Probability in the Engineering and Informational Sciences, **14**, 2000, 151–160. Printed in the U.S.A.

A GOLD-MINING PROBLEM

OPTIMAL BACKUP STRATEGY IN COMPUTER PROGRAMS

Minoru Sakaguchi
3-26-4 Midorigaoka
Toyonaka, Osaka 560-0002, Japan

Toshio Hamada
Department of Management Sciences
Kobe University of Commerce
Nishi-ku, Kobe 651-2197, Japan
E-mail: hamada@kobeuc.ac.jp

We study an example of R. Bellman's gold-mining problem related to a programming job on the computer. The problem is formulated by dynamic programming and the optimal strategy is explicitly derived. The Bayesian version when the parameter involved is unknown is also solved by the same method. It is shown that the optimal strategy in each of two versions has the "no-island" (or, in other words, "control-limit") property.

1. THE PROBLEM

There are n identical items each of which has the probability of failure $p \in (0, \frac{1}{2})$. When we use these items to construct a "system" (i.e., a series connection of "units"), the $\left\{ \begin{matrix} 1 \\ 2 \end{matrix} \right\}$-item unit works "on" with probability $\left\{ \begin{matrix} 1-p \\ 1-p^2 \end{matrix} \right\}$, and "off" with probability $\left\{ \begin{matrix} p \\ p^2 \end{matrix} \right\}$. We have to choose, one-by-one sequentially, either one of the two kinds of the units, with the objective of

$$\text{E[length of run of "on" units until END]} \to \max \tag{1.1}$$

https://doi.org/10.1017/S0269964800142020 Published online by Cambridge University Press

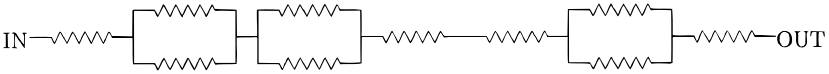


Figure 1. A system with $n = 10$.

subject to

$$2(\text{Number of 2-item units used}) + (\text{Number of 1-item units used}) \leq n, \qquad \textbf{(1.2)}$$

where END means the event in which either some unit works off or there remains no item, whichever occurs first.

Figure 1 shows a system with $n = 10$, and the length of run of "on" units in this system has the expected value

$$q\{p^2 + 2(1-p^2)p^2 + 3(1-p^2)^2 p + 4(1-p^2)^2 qp + \cdots\},$$

where $q = 1 - p$.

The problem is formulated by dynamic programming and the optimal strategy is explicitly derived as a function of $p \in (0, \frac{1}{2})$. This work was motivated by an article by Hamada [4], related to a programming job on the computer. In [4], the author investigates the same problem as in the present article by trying another approach in which the optimality equations are recursively solved in detail.

2. DYNAMIC PROGRAMMING

Let F_n be the maximum expected reward when there are n items available. Then, the Optimality Equation is evidently

$$F_n = \{q(1 + F_{n-1})\} \vee \{(1-p^2)(1 + F_{n-2})\} \qquad (n = 2,3,\ldots; F_0 = 0, F_1 = q), \qquad \textbf{(2.1)}$$

where $q = 1 - p$.

We prove the following:

Theorem 1: *Let $n_0 = [\log 2/(-\log q)]$ and let δ be the strategy: Use a 2-item unit as long as $n > n_0$, and switch to a 1-item unit as soon as $n \leq n_0$. (δ is denoted by $2^{[(n-n_0)/2]}1^{n_0}$.) Then, δ is optimal [nearly optimal] if $n - n_0$ is even [odd]. "Nearly optimal" means that either δ or $2^{[(n-n_0)/2]}1^{n_0+1}$ is optimal.*

Proof: Denote by $12*$ if $n \geq 3$, the strategy of using a 1-item unit first, a 2-item unit second, and the optimal continuation third, fourth, and so on. Let F_n^{12*} be the expected reward obtained by employing the strategy $12*$. Then, for $n \geq 3$, from (2.1) we have

$$F_n^{12*} = qp^2 + q(1-p^2)(2 + F_{n-3}),$$
$$F_n^{21*} = (1-p^2)p + (1-p^2)q(2 + F_{n-3}),$$

and hence

$$F_n^{12*} - F_n^{21*} = q\{p^2 + 2(1-p^2)\} - (1-p^2)(p + 2q) = -pq < 0. \qquad \textbf{(2.2)}$$

https://doi.org/10.1017/S0269964800142020 Published online by Cambridge University Press

For $n = 2$, we have

$$F_n^2 - F_n^{11} = 1 - p^2 - (qp + 2q^2) = q(1 - 2q) < 0, \tag{2.3}$$

because $q > \frac{1}{2}$.

