# BAYESIAN SEQUENTIAL TESTS FOR COMPARING THE NUMBER OF SPECIES IN TWO POPULATIONS[1]

Jack C. Yue

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

Murray K. Clayton

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

We study a problem of sequentially comparing the number of species in two populations. We consider a sequential Bayesian approach which incorporates a sampling cost and a misclassification loss, and examine optimal and sub-optimal stopping rules. The optimal stopping rule is shown to be truncated when the prior distribution of the number of species is bounded, or when the prior is unbounded and satisfies certain conditions.

## 1. INTRODUCTION

In ecology and biology, the comparison of two or more populations based on sampling information is often of interest. In past studies, the number of species

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

and proportions of species in the population, and functions of these, are used to measure species diversity and determine whether two populations are the same. An important measure for describing the evenness of the population is *Shannon's Index* (or the Shannon–Wiener Index), which is defined as $-C\sum_1^s p_i \log p_i$, where $s$ is the number of species, $p_i$ is the proportion of $i$th species, and $C$ is a positive constant. Another frequently used measure is *Simpson's Index*, which is defined as $\sum_1^s p_i^2$ and is the probability that two observations taken randomly belong to the same species. Discussions of these and other diversity measures can be found in Pielou (1975) or Engen (1978). The distinctness, or *complementarity*, of two populations is another way to compare populations. For example, *Marczewski-Steinhau distance*, defined as the ratio of the number of species not in common to the total number of different species in two populations, can be used to measure the complementarity of two or more populations. Examples and discussion of complementarity can be seen in Colwell and Coddington (1994).

The conditional probability of observing a new species in the next observation has received almost as much attention in past studies. This probability is related to *sample coverage* and can be expressed as $u(n) \equiv \sum_1^s p_i I\{\text{species } i \text{ does not appear in first } n \text{ observations}\}$, where $I\{A\}$ is an indicator function. Reviews of past work related to sample coverage can be seen in, for example, Chao (1981), Clayton and Frees (1987), and Lee (1989). In this study, we will focus on the number of species, which will be used to compare populations. This serves as a beginning for the more complex comparison of communities based on diversity measures. One example of comparing species number can be seen in microbial ecology. For example, microbial ecologists usually are interested in comparing the number of species in different communities at the same time or in seeing the change in the number of species in a community over time. See, for example, Kinkel et al. (1992), Andrews et al. (1987) or Pielou (1979).

In early studies, the estimation of the number of species was studied mainly in the equiprobable population case; see for example, Lewontin and Prout (1956), and Harris (1968). Most researchers have used a parametric approach when the equiprobable assumption is relaxed; for example, Sichel (1986) considered the case when the species proportions are from the zero–truncated inverse Gaussian distribution. There have been Bayesian studies based on a parametric assumption as well, such as Hill (1979) and Lewins and Joanes (1984). Several nonparametric approaches also have been taken for estimating the number of species. For exam-

ple, Burnham and Overton (1978) used a jackknife technique; Smith and Van Belle (1984) considered bootstrap procedures; Chao (1984) developed an estimator based on the number of rare species; Chao and Lee (1992) proposed an estimator using the sample coverage. For a more complete review of estimation methods, see Bunge and Fitzpatrick (1993).

While most of the previous work focuses on the nonsequential estimation of the number of species, a sequential method incorporating the sampling cost also has been proposed. Rasmussen and Starr (1979) first introduced a sequential approach with the assumptions that a reward is given for every new species discovered and a sampling cost is collected for every observation sampled. Lee (1989) continued this sequential approach. In addition, Lee also considered a decision approach in a two–stage setting, in which the sampling cost and the reward of discovering new species are different in two stages.

In past studies, the estimation of the number of species in the population was not used to compare the similarity of two or more populations. Note that although the jackknife procedure can be applied to sequentially estimate the number of species, it is not typically used to compare populations. The present paper is focused on comparing the number of species of two populations. Like Rasmussen and Starr, we consider a cost model which consists of a sampling cost and a decision loss. Also, we take a Bayesian approach, similar to Lewins and Joanes (1984). One aspect we will focus on is whether a given Bayesian sequential strategy is truncated. This has been an issue that has received attention for numerous other Bayesian sequential problems. For example, Ray (1965) derived sufficient conditions for truncation when sampling from an exponential family. Clayton (1985) considered a related problem when the probability of observations follows a Dirichlet process.

In the following, in order to proceed with the Bayesian approach, posterior distributions of the number of species and the marginal probability of the samples are required; these are derived in Section 2. In Section 3, we discuss optimal and suboptimal strategies when observations are drawn one from each population at a time. The group sampling case is in Section 4.

## 2. MODEL FOR THE NUMBER OF SPECIES

In this section we give a general introduction to the model of Lewins and Joanes which we will apply to both populations. We start with a model for a single population. 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 $s, p_1, p_2, \ldots, p_s$ be

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

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

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

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 $s'$ observed species and $K = 1$ can be shown to be

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

Equation (2) is the same as Lewis and Joanes' equation (2.2).

To proceed with a Bayesian sequential approach, we also need to calculate the marginal probabilities of observations. Let $s'(m)$ be the number of species observed in $m$ observations, for arbitrary $m$. Then

$$P^*(s'(n+m) - s'(n) = k \mid n, s'(n))$$

denote the marginal probability of discovering $k$ additional new species in $m$ new observations, after $n$ observations are taken and $s'(n)$ is given, and averaging over all $s$. When there is no danger of confusion, we use $s'$ to denote $s'(n)$. This marginal probability can be expressed as

$$\begin{aligned} & P^*(s'(n+m) - s'(n) = k \mid n, s'(n)) \\ & \quad = \sum_s \pi(s \mid n, s') \times P(\ k \text{ new species from } m \text{ new observations} \mid n, s, s'). \end{aligned} \tag{3}$$

In particular, if $m = 1$ then the probability of discovering a new species is the posterior expected value of the unobserved species proportion. Thus, since the posterior expected value of each unobserved species proportion $p_i$ is $1/(n+s)$, given $n$ observations and $s$ species, this probability will be

$$E[u(n) \mid n, s, s'] = \frac{s - s'}{s + n}. \tag{4}$$

This form of marginal probability in the case of $m = 1$ is the same as in Hill (1979). In general, we have the following result.

**Theorem 1** *The marginal probability of observing $k$ new species in $m$ additional observations, after $n$ observations are taken, is given by*

$$P^*(s'(m+n) - s'(n) = k \mid n, s'(n))$$

$$= \begin{pmatrix} m \\ k \end{pmatrix} \sum_s \frac{\prod_{i=0}^{k-1}(s - s' - i)\prod_{i=k}^{m-1}(s' + n + i)}{\prod_{i=0}^{m-1}(s + n + i)} \pi(s \mid n, s') \qquad (5)$$

*where* $0 \le k \le m$.

**Proof:** The proof follows by induction on $m$. The case $m = 1$ is verified by (4). Suppose that (5) is true when the number of additional observations is not larger than $m$. The marginal probability of observing $k$ ($0 \le k \le m+1$) new species from $m+1$ new observations can be divided into two cases:

- Observing $k$ new species from the first $m$ observations, and none from the $(m+1)$st observation.
- Observing $k-1$ new species from the first $m$ observations, and one from the $(m+1)$st observation.

The corresponding marginal probabilities can be expressed using the induction hypothesis, as

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

and

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

where $y_n$ is the $n$th observation and $y_n = 0$ or 1 if the $n$th observation is an observed or newly discovered species, respectively.

