Ann. Inst. Statist. Math.
Vol. 52, No. 2, 397–405 (2000)

SOME OPTIMAL STRATEGIES FOR BANDIT PROBLEMS WITH BETA PRIOR DISTRIBUTIONS

CHIEN-TAI LIN[1] AND C. J. SHIAU[2]

[1] *Department of Mathematics, Tamkang University, Tamsui, Taiwan 251, R.O.C.*
[2] *Institute of Mathematical Statistics, National Chung-Cheng University, Chia-Yi, Taiwan 621, R.O.C.*

(Received August 3, 1998; revised November 30, 1998)

Abstract. A bandit problem with infinitely many Bernoulli arms is considered. The parameters of Bernoulli arms are independent and identically distributed random variables from a common distribution with $beta(a, b)$. We investigate the k-failure strategy which is a modification of Robbins's stay-with-a-winner/switch-on-a-loser strategy and three other strategies proposed recently by Berry *et al.* (1997, *Ann. Statist.*, **25**, 2103–2116). We show that the k-failure strategy performs poorly when b is greater than 1, and the best strategy among the k-failure strategies is the 1-failure strategy when b is less than or equal to 1. Utilizing the formulas derived by Berry *et al.* (1997), we obtain the asymptotic expected failure rates of these three strategies for beta prior distributions. Numerical estimations and simulations for a variety of beta prior distributions are presented to illustrate the performances of these strategies.

Key words and phrases: Bandit problems, sequential experimentation, dynamic allocation of Bernoulli processes, staying-with-a-winner, switching-on-a-loser, k-failure strategy, m-run strategy, non-recalling m-run strategy, N-learning strategy.

1. Introduction

A bandit problem consists of a series of choices from a set of Bernoulli stochastic processes, or arms with unknown prior parameters that have to be made. At each decision stage, the decision maker will choose an arm for observation. The choices are sequential in the sense that they can depend on which arms were chosen previously and on the resulting observations. The field of bandit problems is a very fascinating area with wide variety of applications in various branches of sciences. A complete account of these applications can be found in the paper of Banks and Sundaram (1992) and the references contained therein.

Two types of strategies have given rise to interesting decision problems. One is to discount future observations and minimize the expected discounted number of failures. Berry and Fristedt (1985), Gittins (1989), and Banks and Sundaram (1992) have discussed several of these strategies. The other variation is that of Robbins (1952), who considered minimizing the expected long run failure rate (failure proportion). More recently, Herschkorn *et al.* (1995) and Berry *et al.* (1997) have proposed some strategies for the bandit problems with infinitely many arms. This paper can be interpreted as a detailed study of the work of Berry *et al.* (1997) when the parameters of Bernoulli arms are independent and identically distributed random variables from a beta distribution with $a, b > 0$. We will show that the k-failure strategy for the bandit problems with infinitely many arms performs poorly for $b > 1$, and the best strategy among the k-failure strategies is the 1-failure strategy when $0 < b \leq 1$.

In addition to the 1-failure strategy for the use in the bandit problems when $0 < b \leq 1$, the possible competitors are the three strategies proposed by Berry *et al.* (1997). The goal of this article is to compare the asymptotic expected failure rates by using these four strategies. The article is organized as follows. Section 2 shows the claims that the k-failure strategy performs poorly when $b > 1$, and the best strategy among the k-failure strategies is the 1-failure strategy when $0 < b \leq 1$. We then derive the asymptotic expected failure rates by using the 1-failure strategy and the other three strategies proposed by Berry *et al.* (1997). A lower bound for the expected failure proportion over all strategies derived by Berry *et al.* (1997) is also included. Finally, we present the asymptotic estimated expected failure rates based on the formulas in Berry *et al.* (1997) for various beta distributions. We also provide the asymptotic expected failure rates through simulation. Tables are given to illustrate the performance of these strategies and compare them with the lower bound.