Combining (2.2) and (2.3), we find that the optimal strategy has the form of $2^{(n-n_0)/2}1^{n_0}$, for any $n \geq 2$.

Our next step is to determine n_0 as a function of p. Let us compare the two strategies 21^x and 1^{x+2}, for any integer $x \geq 1$. The expected reward obtained by employing the strategy 1^x, for $x = 1,2,\ldots,$ is

$$g(1^x) = p\sum_{k=1}^{x-1} kq^k + xq^x = \left(\frac{q}{p}\right)(1 - q^x), \tag{2.4}$$

and also the expected reward by the strategy 2^y1^x, for $y = 0,1,2,\ldots,$ is

$$\begin{aligned} g(2^y1^x) &= p^2\sum_{k=1}^{y-1} k(1-p^2)^k + (1-p^2)^y\{y + g(1^x)\} \\ &= (p^{-2} - 1)\{1 - (1-p^2)^y\} + (1-p^2)^y g(1^x). \end{aligned} \tag{2.5}$$

Therefore, we have, from (2.4) and (2.5), after some algebra,

$$\begin{aligned} g(21^x) - g(1^{x+2}) &= (1 - p^2)\{1 + g(1^x)\} - \left(\frac{q}{p}\right)(1 - q^{x+2}) \\ &= q(1 - 2q^{x+1}). \end{aligned} \tag{2.6}$$

Equating (2.6) to zero, we obtain

$$q^{x+1} = \frac{1}{2} \qquad \left(\text{i.e., } x + 1 = \frac{\log 2}{(-\log q)}\right). \tag{2.7}$$

Let n_0 be the positive integer such that

$$q^{n_0+1} \leq \tfrac{1}{2} < q^{n_0}. \tag{2.8}$$

Then, because (2.6) is increasing in x, we have

$$g(21^{n_0-1}) - g(1^{n_0+1}) < 0 \leq g(21^{n_0}) - g(1^{n_0+2}).$$

This completes the proof of the theorem. ■

Table 1 gives the values of $n_0 = [\log 2/(-\log q)]$ for some small values of p. From Theorem 1 and Table 1, the optimal system with $n = 10$ when $p = 0.1$ is given by Figure 2, and the system shown by Figure 1 is not optimal.

Some remarks around Theorem 1 are given in Remarks 1, 2, and 3 of Section 4.

TABLE 1. Values of n_0

$p =$	0.01	0.015	0.02	0.03	0.05	0.1	0.15	0.2	0.25–0.29
$n_0 =$	68	45	34	22	13	6	4	3	2

https://doi.org/10.1017/S0269964800142020 Published online by Cambridge University Press

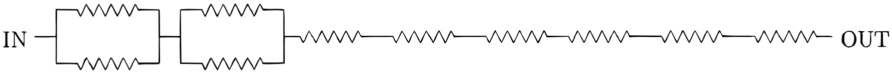


Figure 2. A system with $n = 10$.

3. BAYESIAN DYNAMIC PROGRAMMING

Consider the case where the values of p are unknown. Suppose that there is the prior information that p is a random variable with the distribution beta(α,β); that is, its probability density function (pdf) is

$$f(p|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}q^{\beta-1}I(0<p<1), \qquad \alpha,\beta \geq 1.$$

We define a state $(\alpha,\beta|n)$ to mean that (1) there are n items available and (2) the current information about the unknown value of p is that it is a random variable (r.v.) distributed as beta(α,β).