Thus, combining the above two marginal probabilities

$$
\begin{aligned}
& P^*(s'(m+n+1) - s'(n) = k \mid n, s') \\
&= \sum_s \frac{\prod_{i=0}^{k-1}(s - s' - i)\prod_{i=k}^{m-1}(s' + n + i)}{\prod_{i=0}^{m}(s + n + i)} \times \\
&\quad \left[ \begin{pmatrix} m \\ k \end{pmatrix} \cdot (s' + n + m + k) + \begin{pmatrix} m \\ k-1 \end{pmatrix} \cdot (s' + n + k - 1) \right] \cdot \pi(s \mid n, s') \\
&= \begin{pmatrix} m+1 \\ k \end{pmatrix} \sum_s \frac{\prod_{i=0}^{k-1}(s - s' - i)\prod_{i=k}^{m}(s' + n + i)}{\prod_{i=0}^{m}(s + n + i)} \pi(s \mid n, s'),
\end{aligned}
$$

which yields the expression (5).

In this study, 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 population 1 and 2, respectively. Accordingly, we use the 0–1 loss function:

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

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

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

where $s_1'$ and $s_2'$ are the numbers of observed species from $n$ pairs of observations in two populations and $\pi^n$ is the joint posterior distribution of $s_1$ and $s_2$. For convenience, we also define

$$f(n, s_1', s_2') \equiv \rho(n, s_1', s_2', a_1) = \rho(\pi^n, a_1).$$

Since $L(\theta, a_0) = 1 - L(\theta, a_1)$,

$$\rho(\pi^n, a_0) = 1 - \rho(\pi^n, a_1).$$

Let $\rho_o(\pi^n)$ denote the expected posterior loss of the best action taken when $n$ pairs of samples are collected. Then

$$\rho_o(\pi^n) = \min\{\rho(\pi^n, a_0), \rho(\pi^n, a_1)\}.$$

Also, let $E^*(\rho_o(\pi^n) \mid y_{n+1})$ be the expected posterior loss of the best action when one more pair of observations is to be gathered. Although $y_n$ is used to indicate the $n$th observation previously, in the following discussion it will be used to denote the $n$th pair of observations. Define

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

Then $\Delta_n(\pi^n)$ is the expected reduction in risk due to the information contributed by the $(n+1)$st pair of observations. Naturally, one would expect that $\Delta_n(\pi^n) \geq 0$ for all $n$ since more observations lead to less uncertainty and thus smaller loss, i.e. $E^*(\rho_o(\pi^n) \mid y_{n+1}, \ldots, y_{n+m})$ is a non-increasing function of $m$, which will be proven in the next section.

## 3. SEQUENTIAL DECISION RULES

In this section, it is assumed that one observation is drawn from each population at a time. Most notation used here has the same definition as in Berger (1985).

Our goal is to find the optimal stopping rule, the decision rule which has the smallest Bayes risk. If the optimal stopping rule takes at most a finite number of observations, i.e. it is truncated, then *backward induction* can be used to determine this optimal rule. However, the truncation bound of the optimal stopping rule is usually difficult to find. Often the truncated procedure and the look–ahead procedure are used as alternative suboptimal decision rules. The truncated procedure is a truncated approximation to the optimal stopping rule; this procedure can also be constructed using backward induction. For example, suppose the $m$–truncated rule (at most $m$ pairs of observations will be taken under this rule) is of interest. Then the Bayes risk of the $m$–truncated rule at stage $n$ can be determined recursively using

$$r_m(\pi^n) = \min\{r_0(\pi^n), E^*[r_{m-1}(\pi^n \mid y_{n+1})]\}, \tag{6}$$

where $r_m(\pi^n)$ is the Bayes risk of the $m$–truncated rule at stage $n$, and stage $n$ is the time when $n$ pairs of observations are sampled. Under the $m$–truncated procedure sampling is terminated for the first $i$ $(0 \leq i \leq m)$ for which

$$r_0(\pi^{n+i}) = r_{m-i}(\pi^{n+i}).$$

Note that since

$$r_0(\pi^n) = \rho_0(\pi^n) + nc,$$

where $c$ is the sampling cost, equation (6) can also be expressed as

$$\rho_m(\pi^n) = \min\{\rho_0(\pi^n), E^*[\rho_{m-1}(\pi^n \mid y_{n+1})] + c\}. \tag{7}$$

With the one–step look–ahead rule, also known as the *myopic* rule, sampling stops when $\Delta_n(\pi^n) \leq c$, or from (7), $\rho_1(\pi^n) = \rho_0(\pi^n)$. Similarly, with the $m$–step rule sampling stops when $\rho_m(\pi^n) = \rho_0(\pi^n)$.

Recall we mentioned in the previous section that $\Delta_n(\pi^n) \geq 0$ for all $n$. Thus, under the one–step rule, we will take as many pairs of observations as possible if $c = 0$. The following Theorem states that $\Delta_n(\pi^n) \geq 0$. The proof can be shown by Doob's process and is omitted.

**Theorem 2** *If the decision loss is 0–1 loss, then* $\Delta_n(\pi^n) \equiv \rho_o(\pi^n) - E^*\rho_o(\pi^n) \geq 0$.

The one–step rule can be used to determine the truncation bound of the optimal stopping rule. If there exists $N > 0$ such that $\Delta_n(\pi^n) \leq c$ for all $n \geq N$ and all $\pi^n$, and $\lim_{m\to\infty} r_m(\pi) = r(\pi)$ ($r(\pi)$ is the Bayes risk of the optimal rule), then the optimal stopping rule is truncated at $N$. This result is Theorem 7.4.4 in Berger (1985). Given this theorem, it is sufficient to show that $\Delta_n(\pi^n) \leq c$ and $\lim_{m\to\infty} r_m(\pi) = r(\pi)$.

Usually people use Theorem 7.4.5 in Berger to show that $\lim_{m\to\infty} r_m(\pi) = r(\pi)$ by showing

$$\lim_{n\to\infty} \rho_0(\pi^n) = 0. \tag{8}$$

Although $\rho_0(\pi^n) \leq 1/2$ for all $\pi^n$ and $c > 0$ in the case of the 0–1 loss, the following example shows that (8) is not always true, and Berger's theorem cannot be applied.

**Example 1** Let $\pi(s) = 0$ for $s \neq 2^m$, $m = 0, 1, 2, \ldots$. Define $\pi(1) = \pi(2) = b$, where $b$ is a constant that lets $\sum \pi(s) = 1$, and we will show the existence of such constant

$b$ later. For $s = 2^{m+1}$, define $\pi(2^{m+1})$ recursively by setting

$$\frac{\pi(2^{m+1})}{\pi(2^m)} = \prod_{i=1}^{2^m} \frac{i(2^{m+1}+i-1)}{(2^m+i-1)(2^m+i)}. \tag{9}$$

Combining with (2), the definition in (9) leads to the identity:

$$\pi(2^m \mid n = s^{'} = 2^m) = \pi(2^{m+1} \mid n = s^{'} = 2^m).$$

In addition, we also have $\pi(2^{m+1})/\pi(2^m) < 1$, since for all $i$

$$\begin{aligned} & i(2^{m+1}+i-1) - (2^m+i-1)(2^m+i) \\ = \;& i^2 + (2^{m+1}-1)i - i^2 - (2^{m+1}-1)i - 2^{2m} + 2^m \\ = \;& -2^{2m} + 2^m < 0. \end{aligned}$$

Therefore, $\pi(2^m)$ decreases as $m$ increases.

