“BANDIT PROBLEMS WITH INFINITELY MANY ARMS”

DONALD A. BERRY
Institute of Statistics & Decision Sciences, Duke University

ROBERT W. CHEN AND ALAN ZAME
Department of Mathematics & Computer Science
University of Miami

DAVID E. HEATH
School of Operations Research & Industrial Engineering
Cornell University

LARRY A. SHEPP
AT&T Bell Laboratories, Murray Hill, NJ

DP #95-28
Bandit Problems with Infinitely Many Arms

Donald A. Berry*
Institute of Statistics & Decision Sciences
Duke University, Durham, NC 27708–0251

Robert W. Chen and Alan Zame
Department of Mathematics and Computer Science
University of Miami, Coral Gables, FL 33214

David E. Heath
School of Operations Research and Industrial Engineering
Cornell University, Ithaca, NY 14853

Larry A Shepp
AT&T Bell Laboratories, Murray Hill, NJ 07974

ABSTRACT

We consider a bandit problem consisting of an sequence of $n$ choices from an infinite number of Bernoulli arms, with $n \to \infty$. The objective is to minimize the long-run failure rate. The Bernoulli parameters are independent observations from a distribution $F$. We first assume $F$ to be the uniform distribution on $(0, 1)$ and consider various extensions. In the uniform case we show that the best lower bound for the expected failure proportion is between $\sqrt{2}/\sqrt{n}$ and $2/\sqrt{n}$ and we exhibit classes of strategies that achieve the latter.

Key words and phrases: Bandit problems, sequential experimentation, dynamic allocation of Bernoulli processes, staying with a winner, switching with a loser.

* Research supported in part by the National Science Foundation under grant DMS 94-04200.
1 Introduction

A bandit problem consists of a series of choices from among a set of stochastic processes, or arms. At each decision epoch the decision maker chooses an arm for observation. The problem is sequential in that the choices can depend on arms chosen previously and on the resulting observations. Time is either discrete or continuous. The processes can be of arbitrary type. The decision problem is easy unless some of the available processes have unknown characteristics about which information can be gained through observation. The objective in bandit problems is to maximize the expected value of some function of the observations. One such function is the sum (or integral) over a finite-time horizon.

When time is discrete and the horizon decreases by one with each observation and information accrues about the arms observed. The character of optimal strategies depends greatly on the horizon $n$. When $n$ is large, the decision maker is willing to sacrifice immediate gain while testing arms attempting to find one that has a large mean and so can be used to produce a substantial long-term benefit. But when $n$ is small, information has less value and it might properly be eschewed in favor of using an arm that has a large mean. Therefore, during the course of an experiment and as the horizon nears, arms having a greater mean will be more appealing even if they have less potential for providing information.
Bandit problems have applications in clinical trials and in on-line industrial experimentation. In these and in other applications, current observations are usually regarded as more important than those in the distant future. So future observations may reasonably be discounted in some fashion (see Berry and Fristedt 1985, especially Chapter 3). Assuming a finite horizon is a special type of discounting. The most commonly assumed type of discounting is geometric: an observation $j$ stages in the future is worth only $\beta^{j-1}$ as much as the current observation, for some $\beta \in (0,1)$.

Berry and Fristedt (1985) review the early literature of bandit problems. Thompson (1933) and Robbins (1952) make important contributions. Both consider two Bernoulli processes with unknown parameters $p_1$ and $p_2$. Thompson takes a Bayesian approach and assumes that the unknown parameters are independent \textit{a priori} and have beta distributions. Taking the objective to be maximizing the expected number of successes in $n$ trials, he investigates the performance of randomized strategies in which Arm 1 is selected with probability equal to that of $p_1 > p_2$. This probability is updated using Bayes theorem as information accumulates on the two arms and the current probability is used in making each selection.

Robbins (1952) considers maximizing the long-run success proportion. He shows that basing selections on only the immediately preceding observation—staying with a winner and switching on a loser—dominates strategies that do not depend on the accumulating data. He also exhibits selection strategies that are asymptotically optimal.
in the sense that long-run success proportion is \( \max\{p_1, p_2\} \).

