

BAYESIAN SEQUENTIAL TESTS FOR COMPARING THE NUMBER
OF SPECIES IN TWO POPULATIONS: A SIMULATION STUDY ¹

Jack C. Yue

Murray K. Clayton

Department of Statistics
National Chengchi University
Taipei, Taiwan, R.O.C.

Department of Statistics
University of Wisconsin–Madison
Madison, WI 53706

Key Words and Phrases: Sampling species; Species diversity; Number of species; Sequential sampling; Bayesian decision making; Optimal stopping rule; Group sampling

ABSTRACT

In Yue and Clayton (1996), a sequential Bayesian approach was used to compare the number of species in two populations based on a cost structure, which consists of a sampling cost and a misclassification loss. We discuss the search for good alternative decision rules when the optimal stopping rule is difficult to apply. Different types of decision rules, including look-ahead rules, are examined by means of backward induction and simulations.

1. INTRODUCTION

Two populations, each consisting of unknown number of species, are compared to determine if they have the same number of species. Observations will be drawn sequentially from both populations until we can conclude whether the number of species is the same or not. If there is not enough information, sampling will be

¹This research was supported in part by a USDA–Hatch grant, project No. 3022, UW-Madison.

continued. We are likely to have a better understanding of these populations if more observations are drawn from each; however, the sampling cost also increases accordingly. The objective is to minimize the expected loss, which incorporates a sampling cost and a decision loss.

The following model assumption is the same as that of Lewins and Joanes (1984), and is applied to both populations. Let s denote the number of species and p_i ($1 \leq i \leq s$) denote the proportion of species i in the population. Also, let K be a prior parameter and let the joint prior distribution of s, p_1, p_2, \dots, p_s be

$$\pi(s) \times f(p_1, p_2, \dots, p_s \mid s, K),$$

where $\pi(s)$ is the prior distribution for s . Lewins and Joanes take

$$f(p_1, \dots, p_s \mid s, K) = \frac{\Gamma(Ks)}{[\Gamma(K)]^s} \prod_{i=1}^s p_i^{K-1}, \quad (1)$$

where $K \geq 0$ and $\Gamma(\cdot)$ is the gamma function. We will use $K = 1$, which corresponds to the uniform density case, for mathematical simplicity.

Using the nonsymmetric Dirichlet integral, the posterior distribution of s given $s'(n)$ observed species and $K = 1$ can be shown to be

$$\pi(s \mid n \text{ observations}, s'(n)) \propto \pi(s) \binom{s}{s'(n)} \binom{s+n-1}{n}^{-1} \text{ if } s \geq s'(n). \quad (2)$$

Equation (2) is the same as Lewis and Joanes' Equation (2.2). Also, if there is no danger of confusion, we will use s' to replace $s'(n)$.

The decision loss used is 0–1 loss since we are interested in knowing if the number of species is the same in two populations, i.e. whether H_0 or H_1 is correct, where

$$H_0 : s_1 = s_2 \text{ and } H_1 : s_1 \neq s_2,$$

and s_1 and s_2 are the number of species in Populations 1 and 2, respectively. Accordingly,

$$L(\theta, a_i) = \begin{cases} 0 & \text{if } H_i \text{ is true} \\ 1 & \text{otherwise} \end{cases}$$

where $i = 0$ or 1 , and a_i is the action accepting H_i .

Our goal is to find the optimal stopping rule, the decision rule that has the smallest Bayes risk. In terms of notation, if we let $r(\pi, \delta)$ be the Bayes risk of a

decision rule δ when π is the prior distribution, then the optimal stopping rule δ^* satisfies

$$\begin{aligned} r(\pi, \delta^*) &= \inf_{\delta} E^{\pi} E_{\theta}^{X_n} [L(\theta, \delta) + nc] \\ &= \inf_{\delta} E^{X_n} E^{\pi^n} [L(\theta, \delta) + nc], \end{aligned} \quad (3)$$

where X_n represents n pairs of observations and π^n is the posterior distribution given X_n . The notation used here is same as in Berger (1985). Throughout this study, Equation (4) will be used to compare decision rules.

Note that the expected posterior loss of taking the action a_1 after n pairs of observations are sampled, $\rho(\pi^n, a_1)$, can be expressed as

$$\rho(\pi^n, a_1) = E^{\pi^n} L(\theta, a_1) = \sum_{s \geq \max\{s'_1, s'_2\}} \pi_1(s_1 = s | n, s'_1) \cdot \pi_2(s_2 = s | n, s'_2),$$

where s'_1 and s'_2 are the numbers of observed species from n pairs of observations in the first and second populations, respectively, and π^n is the joint posterior distribution of s_1 and s_2 . Since $L(\theta, a_0) = 1 - L(\theta, a_1)$, $\rho(\pi^n, a_0) = 1 - \rho(\pi^n, a_1)$. Define $\rho_0(\pi^n)$ to be the expected posterior loss of the best action taken when n pairs of samples are collected. Then $\rho_0(\pi^n) = \min\{\rho(\pi^n, a_0), \rho(\pi^n, a_1)\}$.

Suppose we are allowed to take at most a finite and known number of pairs of observations; then backward induction can be used to find the best decision rule under the given restriction. For example, if we can take at most m more pairs of observations when n pairs of observations have already been sampled, then we stop sampling for the first i ($0 \leq i \leq m$) for which $\rho_0(\pi^{n+i}) = \rho_{m-i}(\pi^{n+i})$, where $\rho_i(\pi^n)$ is the expected posterior loss when n pairs of observations are sampled and at most i more pairs of observations are allowed,

$$\rho_m(\pi^n) = \min\{\rho_0(\pi^n), E^*[\rho_{m-1}(\pi^n | y_{n+1})] + c\}, \quad (5)$$

and y_{n+1} is the $(n+1)$ th pair of observations. The best decision rule which takes at most m pairs of observations is called the *m-truncated rule*. The truncated rule is easy to use but change of the truncation bound may cause the truncated rule to become very suboptimal. Detailed discussions of truncated rules can be seen in Berger.

Another frequently used decision rule is the *look-ahead* rule. As an example, the *one-step look-ahead* rule, or the *myopic* rule, is to stop when $\Delta_n(\pi^n) \leq c$, where

$$\Delta_n(\pi^n) \equiv \rho_0(\pi^n) - E^*(\rho_0(\pi^n) | y_{n+1}).$$

Note that $\Delta_n(\pi^n) \leq c$ is equivalent to $\rho_0(\pi^n) = \rho_1(\pi^n)$, i.e. the expected gain of information is equal to the sampling cost. Unlike the truncated rule, the one-step rule provides a good approximation to the optimal stopping rule in several common exponential family-conjugate prior cases (Berger, 1985). In our previous study (Yue, 1994), the one-step rule is shown to be a good approximation of the optimal rule when squared-error loss is used.

In order to proceed with the discussion of the optimal and suboptimal rules, results from Yue and Clayton (1996) are listed as follow:

Theorem 1 *Suppose n observations have been taken and $s'(n)$ is given. Then the marginal probability of observing k new species in m additional observations is*

$$\begin{aligned} & P^*(s'(m+n) - s'(n) = k \mid n, s'(n)) \\ &= \binom{m}{k} \sum_s \frac{\prod_{i=0}^{k-1} (s - s' - i) \prod_{i=k}^{m-1} (s' + n + i)}{\prod_{i=0}^{m-1} (s + n + i)} \pi(s \mid n, s') \end{aligned} \quad (6)$$

where $0 \leq k \leq m$.

Note that given n observations and s' observed species, when $m = 1$ the marginal probability of observing a new species is $(s - s')/(s + n)$, which is the same as in Hill (1979).

Theorem 2 $\Delta_n(\pi^n) \geq 0$ in the case of 0-1 loss.

Theorem 3 *If there exists a decision rule which has finite Bayes risk and $\rho_0(\pi^n) \leq M_0 (M_0 > 0)$, then $\lim_{m \rightarrow \infty} r_m(\pi^n) = r(\pi)$ and the optimal stopping rule exists.*

Thus, the optimal rule exists since $\rho_0(\pi) \leq 1/2$.

Theorem 4 *If H_0 is rejected at stage n and $n + 1$ regardless of the data at stage $n + 1$, then $\Delta_n(\pi^n) = 0$, and the one-step rule stops sampling at stage n .*

Using Theorem 4, we proved the following two results:

Theorem 5 *If the prior distributions of s_1 and s_2 are bounded, then the optimal rule exists and is truncated.*

Theorem 6 *If the prior distributions of s_1 and s_2 are unbounded, if $\pi(s) > 0$ and $\pi(s+1)/\pi(s)$ is non-increasing except for a finite number of $\pi(s)$'s and*

$$\frac{\pi(s)\pi(s+2)}{\pi(s+1)^2} \rightarrow 1 \text{ as } s \rightarrow \infty.$$

then the optimal rule exists and is truncated.