2. Main results

Under the assumption that the common prior distribution F is a beta distribution ($beta(a, b)$ with $a, b > 0$), the goal of this section is to investigate some of the results among the k-failure strategies, and three other strategies which are proposed by Berry *et al.* (1997). Some of our results generalize the findings of Berry *et al.* (1997), who have shown a number of results when the prior distribution is uniform $(0, 1)$. In particular, Theorems 1, 2, and 8 generalize Theorems 1, 2, and 3 of theirs respectively. Also, Theorems 6 and 9 extend their Theorem 4, Theorem 7 extends their Theorem 5, and Theorems 5 extends their Theorem 6.

Let us begin by introducing 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 recalling arms that have yielded failures). With the possible exception of the arm being used when the horizon n is reached, every arm yielded exactly k failures. In particular, the 1-failure strategy is a modification of Robbins's stay-with-a-winner/switch-on-a-loser strategy to the infinite-arm setting. The failure rate (failure proportion) of this strategy in n trials, when F is a $beta(a, b)$, is asymptotically equal to

$$\frac{\beta(a,b) - \beta(a+n,b)}{\sum_{j=0}^{n-1} \beta(a+j,b)},$$

where $\beta(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$.

Let $N(n, k, a, b)$ denote the expected number of tosses to the k-th failure or the n-th trial. Since F is a $beta(a, b)$ distribution and $k \leq n$, we have

$$\begin{aligned}
&N(n,k,a,b)\\
&= \int_0^1 \sum_{j=k}^{n} j \binom{j-1}{j-k} \alpha^{j-k}(1-\alpha)^k dF(\alpha) + n \int_0^1 \sum_{j=0}^{k-1} \binom{n}{j} \alpha^{n-j}(1-\alpha)^j dF(\alpha)\\
&= \sum_{j=k}^{n} k \binom{j}{k} \frac{\beta(a+j-k, b+k)}{\beta(a,b)} + n \sum_{j=0}^{k-1} \binom{n}{j} \frac{\beta(a+n-j, b+j)}{\beta(a,b)}\\
&= k \sum_{j=0}^{n-k} \frac{\beta(a+j, b+k)}{(k+j+1)\beta(a,b)\beta(j+1,k+1)}
\end{aligned}$$

$$+\frac{n}{n+1}\sum_{j=n-k+1}^{n}\frac{\beta(a+j,b+n-j)}{\beta(a,b)\beta(j+1,n-j+1)}.$$

Notice that if $b>1$, $N(n,k,a,b)$ is bounded for any $n \geq k$. Hence the expected failure rate $k/N(n,k,a,b)$ of the k-failure strategy does not converge to 0 as $n \to \infty$ for any fixed k, i.e., when $b>1$, the k-failure strategy is a very poor strategy.

On the other hand, for $0<b\leq 1$

$$\frac{N(n,k,a,b)}{k}=\frac{n}{k(n+1)}\sum_{j=n-k+1}^{n}\frac{\beta(a+j,b+n-j)}{\beta(a,b)\beta(j+1,n-j+1)} + \sum_{j=0}^{n-k}\frac{\beta(a+j,b+k)}{(k+j+1)\beta(a,b)\beta(j+1,k+1)}$$

is decreasing in k. Hence the expected failure rate $k/N(n,k,a,b)$ of the k-failure strategy is increasing. Therefore, we have the following theorem.

THEOREM 1. *If F is a beta(a,b) distribution and $0<b\leq 1$. Then the best strategy among k-failure strategies is the 1-failure strategy asymptotically.*

Note that $N(n,1,a,1)\approx\frac{1}{a\ln((n+a)/a)}$, and $N(n,1,a,b)\approx\frac{\Gamma(a+b)n^{1-b}}{\Gamma(a)(1-b)}$ for $0<b<1$. Thus, with different value of b, we have the following two results.

