部　　玉置　光司 (Mitsushi Tamaki)
Faculty of Business Administration, Aichi University

1   Introduction and Summary

We first review the gambler's ruin problem. Consider a gambler who at each play of the game has probability p of winning one unit and probability $q = 1 - p$ of losing one unit. Successive plays of the game are assumed to be independent and the gambling process continues until the gambler having initial fortune i will attain his goal of N or go broke. It is well known (see, e.g., Ross[1]) that P_i, the probability that this gambler achieves his goal before he becomes broke, is given as

$$P_i = \begin{cases} \dfrac{1 - \theta^i}{1 - \theta^N}, & \text{if } \theta \neq 1 \\[2mm] \dfrac{i}{N}, & \text{if } \theta = 1, \end{cases} \tag{1.1}$$

where $\theta = q/p$. For general use, it is more convenient to introduce the notation $P_i(a,b)$ the probability that the gambler's fortune will reach $i + a$ before reaching $i - b$ for $a, b > 0$;

$$P_i(a,b) = \begin{cases} \dfrac{1 - \theta^b}{1 - \theta^{a+b}}, & \text{if } \theta \neq 1 \\[2mm] \dfrac{b}{a+b}, & \text{if } \theta = 1. \end{cases} \tag{1.2}$$

Note that $P_i(a,b)$ dose not depend on the initial fortune i (of course $P_i = P_i(N - i, i)$).

The problem we consider in Section 2 is concerned with finding an optimal stopping rule which maximizes the probability that the gamble quits the casino with the largest possible fortune. If X_n be the state of the gambling process at time n, i.e., the gambler's fortune after n plays, then the above problem can be described, in terms of the stopping time, as that of seeking an optimal stopping time σ^* such that

$$P_r\{X_{\sigma^*} = \max_{0 \leq n \leq T} X_n \mid X_0 = i\} = \max_{\sigma \in C} P_r\{X_\sigma = \max_{0 \leq n \leq T} X_n \mid X_0 = i\},$$

where C is the class of all stopping times and T is the time that the process $\{X_n; n \geq 0\}$ first reaches state 0 or N, namely,

$$T = \min\{n \geq 0; X_n = 0 \text{ or } X_n = n\}.$$

In Section 3, this problem is generalized to the one where the gambler does not persist in the very largest, but rather feels satisfaction provided that his fortune upon stopping is within m units from the largest. More specifically, when the allowance measure is m (m is a given positive integer), the gambler seeks an optimal stopping time σ_m^* such that

$$P_r\{X_{\sigma_m^*} \geq \max_{0 \leq n \leq T} X_n - (m-1) \mid X_0 = i\} = \max_{\sigma \in C} P_r\{X_\sigma \geq \max_{0 \leq n \leq T} X_n - (m-1) \mid X_0 = i\}. \quad (1.3)$$

Obviously when $m = 1$, this problem is reduced to the one considered in Section 2. It can be shown that there exists an optimal stopping time σ_m^* of the following form;

$$\sigma_m^* = \min\left\{n : X_n = \max_{0 \leq j \leq n} X_j - (m-1), \quad X_n < k_m^*\right\},$$

where k_m^* is a positive integer defined as

$$k_m^* = \min\left\{k : \frac{1 - \theta^k}{1 - \theta^{N-m+1}} \geq \frac{\theta^k(1 - \theta^m)}{1 - \theta^{k+m}}\right\}.$$

2 Stopping on the largest

Suppose that we enter a casino having an initial fortune of x units and bet one unit each time. Then we either win one unit with probability p or lose one unit with probability $q = 1 - p$, where $0 < p < 1$. The gambling process $\{X_n; n \geq 0\}$, a Markov chain, continues until it reaches its absorbing states 0 or N, but we can stop the process before absorption, if we like. If we decide to stop at time n, then we are said to succeed if $X_n = \max_{0 \leq j \leq T} X_j$. The objective of the problem is to find a stopping policy that will maximize the probability of success.

Suppose that we have observed the values of $X_0, X_1, \cdots, X_n$. Then serious decision of either stop or continue takes place at time n when X_n is the maximum value observed so far, that is, $X_n = \max(X_0, X_1, \cdots, X_n)$. This decision epoch is simply referred to as *state* x if $X_n = x$, because, as a bit of consideration shows, the decision depends neither on the values of $X_0, X_1, \cdots, X_{n-1}$ nor on the number of plays already made.