Note that the negative binomial distribution and the Poisson distribution are two unbounded prior distributions that satisfy the conditions in Theorem 6. But if either of the above two conditions is not satisfied, there is no guarantee that there exists an $M > 0$ such that $\Delta_n(\pi^n) \leq c$ for $n \geq M$.

In the following, we will use simulations to compare different decision rules, including the one-step and two-step rules, and the exact Bayes risks are computed if possible. We will also consider group sampling, where the sampling cost is 0.001 or 0.0002 per pair of groups (one group of equal number of observations from each population). We assume that there is no special cost added to individual observations. Discussion of the case when there is a special cost applied to each observation in group sampling can be found in Yue (1994). In addition, the upper bounds of s_1 and s_2 are set to be small (not larger than 20). Since the optimal rule is bounded in the case of binomial prior, the optimal stopping rule can be derived from backward induction. Computations and simulations in this report were based on a DEC 5000 owned and operated by the Department of Statistics at the University of Wisconsin–Madison. The simulations were based on the National Bureau of Standard’s Core Math Library (CMLIB) pseudouniform random number generator UNI.

The comparison of Bayes risk between the one-step rule (and two-step rule) and the optimal stopping rule is considered in Section 2. Proposed decision rules are considered and their simulations are examined in Section 3 & 4.

2. BAYES RISK: THE LOOK-AHEAD RULES

In this section, the Bayes risks of the one-step and two-step rules are computed and compared to the Bayes risk of the optimal stopping rule. The notation $\pi_1 \times \pi_2$ is used to denote the joint prior distribution of (s_1, s_2) . The prior distributions considered are:

- $\pi_1 \times \pi_2 = B(M, \theta_1) \times B(M, \theta_2)$,

where $B(M, \theta)$ is the binomial distribution with parameters M and θ , $\theta_1 = \theta_2 = 0.2, 0.5, 0.8$, and $M = 4, 5, \dots, 10$. The sampling cost is $c = 0.001$ per pair of groups, and we will use g to denote the size of each group, or the number of observations in each group. Readers who are interested in the group sampling case can find a detailed discussion in Yue (1994). Because we assume that there is at least one species in each population, the prior probability that the number of species equals 0 will be set to 0. Likewise, because the first observation is always a new species,

the one-step and two-step rules in this study will take at least one pair of groups of observations. Note that by Theorem 5, the optimal stopping rule is truncated when the prior distributions of s_1 and s_2 are bounded and the species proportions are from a Dirichlet distribution, or Dirichlet(1, 1, ..., 1). Thus, we can use backward induction to construct the optimal stopping rule.

The case with $\theta_1 = \theta_2 = 0.2$ and $g = 1$ is shown in the top graph of Figure 1. Note that when θ_1 and θ_2 are small (e.g. 0.2), a prior weight toward smaller values of s_1 and s_2 is suggested. Thus, a few new species in new observations will change the entire posterior structure of s_1 and s_2 , and possibly reduce the expected posterior loss. For example, in the case of $s_1 \sim B(6, 0.2)$, the prior mode is $s_1 = 1$; while $\pi_1(s_1 = 2)$ is about half of $\pi_1(s_1 = 1)$ and $\pi_1(s_1 = i)$ is almost negligible for $i = 3, 4, 5, 6$. If s_2 is also from the same prior, $B(6, 0.2)$, then $\rho_0(\pi^1)$ is roughly 0.4077. When there is a new species discovered in the second observation, the posterior mode will switch to $s_1 = 2$ with $\pi_1(s_1 \mid n = 2 = s'_1) = 0$ and $\rho_0(\pi^2 \mid s'_1 = 2, s'_2 = 1) = \rho_0(\pi^2 \mid s'_1 = 1, s'_2 = 2)$ is approximately 0.18. On the other hand, if no new species is discovered then $s_1 = 1$ is still the posterior mode and $\rho_0(\pi^2 \mid s'_1 = 1, s'_2 = 1)$ is close to 0.41. Thus, the expected posterior loss of taking one more pair of observations at stage 1 would be smaller than that of stopping at stage 1, and the one-step rule is likely to continue the sampling.

Similarly, if $s_1 = 2$ is the posterior mode after 2 observations are sampled then a new observed species can also switch the posterior mode to $s_1 = 3$ with $\pi_1(s_1 = 2 \mid n = 3, s'_1 = 3) = 0$, which implies that $\rho_0(\pi^3 \mid s'_1 = 3, s'_2 = 2) = \rho_0(\pi^3 \mid s'_1 = 2, s'_2 = 3)$ are close to 0; while $\rho_0(\pi^3 \mid s'_1 = 2, s'_2 = 2)$ is close to 0.5. The cases when $s_1 = 3, 4$, or 5 is the posterior mode also follow. So, if the one-step rule is used then many observations are likely to be taken, similar to the optimal stopping rule. The two-step rule is similar, in which case we consider the possibility of taking one or two more observations. But the prior distribution has less prior weight on smaller values of numbers of species as M increases, and the effect of observing a new species in the next observation is also smaller. This explains that the one-step rule is a good approximation to the optimal rule when M is small, but not as good when $M \geq 7$ (since the prior mode of $B(M, 0.2)$ is no longer at 1 for $M \geq 7$).

The case with $\theta_1 = \theta_2 = 0.2$ and 5 observations per group is shown in the middle graph of Figure 1. As mentioned above, the prior weight is on the smaller values of s_1 and s_2 when $\theta_1 = \theta_2 = 0.2$. If a large number of observations is taken each time, the chance that new species are discovered in new observations is larger and

the posterior distribution is also likely to be altered. Therefore, the one-step and the two-step rules also are good alternative decision rules when the group size is 5. Note that using the one-step rule with group size being 5 would be similar to considering the five-step rule with group size being 1.

When θ_1 and θ_2 are larger, the prior weight is switching to larger values of numbers of species and the posterior distribution of s_1 and s_2 will not be influenced by a few new discovered species. So, as shown in the bottom graph of Figure 1, even when the group size is increased to 5, no observations will be taken under the one-step and the two-step rules if M is large and hence these two decision rules are not good approximations to the optimal rule.

Other cases of $\pi_1 \times \pi_2$ combining different sampling cost and group size structures can be found in Yue (1994). In general, we found that the one-step and two-step rules are good approximations to the optimal stopping rule only when the prior distributions have larger weight on smaller values of s_1 and s_2 , i.e. $\theta_1 = \theta_2 = 0.2$. However, even when $\theta_1 = \theta_2 = 0.2$, the one-step rule and the two-step rules become less effective as M increases. This is due to the fact that there is less prior weight on smaller values of numbers of species as M increases. We do not suggest using the one-step and the two-step rules when θ_1 or θ_2 is large, or when M is large. Unfortunately, the number of species in a natural population tends to be large and the prior distribution of the number of species does not necessarily concentrate on smaller values. Thus the search for good alternative decision rules is essential.

3. ALTERNATIVE DECISION RULES

Simulations in this section were performed under the same condition 1000 times, which reduces the standard error to approximately 0.0316 times that of an individual simulation. Two different classes of simulation assumptions are studied: in one the species of populations are geometric distributions and in the other the species populations are Dirichlet distributions. Reasons why geometric and Dirichlet distributions are used will follow.

Geometric prior distribution: In biological populations, the species proportions p_i 's sometimes follow a (truncated) geometric distribution (see Pielou, 1979, for a more detailed discussion), namely $Geom(\alpha)$, or $p_i \propto (1 - \alpha)^i$, for $i = 1, 2, \dots, s$, where $0 < \alpha < 1$ and s is the number of species. If the parameter α is close to 1, then the species with species proportion p_1 will dominate the whole population and

all other species are rare in this population. On the other hand, a population with $\alpha \rightarrow 0$ implies that this population is close to equiprobable. In our simulations, two cases of geometric distributions are considered: $Geom(0.5)$ and $Geom(0.9)$. Each indicates a more even and uneven population, respectively.

Dirichlet prior distribution: In the previous section, the optimal stopping rule was derived based on the assumption that the prior distribution of the species proportions is a $Dirichlet(1, 1, \dots, 1)$. It would be appropriate to compare the look-ahead rule (and other decision rules) with the optimal stopping rule under this species proportion assumption.

Our proposed approach is similar to Rasmussen and Starr (1979), and is to use an estimate of the conditional probability of discovering new species to form the stopping rule. Specifically, we stop the sampling if

$$\hat{u}_1(n) + \hat{u}_2(n) \leq c$$

where $\hat{u}_1(n)$ and $\hat{u}_2(n)$ are estimates of the conditional probabilities of discovering a new species in Populations 1 and 2. Readers who are interested in other stopping rules can find a detailed discussion in Yue (1994).