Let $F_n(\alpha,\beta)$ be the expected reward obtained by employing the optimal strategy under the Bayesian learning starting from state $(\alpha,\beta|n)$. Then, by the well-known manner of Bayesian learning, for Bernoulli/beta (see, e.g., DeGroot [2; Sects. 5.9 and 6.3]), we have the Optimality Equation

$$F_n(\alpha,\beta) = F_n^1(\alpha,\beta) \vee F_n^2(\alpha,\beta)$$
$$(n \geq 2;\, F_0(\alpha,\beta) = 0,\, F_1(\alpha,\beta) = \beta/(\alpha+\beta)), \tag{3.1}$$

where

$$\begin{aligned} F_n^1(\alpha,\beta) &= \int_0^1 (1-p)(1+F_{n-1}(\alpha,\beta+1))f(p|\alpha,\beta)\,dp \\ &= \frac{\beta}{\alpha+\beta}(1+F_{n-1}(\alpha,\beta+1)) \end{aligned} \tag{3.2}$$

and

$$\begin{aligned} F_n^2(\alpha,\beta) = \int_0^1 &\{(1-p)^2(1+F_{n-2}(\alpha,\beta+2)) \\ &+ 2p(1-p)(1+F_{n-2}(\alpha+1,\beta+1))\}f(p|\alpha,\beta)\,dp \\ = {}& \frac{\beta(\beta+1)}{(\alpha+\beta)(\alpha+\beta+1)}(1+F_{n-2}(\alpha,\beta+2)) \\ &+ \frac{2\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}(1+F_{n-2}(\alpha+1,\beta+1)). \end{aligned} \tag{3.3}$$

https://doi.org/10.1017/S0269964800142020 Published online by Cambridge University Press

Note that if a 1-item unit is used at state $(\alpha, \beta|n)$, then the state is transferred to

state $(\alpha, \beta+1|n-1)$ with probability $\beta/(\alpha+\beta)$
END with probability $\alpha/(\alpha+\beta)$;

and if a two-item system is used at state $(\alpha, \beta|n)$, then the state is transferred to

state $(\alpha, \beta+2|n-2)$ with probability $\beta(\beta+1)/(\alpha+\beta)(\alpha+\beta+1)$
state $(\alpha+1, \beta+1|n-2)$ with probability $2\alpha\beta/(\alpha+\beta)(\alpha+\beta+1)$
END with probability $\alpha(\alpha+1)/(\alpha+\beta)(\alpha+\beta+1)$.

By using the same method as in Theorem 1, we prove the following:

Theorem 2:

i. *For the Bayesian Bernoulli/beta version* (3.1)–(3.3), *there exists a function* $n_0(\alpha, \beta)$ *such that the optimal strategy in state* $(\alpha, \beta|n)$ *is as follows: Use a* 2-*item* [1-*item*] *unit, if* $n >[\le] n_0(\alpha, \beta)$.
ii. $n_0(\alpha, \beta)$ *is determined by the positive integer* n_0 *satisfying*

$$\sum_{k=2}^{n_0-1} k \frac{(\alpha)_2(\beta)_k}{(\alpha+\beta)_{k+2}} + \{(\alpha-1)n_0 - \beta\} \frac{(\beta)_{n_0}}{(\alpha+\beta)_{n_0+1}}$$
$$< \frac{\alpha\beta}{(\alpha+\beta)_2}$$
$$\le \sum_{k=2}^{n_0} k \frac{(\alpha)_2(\beta)_k}{(\alpha+\beta)_{k+2}} + \{(\alpha-1)(n_0+1) - \beta\} \frac{(\beta)_{n_0+1}}{(\alpha+\beta)_{n_0+2}}, \quad \textbf{(3.4)}$$

where $(m)_k = m(m+1)\cdots(m+k-1)$, *and the empty sum is meant by zero*.