In this paper, we consider discrete-time bandit problems in which there are an infinite number of Bernoulli arms. The objective is the same as that of Robbins (1952): Maximize the long-run success proportion. Arm \( i \) has parameter \( p_i \). We assume that \( p_1, p_2, \ldots \) are themselves independent observations from a known distribution \( F \). Therefore, the arms are exchangeable before making any observations. In view of exchangeability we can restrict consideration to strategies that call for using Arm 1 first and whenever a new arm (one not previously used) is selected, it is the smallest numbered arm. So a nonrecalling strategy (one that always uses a new arm when switching) uses Arm 1 for a period of time and then Arm 2 for a period of time and then Arm 3 and so on.

Once Arm 1 has produced an observation, \( p_1 \) is no longer exchangeable with the other \( p_i \). In particular, the distribution of \( p_1 \) becomes \( pF(dp)/\int pF(dp) \) after a success and \( (1 - p)F(dp)/\int (1 - p)F(dp) \) after a failure, and the distribution of the other \( p_i \) continues to be \( F(dp) \). It seems reasonable to expect (and it is true) that Arm 1 should be used again after an immediate success and not after an immediate failure. (Moreover, if the first observation on any arm is a failure, that arm can be discarded and never used again.) But suppose Arm 1 yields \( s \) immediate successes and then a failure, should one switch to Arm 2 or stay with Arm 1? The answer depends on \( s \) and on the horizon \( n \) and is not easy to find. It is even harder to say which arm is
best when sample information becomes available about many of the $p_i$.

In Section 2, we suppose that $F$ is beta(1,1), the uniform distribution on (0,1). Section 3 extends Section 2 to the case in which $F$ is a uniform distribution on a subset of (0,1). We consider the case in which $F$ is an arbitrary distribution in Section 4. We do not consider the interesting and difficult hierarchical setting in which $F$ is itself unknown and observations on Arm $i$ give information about $F$ and therefore also about Arm $j$ for $j \neq i$.

2 Uniform Distribution on (0,1)

In this section, we study the case that the distribution $F$ is uniform over the interval (0,1)—that is, beta(1,1). We show that an excellent strategy is to discard an arm whenever it yields a failure, never using it again. However, this strategy can be improved. We find a lower bound for the expected failure proportion and exhibit several strategies that have the same order of magnitude of this lower bound.

We start with some notation and definitions. For each positive integer $k$, a strategy is called a $k$-failure strategy if it calls for using the same arm until that arm produces $k$ failures, and when this happens, it calls for switching to a new arm (never returning to arms that have yielded failures). With the possible exception of the arm being used when the horizon $n$ is reached, every arm yields a total of $k$ failures.
For each constant $\alpha$ in $[0, 1)$, a strategy is called an $\alpha$-rate strategy if it stays on the same arm until that arm has produced a failure rate greater than $\alpha$, and when this happens discarding it and switching to a new arm. Again, discarded arms are not reused.

A 1-failure strategy and a 0-rate strategy are the same. It is a modification of Robbin's stay-with-a-winner/switch-on-a-loser strategy to the infinite-arm setting. The failure proportion of this strategy in $n$ trials is asymptotically $1/\ln(n)$. To see this consider the number $S$ of immediate successes with any particular arm having parameter $p \sim F$. For $s = 1, 2, \ldots, n$, the probability of at least $s$ immediate successes on this arm is $P(S \geq s) = \int p^s F(dp) = 1/(s+1)$. Hence

$$E(S) = \sum_{s=1}^{n} \frac{1}{s+1} \approx \ln(n).$$

Thus, when following the 1-failure strategy, the expected number of failures in the first $n$ trials, which is within one of the expected number of arms used, is approximately $n/\ln(n)$. Hence the expected proportion of failures in $n$ trials is asymptotically $1/\ln(n)$.

Consider the $k$-failure strategy. Let $N(n, k)$ be the expected number of trials until the $k$th failure. For $n \leq k$, $N(n, k) = n$. The more interesting case is $n > k$:

$$N(n, k) = \int_0^1 \sum_{j=0}^{k-1} \binom{n}{j} u^{n-j} (1-u)^j du + \int_0^1 \sum_{j=k}^{n} \binom{j-1}{k-1} u^{j-k} (1-u)^k du.$$
\[
= \sum_{j=0}^{k-1} \frac{n}{n+1} + \sum_{j=k}^{n} \frac{k}{j+1} = k \left[ 1 + \sum_{j=k}^{n-1} \frac{1}{j+1} \right].
\]

This applies for fixed \(k\) and asymptotically as \(n \to \infty\).