We first show that $\sum_{s=1}^{\infty} \pi(s) < \infty$ and thus we can choose an appropriate $b > 0$ which satisfies $\sum_s \pi(s) = 1$. From (9), for $m \geq 2$

$$\begin{aligned} \pi(2^{m+1}) &= \pi(2^m) \times \frac{1(2^m+2^m)}{2^m(2^m+1)} \times \frac{2(2^m+1+2^m)}{(2^m+1)(2^m+2)} \\ &\qquad \times \cdots \times \frac{2^m(2^{m+1}-1+2^m)}{(2^{m+1}-1)2^{m+1}} \\ &< \pi(2^m) \times \frac{2}{2^m} \times \frac{2}{2^{m-1}} \\ &= \pi(2^m) \times \frac{1}{2^{2m-3}}, \end{aligned} \tag{10}$$

since

$$\frac{1(2^m+2^m)}{2^m(2^m+1)} = \frac{2}{2^m+1} < \frac{2}{2^m},$$

$$\frac{2(2^m+1+2^m)}{(2^m+1)(2^m+2)} < \frac{2(2^{m+1}+4)}{2^m(2^m+2)} = \frac{2}{2^{m-1}},$$

and $i(2^{m+1}+i-1) < (2^m+i-1)(2^m+i)$ for all $i$.

From (9), $\pi(2^2) = 5b/9$. Then it follows from (10),

$$\begin{aligned} \sum_s \pi(s) &= \pi(1) + \pi(2) + \pi(2^2) + \sum_{m=2}^{\infty} \pi(2^m) \frac{\pi(2^{m+1})}{\pi(2^m)} \\ &\leq \pi(1) + \pi(2) + \pi(2^2) + \pi(2) \sum_{m=2}^{\infty} \frac{\pi(2^{m+1})}{\pi(2^m)} \\ &< b[\frac{23}{9} + \sum_{m=2}^{\infty} \frac{1}{2^{2m-3}}] \\ &< \infty. \end{aligned}$$

The first inequality above is due to the fact that $\pi(2^m)$ is decreasing.

The proof that $\lim_{n\to\infty}\rho_0(\pi^n)\neq 0$ is similar and now follows: For $k\geq m+1$, similar to (10), we have

$$\frac{\pi(2^{k+1}\mid n=s'=2^m)}{\pi(2^k\mid n=s'=2^m)}<\frac{1}{2^{k-2}}\cdot\frac{1}{2^{k-3}}=\frac{1}{2^{2k-5}}.$$

Likewise, for $k\geq m+1\geq 3$, we get

$$\frac{\pi(2^{k+1}\mid n=s'=2^m)}{\pi(2^m\mid n=s'=2^m)}\leq\frac{1}{2^{k+m-4}}.$$

This implies that

$$\begin{aligned}\sum_{k=m+1}^{\infty}\pi(2^{k+1}\mid n=s'=2^m) &\leq \sum_{k=m+1}^{\infty}\frac{\pi(2^{k+1}\mid n=s'=2^m)}{\pi(2^m\mid n=s'=2^m)}\\ &\leq \sum_{k=m+1}^{\infty}\frac{1}{2^{k+m-4}}\to 0,\qquad \text{as } m\to\infty. \quad (11)\end{aligned}$$

Since $\pi(2^m\mid n=s^{'}=2^m)=\pi(2^{m+1}\mid n=s^{'}=2^m)$ and, by (11), the sum of remaining terms goes to zero, we have

$$\pi(2^m\mid n=s'=2^m)=\pi(2^{m+1}\mid n=s'=2^m)\to\frac{1}{2}\qquad \text{as } m\to\infty. \quad (12)$$

Suppose that both $s_1$ and $s_2$ are from the prior distribution we have constructed using (9). Fix $\epsilon>0$. Then, from (12), we can choose a positive number $M_0=M_0(\epsilon)$ such that for $m\geq M_0$,

$$\pi_i(2^m\mid n=s_i^{'}=2^m)=\pi_i(2^{m+1}\mid n=s_i^{'}=2^m)\geq\frac{1-\epsilon}{2}$$

where $i=1,2$. So, if $n=s_1^{'}=s_2^{'}=2^m$ for $m\geq M_0$ then

$$\begin{aligned}f(n,s_1',s_2') &\geq 2\cdot(\frac{1-\epsilon}{2})^2=\frac{1}{2}(1-2\epsilon+\epsilon^2)\\ &\geq \frac{1}{2}-\epsilon.\end{aligned}$$

On the other hand, $f(n,{s_1}',{s_2}')\leq\frac{1}{2}$. To see this, note first that

$$\sum_n u_n v_n\leq\frac{1}{2} \quad (13)$$

when $u_n\geq 0$ and $v_n\geq 0$ for all $n$, $\max_n u_n\leq 1/2$, and $\sum_n u_n=\sum_n v_n=1$, and $\max_s\pi(s\mid n=s^{'}=2^m)\leq 1/2$ by $\pi(2^m\mid n=s^{'}=2^m)=\pi(2^{m+1}\mid n=s^{'}=2^m)$. Therefore, for $m\geq M_0$

$$\frac{1}{2}\geq f(2^m,2^m,2^m)\geq\frac{1}{2}-\epsilon.$$

In other words, $\frac{1}{2} \geq \rho_0(\pi^n) \geq \frac{1}{2} - \epsilon$. Letting $\epsilon \to 0$, if $n = s_1^{'} = s_2^{'} = 2^m$ then $\rho_0(\pi^n) \to 1/2$, rather than converging to 0.

In the following, we provide an alternative approach to guarantee the convergence of the Bayes risk in the truncated procedure. The proof of the following theorem is similar to that of Theorem 7.4.5 in Berger and is omitted.

**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*

$$\lim_{m \to \infty} r_m(\pi) = r(\pi)$$

*and the optimal stopping rule exists.*

Note that, if the one–step rule is used and $H_0$ would be accepted at both stage $n$ and $n+1$ regardless of the data, then the sampling is stopped at stage $n$. If the decisions associated with $n$ and $n+1$ pairs of observations are the same, no new information is gained from the $(n+1)$st pair of observations and the sampling cost $c$ is wasted. Thus, if we only look ahead one step, the search should be terminated at stage $n$. A similar conclusion when $H_0$ would be rejected holds at the stage $n$ and $n+1$ under the one–step rule. If the one–step rule is used, this result can help us to reduce the computation effort, especially when the posterior distribution or the marginal probability is difficult to calculate. The following theorem proves this result.

**Theorem 4** *Suppose n pairs of observations have been taken. Recall that* $f(n, s_1', s_2')$ $= \rho(\pi^n, a_1)$. *If* $f(n, s_1', s_2') \leq 1/2$ *and* $f(n+1, s_1'+i, s_2'+j) \leq 1/2$ *for* $0 \leq i, j \leq 1$, *i.e. the action* $a_1$ *is chosen at stage n and* $n+1$ *regardless of the data at stage* $n+1$, *the one–step rule stops sampling at stage n.*

**Proof:** If the hypotheses of the theorem hold, it can be shown directly that $E^* \rho_0(\pi^n \mid y_{n+1}) = \rho_0(\pi^n)$ by (5).

The result of Theorem 4 can be extended to the $m$–truncated rule. Since the proof of Theorem 4 indicates that

$$f(n, s_1^{'}, s_2^{'}) = E^*[f(n, s_1^{'}, s_2^{'}) \mid y_{n+1}]$$

if the action $a_1$ is chosen at stage $n$ and $n+1$ regardless of the data at stage $n+1$, it follows by induction that for $m > 0$

$$f(n, s_1', s_2') = E^*[f(n, s_1', s_2') \mid y_{n+1}, \ldots, y_{n+m}]. \tag{14}$$

Thus, if $f(j, s_1'(j), s_2'(j)) \leq 1/2$ for $j = n, n+1, \ldots, n+m$, then

$$\rho_0(\pi^n) = E^*[\rho_0(\pi^n) \mid y_{n+1}, \ldots, y_{n+k}]$$

for $k = 1, \ldots, m$; and using the $m$–truncated rule would stop sampling after $n$ pairs of observations are taken.