THEOREM 2. *For any fixed k, the expected failure rate of the k-failure strategy is asymptotically equal to $\frac{1}{a\ln((n+a)/a)}$ if F is a beta$(a,1)$ distribution.*

PROOF. For fix k and $k\leq n$, we use the Stirling's expansion to have

$$\begin{aligned}
&N(n,k,a,1)\\
&\quad= an\sum_{j=n-k+1}^{n}\frac{\Gamma(n+1)\Gamma(a+j)}{\Gamma(n+a+1)\Gamma(j+1)}+ak\sum_{j=0}^{n-k}\frac{\Gamma(k+j+1)\Gamma(a+j)}{\Gamma(j+1)\Gamma(a+j+k+1)}\\
&\quad\approx an\sum_{j=n-k+1}^{n}\frac{(a+j-1)^{a-1}}{(n+a)^{a}}+ak\sum_{j=0}^{n-k}\frac{(k+j)^k}{(a+j+k)^{k+1}}\\
&\quad\approx ak\sum_{j=k}^{n}\frac{j^k}{(a+j)^{k+1}}.
\end{aligned}$$

Then the expected failure rate of the k-failure strategy is asymptotically equal to

$$\frac{1}{a\sum_{j=k}^{n}\frac{j^k}{(a+j)^{k+1}}}\approx\frac{1}{a\int_{k/(a+k)}^{n/(a+n)}\frac{u^k}{1-u}du}\approx\frac{1}{a\ln\left(\frac{a+n}{a}\right)}.$$

THEOREM 3. *If $0<b<1$ and F is a beta(a,b) distribution. Then the expected failure rate of the 1-failure strategy is asymptotically equal to $\frac{(1-b)\Gamma(a)}{\Gamma(a+b)n^{1-b}}$.*

From previous discussions we know the k-failure strategy performs poorly under the prior distribution $beta(a,b)$ with $b>1$, and the best strategy among the k-failure strategies is the 1-failure strategy when $0 < b \leq 1$. Also, it may occur that the asymptotic expected failure proportion of the 1-failure strategy is not good when $\sum_{j=0}^{\infty} \beta(a+j,b)/\beta(a,b) < \infty$. Berry *et al.* (1997) have proposed three strategies and pointed out that their expected failure rates are very close to the lower bound given in the following theorem.

THEOREM 4. (Berry *et al.* (1997) Theorem 11) *For* $1 < c_n < n$ *and* $G(c_n) = \min_{1\leq c\leq n} G(c)$,

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

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

However, they did not provide a detail investigation about these three strategies when the prior distributions F are $beta(a,b)$. It is the motivation of this paper, among other results, to derive the asymptotically expected failure rates of these three strategies by following the results of Berry *et al.* (1997).

Before getting into the details we must introduce these three strategies and their corresponding asymptotically expected failure rates first.

- A strategy is called *an m-run strategy* if it 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 obtained, then the current arm is used for the all remaining trials. If the latter obtained, then the arm with lowest failure proportion among the m arms used so far is used for the all 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 whole remaining trials. Thus, the expected number of failures produced by the m-run strategy will be asymptotically less than or equal to

$$H(n,m) = m + (n-m)\int_0^1 F^m(\alpha)d\alpha.$$

For each n, there exists a k_n such that $H(n,k_n) = \min_{1\leq m\leq n} H(n,m)$.

- A strategy is called *a non-recalling 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 the all remaining trials. If no arm produces a success run of length m, the 1-failure strategy is used for all n trials. Then, the expected number of failures produced by the non-recalling m-run strategy will be asymptotically equal to

$$N(n,m) = \left(\frac{\beta(a,b)}{\beta(a+m,b)} - 1\right) + \left(n - \left(\frac{\beta(a,b)}{\beta(a+m,b)} - 1\right)\sum_{j=0}^{n-1}\frac{\beta(a+j,b)}{\beta(a,b)}\right)\left(1 - \frac{\beta(a+m+1,b)}{\beta(a+m,b)}\right).$$

For each n, we can find u_n such that $N(n,u_n) = \min_{1\leq m\leq n} N(n,m)$.

- A strategy is called *an N-learning strategy* ($N \leq n$) if it follows the 1-failure strategy for the first N trials (the arm used at the Trial N will be used until such time that it yields a failure), and then it calls for using the arm that has performed best during the learning period for the all remaining trials. Under this strategy the expected number of failures will be asymptotically less than or equal to

$$L(n,m) = m + (n-N)\int_0^1 F^m(\alpha)d\alpha \quad \text{where} \quad m = \frac{N\beta(a,b)}{\sum_{j=0}^{n-1}\beta(a+j,b)}.$$

For each n, there exists an m_n such that $L(n,m_n) = \min_{1\leq m\leq N} L(n,m)$. Therefore, we have the following theorem.