The next result indicates that asymptotically the best strategy among \(k\)-failure strategies is to take \(k = 1\).

**Theorem 1** As \(n \to \infty\) the expected failure proportion for \(k\)-failure strategies is increasing in \(k\).

*Proof:* Asymptotically, the expected failure proportion when following a \(k\)-failure strategy is

\[
\frac{k}{N(n,k)} = \frac{1}{1 + \sum_{j=k}^{n-1} \frac{1}{j+1}}.
\]

Since \(\sum_{j=k}^{n-1} \frac{1}{j+1}\) is decreasing in \(k\), when \(n\) is large, \(k/N(n,k)\) is increasing in \(k\). So, asymptotically, the expected failure proportion is increasingly in \(k\). \(\diamondsuit\)

Theorem 1 indicates that the 1-failure strategy has the smallest asymptotic expected failure rate among \(k\)-failure strategies. The next theorem indicates that, asymptotically, the advantage of \(k = 1\) is not great: All \(k\)-failure strategies have asymptotic failure rate equal to \(1/\ln(n)\).

**Theorem 2** For any fixed \(k\) the expected failure proportion of the \(k\)-failure strategy is asymptotically \(1/\ln(n)\).
Proof: For \( n = 1, 2, \ldots \) let \( \phi(n, k) \) be the expected number of failures in \( n \) trials produced by the \( k \)-failure strategy. Fix \( k \in \{1, 2, \ldots \} \). For \( n \leq k \), \( \phi(n, k) = n/2 \). For \( n \geq k \), \( \phi(n, k) \) can be found recursively as follows, where \( \phi(0, k) = 0 \):

\[
\phi(n, k) = \int_0^1 \sum_{j=1}^{k-1} j \binom{n}{j} u^{n-j} (1-u)^j \, du + \int_0^1 \sum_{j=k}^{n} [k + \phi(n-j, k)] \binom{j-1}{k-1} u^{j-k} (1-u)^k \, du
\]

\[
= \sum_{j=1}^{k-1} j \binom{n-j}{j} \frac{(n-j)!j!}{(n+1)!} + \sum_{j=k}^{n} [k + \phi(n-j, k)] \binom{j-1}{k-1} \frac{(j-k)!k!}{(j+1)!} + \sum_{j=k}^{n} \phi(n-j, k) \binom{j-1}{k-1} \frac{(j-k)!k!}{(j+1)!}
\]

\[
= k - \frac{1}{n+1} \binom{k+1}{2} + k \sum_{j=k}^{n} \phi(n-j, k) \frac{1}{j(j+1)}.
\]

Let \( G_k(t) \) be the generating function of \( \{\phi(n, k)\}_{n \geq 1} \); that is,

\[
G_k(t) = \sum_{n=0}^{\infty} \phi(n, k) t^n.
\]

Then

\[
G_k(t) = \sum_{n=0}^{k} \frac{n}{2} t^n + \sum_{n=k+1}^{\infty} \left\{k - \frac{1}{n+1} \binom{k+1}{2} \right\} t^n + k \sum_{n=k+1}^{\infty} \sum_{j=k}^{n} \phi(n-j, k) \frac{1}{j(j+1)} t^n
\]

\[
= \sum_{n=1}^{k} \frac{n}{2} t^n + k \frac{t^{k+1}}{1-t} - \binom{k+1}{2} \frac{1}{1-t} \int \frac{t^{k+1}}{1-t} dt
\]

\[
+ k \sum_{j=k}^{\infty} \sum_{n=j}^{\infty} \phi(n-j, k) \left\{\frac{1}{j} t^j - \frac{1}{j+1} t^j \right\} t^{n-j}.
\]

Hence

\[
G_k(t) \left\{1 - k \int \frac{1}{1-t} t^{k-1} \, dt + \frac{k}{t} \int \frac{1}{1-t} t^k \, dt \right\}
\]

8
\[
\frac{k}{2} \int_{0}^{1} t^{n} + \frac{k}{1 - t} t^{k+1} - \frac{1}{t} \left( \frac{k+1}{2} \right) \int_{0}^{1} t^{k+1} \frac{dt}{1 - t}.
\]

Alternatively,

\[
G_{k}(t) \left\{ 1 + \frac{k}{t} (1 - t) \int_{0}^{1} \frac{1}{1 - t} t^{k-1} \frac{dt}{1 - t} \right\}
\]

\[
= \left\{ \sum_{n=1}^{k} \frac{n}{2} t^{n} + \frac{k}{1 - t} t^{k+1} - \frac{1}{t} \left( \frac{k+1}{2} \right) \int_{0}^{1} t^{k+1} \frac{dt}{1 - t} \right\}.
\]

Using the Tauberian theorem it follows that for any fixed \( k \), \( \phi(n, k)/n \sim 1/\ln(n) \).