This stopping rule is independent of the prior assumptions: both the prior distributions of the number of species and the species proportions. After sampling is terminated, we accept $H_0 : s_1 = s_2$ if

$$\left| \frac{\hat{s}_1(n) - \hat{s}_2(n)}{\sqrt{Var(\hat{s}_1(n)) + Var(\hat{s}_2(n))}} \right| \leq 2,$$

where $\hat{s}_1(n)$ and $\hat{s}_2(n)$ are the first-order jackknife estimators of s_1 and s_2 (Burnham and Overton, 1978), or

$$\hat{s}_1(n) = s'_1(n) + \frac{n-1}{n}N_1^n(1) \quad \text{and} \quad \hat{s}_2(n) = s'_2(n) + \frac{n-1}{n}N_1^n(2),$$

and $N_1^n(1)$ and $N_1^n(2)$ are the species observed exactly once in the sample from Populations 1 and 2, respectively. The estimates of $u_1(n)$ and $u_2(n)$ chosen in this study are Turing's estimate (Rasmussen and Starr, 1979) and NPMLE (Clayton and Frees, 1987, and Lee, 1989). For example, the corresponding Turing's estimate and NPMLE of $u_1(n)$ are

$$\hat{u}_1(n) = \frac{N_1^n(1)}{n} \quad \text{and} \quad \hat{u}_1(n) = \sum_{i=1}^{s_1} \hat{p}_i(1 - \hat{p}_i)^n,$$

where \hat{p}_i is the MLE of p_i based on n observations. In addition, because Turing's estimate depends on $N_1^n(1)$ and $N_2^n(2)$, it is likely that the decision rule using Turing's estimate stops sampling too early when the number of species is small. Therefore, we will require that at least 10 groups of observations are taken from each population. In the following, we will use the phrase "Turing jackknife" and "NPMLE jackknife" to denote these two sampling rules.

Intuitively, we should terminate sampling when fewer species are observed from one population and there is no or little chance of observing more species from this population. Therefore we also modify the stopping rule: sampling can also be terminated if

$$s'_1 < s'_2 \quad \text{and} \quad N_1^n(1) = 0 \quad (7)$$

or

$$s'_2 < s'_1 \quad \text{and} \quad N_1^n(2) = 0. \quad (8)$$

In other words, the sampling also will be terminated if the number of observed species in Population 1 is smaller than that of Population 2 and the chance of discovering new species in Population 1 is very small, i.e. we tend to believe that $s_2 > s_1$ and thus H_0 is rejected. Similar argument holds if $s'_2 < s'_1$ and $N_1^n(2) = 0$. In order to further compare the effect of adapting earlier stopping by using (7) and (8), we also consider two other stopping rules in which 3 and 5 consecutive stages of (7) or (8) are satisfied, rather than 1.

Figure 2 shows the comparison of risks for the proposed Turing and NPMLE jackknife rules and the optimal stopping rule in the case of $\pi_1 \times \pi_2 = B(9, 0.2) \times B(9, 0.2)$ with sampling cost 0.001 per pair of groups and species proportions $Geom(0.5) \times Geom(0.5)$. The risks of both jackknife rules are very close to those of the optimal stopping rule, and the shapes of their risk functions are also similar to that of the optimal stopping rule. The NPMLE jackknife rule is slightly better than the Turing jackknife rule. In addition, risks of the NPMLE jackknife rule decrease as the number of consecutive stages with the early stopping from (7) and (8) increases. Since NPMLE always performs better than Turing's estimate in reducing risk, the following discussion will be mainly focused on NPMLE.

Figure 3 shows the case with $g = 5$, where the species are sampled from populations given by $Geom(0.9) \times Geom(0.9)$, and $\pi_1 \times \pi_2 = B(9, 0.2) \times B(9, 0.2)$. The risks of the NPMLE jackknife rule are also very close to that of the optimal stopping rule, except for very small s_1 's. Figure 4 shows the comparison be-

tween the NPMLE rule and the optimal rule in the case of the Dirichlet prior when $\pi_1 \times \pi_2 = B(9, 0.2) \times B(9, 0.2)$. Note that the optimal stopping rule is originally derived based on the Dirichlet prior distribution and thus risks of the optimal stopping rule are very small, especially when the group size is 5. From these two graphs, the NPMLE jackknife rule has a noticeable decrease in the risk when the group size increases from 1 to 5. In addition, combined with the results in Figure 2, we would suggest using NPMLE with 3 stages of (7) or (8) if $\pi_1 \times \pi_2 = B(9, 0.2) \times B(9, 0.2)$.

Because the jackknife rule is independent of the assumption of the prior, the risks of the NPMLE jackknife rule are unchanged when $\pi_1 \times \pi_2 = B(9, 0.5) \times B(9, 0.5)$ or $\pi_1 \times \pi_2 = B(9, 0.8) \times B(9, 0.8)$. The simulation results of these two cases are in Tables 1, 2, and 3. The differences in risks between these two jackknife rules and the optimal stopping rule are very similar to those shown in the case of $\pi_1 \times \pi_2 = B(9, 0.2) \times B(9, 0.2)$.

Due to computer memory limits, we were unable to find the optimal stopping rule under the prior assumption $\pi_1 \times \pi_2 = B(20, 0.2) \times B(20, 0.2)$. Instead, the 210-truncated rule was compared with the jackknife rules. Figure 5 shows the case of $Geom(0.5) \times Geom(0.5)$, and it is obvious that the risks of both jackknife rules are very close to those of the 210-truncated rule and they all have a significant decrease in the risk from group size 1 to 5. However, when the group size is 5, risks of the Turing jackknife rule are slightly larger than those of the 210-truncated rule, while risks of the NPMLE jackknife rule are slightly smaller than those of the 210-truncated rule. Except when s_1 is very small (e.g. 1 or 2), this similar pattern also appears in the case of $Geom(0.9) \times Geom(0.9)$, which can be seen in Figure 6. For the case of a Dirichlet prior (Figure 7), the risks are different from the geometric distribution: the 210-truncated rule is apparently superior to the NPMLE rule, especially when the group size is 5. This result is similar to that of the case $B(9, 0.2) \times B(9, 0.2)$, in which case the optimal stopping rule is also superior to the jackknife rules because the optimal rule takes into account the prior assumption. Similar pattern can be seen in the cases of $\pi_1 \times \pi_2 = B(20, 0.5) \times B(20, 0.5)$ and $B(20, 0.8) \times B(20, 0.8)$; results are shown on Tables 4, 5, and 6.

Note that stopping when $h(u(n)) \leq c$ is proven to be optimal in the Rasmussen and Starr's payoff structure, i.e. the payoff function $w(n) = h(s'(n)) - cn$ and $h(k+1) - h(k)$ is a nonincreasing function of k . Similar to their assumption, we

can show that stopping when $u_1(n) + u_2(n) \leq c$ is also optimal if the payoff is

$$w(n) = s'_1(n) + s'_2(n) - cn. \quad (9)$$

The proof is similar to that in Rasmussen and Starr.

Theorem 7. *If $w(n)$ is defined as in (9), then it is optimal to terminate the sampling at time*

$$n_0 = \inf\{n \geq 0 : u_1(n) + u_2(n) \leq c\}.$$

Although this theorem generalizes the result of Rasmussen and Starr, it is not a complete generalization. Ideally, we would be able to deal with a more general reward function $w(n) = h(s'_1(n) + s'_2(n)) - cn$ for monotone h . However, we can construct counterexamples to the natural generalization of Theorem 7 for this more general reward function.

It is not surprising that the stopping rule in Theorem 7 is not optimal for the 0-1 loss function that we focus on in this paper. Nonetheless, based on our simulations, it appears that an ad hoc rule based on the stopping rule in Theorem 7, coupled with a suitable terminal decision rule, does provide a good approximation in the cases of a geometric prior or a Dirichlet prior. In particular, the NPMLE jackknife rule is especially good in approximating the optimal stopping rule. This may be due to the fact that the NPMLE jackknife rule used in this study combines both the Turing's estimate and NPMLE, in which case the disadvantage of using a discrete estimator in the Turing's estimate ($N_1^n(1)$ and $N_1^n(2)$) is compensated by using NPMLE. Similar results hold in the case of $B(20, \theta) \times B(20, \theta)$ when jackknife rules are compared to the 210-truncated rule. We suggest using the NPMLE jackknife rule when the number of species tends to be large because this decision rule performs extremely well in our simulations.