Proof of (i): We use the same notations as used in the proof of Theorem 1. In state $(\alpha, \beta|n)$ with $n \ge 3$, we have, from (3.1)–(3.3),

$$F_n^{12*} = \frac{\beta}{(\alpha+\beta)_3}[\alpha(\alpha+1) + (\beta+1)(\beta+2)\{2 + F_{n-3}(\alpha, \beta+3)\}$$
$$+ 2\alpha(\beta+1)\{2 + F_{n-3}(\alpha+1, \beta+2)\}],$$
$$F_n^{21*} = \frac{\beta}{(\alpha+\beta)_3}[\alpha + 2\alpha(\alpha+1) + (\beta+1)(\beta+2)\{2 + F_{n-3}(\alpha, \beta+3)\}$$
$$+ 2\alpha(\beta+1)\{2 + F_{n-3}(\alpha+1, \beta+2)\}]$$

and, hence,

$$F_n^{12*} - F_n^{21*} = -\frac{2\alpha\beta}{(\alpha+\beta)_2} < 0. \quad \textbf{(3.5)}$$

Note that the terms involving $F_{n-3}(\alpha, \beta+3)$ and $F_{n-3}(\alpha+1, \beta+2)$ disappear, and (3.4) is valid independently of $n \ge 3$.

https://doi.org/10.1017/S0269964800142020 Published online by Cambridge University Press

For $n = 2$, we have

$$F_2^2 - F_2^{11} = \left\{1 - \frac{\alpha(\alpha+1)}{(\alpha+\beta)_2}\right\} - \frac{\beta}{(\alpha+\beta)_2}(\alpha + 2(\beta+1))$$
$$= \frac{-\beta(\beta+1-\alpha)}{(\alpha+\beta)_2}, \tag{3.6}$$

implying that $F_2^2 < F_2^{11} \Leftrightarrow \beta > \alpha - 1$. The condition $\beta > \alpha - 1$ in state $(\alpha, \beta|2)$ is usually not restrictive, because we start from state $(\alpha, \beta|n)$ with $\alpha \leq \beta$ and $n \geq 3$.

Combining (3.5) with (3.6) we find that the optimal strategy has the form $2^{(n-n_0)/2}$, for any $n \geq 2$.

Proof of (ii): We have to determine $n_0(\alpha, \beta)$. We compare the two strategies 21^x and 1^{x+2} in state $(\alpha, \beta|x+2)$. Denote by $G(21^x|\alpha, \beta)$ the expected reward obtained by following the strategy 21^x in state $(\alpha, \beta|x+2)$. We can find, after some algebra, based on (3.2)–(3.3), that

$$G(1^{x+2}|\alpha,\beta) = \alpha \sum_{k=1}^{x+1} k \frac{(\beta)_k}{(\alpha+\beta)_{k+1}} + (x+2)\frac{(\beta)_{x+2}}{(\alpha+\beta)_{x+2}} \tag{3.7}$$

and also

$$G(21^x|\alpha,\beta) = \left[\alpha \sum_{k=1}^{x-1}(k+1)\frac{(\beta)_{k+2}}{(\alpha+\beta)_{k+3}} + (x+1)\frac{(\beta)_{x+2}}{(\alpha+\beta)_{x+2}}\right]$$
$$+ \left[2\alpha(\alpha+1)\sum_{k=1}^{x-1}(k+1)\frac{(\beta)_{k+1}}{(\alpha+\beta)_{k+3}} + 2\alpha(x+1)\frac{(\beta)_{x+1}}{(\alpha+\beta)_{x+2}}\right], \tag{3.8}$$

where the first [second] part is due to the strategy starting from state $(\alpha, \beta+2|x)$ $[(\alpha+1, \beta+1|x)]$, which is left after the first choice of a 2-item unit. Subtracting (3.7) and (3.8) we obtain, after some algebra,

$$\varphi(x|\alpha,\beta) \equiv G(21^x|\alpha,\beta) - G(1^{x+2}|\alpha,\beta)$$
$$= \sum_{k=1}^{x} k \frac{(\alpha)_2(\beta)_k}{(\alpha+\beta)_{k+2}} - \frac{\alpha\beta}{(\alpha+\beta)_2} - \frac{(\alpha)_2\beta}{(\alpha+\beta)_3}$$
$$+ \{(\alpha-1)(x+1) - \beta\}\frac{(\beta)_{x+1}}{(\alpha+\beta)_{x+2}}, \tag{3.9}$$

which, we can find, is increasing in x. Because if we consider

$$h(x) = \sum_{k=2}^{x} k \frac{(\alpha)_2(\beta)_k}{(\alpha+\beta)_{k+2}} + (\alpha-1)(x+1)\frac{(\beta)_{x+1}}{(\alpha+\beta)_{x+2}},$$

https://doi.org/10.1017/S0269964800142020 Published online by Cambridge University Press

then

$$h(x) - h(x-1) = (2\alpha - 1)(\alpha + 1)x \frac{(\beta)_x}{(\alpha+\beta)_{x+2}} + (\alpha - 1) \frac{(\beta)_{x+1}}{(\alpha+\beta)_{x+2}} > 0.$$