THEOREM 5. *If F is a beta(a,b) distribution. Then the expected failure rate of the non-recalling $(cn)^{1/(1+b)}$-run strategy is less than or equal to $(1+b)(cn)^{-1/(1+b)}$ asymptotically, where $c = \Gamma(a+b)/\Gamma(a)$.*

PROOF. Using the Stirling's expansion $\Gamma(x+1) \approx (2\pi)^{1/2}x^{x+1/2}e^{-x}$, the expected number of failures produced by non-recalling m-run strategy can be easily calculated and is asymptotically less than or equal to

$$(a+b+m-1)^b\frac{\Gamma(a)}{\Gamma(a+b)} + \frac{nb}{a+b+m}.$$

We now want to find the value of m that minimizes the equation above. From the differentiation and simplification of the solution we find that m is asymptotically equal to $(cn)^{1/(b+1)}$ with $c = \Gamma(a+b)/\Gamma(a)$, and the corresponding expected failure proportion of non-recalling m-run strategy is asymptotically less than or equal to

$$c^{-1/(b+1)}n^{b/(b+1)}(1+b)/n = (1+b)(cn)^{-1/(b+1)}.$$

THEOREM 6. *If F is a beta$(a,1)$ distribution. Then the expected failure rate of the $\sqrt{n/a}$-run strategy is asymptotically less than or equal to $2/\sqrt{an}$.*

PROOF. Since $F \sim$ *beta*$(a,1)$, we have $F(\alpha) = \alpha^a$ and the expected number of failures produced by the m-run strategy is asymptotically less than or equal to

$$m + n\int_0^1 \alpha^{am}d\alpha = m + \frac{n}{am+1}.$$

Taking the differentiation with respect to m and then setting it equal to zero, we thus have the minimum expected failure proportion of the m-run strategy is asymptotically equal to $2/\sqrt{an}$, and m is asymptotically equal to $\sqrt{n/a}$. It completes the proof.

THEOREM 7. *If F is a beta$(a,1)$ distribution. Then the expected failure rate of the $\sqrt{an}\ln(\frac{a+n}{a})$-learning strategy is asymptotically less than or equal to $2/\sqrt{an}$.*

PROOF. Following the argument in the proof of Theorem 6 with $N \approx am\ln(\frac{n+a}{a})$, we have the expected number of failures produced by the N-learning strategy is asymptotically less than or equal to $m + \frac{n}{am+1}$. Applying the same procedure we therefore reach the conclusion.

THEOREM 8. *If F is a beta$(a,1)$ distribution. Then $\frac{2}{\sqrt{a(a+1)n}}$ is a lower bound for all strategies asymptotically.*

PROOF. From Theorem 4,

$$G(c_n) = c_n \int_0^1 \alpha^a d\alpha + (n - c_n) \int_0^1 \alpha^{ac_n} d\alpha$$
$$= \frac{c_n}{a+1} + \frac{n-c_n}{ac_n+1} \approx \frac{c_n}{a+1} + \frac{n}{ac_n}.$$

Setting $dG(c_n)/dc_n = 0$ and solving, we obtain $c_n = \sqrt{n(a+1)/a}$ and thus $G(c_n)/n \approx 2/\sqrt{a(a+1)n}$.