**Corollary 1** *If, at a given stage, the sum of posterior probability products are all greater (or all smaller) than* $1/2$ *for* $m+1$ *consecutive stages, regardless of the data, then the* $m$*–step rule would stop sampling at the current stage.*

By Theorem 3 and Theorem 7.4.4 in Berger, we can show that the optimal rule exists, and if there exists $N > 0$ such that $\Delta_n(\pi^n) \leq c$ for all $\pi^n$ and $n \geq N$ then it is also truncated. We now show that this condition on $\Delta_n(\pi^n)$ holds. We start the discussion assuming that the prior distributions of $s_1$ and $s_2$ are bounded.

**Theorem 5** *If the prior distributions of* $s_1$ *and* $s_2$ *are bounded, i.e.* $P(s_1 \leq N_0) = P(s_2 \leq N_0) = 1$ *for some* $N_0 > 0$*, then there exists a positive number* $N$*, such that* $\Delta_n(\pi^n) \leq c$ *for* $n \geq N$*, for all possible* $\pi^n$*.*

**Proof:** Let

$$m_{1,n} \equiv P((n+1)\text{st observation is not new in Population } 1 \mid n)$$

and

$$m_{2,n} \equiv P((n+1)\text{st observation is not new in Population } 2 \mid n). \qquad (15)$$

If for $\epsilon > 0$, we can show that $m_{1,n} > 1 - \epsilon$, $m_{2,n} > 1 - \epsilon$, and

$$\left|\rho_0(\pi^n \mid s_1', s_2') - \rho_0(\pi^{n+1} \mid s_1'(n+1) = s_1', s_2'(n+1) = s_2')\right| \; < \; 2\,\epsilon$$

for sufficiently large $n$ and all $\pi^n$, then as $n \to \infty$

$$\begin{aligned}
\sup_n \Delta_n(\pi^n) &= \sup_n [\rho_0(\pi^n) - E^*(\rho_0(\pi^n) \mid y_{n+1})] \\
&= \sup_n \sum_{0 \leq i,j \leq 1} P(s_1'(n+1) = s_1' + i, s_2'(n+1) = s_2' + j \mid n, s_1', s_2') \\
&\quad \times \left[\rho_0(\pi^n) - \rho_0(\pi^{n+1} \mid s_1'(n+1) = s_1' + i, s_2'(n+1) = s_2' + j)\right] \quad (16) \\
&\leq \sup_n \; m_{1,n} m_{2,n} \left|\rho_0(\pi^n) - \rho_0(\pi^{n+1} \mid s_1'(n+1) = s_1', s_2'(n+1) = s_2')\right| \\
&\quad + m_{1,n}(1 - m_{2,n}) + (1 - m_{1,n}) m_{2,n} + (1 - m_{1,n})(1 - m_{2,n}) \\
&\leq \sup_n \left|\rho_0(\pi^n) - \rho_0(\pi^{n+1} \mid s_1'(n+1) = s_1', s_2'(n+1) = s_2')\right| + 3\epsilon \\
&\leq 5\epsilon \to 0.
\end{aligned}$$

The first inequality is true because of the triangle inequality and $\rho_0(\pi^n) \leq 1/2$.

From (3), since

$$\pi_1(s_1 \mid n, s_1^{'}) \propto \pi_1(s_1) \times \frac{(s_1 - s_1^{'} + 1) \cdots (s_1 - 1)s_1}{s_1(s_1 + 1) \cdots (s_1 + n - 1)}$$

for $s_1 = s_1^{'}, s_1^{'} + 1, \ldots, N_0$, if $\pi_1(s_1) \neq 0$ then we can define

$$a_{s_1}(n, s_1^{'}) \equiv \frac{\pi_1(s_1 + 1 \mid n, s_1^{'})}{\pi_1(s_1 \mid n, s_1^{'})} = \frac{s_1(s_1 + 1)}{(n + s_1)(s_1 - s_1^{'} + 1)} \cdot \frac{\pi_1(s_1 + 1)}{\pi_1(s_1)}.$$

Fix $s_1^{'}$. Because $s_1^{'} \leq s_1 \leq N_0$, it follows that $a_{s_1}(n, s_1^{'})$ is a decreasing function of $n$ for every $s_1$ and $a_{s_1}(n, s_1^{'}) \to 0$ as $n \to \infty$. Also, by the definition of $a_{s_1}(n, s_1^{'})$, we also have the identity

$$\frac{\pi_1(s_1 + k \mid n, s_1^{'})}{\pi_1(s_1 \mid n, s_1^{'})} = \prod_{i=s_1}^{s_1+k-1} a_i(n, s_1^{'}).$$

Thus, $\pi_1(s_1^{'} \mid n, s_1^{'}) \to 1$ as $n \to \infty$ for fixed $s_1^{'}$, since as $n \to \infty$

$$\begin{aligned}
\sum_{s_1=s_1'+1}^{N_0} \pi_1(s_1 \mid n, s_1') &= \pi_1(s_1' \mid n, s_1') \times \sum_{s_1'+1}^{N_0} \prod_{i=s_1'}^{s_1-1} a_i(n, s_1') \\
&\leq \sum_{s_1'+1}^{N_0} a_{s_1'}(n, s_1') \\
&= (N_0 - s_1') a_{s_1'}(n, s_1') \to 0.
\end{aligned}$$

The above inequality holds since $a_{s_1}(n, s_1^{'}) \to 0$. Note that we have assumed that $\pi_1(s_1) > 0$ for all $s_1$. The case in which there are some $s_1$'s such that $\pi_1(s_1) = 0$ is similar.

Therefore, for $\epsilon > 0$, there exists a $N = N(s_1^{'}, \epsilon)$ such that $\pi_1(s_1 = s_1^{'} \mid n, s_1^{'}) > 1 - \epsilon$ when $n \geq N(s_1^{'}, \epsilon)$. Since $P(s_1 \leq N_0) = 1$, we can define

$$N^* = \max_{1 \leq s_1^{'} \leq N_0} N(s_1{}', \epsilon)$$

So, if $n \geq N^*$ then $\pi_1(s_1 = s_1^{'} \mid n, s_1^{'}) > 1 - \epsilon$ for all $\pi^n$ and all $s_1^{'}$.

Because

$$\begin{aligned}
m_{1,n} &= \sum_{s_1=s_1'}^{N_0} \frac{s_1' + n}{s_1 + n} \cdot \pi_1(s_1 \mid n, s_1') \\
&\geq \frac{s_1' + n}{s_1' + n} \pi_1(s_1' \mid n, s_1') \\
&= \pi_1(s_1' \mid n, s_1'),
\end{aligned}$$

if $n \geq N^*$ then for all $\pi^n$ and all $s_1^{'}$

$$1 - m_{1,n} \leq 1 - \pi_1(s_1^{'} \mid n, s_1^{'}) < \epsilon,$$

by construction of $N^*$. Similarly, $1 - m_{2,n} < \epsilon$ when $n \geq N^*$.