The 1–failure strategy indicates a switch to a new arm whenever the current arm produces a failure. Discarding an arm that has given a very large number of successes and a single failure seems wasteful. One might expect an \( \alpha \)-rate strategy to do better for some \( \alpha \). However, the following argument shows that for any \( \alpha > 0 \), the \( \alpha \)-rate strategy has a positive expected failure proportion asymptotically.

Since \( F \) is uniform on \((0, 1)\), for any \( \alpha > 0 \) there is a positive probability that any particular arm has parameter \( p \) such that

\[
1 - \frac{3}{4} \alpha < p < 1 - \frac{1}{4} \alpha.
\]

By the strong law of large numbers, the failure rate produced by this arm is between \( \alpha/2 \) and \( \alpha \) with positive probability. Hence the proportion of failures when following the \( \alpha \)-rate strategy is at least \( \alpha/2 \) with positive probability. Therefore, the expected failure proportion of the \( \alpha \)-rate strategy is greater than a positive constant as \( n \to \infty \).
for any $\alpha > 0$. Therefore, for any $\alpha > 0$ and sufficiently large $n$, the $\alpha$–rate strategy is inferior to the 1–failure strategy.

Is $1/\ln(n)$, the asymptotic expected failure proportion of the 1–failure strategy, the best possible? We seek the best lower bound for the expected failure proportion. To this end, define an $m$–run strategy as one that follows the 1–failure strategy until either the current arm has produced a success run of length $m$ or Arm $m$ is used. If the former obtains then the current arm is used for the remaining trials. If the latter obtains then the arm with highest proportion of success among the $m$ arms used so far is used for the remaining trials. So an $m$–run strategy uses at most $m$ arms; if it uses $m$ arms then the best performing arm is recalled and used for the duration.

The next two theorems show that $\sqrt{2/n}$ is a lower bound for the expected failure proportion over all strategies and the failure rate of the $\sqrt{n}$–run strategy has the same order of magnitude as this lower bound.

**Theorem 3** $\sqrt{2/n}$ is a lower bound for the expected failure proportion over all strategies.

**Proof:** Let $C$ be the number of arms actually used when following a particular strategy. Given $C = c$, and since each arm is used at least once, the conditional expected number of failures is greater than or equal to

$$\frac{c}{c+1} + \frac{c-1}{c+1} + \cdots + \frac{1}{c+1} + \frac{n-c}{c+1}$$

(imagine that each arm is used once, then the best arm of these $c$ arms is used for
the remaining $n - c$ trials). Hence, by the Jensen inequality, the expected number of failures

$$\geq \frac{E(C + 1)}{2} + \frac{n + 1}{E(C + 1)} - \frac{3}{2}.$$ 

Since $\frac{a^2}{2} + \frac{x}{a} \geq \sqrt{2x}$, the expected failure proportion

$$\geq \frac{\sqrt{2(n + 1)}}{n} - \frac{3}{2n} \sim \sqrt{\frac{2}{n}}$$

for any particular strategy. Therefore, $\sqrt{n}$ is a lower bound for the expected failure proportion over all strategies.

**Theorem 4** The expected failure proportion for the $\sqrt{n}$-run strategy is less than or equal to $\frac{2}{\sqrt{n}}$.

**Proof:** Let $C$ be the number of arms used in the $\sqrt{n}$-run strategy and let $T$ be the corresponding number of failures produced. It is easy to see that

$$E(T) \leq \sqrt{n} + \frac{n}{\sqrt{n} + 1} \leq 2\sqrt{n}.$$ 

Therefore, the expected failure proportion for the $\sqrt{n}$-run strategy is $E(T)/n \leq 2/\sqrt{n}$.

The $\sqrt{n}$-run strategy is not the only strategy that has the expected failure proportion less than or equal to $2/\sqrt{n}$. The following is another.