THEOREM 9. *If F is a beta$(1,b)$ distribution. Then the expected failure rate of the $(Dn)^{bk}$-run strategy is asymptotically less than or equal to C/n^k, where $k = 1/(1+b)$, C is a function of b, and $D = \Gamma(1+\frac{1}{b})/b$.*

PROOF. Since $F \sim beta(1,b)$, we have $F(\alpha) = 1 - (1-\alpha)^b$. Thus, the expected number of failure produced by the m-run strategy is asymptotically less than or equal to

$$m + (n-m) \int_0^1 (1-(1-\alpha)^b)^m d\alpha = m + \frac{n-m}{b} \beta\left(\frac{1}{b}, m+1\right).$$

Using the Stirling's expansion, we get

$$m + (n-m)\Gamma\left(1+\frac{1}{b}\right) \Big/ \left(m+\frac{1}{b}\right)^{1/b} \approx m + n\Gamma\left(1+\frac{1}{b}\right)/m^{1/b}.$$

Again, we take the differentiation to $m + n\Gamma(1+\frac{1}{b})/m^{1/b}$ with respect to m and then set it equal to 0 to find the solution $m = (Dn)^{b/(1+b)}$ with $D = \Gamma(1+\frac{1}{b})/b$ will have a minimum expected failure proportion

$$\frac{(Dn)^{b/(1+b)}}{n} + \frac{\Gamma\left(1+\frac{1}{b}\right)}{(Dn)^{1/(1+b)}} = \frac{D^{b/(1+b)}(1+b)}{n^{1/(1+b)}}.$$

Hence, the proof is completed by simply taking $k = 1/(1+b)$ and $C = D^{bk}(1+b)$.

THEOREM 10. *If F is a beta$(1,b)$ distribution. Then the expected failure rate of the $(Dn)^{bk}\Gamma(1+b)n^{1-b}/(1-b)$-learning strategy is asymptotically less than or equal to C/n^k, where $k = 1/(1+b)$, C is a function of b, and $D = \Gamma(1+\frac{1}{b})/b$.*

PROOF. Using the reasoning as for Theorem 9 with $N \approx m\Gamma(1+b)n^{1-b}/(1-b)$, the expected number of failures produced by the N-learning strategy is asymptotically less than or equal to $m + n\Gamma(1+\frac{1}{b})/m^{1/b}$, and then the result follows directly.

THEOREM 11. *If F is a beta$(1,b)$ distribution. Then A/n^k is a lower bound for all strategies asymptotically, where A is a function of b and $k = 1/(1+b)$.*

PROOF. From $F(\alpha) = 1 - (1-\alpha)^b$ and Theorem 4,

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

Table 1. Estimated and simulated expected failure rates for various distributions.