4. DISCUSSION

In the above our approach has been to use $\hat{u}_1(n)$ and $\hat{u}_2(n)$ to form a stopping rule. This is based on the intuition underlying Theorem 7, namely, that we have obtained much of the information about species in the populations under study if $\hat{u}_1(n) + \hat{u}_2(n)$ is small. On the other hand, other stopping rules could be constructed based in other considerations. We review some possibilities in this section. In particular, we consider a rule where the sampling is terminated when

$$\sqrt{\text{Var}(s_1 \mid n \text{ obs.})} + \sqrt{\text{Var}(s_2 \mid n \text{ obs.})} \leq 2, \quad (10)$$

or when the sum of the posterior standard deviations is smaller than 2. It should be noted that this stopping rule is not directly related to 0-1 loss and is not necessarily a good approximation of the optimal rule. Readers who are interested in other stopping rules can find a more detailed discussion in Yue (1994).

In the following, we propose three other decision rules based on the stopping rule from (10). After the sampling is terminated, the action taken by these three decision rules is determined by the idea similar to testing a hypothesis, i.e. accept $H_0 : s_1 = s_2$ if

$$\left| \frac{\hat{s}_1(n) - \hat{s}_2(n)}{\sqrt{\text{Var}(\hat{s}_1(n)) + \text{Var}(\hat{s}_2(n))}} \right| \leq 2.$$

The first decision rule is adapted from Lewins and Joanes (1984), in which $\hat{s}_1(n)$ and $\hat{s}_2(n)$ are the posterior modes of s_1 and s_2 , and the variances of $\hat{s}_1(n)$ and $\hat{s}_2(n)$ are the posterior variances of s_1 and s_2 . The second decision rule is the first-order jackknife estimators (see for example Burnham and Overton, 1978, for higher-order jackknife estimators). The variance of the first-order jackknife estimator $\hat{s}_1(n)$ is

$$\text{Var}(\hat{s}_1(n)) = \frac{n-1}{n} N_1^n(1) - \frac{n-1}{n^2} (N_1^n(1))^2$$

and $\text{Var}(\hat{s}_2(n))$ is defined similarly. The third decision rule is based on using $\hat{s}_1(n) = s'_1(n)$ and $\hat{s}_2(n) = s'_2(n)$. Also the variances are calculated as the posterior variances of s_1 and s_2 .

However, these three decision rules are fully dictated by the stopping rule and the stopping rule is independent of the sampling cost. In our simulation, we found that these three decision rules are not good approximations to the optimal stopping rule, and we do not recommend these decision rules.

BIBLIOGRAPHY

- Berger, J. O. (1985) *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York.
- Burnham, K. P. and Overton, W. S. (1978) Estimation of The Size of a Closed Population When Capture Probabilities Vary Among Animals, *Biometrika*, 65, 625-633.
- Clayton, M. K. and Frees, E. W. (1987) Nonparametric Estimation of the Probability of Discovering a New Species, *Journal of the American Statistical Association*, 82, 305-311.

- Lee, J. (1989) *On Asymptotics for the NPMLE of the Probability of Discovering a New species and an Adaptive Stopping rule in Two-Stage Searches*, Ph.D. Thesis, Department of Statistics, University of Wisconsin–Madison.
- Lewins, W. A. and Joanes, D. N. (1984) Bayesian Estimation of the Number of Species, *Biometrics*, 40, 323-328.
- Pielou, E. C. (1979) *Ecological Diversity*, Wiley, New York.
- Rasmussen, S. L. and Starr, N. (1979) Optimal and Adaptive Stopping in the Search for New Species, *Journal of the American Statistical Association*, 74, 661-667.
- Yue, Jack C.S. (1994) *Bayesian Sequential Tests for Comparing the Species Richness of Two Populations*, Ph.D. Thesis, Department of Statistics, University of Wisconsin–Madison.
- Yue, Jack C. and Clayton, M. K. (1996) Bayesian Sequential Tests for Comparing the Number of Species in Two Populations, *Sequential Analysis*, 15, 185-210.

Figure 1. Bayes Risks of the one-step, two-step, and optimal rules, when the prior is $B(M, \theta) \times B(M, \theta)$ and sampling cost is 0.001 per pair of groups.

o: optimal stopping rule, **1**: one-step rule, and **2**: two-step rule.

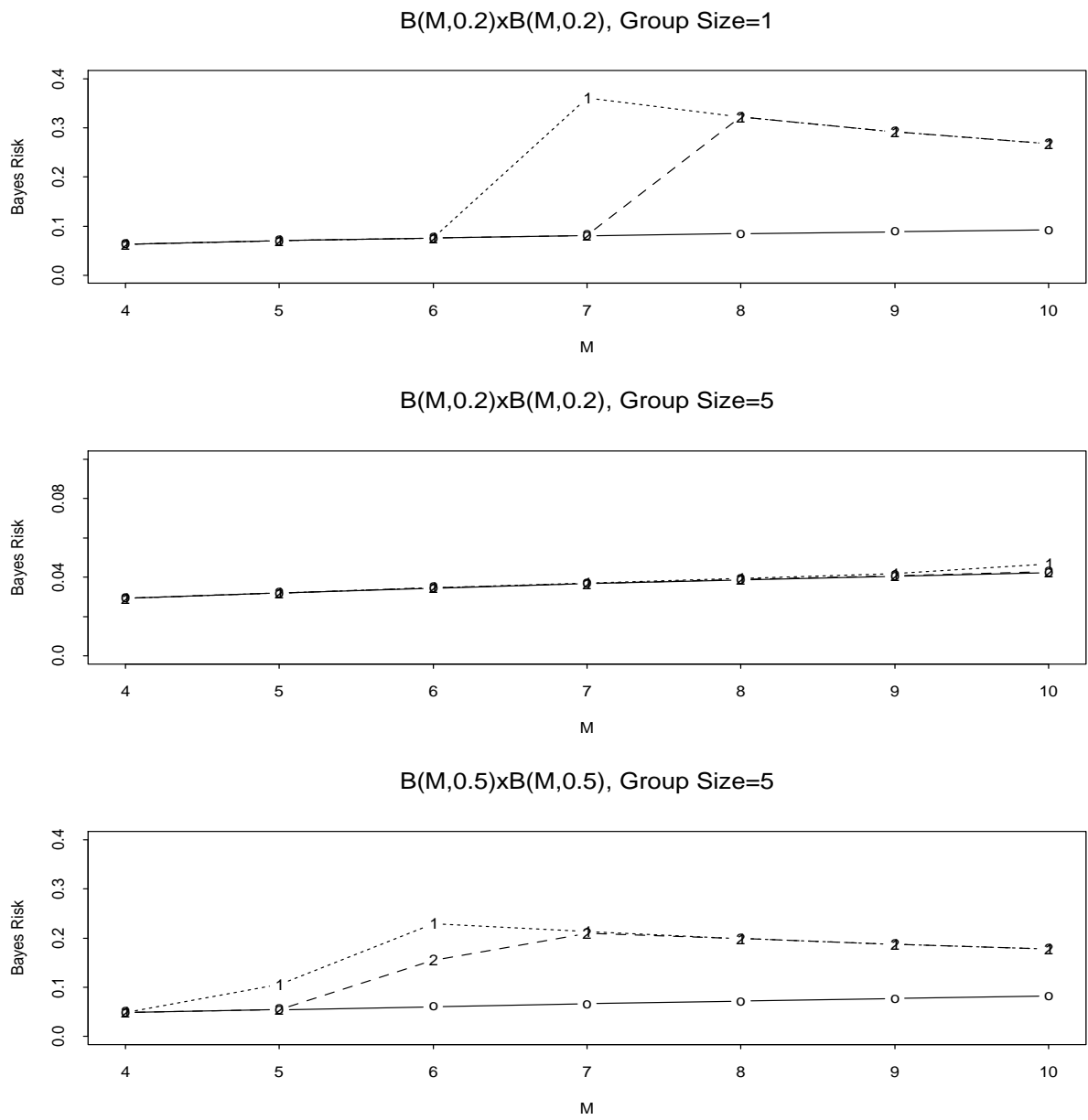


Figure 2. The risks of the optimal, Turing jackknife, and NPMLE jackknife rules in the case of $B(9, 0.2) \times B(9, 0.2)$, sampling cost is 0.001 per pair of groups, $s_2 = 5$, and the species are sampled from populations given by $Geom(0.5) \times Geom(0.5)$: **o** is the optimal stopping rule. **a, b, c** and **A, B, C** are Turing jackknife rule and NPMLE jackknife rule with early stopping if one, three, five consecutive stages of (7) or (8) occur, respectively.