A strategy is called an $m$-learning strategy if it follows the 1-failure strategy for the first $m$ trials (with the arm selected at trial $m$ used until such time that it yields a failure), and then it calls for using the arm that performed best during the first $m$ trials for the remaining trials. The next theorem shows that a $\ln(n)\sqrt{n}$-learning strategy has expected failure proportion less than or equal to $2/\sqrt{n}$.  

11
Theorem 5  The $\ln(n)\sqrt{n}$–learning strategy has the expected failure proportion less than or equal to $2/\sqrt{n}$.

Proof: Since the expected number of trials to the first failure of each new arm is asymptotically equal to $\ln(n)$, the expected number of arms used during the learning period will be $\ln(n)\sqrt{n}/\ln(n)$. Also, the expected number of failures and the expected number of arms used during this learning period is $\sqrt{n}$. The (expected) probability of failure on the best of these $\sqrt{n}$ arms is $1/(\sqrt{n} + 1)$. So the expected number of failures is less than or equal to

$$\sqrt{n} + (n - \ln(n)\sqrt{n})\frac{1}{\sqrt{n} + 1} \leq \sqrt{n} + \sqrt{n} = 2\sqrt{n}.$$ 

Therefore, the expected failure proportion of the $\ln(n)\sqrt{n}$–learning strategy is asymptotically less than or equal to $2/\sqrt{n}$ and the proof of Theorem 5 now is complete. \* 

The $\sqrt{n}$–run strategy of Theorem 4 and the $\ln(n)\sqrt{n}$–learning strategy of Theorem 5 are recalling strategies. Is there a nonrecalling strategy that is asymptotically as good as these two recalling strategies? The next theorem gives an affirmative answer.

A strategy is a nonrecalling $m$–run strategy if it uses the 1–failure strategy until an arm produces a success run of length $m$ at which this arm is used for all remaining trials; if no arm produces a success run of length $m$ the 1–failure strategy is used for all $n$ trials. The following theorem says the the nonrecalling $\sqrt{n}$–run strategy has
expected failure proportion asymptotically less than or equal to $2/\sqrt{n}$.

**Theorem 6** The expected failure proportion of the nonrecalling \( \sqrt{n} \)-run strategy is less than or equal to $2/\sqrt{n}$ asymptotically.

**Proof:** Let \( B \) be the number of arms tried until finding one that produces a success run of length \( \sqrt{n} \). If an arm has produced a success run \( \sqrt{n} \), then this arm is expected to produce no more than \( n(1/\sqrt{n}) \) failures. Hence the expected number of failures produced by the nonrecalling \( \sqrt{n} \)-run strategy will be less than or equal to \( \sqrt{n} + E(B) \).

Now

\[
E(B) = \sum_{j=1}^{n} jP(B = j) = \sum_{j=1}^{n} P(B \geq j) \leq 1 + \sum_{j=1}^{n} P(B > j).
\]

\( P(B > j) \) is the probability that none of the first \( j \) arms has produced a success run of length \( \sqrt{n} \), which is the \( j \)th power of the probability that any particular arm has not produced a success run of length \( \sqrt{n} \). That is,

\[
P(B > j) = \{ \int_0^1 [(1 - u) + u(1 - u) + \ldots + u^{\sqrt{n}-1}(1 - u)]du \}^j
\]

\[
= \{ \int_0^1 (1 - u^{\sqrt{n}})du \}^j = (1 - \frac{1}{\sqrt{n} + 1})^j = (\frac{\sqrt{n}}{\sqrt{n} + 1})^j.
\]

Therefore,

\[
\sum_{j=1}^{n} P(B > j) = \sum_{j=1}^{n} (\frac{\sqrt{n}}{\sqrt{n} + 1})^j \leq \sum_{j=1}^{\infty} (\frac{\sqrt{n}}{\sqrt{n} + 1})^j = \sqrt{n}.
\]

Therefore, the expected number of failures produced by the nonrecalling \( \sqrt{n} \)-run
strategy is less than $2\sqrt{n} + 1$ and the expected failure proportion is less than or equal to $2/\sqrt{n}$ asymptotically. ◦

Based on Theorems 3, 4, 5, and 6, we have the following corollary.