			lower	Proc. I			Proc. II			Proc. III			Proc. IV	
	n	c_n	bound	k_n	E	S	m_n	E	S	u_n	E	S	E	S
beta(1, 1)	100	13	0.127	9	0.181	0.179	9	0.143	0.168	9	0.165	0.160	0.191	0.221
	200	19	0.093	13	0.132	0.145	13	0.109	0.128	13	0.122	0.115	0.169	0.198
	300	24	0.077	16	0.109	0.128	17	0.092	0.108	16	0.102	0.096	0.159	0.183
	400	27	0.067	19	0.095	0.102	19	0.082	0.096	19	0.089	0.088	0.152	0.169
	500	31	0.060	21	0.086	0.100	22	0.074	0.089	21	0.081	0.076	0.147	0.163
	600	34	0.055	24	0.078	0.091	24	0.069	0.080	23	0.074	0.071	0.143	0.159
	700	36	0.051	25	0.073	0.086	26	0.064	0.075	25	0.069	0.067	0.140	0.159
	800	39	0.048	27	0.068	0.080	27	0.061	0.073	27	0.065	0.063	0.138	0.155
	900	41	0.046	29	0.064	0.079	29	0.058	0.068	29	0.061	0.060	0.135	0.153
	1,000	44	0.043	31	0.061	0.074	31	0.055	0.066	30	0.058	0.056	0.133	0.154
beta(2, 1)	100	12	0.075	7	0.132	0.132	7	0.097	0.117	12	0.113	0.104	0.117	0.136
	200	17	0.054	10	0.095	0.101	10	0.074	0.090	18	0.084	0.079	0.101	0.119
	300	21	0.045	12	0.078	0.082	12	0.063	0.075	23	0.070	0.070	0.094	0.112
	400	24	0.039	14	0.068	0.072	14	0.056	0.066	26	0.062	0.060	0.089	0.102
	500	27	0.035	15	0.061	0.066	15	0.051	0.060	30	0.056	0.054	0.086	0.098
	600	30	0.032	17	0.056	0.060	17	0.047	0.057	33	0.051	0.052	0.083	0.095
	700	32	0.030	18	0.052	0.055	18	0.044	0.051	36	0.048	0.047	0.081	0.095
	800	34	0.028	20	0.049	0.053	20	0.042	0.050	38	0.045	0.044	0.080	0.090
	900	36	0.026	21	0.046	0.052	21	0.040	0.050	41	0.043	0.042	0.078	0.091
	1,000	38	0.025	22	0.044	0.046	22	0.038	0.044	43	0.041	0.039	0.077	0.089
beta(1, 2)	100	18	0.288	13	0.338	0.377	14	0.307	0.371	4	0.399	0.361	0.505	0.519
	200	28	0.235	21	0.275	0.330	22	0.255	0.310	5	0.332	0.302	0.502	0.510
	300	36	0.209	27	0.243	0.303	29	0.228	0.288	6	0.297	0.277	0.502	0.508
	400	44	0.191	33	0.222	0.281	34	0.210	0.275	7	0.274	0.257	0.501	0.502
	500	51	0.179	38	0.208	0.268	40	0.197	0.260	8	0.257	0.241	0.501	0.502
	600	57	0.169	43	0.196	0.258	45	0.186	0.252	9	0.244	0.226	0.501	0.504
	700	63	0.161	47	0.187	0.250	49	0.178	0.236	9	0.233	0.218	0.501	0.502
	800	69	0.154	52	0.179	0.243	54	0.171	0.233	10	0.225	0.210	0.501	0.506
	900	74	0.149	56	0.173	0.243	58	0.165	0.222	10	0.217	0.204	0.501	0.504
	1,000	79	0.144	60	0.167	0.229	62	0.160	0.221	10	0.211	0.197	0.501	0.503
beta(2, 2)	100	16	0.192	10	0.252	0.279	11	0.219	0.262	6	0.281	0.256	0.340	0.357
	200	24	0.155	15	0.202	0.243	17	0.181	0.222	8	0.232	0.214	0.337	0.346
	300	32	0.137	20	0.177	0.215	21	0.161	0.210	9	0.208	0.195	0.336	0.341
	400	38	0.125	24	0.162	0.206	25	0.149	0.190	11	0.191	0.183	0.335	0.340
	500	44	0.116	27	0.150	0.192	29	0.139	0.185	12	0.179	0.173	0.335	0.337
	600	49	0.110	31	0.142	0.178	33	0.132	0.174	13	0.170	0.162	0.334	0.337
	700	55	0.104	34	0.135	0.178	36	0.126	0.166	13	0.163	0.151	0.334	0.337
	800	59	0.100	37	0.129	0.175	39	0.121	0.163	14	0.156	0.149	0.334	0.337
	900	64	0.096	40	0.124	0.168	42	0.116	0.157	15	0.151	0.144	0.334	0.334
	1,000	69	0.093	43	0.120	0.162	45	0.113	0.156	15	0.146	0.141	0.334	0.336

Applying the Stirling's expansion to get the approximation $c_n + n\Gamma(1+\frac{1}{b})/c_n^{1/b}$. Setting $dG(c_n)/dc_n = 0$ and solving, we get $c_n = (n\Gamma(1+\frac{1}{b})/b)^{b/(b+1)}$, and then

$$\frac{G(c_n)}{n} \approx \frac{\left(\Gamma\left(1+\frac{1}{b}\right)/b\right)^{b/(b+1)} + \left(\Gamma\left(1+\frac{1}{b}\right)\right)^{b/(b+1)} b^{1/(b+1)}}{n^{1/(1+b)}}.$$

We take $k = \frac{1}{1+b}$ and $A = (\Gamma(1+\frac{1}{b})/b)^{b/(b+1)} + (\Gamma(1+\frac{1}{b}))^{b/(b+1)} b^{1/(b+1)}$ to have the desired result.