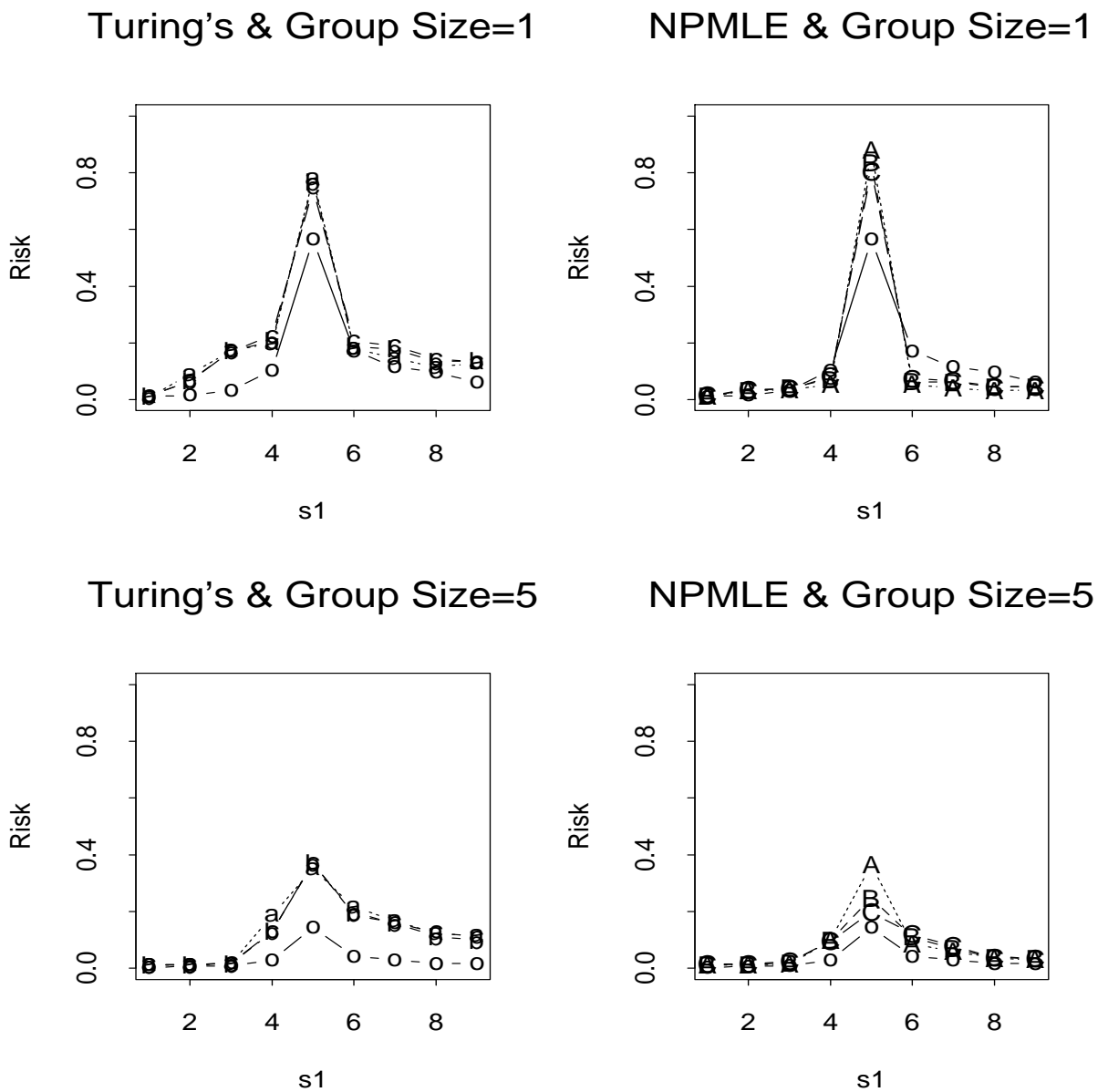


Figure 3. The risks of the optimal rule and NPMLE jackknife rule in the case of $B(9, 0.2) \times B(9, 0.2)$, sampling cost is 0.001 per pair of groups, $s_2 = 5$, and the species are sampled from populations given $Geom(0.9) \times Geom(0.9)$:
o is the optimal rule. **A**, **B**, and **C** are NPMLE jackknife rules with early stopping if one, three, and five stages of (7) or (8) occur

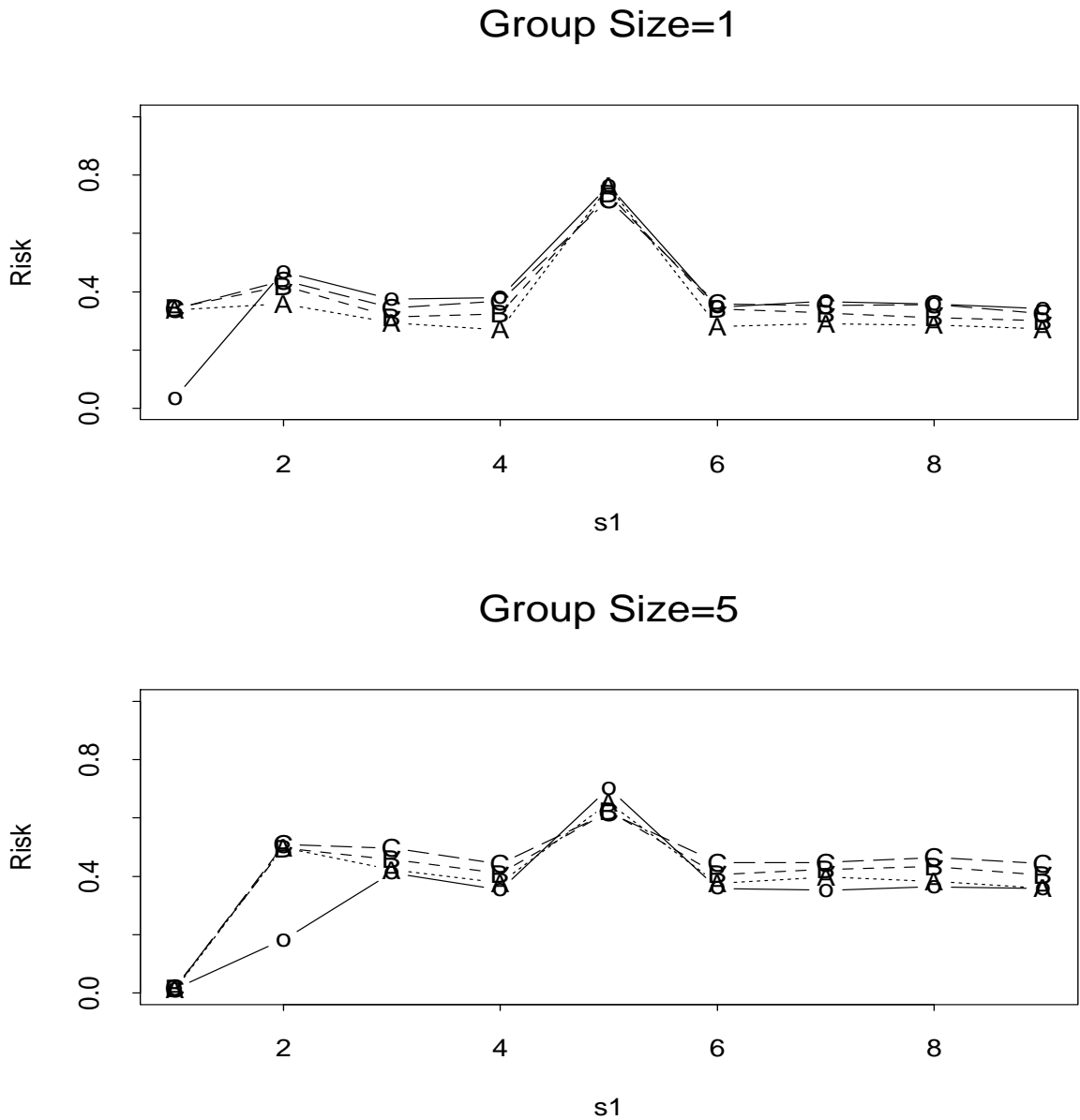


Figure 4. The risks of the optimal rule and the jackknife rule (NPMLE) in the case of $B(9, 0.2) \times B(9, 0.2)$, sampling cost is 0.001 per pair of groups, $s_2 = 5$, and the species are sampled from populations given by Dirichlet distribution: \circ is the optimal rule. **A**, **B**, and **C** are NPMLE jackknife rules with early stopping if one, three, and five stages of (7) or (8) occur.