**Corollary** The best lower bound for the expected failure proportion over all strategies is between $\sqrt{2/n}$ and $2/\sqrt{n}$.

**Remark:** We suspect that the best lower bound for the expected failure proportion over all strategies is $2/\sqrt{n}$. However, we do not have a proof.

### 3 Uniform Distribution on $[a, b]$.

We now investigate the situation when the prior distribution $F$ is uniform over a subinterval $[a, b]$ of the unit interval.

Using the same argument as for Theorem 3 above, we can show that the expected failure proportion is greater than or equal to $(1-b) + (b-a)(\sqrt{2/n} - \frac{3}{2n})$. Therefore, $(1-b) + (b-a)\sqrt{2/n}$ is a lower bound for the expected failure proportion over all strategies under this setting.

**Theorem 7** If the distribution $F$ is uniform over the subinterval $[a, b]$ of the unit interval, then $(1-b) + (b-a)\sqrt{2/n}$ is a lower bound for the expected failure proportion over all strategies.
With little modification, the arguments for Theorems 4, 5, 6, above are valid for the current situation. Hence we have the following theorems.

**Theorem 8** The expected failure proportion of the $\sqrt{n(b - a)}$-run strategy is less than or equal to $(1 - b) + 2\sqrt{(b - a)/n}$ asymptotically.

**Theorem 9** The expected failure proportion of the $\sqrt{n(b - a)\ln(n(b - a))}$-learning strategy is less than or equal to $(1 - b) + 2\sqrt{(b - a)/n}$ asymptotically.

**Theorem 10** The expected failure proportion of the nonrecalling $\sqrt{n(b - a)}$-run strategy is less than or equal to $(1 - b) + 2\sqrt{(b - a)/n}$.

**Remark:** We suspect that the best lower bound for the expected failure proportion is $(1 - b) + 2\sqrt{(b - a)/n}$. However, we do not have a proof.

### 4 Arbitrary Prior Distributions

In this section, we will briefly discuss the case in which $F$ is an arbitrary distribution on the interval $[0, 1]$. We assume that $F$ is continuous, $F(0) = 0$, and $F(1) = 1$.

Suppose that the number of arms used over the course of the trial is $C$. Then, given $C = c$, there will be at least $G(c) = c\int_0^1 \alpha dF(\alpha) + (n - c)\int_0^1 F^c(\alpha)d\alpha$ (on the conditional space) expected failures (imagine that each arm is used only once, then an oracle tells us which of these $c$ arms is the best arm and we use this best arm for
the remaining $n - c$ trials. Since $G''(c) = (n - c) \int_0^1 F_c^c(\alpha) [\ln F(\alpha)]^2 d\alpha > 0$ if $c < n$, $G(C)$ is a convex function. By the Jensen inequality,

$$E\{G(C)\} \geq G(E(C)) = E(C) \int_0^1 \alpha dF(\alpha) + (n - E(C)) \int_0^1 F^E(C)(\alpha) d\alpha.$$ 

Since $G'(0) < 0$, $G(0) = G(n) = n$, and $G''(c) > 0$ if $c < n$, there exists a positive integer $c_n$ such that $1 \leq c_n < n$ and $G(c_n) = \min_{0 \leq c \leq n} G(c)$. Therefore, we have the following theorem.

**Theorem 11**

$$\frac{G(c_n)}{n} = \frac{1}{n} \left\{ c_n \int_0^1 \alpha dF(\alpha) d\alpha + (n - c_n) \int_0^1 F^{c_n}(\alpha) d\alpha \right\}$$

is a lower bound for the expected failure proportion over all strategies.

For $k = 0, 1, 2, \ldots, n$, let $H(k) = k + (n - k) \int_0^1 F^k(\alpha) d\alpha$. It follows that $H(0) = H(n) = n$,

$$H'(0) = n \int_0^1 \ln(F(\alpha)) d\alpha < 0, \quad H'(n) = 1 - \int_0^1 F^n(\alpha) d\alpha > 0,$$

$H''(k) > 0$ for $0 \leq k < n$, and there exists a positive integer $k_n$ such that $H(k_n) = \min_{0 \leq k \leq n} H(k)$.