For any $0 < b < 1$ we find $1/(1+b) > 1-b$, and therefore, from the results of Theorems 3 and 5, the 1-failure strategy is inferior to the non-recalling strategy asymptotically.

Table 2. Estimated and simulated expected failure rates for various distributions.

			lower	Proc. I			Proc. II			Proc. III			Proc. IV	
	n	c_n	bound	k_n	E	S	m_n	E	S	u_n	E	S	E	S
$beta\left(\frac{1}{2},\frac{1}{2}\right)$	100	11	0.082	9	0.126	0.128	7	0.083	0.093	14	0.084	0.079	0.084	0.099
	200	14	0.054	11	0.084	0.089	9	0.057	0.066	23	0.055	0.054	0.060	0.076
	300	16	0.042	12	0.065	0.069	11	0.045	0.054	30	0.043	0.043	0.050	0.061
	400	18	0.035	14	0.054	0.060	12	0.038	0.047	36	0.036	0.036	0.043	0.054
	500	20	0.030	15	0.047	0.050	13	0.034	0.040	42	0.031	0.031	0.039	0.048
	600	21	0.027	16	0.042	0.046	14	0.030	0.035	48	0.028	0.028	0.035	0.044
	700	22	0.024	17	0.038	0.040	15	0.028	0.035	53	0.025	0.025	0.033	0.041
	800	23	0.022	18	0.035	0.038	16	0.026	0.031	58	0.023	0.023	0.031	0.037
	900	24	0.021	19	0.032	0.035	17	0.024	0.031	63	0.021	0.022	0.029	0.038
	1,000	25	0.019	20	0.030	0.034	17	0.023	0.028	67	0.020	0.020	0.028	0.036
$beta(1,3)$	100	20	0.410	15	0.453	0.515	17	0.425	0.503	2	0.574	0.517	0.667	0.670
	200	32	0.355	25	0.390	0.467	27	0.370	0.449	3	0.507	0.463	0.667	0.668
	300	43	0.325	33	0.356	0.442	36	0.340	0.440	4	0.470	0.437	0.667	0.666
	400	53	0.305	41	0.334	0.417	44	0.320	0.413	4	0.446	0.409	0.667	0.667
	500	62	0.290	48	0.317	0.407	52	0.305	0.399	5	0.427	0.394	0.667	0.665
	600	70	0.278	55	0.304	0.393	59	0.293	0.389	5	0.411	0.374	0.667	0.666
	700	79	0.269	61	0.293	0.392	66	0.283	0.380	5	0.400	0.369	0.667	0.667
	800	87	0.261	68	0.285	0.380	72	0.275	0.369	6	0.389	0.358	0.667	0.668
	900	94	0.254	74	0.277	0.374	78	0.268	0.359	6	0.380	0.350	0.667	0.667
	1,000	102	0.248	79	0.270	0.367	84	0.262	0.366	6	0.372	0.342	0.667	0.667
$beta(2,3)$	100	18	0.295	11	0.351	0.402	13	0.319	0.384	4	0.429	0.383	0.500	0.506
	200	29	0.252	19	0.298	0.367	21	0.276	0.347	5	0.375	0.348	0.500	0.504
	300	38	0.229	25	0.270	0.339	28	0.252	0.325	6	0.347	0.316	0.500	0.501
	400	46	0.215	30	0.252	0.318	34	0.237	0.317	6	0.328	0.307	0.500	0.500
	500	54	0.203	36	0.238	0.313	39	0.225	0.306	7	0.313	0.289	0.500	0.501
	600	62	0.195	40	0.228	0.304	44	0.216	0.293	8	0.302	0.280	0.500	0.502
	700	69	0.188	45	0.219	0.298	49	0.208	0.291	8	0.292	0.278	0.500	0.502
	800	75	0.182	50	0.212	0.282	54	0.202	0.282	8	0.284	0.271	0.500	0.500
	900	82	0.177	54	0.206	0.279	59	0.197	0.272	9	0.277	0.258	0.500	0.501
	1,000	88	0.173	58	0.201	0.277	63	0.192	0.269	9	0.271	0.264	0.500	0.500
$beta(3,3)$	100	17	0.233	9	0.294	0.337	11	0.261	0.321	5	0.348	0.318	0.400	0.411
	200	27	0.198	15	0.247	0.302	18	0.225	0.284	6	0.304	0.282	0.400	0.406
	300	36	0.179	20	0.223	0.273	23	0.205	0.260	7	0.280	0.263	0.400	0.403
	400	43	0.167	25	0.207	0.267	28	0.192	0.261	8	0.264	0.257	0.400	0.401
	500	51	0.158	29	0.196	0.261	33	0.183	0.252	9	0.252	0.235	0.400	0.400
	600	57	0.152	33	0.187	0.250	37	0.175	0.245	10	0.243	0.224	0.400	0.400
	700	64	0.146	37	0.180	0.243	41	0.169	0.241	10	0.235	0.220	0.400	0.400
	800	70	0.141	41	0.174	0.229	45	0.164	0.228	11	0.229	0.217	0.400	0.400
	900	76	0.137	44	0.169	0.235	49	0.159	0.226	11	0.223	0.208	0.400	0.401
	1,000	82	0.134	48	0.164	0.228	52	0.155	0.223	11	0.218	0.210	0.400	0.400