Let $v(x)$ be the probability of success starting from state x, and let $s(x)$ and $c(x)$ be respectively the probability of success when we stop in state x and we continue gambling in an optimal manner in state x; then from the principle of optimality.

$$v(x) = \max\{s(x), c(x)\}, \quad 0 < x < N \quad (2.1)$$

where

$$s(x) = 1 - P_x(1, x) \quad (2.2)$$

$$c(x) = pv(x+1) + q(1 - s(x-1))v(x), \quad 0 < x < N \quad (2.3)$$

with the boundary condition $v(0) = v(N) = 1$. (2.3) is immediate from (1.2). (2.4) follows because the next possible state visited (after leaving state x) is either $x + 1$ or x depending on whether we win one unit at the next play or lose one unit at the next

play but attain state x again before going broke.To solve the optimality equation (2.2)-(2.4), we invoke the result of positive dynamic programming (see, for detail, chapter IV of Ross[2]). The main result (Theorem 1.2, p.75 of Ross) is that if the problem fits the framework of the positive dynamic programming, then a given stationary policy is optimal if its value function satisfies the optimality equation. This gives us a method of checking whether or not a guessed policy can be optimal and is particularly useful in cases in which there exists an obvious optimal policy. Our problem in fact fits the framework of positive dynamic programming with the state being the gambler's fortune, since if we suppose that a reward of 1 is earned if we attain the largest over all and all other rewards are zero, then the expected total reward equals the probability of success.

Theorem 2.1 Let f be a stationary policy which, when the decision process is in state x, chooses to stop if and only if $x < x^*$, where

$$x^* = \min\left\{x : \frac{1 - \theta^x}{1 - \theta^N} \geq \frac{\theta^x - \theta^{x+1}}{1 - \theta^{x+1}}\right\}. \tag{2.4}$$

Then f is optimal.

Proof Let $v_f(x)$ denote the probability of success starting from state x when policy f is employed. Once the decision process leaves state $x(\geq x^*)$, it never stops until absorption takes place under f. Thus we soon obtain

$$v_f(x) = \begin{cases} s(x), & \text{if } x < x^* \\ P_x(N - x, x), & \text{if } x \geq x^*, \end{cases} \tag{2.5}$$

and it is easy to check that $s(x)$ is decreasing in x, while $P_x(N - x, x)$ is increasing in x and that x^* can be described as

$$x^* = \min\{x : s(x) \leq P_x(N - x, x)\}.$$

Therefore, to prove f is optimal, it suffices to show that, from (2.2)-(2.4), $v_f(x)$ given in (2.5) satisfies

$$v_f(x) = \max\{s(x), pv_f(x + 1) + q(1 - s(x - 1))v_f(x)\}, \quad 0 < x < N. \tag{2.6}$$

We now show this distinguishing among three cases.

Case(i): $x < x^* - 1$
Since $v_f(x) = s(x)$, to show the validity of (2.6), it suffices to show that

$$v_f(x) \geq pv_f(x + 1) + q(1 - s(x - 1))v_f(x),$$

or

$$s(x) \geq ps(x + 1) + q(1 - s(x - 1))s(x),$$

or equivalently

$$\{s(x) - s(x+1)\} + \theta s(x+1)s(x) \geq 0,$$

which is immediate since $s(x)$ is non-negative and decreasing in x.

Case(ii): $x \geq x^*$

Since $v_f(x) = P_x(N - x, x) \geq s(x)$, it suffices to show that

$$v_f(x) \geq pv_f(x+1) + q(1 - s(x-1))v_f(x),$$

or equivalently

$$P_x(N - x, x) \geq pP_{x+1}(N - x - 1, x + 1) + qP_{x-1}(1, x + 1)P_x(N - x, x),$$

which in fact holds with equality, because straightforward calculation from (1.2) leads to the following identity;

$$P_x(N - x, x) = pP_{x+1}(N - x - 1, x + 1) + qP_{x-1}(1, x - 1)P_x(N - x, x). \qquad (2.7)$$

It is easy to interpret this identity probabilistically by conditioning on the result of the first play after leaving state x.