Thus, $h(x)$, and hence (3.9), is increasing in $x \geq 1$. Moreover,

$$\begin{aligned} \varphi(1|\alpha,\beta) &= -\frac{\alpha\beta}{(\alpha+\beta)_2} + (2\alpha - \beta - 2) \frac{(\beta)_2}{(\alpha+\beta)_3} \\ &= -\frac{\beta}{(\alpha+\beta)_3} \{(\alpha-\beta)^2 + \alpha\beta + 3\beta + 2\} < 0 \end{aligned}$$

and $\varphi(x|\alpha,\beta)$ becomes positive for some large x. Therefore, we have

$$\varphi(n_0(\alpha,\beta) - 1|\alpha,\beta) < 0 \leq \varphi(n_0(\alpha,\beta)|\alpha,\beta)$$

for some $n_0(\alpha,\beta)$, which is equivalent to (3.6). This completes the proof of Theorem 2(ii). ■

Note that Theorem 2 indicates the following: If one uses a 2-item unit and it works "on," then one has to choose either a 2-item unit or a 1-item unit next. If one uses a 1-item unit and it works on, then one must choose a 1-item unit only until END.

More discussions on Theorem 2 are made in Remark 4 of the next section.

4. REMARKS

We present some remarks.

1. The problem discussed in the present work is an example of Bellman's gold-mining problem [1, pp. 61–80]. Theorems 1 and 2 show that the optimal strategy in each of two versions of the problem has the "no-island" property (i.e., every optimal decision region is a connected set).

Ross [6] and others (Monahan [5], for example) found that a counterintuitive strategy, with a disconnected decision region, is optimal, by a model of a partially observable Markov decision process.

2. Concerning Theorem 1, the expected reward obtained by employing the strategy 2^y1^x is given by (2.4) and (2.5). Therefore, the optimal system with $n = 10$ and $p = 0.1$, shown by Figure 2, has the expected reward

$$\begin{aligned} g(2^21^6) &= (p^{-2} - 1)(1 - (1-p^2)^2) + (1-p^2)^2 g(1^6) \\ &= (1-p^2)\{1 + (1-p^2)p^{-1}(1-q^7)\}, \end{aligned}$$

giving 6.1032 if $p = 1 - q = 0.1$ is substituted.

For the nonoptimal system 21^8, it is

$$\begin{aligned} g(21^8) &= (p^{-2} - 1)p^2 + (1-p^2)g(1^8) \\ &= (1-p^2)p^{-1}(1-q^9), \end{aligned}$$

giving 6.0645 if $p = 0.1$ is substituted.

https://doi.org/10.1017/S0269964800142020 Published online by Cambridge University Press

For another nonoptimal system, 1^{10}, it is

$$g(1^{10}) = \left(\frac{q}{p}\right)(1 - q^{10}) = 5.8619 \quad \text{if } p = 0.1.$$

Moreover, the specific number $n_0 \equiv [\log 2/(-\log q)]$ appeared in the past literature in an article by Domansky [3] related to certain optimal stopping game connected with an i.i.d. sequence of Bernoulli r.v.'s.

3. Two extended problems will arise.

Problem A. We can form two kinds of units: U_1 and U_2. For each $i = 1$ and 2, unit U_i works "on" with probability q_i and "off" with probability $p_i = 1 - q_i$ with operating cost c_i. Assume that $q_1 < q_2$ and $c_1 < c_2$. Let C be the total budget available. Then, the problem is (1.1) with (1.2) replaced by

$$c_1 \text{ (Number of } U_1 \text{ used)} + c_2 \text{ (Number of } U_2 \text{ used)} \leq C.$$

Find the condition on p_i and c_i, $i = 1, 2$, under which the optimal strategy has the "no-island" property.