Let $h(0) = 0$, and for each $\ell = 1, 2, \ldots, n$, let

$$e(\ell) = \sum_{j=0}^{n-h(\ell-1)} \int_0^1 (1 - \alpha)^j dF(\alpha)$$

16
and \( h(\ell) = h(\ell - 1) + e(\ell) \). We define \( M(\ell) = \ell + (n - h(\ell)) \int_0^1 F^e(\alpha) d\alpha \) for all \( \ell = 0, 1, 2, \ldots \) and \( h(\ell) \leq n \). Notice that \( h \) is increasing and \( e(\ell) \) is decreasing. It is easy to see that \( M(\ell) \) is a convex function. Since \( M(0) = n \) and \( M(\ell^*) = n \) if \( h(\ell^*) = n \), there exists a positive integer \( \ell_n \) such that \( M(\ell_n) = \min_{0 \leq \ell, h(\ell) \leq n} M(\ell) \).

When \( n \) is relatively large, \( \ell_n \) is relatively small, \( e(1) \approx e(2) \approx \ldots \approx e(\ell_n) \). So we can re-define \( M(\ell) \) as \( M^*(\ell) = \ell + (n - \ell e(1)) \int_0^1 F^e(\alpha) d\alpha \). Then we find \( \ell_n^* \) such that \( M(\ell_n^*) = \min_{0 \leq \ell \leq n/e(1)} M^*(\ell) \).

With little modification of the proofs of Theorem 4 and 5, we have the following theorems.

**Theorem 12** The expected failure proportion for the \( k_n \)-run strategy is less than or equal to \( H(k_n)/n \) asymptotically.

**Theorem 13** The \( e(1)\ell_n^* \)-learning strategy has expected failure proportion less than or equal to \( M^*(\ell_n^*)/n \) asymptotically.

By a slight modification of Theorem 6, we have the following theorem.

**Theorem 14** The expected failure proportion of the nonrecalling \( u_n \)-run strategy is less than or equal \( N(u_n)/n \) asymptotically.
5 Numerical Examples

It is difficult to find the minimum values of $G, H, M^*$, and $N$ functions analytically. However, finding $c_n, G(c_n), k_n, H(k_n), \ell_n^*, M^*(\ell_n^*), u_n$, and $N(u_n)$ is straightforward numerically using "Maple" or "Mathematica". Tables 1 through 4 give examples of $c_n, \frac{G(c_n)}{n}, k_n, \frac{H(k_n)}{n}, \ell_n^*, \frac{M(\ell_n^*)}{n}$, and $u_n, \frac{N(u_n)}{n}$ for various distribution functions $F$ and various $n$. Evidently, all three strategies perform well. The asymptotical expected failure proportions are close to the lower bound.

References


Table 1

<table>
<thead>
<tr>
<th>$F(t) = t$ \hspace{3cm} (uniform(0,1))</th>
<th>$F(t) = t^2$ \hspace{3cm} (β(2,1))</th>
<th>$F(t) = 2t - t^2$ \hspace{3cm} (β(1,2))</th>
<th>$F(t) = 3t^2 - 2t^3$ \hspace{3cm} (β(2,2))</th>
<th>$F(t) = \frac{1}{2} + \frac{\sin^{-1}(2t - 1)}{\pi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n=100$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$</td>
<td>$n=200$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$</td>
<td>$n=300$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$</td>
<td>$n=400$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$</td>
<td>$n=500$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$ \hspace{3cm} $c_n$ \hspace{3cm} $\frac{G(c_n)}{n}$</td>
</tr>
<tr>
<td>13 \hspace{3cm} 0.127</td>
<td>8 \hspace{3cm} 0.107</td>
<td>31 \hspace{3cm} 0.212</td>
<td>16 \hspace{3cm} 0.192</td>
<td>11 \hspace{3cm} 0.082</td>
</tr>
<tr>
<td>19 \hspace{3cm} 0.093</td>
<td>12 \hspace{3cm} 0.078</td>
<td>47 \hspace{3cm} 0.176</td>
<td>24 \hspace{3cm} 0.155</td>
<td>14 \hspace{3cm} 0.054</td>
</tr>
<tr>
<td>24 \hspace{3cm} 0.077</td>
<td>15 \hspace{3cm} 0.064</td>
<td>61 \hspace{3cm} 0.158</td>
<td>32 \hspace{3cm} 0.137</td>
<td>16 \hspace{3cm} 0.042</td>
</tr>
<tr>
<td>27 \hspace{3cm} 0.067</td>
<td>17 \hspace{3cm} 0.056</td>
<td>73 \hspace{3cm} 0.145</td>
<td>38 \hspace{3cm} 0.125</td>
<td>18 \hspace{3cm} 0.035</td>
</tr>
<tr>
<td>31 \hspace{3cm} 0.060</td>
<td>19 \hspace{3cm} 0.050</td>
<td>84 \hspace{3cm} 0.136</td>
<td>44 \hspace{3cm} 0.116</td>
<td>20 \hspace{3cm} 0.030</td>
</tr>
</tbody>
</table>
Table 2