3. Numerical estimations and simulations

To illustrate the results of the preceding section, here we present some numerical data using four strategies for various *beta* distributions. In judging the performance of these strategies that we used in this article, we rely heavily on a lower bound of Berry *et al.* (1997).

Tables 1 and 2 give some examples using four strategies to obtain the estimated expected failure rates for 8 *beta* distributions with $a, b > 0$. Here Proc. I is the m-run strategy, Proc. II is the N-learning strategy, Proc. III is the non-recalling m-run strategy, and Proc. IV is the 1-failure strategy (a modification of Robbin's stay-with-a winner/switch-on-a-loser strategy). Using these strategies give the estimation that we call E in our tables. For comparison, we include the simulated values obtained from 1000 iteration, which we refer to as S. The values of c_n, k_n, m_n, and u_n discussed in the previous section are also presented.

Berry *et al.* (1997) have presented a graph to compare the expected failures rates of the first three strategies for 5 different beta distributions. In their example the N-learning strategy tends to do better in the sense that the asymptotic estimated expected failure rates are closer to the lower bound than both m-run strategy and non-recalling m-run strategy. This result matches with our table values when $a, b \geq 1$. In addition, the non-recalling m-run strategy typically improves on the m-run strategy when $b = 1$, but often does worse than the m-run strategy for $b > 1$.

From the tables we have also found only the 1-failure strategy gives close estimated and simulated values. The 1-failure strategy performs poorly when $b > 1$. As such, it is inferior to any other three strategies for any value of a. When $b = 1$, the other three strategies tend to do a little better than the 1-failure strategy. However, for $beta(1/2, 1/2)$, the N-learning strategy, the non-recalling m-run strategy, and the 1-failure strategy are very close competitors. In particular, the non-recalling m-run strategy can be shown to be the best strategy for $beta(1/2, 1/2)$, which also verifies the fact that it is superior to the 1-failure strategy.

References

Banks, J. S. and Sundaram, R. K. (1992). Denumerable-armed bandits, *Econometrica*, **60**, 1071–1096.

Berry, D. A. and Fristedt, B. (1985). *Bandit Problems: Sequential Allocations of Experiments*, Chapman and Hall, London.

Berry, D. A., Chen, R. W., Zame, A., Heath, D. C. and Shepp, L. A. (1997). Bandit problems with infinitely many arms, *Ann. Statist.*, **25**, 2103–2116.

Gittins, J. C. (1989). *Multi-armed Bandit Allocation Indices*, Wiley, New York.

Herschkorn, S. J., Pekoz, E. and Ross, S. M. (1995). Policies without memory for the infinite-armed Bernoulli bandit under the average-reward criterion, *Probab. Engrg. Inform. Sci.*, **10**, 21–28.

Robbins, H. (1952). Some aspects of the sequential design of experiments, *Bull. Amer. Math. Soc.*, **58**, 527–536.