To show $\left|\rho_0(\pi^n \mid s_1^{'}, s_2^{'}) - \rho_0(\pi^{n+1} \mid s_1^{'}(n+1) = s_1^{'}, s_2^{'}(n+1) = s_2^{'})\right| < 2\epsilon$, we separate the discussion into three cases: $s_1^{'} > s_2^{'}$, $s_1^{'} < s_2{}'$, and $s_1^{'} = s_2^{'}$. First, for all $\pi^n$, if $s_1^{'} > s_2^{'}$ and $n \geq N^*$ then

$$\begin{aligned} f(n, s_1', s_2') &= \sum_{s_1 = s_2} \pi_1(s_1 \mid n, s_1')\, \pi_2(s_2 \mid n, s_2') \\ &\leq \sum_{s_1'} \pi_2(s_2 \mid n, s_2') \\ &< \epsilon, \end{aligned}$$

since $\sum_{s_1'} \pi_2(s_2 \mid n, s_2^{'}) \leq \sum_{s_2'+1} \pi_2(s_2 \mid n, s_2^{'}) < \epsilon$. The case that $s_1^{'} < s_2^{'}$ and $n \geq N^*$ is similar. If $s_1^{'} = s_2^{'}$ and $n \geq N^*$ then for all $\pi^n$

$$\begin{aligned} f(n, s_1', s_2') &= \sum_{s_1 = s_2} \pi_1(s_1 \mid n, s_1')\, \pi_2(s_2 \mid n, s_1') \\ &\geq \pi_1(s_1' \mid n, s_1')\pi_2(s_1' \mid n, s_1') \\ &> (1 - \epsilon)^2 > 1 - 2\epsilon. \end{aligned}$$

In all three cases,

$$\begin{aligned} &\left|\rho_0(\pi^n \mid s_1', s_2') - \rho_0(\pi^{n+1} \mid s_1'(n+1) = s_1', s_2'(n+1) = s_2')\right| \\ &= \left|\min\{f(n, s_1', s_2'), 1 - f(n, s_1', s_2')\} - \min\{f(n+1, s_1', s_2'), 1 - f(n+1, s_1', s_2')\}\right| \\ &< 2\epsilon. \end{aligned}$$

Theorem 5 guarantees the truncation of the one–step rule, i.e. $\Delta_n(\pi^n) \leq c$ for large enough $n$. Combining Theorem 5 with Theorem 3 and Berger's Theorem 7.4.4, the truncation of the optimal rule holds.

**Corollary 2** *If the prior distributions of $s_1$ and $s_2$ are bounded, i.e. there exists $N_0 > 0$, such that $P(s_1 \leq N_0) = P(s_1 \leq N_0) = 1$, then the optimal rule exists and is truncated.*

When the prior distributions of $s_1$ and $s_2$ are not bounded, there is no guarantee that the one–step rule is truncated. However, we can demonstrate that the one–step rule is truncated if

**(A.1)** $\pi(s) > 0$ and $\pi(s+1)/\pi(s)$ are non–increasing except for a finite number of $\pi(s)$'s

and

**(A.2)** $\pi(s)\pi(s+2)/\pi(s+1)^2 \to 1$ as $s \to \infty$.

Note that if $\pi(s)$ is non-increasing then $\pi(s+1)/\pi(s) \leq 1$; while if $\pi(s+1)/\pi(s)$ is non-increasing then $\pi(s)\pi(s+2)/\pi(s+1)^2 \leq 1$.

**Theorem 6** *If the prior distributions of $s_1$ and $s_2$ satisfy (A.1) and (A.2), then there exists a number $N_0$, such that $\Delta_n(\pi^n) \leq c$ for $n \geq N_0$ and all $\pi^n$.*

In the following, in order to prove that there exist a positive number $N^*$ such that $\Delta_n(\pi^n) \leq c$ for $n \geq N^*$ and all $\pi^n$, we separate the discussion into two parts. Also, when there is no danger of confusion, we use $s_1^{'}$ and $s_2^{'}$ to denote $s_1^{'}(n)$ and $s_2^{'}(n)$.

**Case (i)**: *If (a) $s_1^{'}$ is not the posterior mode of $\pi_1(s_1 \mid n, s_1')$, (b) $s_2'$ is not the posterior mode of $\pi_2(s_2 \mid n, s_2')$, or both, then $\Delta_n(\pi^n) = 0$ for sufficiently large $s_1'$ and $s_2'$.*

We use Lemma 1, 2 and 3 below to prove the above statement. In Lemma 1, we give a bound on the posterior mode in a particular situation.

**Lemma 1** *Suppose the prior distribution of $s$ satisfies (A.1) and (A.2). If $n$ and $s'$ are large enough then $\max_s \pi(s \mid n, s') \leq 1/2$, whenever $s'$ is not the posterior mode of $\pi(s \mid n, s')$.*

**Proof:** From (3),

$$\pi(s \mid n, s') \quad \propto \quad \pi(s)\frac{s(s-1)\cdots(s-s'+1)}{s(s+1)\cdots(s+n-1)}.$$

Then following a similar definition in the proof of Theorem 5, let

$$a_s(n, s') \quad \equiv \quad \frac{\pi(s+1 \mid n, s')}{\pi(s \mid n, s')} = \frac{s(s+1)}{(s+n)(s-s'+1)} \cdot \frac{\pi(s+1)}{\pi(s)} \tag{17}$$

Assume $s_0$ is the posterior mode of $\pi(s \mid n, s^{'})$. Then $\pi(s_0 \mid n, s^{'}) \geq \pi(s_0+1 \mid n, s^{'})$ and $\pi(s_0 \mid n, s^{'}) \geq \pi(s_0-1 \mid n, s^{'})$. Therefore

$$a_{s_0}(n, s^{'}) = \frac{s_0(s_0+1)}{(s_0+n)(s_0-s^{'}+1)}\frac{\pi(s_0+1)}{\pi(s_0)} \leq 1$$

and

$$a_{s_0-1}(n, s^{'}) = \frac{(s_0-1)s_0}{(s_0+n-1)(s_0-s^{'})}\frac{\pi(s_0)}{\pi(s_0-1)} \geq 1.$$

We wish to show $\pi(s_0 \mid n, s^{'}) \leq 1/2$. This will follow if

$$\pi(s_0+1 \mid n, s^{'}) + \pi(s_0-1 \mid n, s^{'}) \geq \pi(s_0 \mid n, s^{'}).$$

Equivalently, we need to show

$$a_{s_0} + \frac{1}{a_{s_0-1}} \geq 1.$$

Note that it is sufficient to check that

$$\frac{a_{s_0}}{a_{s_0-1}} \geq \frac{1}{4}, \tag{18}$$

since if (18) is true, then

$$(a_{s_0} - \frac{1}{a_{s_0-1}})^2 \geq 0 \quad \text{implies} \quad (a_{s_0} - \frac{1}{a_{s_0-1}})^2 + 4\frac{a_{s_0}}{a_{s_0-1}} \geq 1.$$

And so,

$$(a_{s_0} + \frac{1}{a_{s_0-1}})^2 \geq 1.$$

Note that by (A.2), we can find a number $s^*$, such that for $s \geq s^*$

$$\frac{3}{4} \leq \frac{\pi(s)\pi(s+2)}{\pi(s+1)^2} \leq 1.$$

Also, since $s'$ is not the posterior mode of $\pi(s \mid n, s')$ and $s_0 > s'$, it is obvious that $s_0 - s' \geq 1$ and so for $s_0 > s' \geq s^*$ and $n \geq 3$,

$$\begin{aligned}
\frac{a_{s_0}}{a_{s_0-1}} &= \frac{s_0}{s_0-1} \cdot \frac{s_0+n-1}{s_0+n} \cdot \frac{s_0-s'}{s_0-s'+1} \cdot \frac{\pi(s_0-1)\pi(s_0+1)}{\pi(s_0)^2} \\
&\geq \frac{s_0+n-1}{s_0+n} \cdot \frac{s_0-s'}{s_0-s'+1} \cdot \frac{\pi(s_0-1)\pi(s_0+1)}{\pi(s_0)^2} \\
&\geq \frac{n-1}{n} \cdot \frac{1}{2} \cdot \frac{\pi(s_0-1)\pi(s_0+1)}{\pi(s_0)^2} \\
&\geq \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{3}{4} = \frac{1}{4}.
\end{aligned}$$

Thus, it follows that $\max_s \pi(s \mid n, s^{'}) \leq 1/2$.