Case(iii): $x = x^* - 1$

Since $v_f(x^* - 1) = s(x^* - 1)$, it suffices to show that

$$v_f(x^* - 1) \geq pv_f(x^*) + q(1 - s(x^* - 2))v_f(x^* - 1),$$

or

$$s(x^* - 1) \geq pP_{x^*}(N - x^*, x^*) + qP_{x^*-2}(1, x^* - 2)s(x^* - 1),$$

or

$$s(x^* - 1) \geq \frac{pP_{x^*}(N - x^*, x^*)}{1 - qP_{x^*-2}(1, x^* - 2)},$$

or equivalently from (2.7)

$$s(x^* - 1) \geq P_{x^*-1}(N - x^* + 1, x^* - 1),$$

which is obvious from the definition of x^*. Thus the proof is complete.

3 Stopping on one of m largest

In this section, we are concerned with stopping on one of m largest of the gambling process. That is, we wish to find an optimal stopping time σ_m^* which satisfies (1.3). Suppose that we have observed the values of $X_0, X_1, \cdots, X_n$. Then serious decision of either stop or continue takes place at time n if $X_n \geq \max_{0 \leq j \leq n} X_j - (m - 1)$. We denote this state by (x, i) if $X_n = x$ and $X_n = \max_{0 \leq j \leq n} X_j - (i - 1), 0 < x < N - m, i = 1, 2, \ldots, m$. Our trial is now regarded as successful if we stop at time n and

$X_n \geq \max_{0 \leq j \leq T} X_j - (m-1)$.Let $v_i(x), 1 \leq i \leq m$, be the probability of success starting from state (x, i), and let $s_i(x)$ and $c_i(x)$ be respectively the probability of successes when we decide to stop and when we decide to continue in an optimal manner in state (x, i), then from the principle of optimality,

$$v_i(x) = \max\{s_i(x), c_i(x)\}, \quad 1 \leq i \leq m, \quad 0 < x < N - m \tag{3.1}$$

where

$$s_i(x) \equiv s(x) = 1 - P_x(m, x) \quad 1 \leq i \leq m \tag{3.2}$$

and

$$c_1(x) = pv_1(x+1) + qv_2(x-1) \tag{3.3}$$

$$c_i(x) = pv_{i-1}(x+1) + qv_{i+1}(x-1), \quad 1 < i < m \tag{3.4}$$

$$c_m = pv_{m-1}(x+1) + qP_{x-1}(1, x-1)v_m(x). \tag{3.5}$$

Assume that we choose to be continue in state $(x, i), 1 \leq i < m$, if $s_i(x) = c_i(x)$. Then as the next lemma shows, we never stop in state $(x, i), 1 \leq i < m$.

Lemma 3.1   For $1 \leq i < m$,

$$c_i(x) \geq s_i(x).$$

Proof.   Since, from the definition,

$$c_i(x) \geq ps(x+1) + qs(x-1), \quad 1 \leq i < m,$$

to show $c_i(x) \geq s_i(x)$, it suffices to show that

$$ps(x+1) + qs(x-1) \geq s(x),$$

or

$$\frac{1-\theta^x}{1-\theta^{x+m}} - \left\{p\frac{1-\theta^{x+1}}{1-\theta^{x+m+1}} + q\frac{1-\theta^{x-1}}{1-\theta^{x+m-1}}\right\} \geq 0. \tag{3.6}$$

Because $p = 1/(1+\theta), q = \theta/(1+\theta)$, after a bit of calculation, the left-hand side of (3.6) becomes

$$\frac{\theta^{2x+m-1}\left(\dfrac{1-\theta^m}{1-\theta}\right)}{\left(\dfrac{1-\theta^{x+m-1}}{1-\theta}\right)\left(\dfrac{1-\theta^{x+m}}{1-\theta}\right)\left(\dfrac{1-\theta^{x+m+1}}{1-\theta}\right)},$$

which is obviously non-negative and proves (3.6).

From Lemma 3.1, we are only concerned with the optimal decision in state (x, m). The following lemma expresses $c_m(x)$ in terms of $v_m(.)$.