Problem B. If we newly introduce the 3-item unit, which works "on" with probability $1 - p^3$ and "off" with probability p^3, we have a conjecture that the strategy $3^z 2^y 1^x$, with $3z + 2y + x = n$, is optimal. Is this conjecture valid? If so, find the optimal x and y as a function of p.

As for Problem B, the argument used in the 2-case can be carried over to the 3-case; that is, we find that using i ($i = 1, 2$) and then 3 is no better than using 3 and then i. So the problem left is to find the optimal y as a function of p. It is interesting to find that the equality (2.7) in the 2-case becomes the equality $y - 1 = \log \frac{2}{3}/\log(1 - p^2)$ in the 3-case, and that if $n = 100$ and $p = 0.1$, then the optimal system is $3^4 2^{41} 1^6$.

4. We present numerical examples of Theorem 2.

Example 1: Let $n = 5$ and $(\alpha, \beta) = (1,4)$. Then, from Theorem 2, we have to consider the three strategies 1^5, 21^3, and $2^2 1$. We compute (4.1) and (4.2) and find that

$$G(1^5|1{,}4) = \sum_{k=1}^{4} k \frac{(4)_k}{(5)_{k+1}} + 5 \frac{(4)_5}{(5)_5} = 2.9825,$$

$$G(21^3|1{,}4) = \left\{2 \frac{(4)_3}{(5)_4} + 3 \frac{(4)_4}{(5)_5} + 4 \frac{(4)_5}{(5)_5}\right\} + \left\{8 \frac{(4)_2}{(5)_4} + 12 \frac{(4)_3}{(5)_5} + 8 \frac{(4)_4}{(5)_5}\right\}$$
$$= 2.7222,$$

and

$$G(2^2 1|1{,}4) = \frac{(4)_2}{(5)_2} (1 + G(21|1{,}6)) + \frac{8}{(5)_2} (1 + G(21|2{,}5)) = 2.3936,$$

https://doi.org/10.1017/S0269964800142020 Published online by Cambridge University Press

because we have, from (3.8),

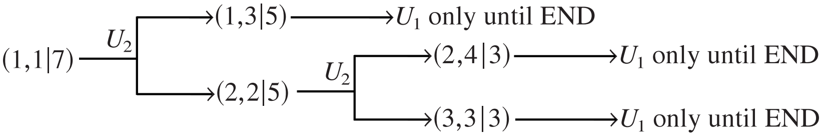
$$G(21|\alpha,\beta) = 2\,\frac{(\beta)_3}{(\alpha+\beta)_3} + 4\alpha\,\frac{(\beta)_2}{(\alpha+\beta)_3}.$$

Hence, we conclude that strategy 1^5 is optimal at state (1,4|5). On the other hand, if we try to compute the Optimality Equation (3.1)–(3.3) starting from

$$F_5(1,4) = \tfrac{4}{5}(1 + F_4(1,5)) \vee \{\tfrac{14}{15} + \tfrac{2}{3}F_3(1,6) + \tfrac{4}{15}F_3(2,5)\},$$

then we arrive, after some steps, at the same result, $F_5(1,4) = 2.9825$.

Example 2: Suppose that we start from state $(\alpha,\beta|n) = (1,1|7)$. Denote a 1-item [2-item] unit by U_1 [U_2]. Then, the optimal strategy is represented as follows:



because (3.4) in Theorem 2 gives $n_0(1,1) = 5$, $n_0(1,3) = 6$, $n_0(2,2) = n_0(2,3) = n_0(2,4) = 4$, and $n_0(3,3) = 3$. The top path means the following: First, use a U_2. If it works "on" and the transferred state is (1,3|5), then use U_1's thereafter until END as long as they work "on."

Finally, the following fact is worth noting.