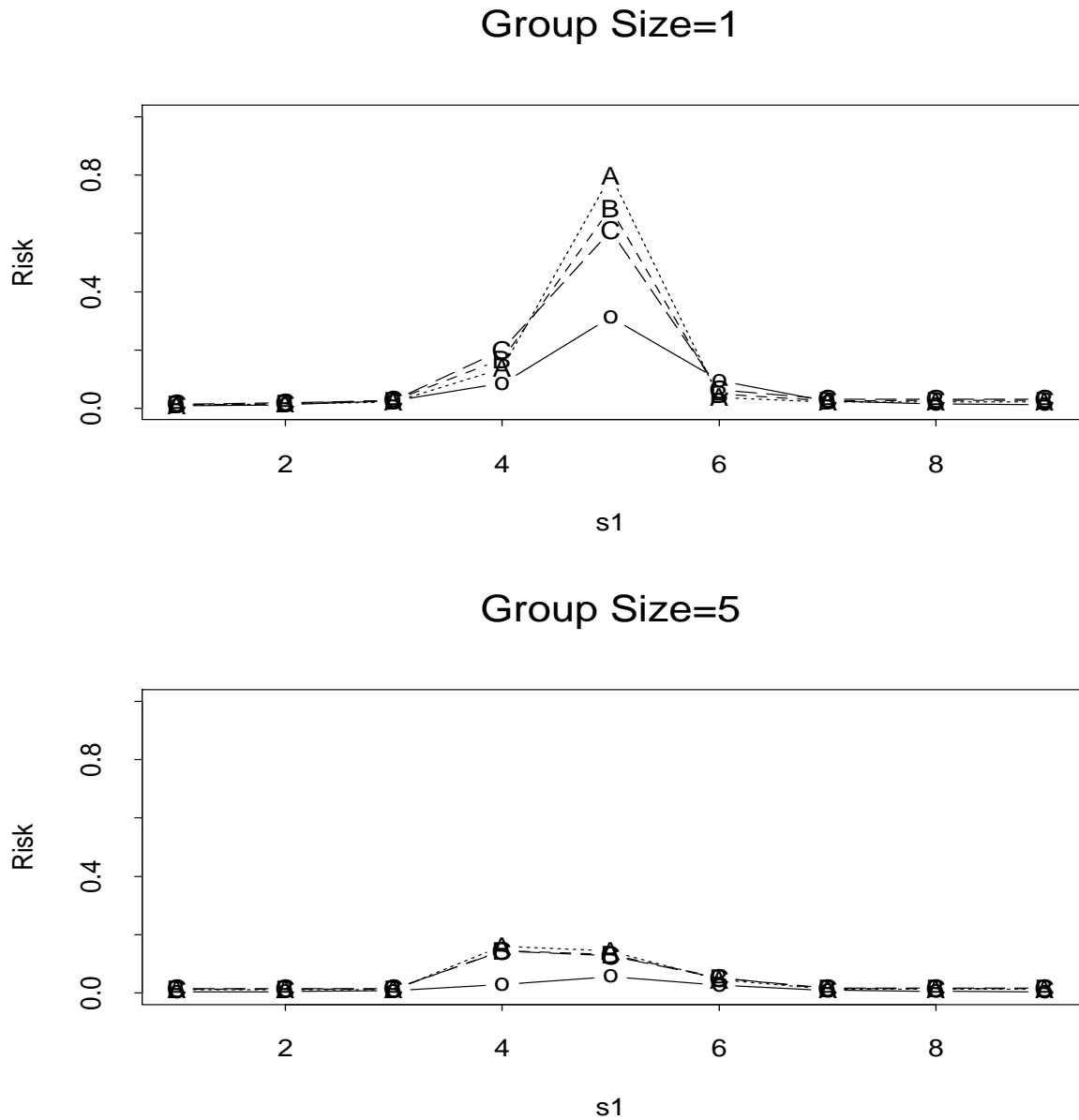


Figure 5. The risks of 210-truncated and NPMLE jackknife rules in the case of $B(20, 0.2) \times B(20, 0.2)$, sampling cost is 0.0002 per pair of groups, $s_2 = 10$, and the species are sampled from populations given by $Geom(0.5) \times Geom(0.5)$: \circ is the optimal rule. **A**, **B**, and **C** are NPMLE jackknife rules with early stopping if one, three, and five stages of (7) or (8) occur.

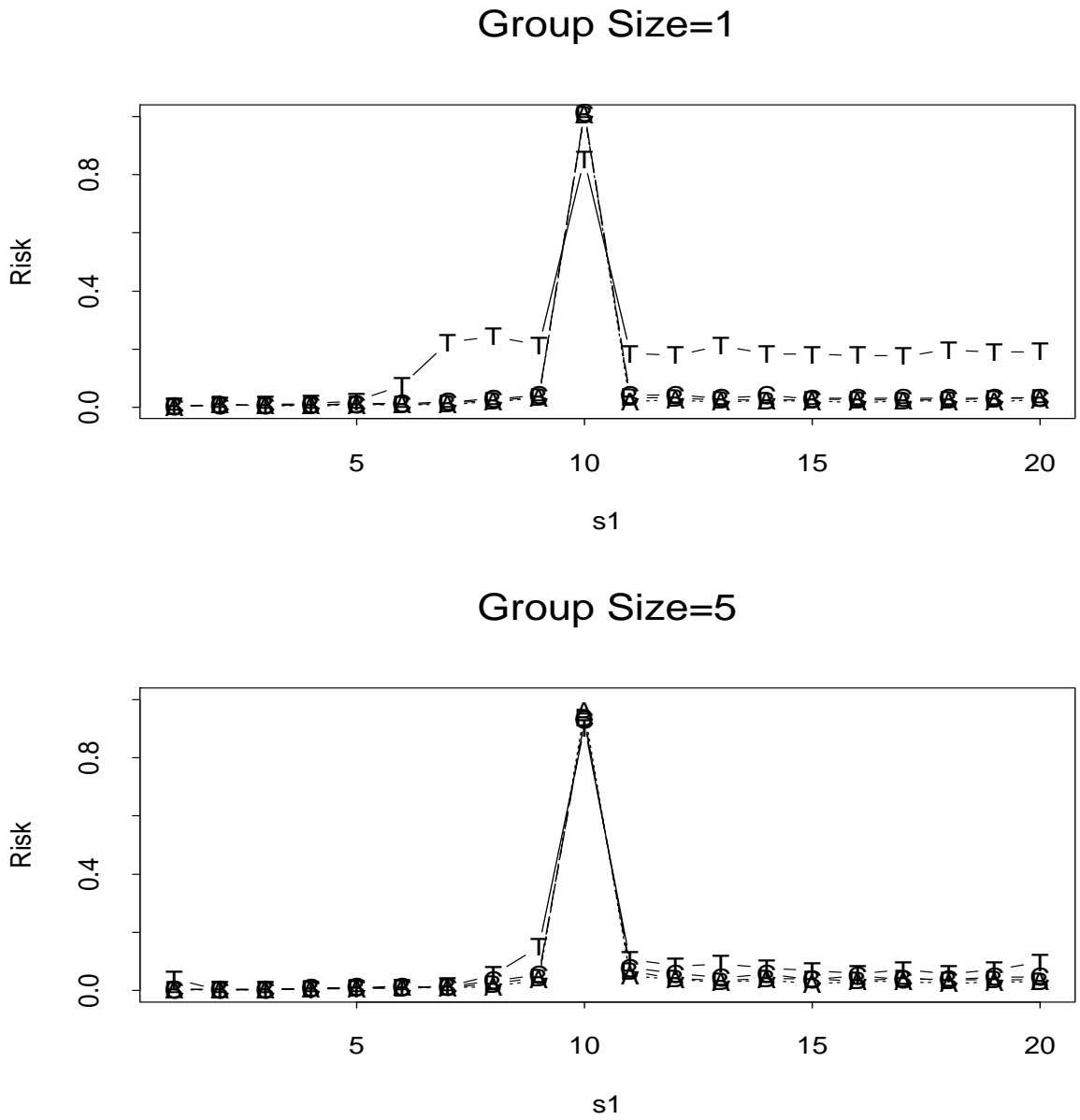


Figure 6. The risks of the optimal and NPMLE jackknife rules in the case of $B(20, 0.2) \times B(20, 0.2)$, sampling cost is 0.0002 per pair of groups, $s_2 = 10$, and the species are sampled from populations given by $Geom(0.9) \times Geom(0.9)$. **o** is the optimal rule. **A**, **B**, and **C** are NPMLE jackknife rules with early stopping if one, three, and five stages of (7) or (8) occur.