Recall from (13), if $u_n \geq 0$ and $v_n \geq 0$ for all $n$, $\max_n u_n \leq 1/2$, and $\sum_n u_n = \sum_n v_n = 1$, then

$$\sum_{n=1}^{\infty} u_n v_n \leq \frac{1}{2}.$$

Thus, the result of Lemma 1 leads to $f(n, s_1^{'}, s_2^{'}) \leq 1/2$, and so $H_0$ is rejected at stage $n$ when either $s_1^{'}$ is not the posterior mode of $\pi_1(s_1 \mid n, s_1^{'})$ or $s_2^{'}$ is not the posterior mode of $\pi_2(s_2 \mid n, s_2^{'})$. Likewise, using Lemmas 2 and 3 below,

$$f(n+1, s_1^{'}+i, s_2^{'}+j) \leq 1/2$$

for $0 \leq i, j \leq 1$ when either $s_1^{'}$ is not the posterior mode of $\pi_1(s_1 \mid n, s_1^{'})$ or $s_2^{'}$ is not the posterior mode of $\pi_2(s_2 \mid n, s_2^{'})$. Therefore, by Theorem 4, it is obvious that $\Delta_n(\pi^n) = 0$. The proof of the following two Lemmas is similar to that in Lemma 1.

**Lemma 2** *Suppose the prior distribution of $s$ satisfies (A.1) and (A.2). If $n$ is large enough, then $\max_s \pi(s \mid n+1, s') \leq 1/2$ whenever $s'$ is not the posterior mode of $\pi(s \mid n, s')$.*

**Lemma 3** *Suppose the prior distribution of $s$ satisfies (A.1) and (A.2). If $n$ is large enough, then $\max_s \pi(s \mid n+1, s'+1) \leq 1/2$ whenever $s'$ is not the posterior mode of $\pi(s \mid n, s')$.*

**Case (ii).** *Suppose both $s_1^{'}$ is the posterior mode of $\pi_1(s_1 \mid n, s_1^{'})$ and $s_2^{'}$ is the posterior mode of $\pi_2(s_2 \mid n, s_2^{'})$. Then $\sup_n \Delta_n(\pi^n) \to 0$ as $n \to \infty$.*

Similar to the case when the prior is truncated, for $\epsilon > 0$, we will show that $m_{1,n} > 1-\epsilon$, $m_{2,n} > 1-\epsilon$, and

$$\left|\rho_0(\pi^n \mid s_1^{'}, s_2^{'}) - \rho_0(\pi^{n+1} \mid s_1^{'}(n+1) = s_1^{'}, s_2^{'}(n+1) = s_2^{'})\right| < 2\epsilon$$

for sufficiently large $n$ and all $\pi^n$ satisfying the assumption of this case. Then as $n \to \infty$, $\sup_n \Delta_n(\pi^n) \leq 5\epsilon \to 0$.

The case $m_{1,n}$ is discussed first. Since the prior distribution of $s_1$ satisfies (A.1), i.e. $\pi_1(s_1+1)/\pi_1(s_1) \leq 1$ is non-increasing, there exists $s_0 > 0$ such that $\pi_1(s_1+1)/\pi_1(s_1) < 1$ for $s_1 \geq s_0$; otherwise, $\sum_{s_1} \pi_1(s_1) = \infty$. Let $p_1 = \pi_1(s_0+1)/\pi_1(s_0) < 1$.

Fix $\epsilon > 0$. First suppose $s_1^{'} \geq s_0$. Since $s_1^{'}$ is the posterior mode, we can use $\pi_1(s_1+1)/\pi_1(s_1) \leq p_1 < 1$ and ideas similar to those in the proof of Theorem 5 to show

$$\begin{aligned} 1 - m_{1,n} &= \sum_{s_1 = s_1^{'}} \frac{s_1 - s_1^{'}}{s_1 + n} \pi_1(s_1 \mid n, s_1\prime) \\ &\leq \sum_{s_1 = s_1^{'}} \frac{s_1 - s_1^{'}}{s_1 + n} \left(\frac{1+p_1}{2}\right)^{s - s_1^{'}} \to 0 \end{aligned} \tag{19}$$

as $n \to \infty$. Note that the result (19) holds as long as $s_1^{'} \geq s_0$. Therefore, we can choose an $N_1^* = N_1^*(\epsilon)$ such that $m_{1,n} > 1 - \epsilon$ when $n \geq N_1^*$ and $s_1^{'} \geq s_0$. For $s_1^{'} \leq s_0$, combining the techniques used in the proof of Theorem 5 and in (19), we can show that there exists an $N^* = N^*(\epsilon, s_0)$ such that $m_{1,n} > 1 - \epsilon$ for $n \geq N^*$. Let $N_0^* = \max\{N^*, N_1^*\}$, then $m_{1,n} > 1 - \epsilon$ for all $s_1^{'}$ when $n \geq N_0^*$.

The case of $s_2$ is similar, i.e. $m_{2,n} > 1 - \epsilon$ for all $s_2^{'}$ when $n \geq N_0^*$. Likewise, the proof of

$$\left|\rho_0(\pi^n \mid s_1^{'}, s_2^{'}) - \rho_0(\pi^{n+1} \mid s_1^{'}(n+1) = s_1^{'}, s_2^{'}(n+1) = s_2^{'})\right| < 2\epsilon$$

also follows, although the details are tedious. A detailed discussion appears in the Appendix. Thus, similar to the proof in Theorem 5, $\sup_n \Delta_n(\pi^n) \to 0$ as $n \to 0$ for all $\pi^n$.

**Corollary 3** *If the prior distributions of $s_1$ and $s_2$ satisfy (A.1) and (A.2), then the optimal rule is truncated.*

**Example 2** The negative binomial distribution, which has probability function

$$P(s \mid r, \alpha) = \begin{pmatrix} r+s-1 \\ s \end{pmatrix} \alpha^r (1-\alpha)^s, \qquad s = 0, 1, 2, \ldots,$$

satisfies :

$$\frac{\pi(s+1)}{\pi(s)} = \frac{r+s}{s}(1-\alpha) \leq 1 \text{ if } s \geq \frac{1-\alpha}{\alpha} r,$$

which implies that $\pi(s)$ is non-increasing if $s \geq (1-\alpha)r/\alpha$. Also for large enough $s$,

$$1 \geq \frac{\pi(s)\pi(s+2)}{\pi(s+1)^2} = \frac{s(r+s+1)}{(s+1)(r+s)} \to 1 \text{ as } s \to \infty,$$

since $s(r+s+1) - (s+1)(r+s) = -r < 0$.

Similar results appear for the Poisson distribution, which has probability function

$$P(s \mid \lambda) = \frac{\lambda^s e^{-\lambda}}{s!}, \qquad \text{for} \quad s \geq 0,$$

$$\frac{\pi(s+1)}{\pi(s)} = \frac{\lambda}{s+1} \leq 1 \text{ if } s \geq \lambda$$

and

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

In the following, we construct a distribution which has infinitely many terms with $\pi(s) = 0$ and thus fails to satisfy (A.1). Fix $M > 0$. Then for sufficiently small $c > 0$, there exists a positive number and $N_1 \geq M$ and a $\pi^{N_1}$ such that $\Delta_{N_1} > c$, and Theorem 7.4.4 in Berger cannot be used to provide the truncation of the optimal stopping rule.