THEOREM 3: *Suppose that* $\beta \geq \alpha + \frac{1}{2}(\sqrt{8\alpha^2+1} - 3)$. *If Bayesian learning* (3.2)–(3.3) *is not made and we exploit the prior knowledge that* $p \sim \text{beta}(\alpha,\beta)$ *only, then* $n_0(\alpha,\beta)$ *is equal to the positive integer that satisfies*

$$\frac{(\beta+1)_{n_0+1}}{(\alpha+\beta+1)_{n_0+1}} \leq \frac{1}{2} < \frac{(\beta+1)_{n_0}}{(\alpha+\beta+1)_{n_0}}. \tag{4.1}$$

PROOF: From (2.6), we have, after some algebra,

$$\begin{aligned} g(21^x|\alpha,\beta) - g(1^{x+2}|\alpha,\beta) &\equiv \int_0^1 \{g(21^x) - g(1^{x+2})\} f(p\,|\alpha,\beta)\,dp \\ &= \int_0^1 (q - 2q^{x+2}) f(p\,|\alpha,\beta)\,dp \\ &= \frac{\beta}{\alpha+\beta}\left\{1 - \frac{2(\beta+1)_{x+1}}{(\alpha+\beta+1)_{x+1}}\right\}. \end{aligned} \tag{4.2}$$

Equating (4.2) to zero, we obtain

$$\frac{(\beta+1)_{x+1}}{(\alpha+\beta+1)_{x+1}} = \frac{1}{2}. \tag{4.3}$$

https://doi.org/10.1017/S0269964800142020 Published online by Cambridge University Press

This equation corresponds to (2.7) in Theorem 1 (i.e., the case where $p = 1 - q$ is known).

Because (4.2) is increasing in $x \geq 1$ and is negative at $x = 1$ if $2(\alpha^2 - 1) < (\beta - \alpha)^2 + 3(\beta - \alpha)$ (i.e., $\beta \geq \alpha + \frac{1}{2}(\sqrt{8\alpha^2 + 1} - 3)$), there exists a $n_0(\alpha, \beta)$ such that

$$g(21^{n_0 - 1} | \alpha, \beta) - g(1^{n_0 + 1} | \alpha, \beta) < g(21^{n_0} | \alpha, \beta) - g(1^{n_0 + 2} | \alpha, \beta),$$

which is equivalent to (4.1). This completes the proof of the theorem. ■

Example 3: Let $(\alpha, \beta | n) = (1,4|5)$, as in Example 1. Because condition (4.1) for $(\alpha, \beta) = (1,4)$ gives $n_0(1,4) = 4$, the optimal strategy, when $n = 5$, is either 21^3 or 1^5 (see Theorem 1). From (2.4) and (2.5), we have

$$g(21^3) = (1 - p^2)(1 + q + q^2 + q^3),$$
$$g(1^5) = q + q^2 + \cdots + q^5$$

and, hence, we obtain

$$g(21^3 | 1,4) \equiv \int_0^1 g(21^3) f(p | 1,4)\, dp = 2.8937,$$

$$g(1^5 | 1,4) \equiv \int_0^1 g(1^5) f(p | 1,4)\, dp = 2.9825$$

because

$$\int_0^1 p^r q^s f(p | \alpha, \beta)\, dp = \frac{(\alpha)_r (\beta)_s}{(\alpha + \beta)_{r+s}} \qquad (r, s = 0,1,2,\ldots).$$

Therefore, 1^5 is optimal, and the expected reward is 2.9825. The result is the same as in Example 1, implying that learning is useless because n is too small.

References

1. Bellman, R. (1957). *Dynamic programming.* Princeton, NJ: Princeton University Press.
2. DeGroot, M.H. (1975). *Probability and statistics.* Reading, MA: Addison-Wesley.
3. Domansky, V.K. (1974). On certain games connected with a sequence of Bernoulli trials. *Engineering Cybernetics* 12: 25–29.
4. Hamada, T. (1999). Optimal sequential backup strategy under constrained resources. *Journal of the Operations Research Society of Japan* 42: 457–470.
5. Monahan, G.E. (1982). Optimal stopping in a partially observable binary-valued Markov chain with costly perfect information. *Journal of Applied Probability* 19: 72–81.
6. Ross, S. (1971). Quality control under Markovian deterioration. *Management Science* 17: 587–596.