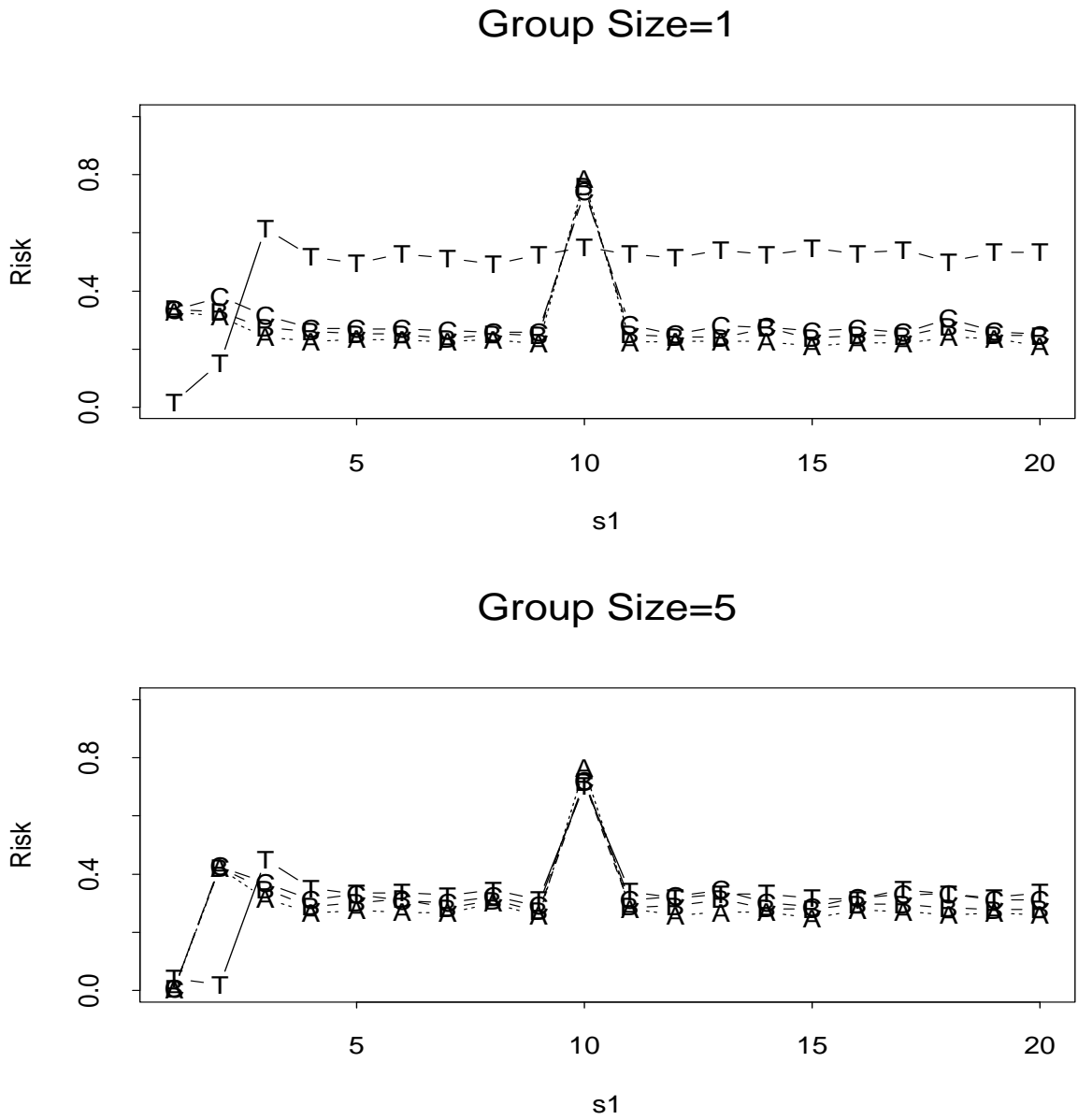
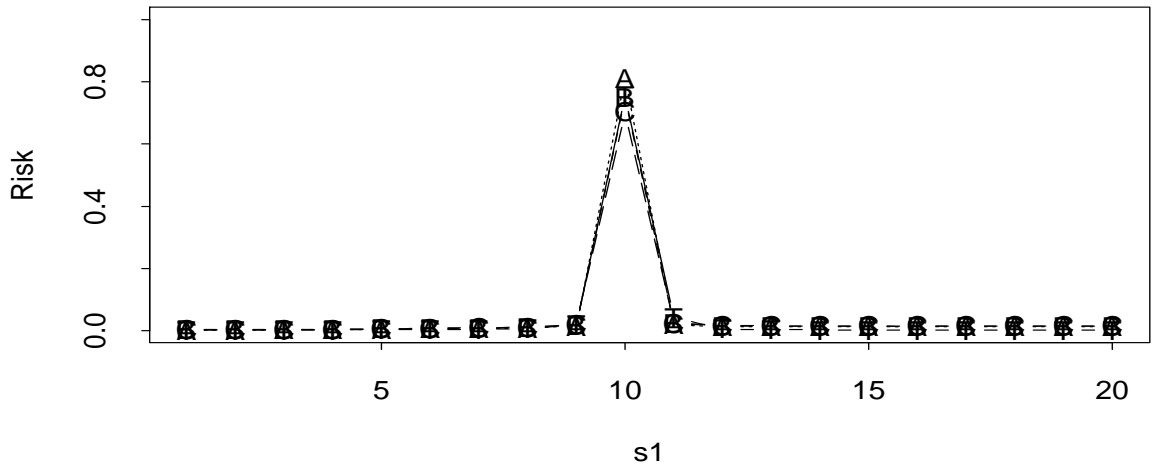


Figure 7. The risks of the optimal rule and NPMLE jackknife rule in the case of $B(9,0.2) \times B(9,0.2)$, sampling cost is 0.0002 per pair of groups, $s_2 = 10$, and the species are sampled from populations given by Dirichlet distribution: \circ is the optimal rule. **A**, **B**, and **C** are NPMLE jackknife rules with early stopping if one, three, and five stages of (7) or (8) occur.

Group Size=1



Group Size=5

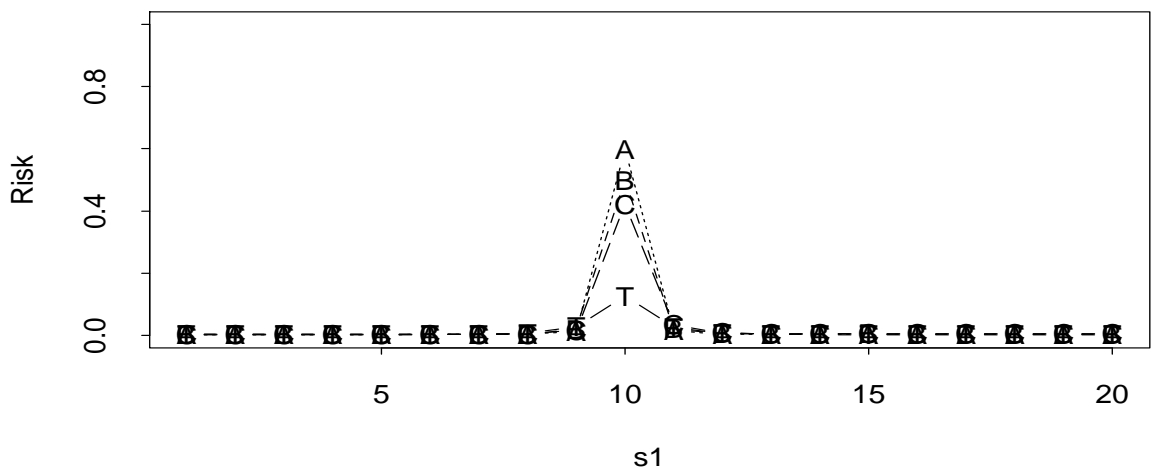


Table 1. The risks of the Turing and NPMLE jackknife rules, and the optimal rules, in the case of $B(9, \theta) \times B(9, \theta)$; sampling cost is 0.001 per pair of groups, and prior is $Geom(0.5) \times Geom(0.5)$.

group size	s_1	stage	Risk (Expected groups)		group size	s_1	Risk (groups)	θ	
			Turing's	NPMLE			Optimal		
1	1	1	.0100 (10.0)	.0100 (10.0)	1	1	.0121 (12.1)	0.5	
		5	.0131 (12.4)	.0139 (12.9)			.0010 (1.0)	0.8	
	5	1	.7779 (24.1)	.8789 (34.8)	5	5	.3586 (148.6)	0.5	
		5	.7453 (30.3)	.8037 (50.2)			1.0010 (1.0)	0.8	
	9	1	1	.1292 (26.5)	.0322 (27.2)	9	9	.0612 (56.2)	0.5
			5	.1322 (34.2)	.0460 (36.0)			.0010 (1.0)	0.8
5	1	1	.0100 (10.0)	.0100 (10.0)	5	1	.0047 (4.7)	0.5	
		5	.0124 (12.4)	.0137 (13.7)			.0080 (8.0)	0.8	
	5	1	.3496 (13.3)	.3651 (21.4)	5	5	.0877 (83.7)	0.5	
		5	.3700 (14.2)	.1979 (25.4)			.1620 (162.0)	0.8	
	9	1	1	.1122 (13.2)	.0283 (13.4)	9	9	.0202 (20.2)	0.5
			5	.1101 (17.0)	.0351 (18.3)			.0272 (27.2)	0.8