<table>
<thead>
<tr>
<th>$F(t) = t$</th>
<th>$F(t) = t^2$</th>
<th>$F(t) = 2t - t^2$</th>
<th>$F(t) = 3t^2 - 2t^3$</th>
<th>$F(t) = \frac{1}{2} + \frac{\sin^{-1}(2t - 1)}{\pi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(uniform(0,1))</td>
<td>($\beta(2,1)$)</td>
<td>($\beta(1,2)$)</td>
<td>($\beta(2,2)$)</td>
<td></td>
</tr>
<tr>
<td>$k_n$</td>
<td>$\frac{H(k_n)}{n}$</td>
<td>$k_n$</td>
<td>$\frac{H(k_n)}{n}$</td>
<td>$k_n$</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>9</td>
<td>0.181</td>
<td>7</td>
<td>0.132</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>13</td>
<td>0.132</td>
<td>10</td>
<td>0.095</td>
</tr>
<tr>
<td>$n = 300$</td>
<td>16</td>
<td>0.109</td>
<td>12</td>
<td>0.078</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>19</td>
<td>0.095</td>
<td>14</td>
<td>0.068</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>21</td>
<td>0.085</td>
<td>15</td>
<td>0.061</td>
</tr>
</tbody>
</table>

20
Table 3

<table>
<thead>
<tr>
<th>$F(t) = t$ (uniform(0,1))</th>
<th>$F(t) = t^2$ ($\beta(2, 1)$)</th>
<th>$F(t) = 2t - t^2$ ($\beta(1, 2)$)</th>
<th>$F(t) = 3t^2 - 2t^3$ ($\beta(2, 2)$)</th>
<th>$F(t) = \frac{1}{2} + \frac{\sin^{-1}(2t - 1)}{\pi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell_n^*$</td>
<td>$M^<em>(\ell_n^</em>)$</td>
<td>$\ell_n^*$</td>
<td>$M^<em>(\ell_n^</em>)$</td>
<td>$\ell_n^*$</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>9</td>
<td>0.162</td>
<td>7</td>
<td>0.113</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>13</td>
<td>0.119</td>
<td>10</td>
<td>0.082</td>
</tr>
<tr>
<td>$n = 300$</td>
<td>16</td>
<td>0.099</td>
<td>12</td>
<td>0.068</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>19</td>
<td>0.087</td>
<td>14</td>
<td>0.060</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>21</td>
<td>0.078</td>
<td>15</td>
<td>0.054</td>
</tr>
</tbody>
</table>
Table 4

<table>
<thead>
<tr>
<th>$F(t) = t$ (uniform(0, 1))</th>
<th>$F(t) = t^2$ ($\beta(2, 1)$)</th>
<th>$F(t) = 2t - t^2$ ($\beta(1, 2)$)</th>
<th>$F(t) = 3t^2 - 2t^3$ ($\beta(2, 2)$)</th>
<th>$F(t) = \frac{1}{2} + \frac{\sin^{-1}(2t - 1)}{\pi}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_n$</td>
<td>$\frac{N(u_n)}{n}$</td>
<td>$u_n$</td>
<td>$\frac{N(u_n)}{n}$</td>
<td>$u_n$</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>9</td>
<td>0.171</td>
<td>8</td>
<td>0.160</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>13</td>
<td>0.125</td>
<td>12</td>
<td>0.118</td>
</tr>
<tr>
<td>$n = 300$</td>
<td>16</td>
<td>0.104</td>
<td>15</td>
<td>0.099</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>19</td>
<td>0.091</td>
<td>18</td>
<td>0.087</td>
</tr>
<tr>
<td>$n = 500$</td>
<td>21</td>
<td>0.082</td>
<td>20</td>
<td>0.079</td>
</tr>
</tbody>
</table>