**Example 3** Let $\pi(s)$ be defined as in Example 1. Fix $M$. From (11) and (12), i.e.

$$\sum_{k=m+1}^{\infty} \pi(2^{k+1} \mid n = s^{'} = 2^m) \to 0 \qquad \text{as} \quad m \to \infty$$

and

$$\pi(2^m \mid n = s^{'} = 2^m) = \pi(2^{m+1} \mid n = s^{'} = 2^m) \to \frac{1}{2} \qquad \text{as} \quad m \to \infty,$$

we can choose a number $M_0$ such that for $m \geq M_0$ and $M_0 > \log_2 M$

$$\pi(2^m \mid n = s^{'} = 2^m) = \pi(2^{m+1} \mid n = s^{'} = 2^m) \geq \frac{511}{1024}.$$

Similarly, we can also show that

$$\sum_{k=m+1}^{\infty} \pi(2^{k+1} \mid n = s^{'} = 2^m + 1) \to 0$$

and $\pi(2^{m+1} \mid n = s^{'} = 2^m + 1) \to 1$ as $m \to \infty$, while $\pi(2^m \mid n = s^{'} = 2^m + 1) = 0$. This also implies that we can find a number $M_0^*$ such that, for $m \geq M_0^*$ ($M_0^* > \log_2 M$)

$$\pi(2^{m+1} \mid n = 2^m + 1 = s^{'}) \geq \frac{1023}{1024}.$$

Suppose the prior distributions of $s_1$ and $s_2$ are of the form (9). Then for $m \geq \max\{M_0, M_0^*\}$,

$$\begin{aligned}
\frac{1}{2} &\geq f(2^m, 2^m, 2^m) \geq \frac{510}{1024} \geq 2(\frac{511}{1024})^2, \\
\frac{1}{2} &\geq f(2^m + 1, 2^m, 2^m) \geq \frac{510}{1024}, \\
\frac{1}{2} &\geq f(2^m + 1, 2^m + 1, 2^m) \geq \frac{1023}{1024} \cdot \frac{511}{1024} \geq \frac{510}{1024}, \\
\frac{1}{2} &\geq f(2^m + 1, 2^m, 2^m + 1) \geq \frac{510}{1024}, \\
1 &\geq f(2^m + 1, 2^m + 1, 2^m + 1) \geq (\frac{1023}{1024})^2 \geq \frac{1022}{1024}.
\end{aligned}$$

Thus, for $i + j \leq 1$

$$\mid \rho_0(\pi^n \mid s_1^{'}, s_2^{'}) - \rho_0(\pi^{n+1} \mid s_1^{'} + i, s_2^{'} + j) \mid \leq \frac{2}{1024};$$

while

$$\rho_0(\pi^n \mid s_1^{'}, s_2^{'}) - \rho_0(\pi^{n+1} \mid s_1^{'}+1, s_2^{'}+1) \geq \frac{508}{1024}.$$

By (5), because the marginal probability of discovering a new species is

$$\begin{aligned} & P((n+1)\text{st observation is a new species } \mid n = 2^m = s') \\ = & \sum_{s=2^m} \frac{s-s'}{s+n} \cdot \pi(s \mid n = s' = 2^m) \\ \geq & \frac{2^{m+1}-2^m}{2^{m+1}+2^m} \cdot \pi(2^{m+1} \mid n = s' = 2^m) \\ \geq & \frac{1}{3} \cdot \pi(2^{m+1} \mid n = s' = 2^m) = \frac{511}{3072}. \end{aligned}$$

Therefore, similar to (16)

$$\begin{aligned} \Delta_n(\pi^n) &= \rho_0(\pi^n) - E\rho_0(\pi^n) \\ &\geq m_{1,n}m_{2,n}\left[\rho_0(\pi^n \mid s_1', s_2') - \rho_0(\pi^{n+1} \mid s_1'+1, s_2'+1)\right] \\ &\quad -\frac{2}{1024}[m_{1,n}(1-m_{2,n}) + (1-m_{1,n})m_{2,n} + (1-m_{1,n})(1-m_{2,n})] \\ &\geq \frac{511}{3072} \cdot \frac{511}{3072} \cdot \frac{508}{1024} - \frac{2}{1024} \\ &= \frac{113775100}{9663676416} > 0.0117 \end{aligned}$$

If the sampling cost is, for example, $c = 0.01$ and if $n = s_1^{'} = s_2^{'} = 2^m$ and $m$ is large enough, then $\Delta_n(\pi^n)$ will always be greater than the sampling cost. This implies that the number $N$ which guarantees $\Delta_n(\pi^n) \leq c$ does not exist in this distribution.

Following the above result, the optimal Bayes rule is truncated when the prior distributions of $s_1$ and $s_2$ are binomial. We give an example to demonstrate the optimal Bayes rule based on simulations.

**Example 4** Suppose that the prior distributions of $s_1$ and $s_2$ are zero-truncated B(9,0.5), and that the species proportions in both populations follow a geometric distribution with parameter 0.5, i.e.

$$p_{ij} = p_{i0} \times (0.5)^{j-1}, \qquad \text{for } 1 \leq j \leq s_i \text{ and } i = 1, 2,$$

where $p_{ij}$ is the species proportion of Species $j$ in Population $i$ and $\sum_{j=1}^{i_1} p_{1j} = \sum_{j=1}^{i_2} p_{2j} = 1$. Assume that the sampling cost is 0.001 per pair of observations and that $s_2 = 5$. Based on 5,000 replicate simulations for each case, we vary $s_1$ from 1

to 9 to estimate the (frequentist) risk. Also, in addition to the optimal Bayes rule, two suboptimal decision rules are considered as well: both rules stop sampling when

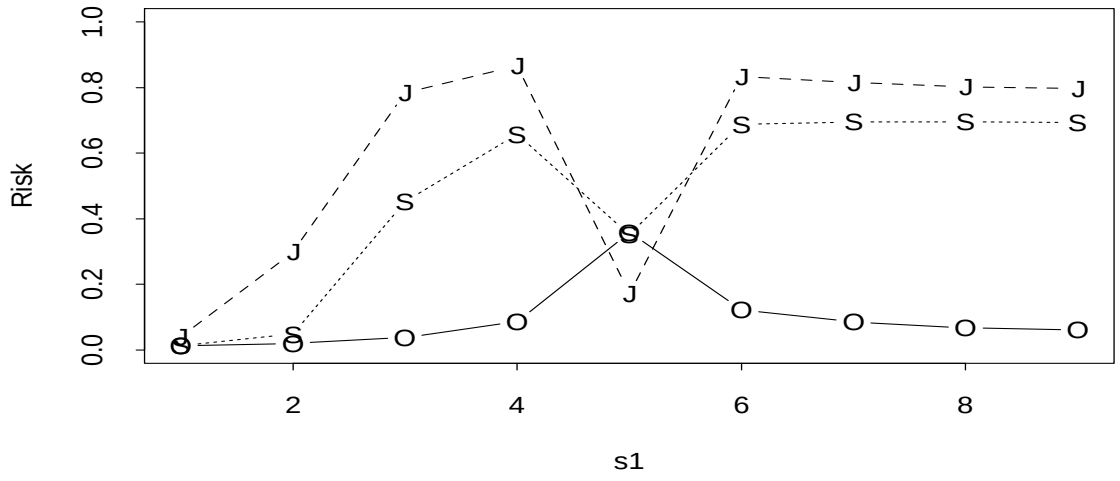
$$\sqrt{Var(s_1 \mid n)} + \sqrt{Var(s_2 \mid n)} \ \leq \ 2,$$

where $Var(s_1 \mid n)$ and $Var(s_2 \mid n)$ are the posterior variance of $s_1$ and $s_2$. Upon stopping, the action taken by these two decision rules is to accept $s_1 = s_2$ if

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

and reject otherwise. We consider 2 estimates of $\hat{s}_i(n)$, specifically, the jackknife estimators (Burnham and Overton, 1978) and the number of observed species. The risks for each decision rule are the sum of sampling cost and the loss of misclassification, and are demonstrated in the following graph:

**Figure 1.** *The risks of the optimal Bayes rule and two decision rules based on confidence interval approach when* $1 \leq s_1 \leq 9$ *and* $s_2 = 5$.