Lemma 3.2   Define $P = (1 - \theta^{m-1})/(1 - \theta^m)$. Then, for $x \leq N - 2m$,

$$c_m(x) = \{p\theta P + qP_{x-1}(1, x-1)\}v_m(x)$$

$$+p(1 - \theta P) \sum_{y=x+1}^{N-2m+1} P^{y-x-1}(1 - P)v_m(y)$$

$$+p(1 - \theta P)P^{N-x-2m-1}.$$

Proof. Let $p(x, y)$ be the transition probability from state (x, m) to state $(y, m), y \geq x$. Then straightforward calculation yields

$$p(x, y) = \begin{cases} qP_{x-1}(1, x-1) + p\left[1 - P_{x+1}(m-1, 1)\right], & \text{if } y = x \\ pP_{x+1}(m-1, 1)\left[\prod_{i=0}^{y-x-2} P_{x+m+i}(1, m-1)\right]\left[1 - P_{y+m-1}(1, m-1)\right], \\ & \text{if } x < y \leq N - 2m + 1 \end{cases}$$

$$= \begin{cases} qP_{x-1}(1, x-1) + p\theta P, & \text{if } y = x \\ p(1 - \theta P)P^{y-x-1}(1 - P), & \text{if } x < y \leq N - 2m + 1. \end{cases}$$

The remaining probability

$$1 - \sum_{y=x}^{N-2m+1} p(x, y) = p(1 - \theta P)P^{N-x-2m+1}$$

corresponds to the probability that the gambling process $\{X_n\}$ attains its absorbing state $N - m + 1$. Thus the proof is complete.

We now have the following identity.

Lemma 3.3 For $m \geq 1$ and $x \leq N - 2m$,

$$P_x(N - m + 1 - x, x) = \{p\theta P + qP_{x-1}(1, x-1)\}P_x(N - m + 1 - x, x)$$

$$+p(1 - \theta P) \sum_{y=x+1}^{N-2m+1} P^{y-x-1}(1 - P)P_y(N - m + 1 - y, y)$$

$$+p(1 - \theta P)P^{N-x-2m-1}.$$

Proof. We can prove this by straightforward calculation from (1.2), but this identity is quite intuitive from Lemma 3.2.

We are now ready to prove the main result.

Theorem 3.4 Let f be a stationary policy which, when the decision process is in state (x, m), chooses to stop if and only if $x < x^*$, where

$$x^* = \min\left\{x : \frac{1 - \theta^x}{1 - \theta^{N-m+1}} \geq \frac{\theta^x - \theta^{x+m}}{1 - \theta^{x+m}}\right\}.$$

Then f is optimal.

Proof Let $v_f(x)$ denote the probability of success starting from state (x, m) when policy f is employed. Once the decision process leaves state (x, m) for $x \geq x^*$, it never stops until absorption takes place under f. Thus we have

$$v_f(x) = \begin{cases} s(x), & \text{if } x < x^* \\ P_x(N - m + 1 - x, x), & \text{if } x \geq x^*, \end{cases} \tag{3.7}$$

and it is easy to check that $s(x)$ is decreasing in x, while $P_x(N - m + 1 - x, x)$ is increasing in x and that x^* can be described as

$$x^* = \min\{x : s(x) \leq P_x(N - m + 1 - x, x)\}.$$

Therefore, to prove f is optimal, it suffices to show that $v_f(x)$ given in (3.7) satisfies

$$v_f(x) = \max\{s(x), c_f(x)\}, \quad 0 < x < N - m + 1,$$

where

$$
\begin{aligned}
c_f(x) = {} & \{p\theta P + qP_{x-1}(1, x-1)\}v_f(x) \\
& + p(1 - \theta P) \sum_{y=x+1}^{N-2m+1} P^{y-x-1}(1 - P)v_f(y) \\
& + p(1 - \theta P)P^{N-x-2m-1}
\end{aligned}
$$

We can show this in a similar way as in Theorem 2.1, appealing to Lemma 3.3.

Finally we give $V(x)$, success probability starting from x units of fortune.

Lemma 3.5 Let P be as defined in Lemma 3.2. Then, for $m - 1 \leq x \leq N - m$,

$$V(x) = \sum_{y=x-m+1}^{N-2m+1} P^{y-x+m-1}(1 - P)v_f(y) + P^{N-m-x+1}.$$

Proof This can be derived easily by conditioning on the first state (y, m) visited.

References

[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.