Table 2. The risks of the Turing and NPMLE jackknife rules, and the optimal rules, in the case of $B(9, \theta) \times B(9, \theta)$; sampling cost is 0.001 per pair of groups, and prior is $Geom(0.9) \times Geom(0.9)$.

group size	s_1	stage	Risk (Expected groups)		group size	s_1	Risk (groups)	θ	
			Turing's	NPMLE			Optimal		
1	1	1	.3370 (10.0)	.3370 (10.0)	1	1	.0478 (47.8)	0.5	
		5	.2667 (12.2)	.3397 (12.7)			.0010 (1.0)	0.8	
	5	1	.6794 (19.2)	.7621 (23.1)	5	5	.8832 (99.2)	0.5	
		5	.7032 (30.0)	.7170 (41.5)			1.0010 (1.0)	0.8	
	9	1	1	.3311 (19.0)	.2715 (23.8)	9	9	.3114 (97.5)	0.5
			5	.3273 (24.4)	.3248 (35.5)			.0010 (1.0)	0.8
5	1	1	.0110 (10.0)	.0110 (10.0)	5	1	.0194 (19.4)	0.5	
		5	.0134 (11.7)	.0144 (12.4)			.0258 (25.8)	0.8	
	5	1	.5941 (16.3)	.6514 (17.4)	5	5	.6944 (60.4)	0.5	
		5	.6319 (20.2)	.6198 (23.9)			.8434 (104.4)	0.8	
	9	1	1	.4041 (15.1)	.3598 (17.0)	9	9	.4344 (60.4)	0.5
			5	.4252 (21.5)	.4439 (25.5)			.3621 (104.0)	0.8

Table 3. The risks of the Turing and NPMLE jackknife rules, and the optimal rules, in the case of $B(9, \theta) \times B(9, \theta)$; sampling cost is 0.001 per pair of groups, and Dirichlet prior.

group size	s_1	stage	Risk (Expected groups)		group size	s_1	Risk (groups)	θ		
			Turing's	NPMLE			Optimal			
1	1	1	.0100 (10.0)	.0100 (10.0)	1	1	.0096 (9.6)	0.5		
		5	.0130 (13.0)	.0140 (14.0)			.0010 (1.0)	0.8		
	5	1	.7181 (20.0)	.7991 (31.9)		5	.2167 (169.7)	0.5		
		5	.6330 (25.2)	.6102 (51.3)			1.0010 (1.0)	0.8		
	9	1	1	.0328 (21.0)		.0210 (21.0)	9	1	.0167 (16.7)	0.5
			5	.0424 (33.4)		.0306 (30.6)			.0010 (1.0)	0.8
5	1	1	.0100 (10.0)	.0100 (10.0)	5	1	.0042 (4.2)	0.5		
		5	.0120 (12.0)	.0132 (13.2)			.0076 (7.6)	0.8		
	5	1	1	.1357 (11.7)		.1429 (18.9)	5	1	.0840 (84.0)	0.5
			5	.1498 (11.6)		.1266 (19.3)			.1620 (162.0)	0.8
	9	1	1	.0115 (11.5)		.0115 (11.5)	9	1	.0056 (5.6)	0.5
			5	.0145 (14.5)		.0161 (16.1)			.0108 (10.8)	0.8

Table 4. The risks of the Turing and NPMLE jackknife rules, and the 210-truncated rule, in the case of $B(20, \theta) \times B(20, \theta)$; sampling cost is 0.0002 per pair of groups, and prior is $Geom(0.5) \times Geom(0.5)$.

group size	s_1	stage	Risk (Expected groups)		group size	s_1	Risk (groups)	θ		
			Turing's	NPMLE			210-truncated			
1	4	1	.1125 (19.0)	.0048 (19.3)	1	4	.0002 (1.0)	0.5		
		5	.1237 (26.6)	.0077 (28.6)			.0010 (1.0)	0.8		
	10	1	1	.9362 (89.3)		1.0067 (94.2)	10	1	1.0002 (1.0)	0.5
			5	.9051 (137.1)		1.0142 (137.2)			1.0002 (1.0)	0.8
	16	1	1	.0819 (87.6)		.0170 (84.9)	16	1	.0002 (1.0)	0.5
			5	.1027 (138.5)		.0318 (143.6)			.0002 (1.0)	0.8
5	4	1	.0380 (10.5)	.0071 (10.6)	5	4	.0056 (27.8)	0.5		
		5	.0345 (13.9)	.0069 (14.6)			.0012 (6.2)	0.8		
	10	1	1	.8836 (55.2)		.9615 (56.9)	10	1	.7147 (158.4)	0.5
			5	.8632 (77.6)		.9314 (88.8)			1.0014 (7.2)	0.8
	16	1	1	.0905 (59.6)		.0308 (65.3)	16	1	.1915 (137.5)	0.5
			5	.1037 (91.3)		.0498 (101.2)			.0014 (7.3)	0.8

Table 5. The risks of the Turing and NPMLE jackknife rules, and the 210-truncated rule, in the case of $B(20, \theta) \times B(20, \theta)$, sampling cost is 0.0002 per pair of groups, and prior is $Geom(0.9) \times Geom(0.9)$.

group size	s_1	stage	Risk (Expected groups)		group size	s_1	Risk (groups)	θ	
			Turing's	NPMLE			210-truncated		
1	4	1	.3358 (20.8)	.2261 (29.3)	1	4	.0002 (1.0)	0.5	
		5	.3336 (29.9)	.2714 (46.0)			.0002 (1.0)	0.8	
	10	1	.6696 (24.4)	.7851 (30.3)	10	10	1.0002 (1.0)	0.5	
		5	.6805 (28.7)	.7424 (42.8)			1.0002 (1.0)	0.8	
	16	1	1	.3306 (23.5)	.2220 (30.3)	16	16	.0002 (1.0)	0.5
			5	.3258 (30.9)	.2706 (48.4)			.0002 (1.0)	0.8
5	4	1	.3867 (18.2)	.2666 (22.5)	5	4	.4934 (202.2)	0.5	
		5	.3858 (25.3)	.3106 (35.3)			.0008 (4.2)	0.8	
	10	1	.6482 (17.4)	.7652 (22.1)	10	10	.5849 (199.5)	0.5	
		5	.6481 (22.8)	.7201 (32.4)			1.0008 (4.2)	0.8	
	16	1	1	.3756 (16.9)	.2753 (20.8)	16	16	.4728 (199.0)	0.5
			5	.3774 (23.8)	.3188 (33.4)			.0008 (4.2)	0.8

Table 6. The risks of the Turing and NPMLE jackknife rules, and the 210-truncated rule, in the case of $B(20, \theta) \times B(20, \theta)$; sampling cost is 0.0002 per pair of groups, and Dirichlet prior.

group size	s_1	stage	Risk (Expected groups)		group size	s_1	Risk (groups)	θ	
			Turing's	NPMLE			210-truncated		
1	4	1	.0077 (13.8)	.0028 (13.8)	1	4	.0002 (10)	0.5	
		5	.0111 (18.8)	.0038 (18.8)			.0002 (1.0)	0.8	
	10	1	.7413 (53.5)	.8098 (89.1)	10	10	1.0002 (1.0)	0.5	
		5	.6721 (67.7)	.7035 (127.8)			1.0002 (1.0)	0.8	
	16	1	1	.0127 (58.5)	.0117 (58.5)	16	16	.0002 (1.0)	0.5
			5	.0168 (75.4)	.0142 (71.3)			.0002 (1.0)	0.8
5	4	1	.0020 (10.0)	.0020 (10.0)	5	4	.0021 (10.3)	0.5	
		5	.0026 (12.9)	.0028 (14.0)			.0016 (8.0)	0.8	
	10	1	.5458 (17.0)	.5997 (31.6)	10	10	.1263 (191.6)	0.5	
		5	.5188 (17.9)	.4214 (42.1)			1.0038 (18.8)	0.8	
	16	1	1	.0043 (16.4)	.0033 (16.4)	16	16	.0022 (11.3)	0.5
			5	.0053 (22.7)	.0046 (22.8)			.0034 (16.8)	0.8