**O** – the optimal stopping rule, **J** – jackknife estimator, **S** – the number of observed species

Even though the Bayes rule is optimal in the Bayesian setting, it is obviously that the risks of the Bayes rule are smaller than those of the other two decision rules, except in the case of $s_1 = 5$. Further discussion of the Bayes rule and other frequentist decision rules can be seen in Yue (1994).

## 4. GROUP SAMPLING

In the previous discussion, we assumed that the observations are sampled in pairs, one from each population. In reality, this is not cost efficient and sometimes not possible. For example, when setting a trap to catch insects, it is difficult to catch exactly one insect. It is appropriate to consider a version of this problem that incorporates group sampling. Also, since the capacity of traps usually has a limit, it is reasonable to assume that the number of observations in a trap is at most $M_0$. Intuitively, given the same number of groups of samples, more observations in a group would give more information about populations. Therefore, the Bayes rule when using group sampling will still be truncated and the truncation bound will depend on $M_0$. Similarly, when the group size is different for the two populations, truncation will still occur if the condition of the previous section are met. For detailed discussion of the group sampling case, see Yue (1994).

## APPENDIX: PARTIAL PROOF OF THEOREM 6

We now show that if Case (ii) is satisfied then

$$\left|\rho_0(\pi^n \mid s_1^{'}, s_2^{'}) - \rho_0(\pi^{n+1} \mid s_1^{'}(n+1) = s_1^{'}, s_2^{'}(n+1) = s_2^{'})\right| \to 0$$

as $n \to 0$. Note that it is equivalent to showing that $\left|f(n, s_1^{'}, s_2^{'}) - f(n+1, s_1^{'}, s_2^{'})\right| \to 0$ as $n \to \infty$.

We begin by defining the following: $a_{s_1} \equiv \pi_1(s_1+1 \mid n, s_1^{'})/\pi_1(s_1 \mid n, s_1^{'}), a^*_{s_1} \equiv \pi_1(s_1+1 \mid n+1, s_1^{'})/\pi_1(s_1 \mid n+1, s_1^{'}) = (s_1+n)a_{s_1}/(s_1+n+1), b_{s_2} \equiv \pi_2(s_2+1 \mid n, s_2^{'})/\pi_2(s_2 \mid n, s_2^{'})$, and $b^*_{s_2} \equiv \pi_2(s_2+1 \mid n+1, s_2^{'})/\pi_2(s_2 \mid n+1, s_2^{'}) = (s_2+n)b_{s_2}/(s_2+n+1)$.

Then by these definitions,

$$\prod_{i=s_1^{'}}^{s_1} a_i^* = \prod_{i=s_1^{'}}^{s_1} \frac{i+n}{i+n+1} a_i = \frac{s_1^{'}+n}{s_1+n+1} \prod_{i=s_1^{'}}^{s_1} a_i \tag{20}$$

and

$$\prod_{i=s_2^{'}}^{s_2} b_i^* = \prod_{i=s_2^{'}}^{s_2} \frac{i+n}{i+n+1} b_i = \frac{s_2^{'}+n}{s_2+n+1} \prod_{i=s_2^{'}}^{s_2} b_i. \tag{21}$$

Therefore by defining

$$c_{1,n} \;=\; 1 + \sum_{s=s_1^{'}}^{\infty} \prod_{i=s_1^{'}}^{s} a_i = \frac{1}{\pi_1(s_1^{'} \mid n, s_1^{'})}$$

$$
\begin{aligned}
c_{2,n} &= 1+\sum_{s=s_2'}^{\infty}\prod_{i=s_2'}^{s} b_i = \frac{1}{\pi_2(s_2' \mid n, s_2')} \\
c_{1,n}^* &= 1+\sum_{s=s_1'}^{\infty}\prod_{i=s_1'}^{s} a_i^* = \frac{1}{\pi_1(s_1' \mid n+1, s_1')} \\
c_{2,n}^* &= 1+\sum_{s=s_2'}^{\infty}\prod_{i=s_2'}^{s} b_i^* = \frac{1}{\pi_2(s_2' \mid n+1, s_2')}
\end{aligned}
$$

and $s' = \max\{s_1', s_2'\}$, it is immediate that

$$
\begin{aligned}
f(n, s_1', s_2') &= \sum_{s=s'}^{\infty} \pi_1(s \mid n, s_1')\pi_2(s \mid n, s_2') \\
&= \pi_1(s_1' \mid n, s_1')\pi_2(s_2' \mid n, s_2') \sum_{s=s'}^{\infty} \frac{\pi_1(s \mid n, s_1')}{\pi_1(s_1' \mid n, s_1')} \cdot \frac{\pi_2(s \mid n, s_2')}{\pi_2(s_2' \mid n, s_2')} \\
&= \frac{1}{c_{1,n}c_{2,n}} \sum_{s=s'}^{\infty} (\prod_{i=s_1'}^{s} a_i)(\prod_{i=s_2'}^{s} b_i)
\end{aligned}
$$

and

$$
\begin{aligned}
& f(n+1, s_1', s_2') = \sum_{s=s'}^{\infty} \pi_1(s \mid n+1, s_1')\pi_2(s \mid n+1, s_2') \\
= \;& \pi_1(s_1' \mid n+1, s_1')\pi_2(s_2' \mid n+1, s_2') \sum_{s=s'}^{\infty} \frac{\pi_1(s \mid n+1, s_1')}{\pi_1(s_1' \mid n+1, s_1')} \cdot \frac{\pi_2(s \mid n+1, s_2')}{\pi_2(s_2' \mid n+1, s_2')} \\
= \;& \frac{1}{c_{1,n}^* c_{2,n}^*} \sum_{s=s'}^{\infty} (\prod_{i=s_1'}^{s} a_i^*)(\prod_{i=s_2'}^{s} b_i^*).
\end{aligned}
$$

Note that by (20), similar to showing $1 - m_{1,n} \to 0$ as $n \to \infty$, we have

$$
\begin{aligned}
\frac{c_{1,n}}{c_{1,n}^*} - 1 &= \frac{c_{1,n} - c_{1,n}^*}{c_{1,n}^*} \\
&= \frac{1}{c_{1,n}^*}[1 + \sum_{s_1=s_1'}^{\infty}\prod_{i=s_1'}^{s_1} a_i - (1 + \sum_{s_1=s_1'}^{\infty}\prod_{i=s_1'}^{s_1} a_i^*)] \\
&= \frac{1}{c_{1,n}^*} \sum_{s_1=s_1'}^{\infty} (1 - \frac{s_1' + n}{s_1 + n + 1}) \prod_{i=s_1'}^{s_1} a_i \\
&= \frac{1}{c_{1,n}^*} \sum_{s_1=s_1'}^{\infty} \frac{s_1 - s_1'}{s_1 + n + 1} \prod_{i=s_1'}^{s_1} a_i \\
&\le \sum_{s_1=s_1'}^{\infty} \frac{s_1 - s_1'}{s_1 + n + 1} \prod_{i=s_1'}^{s_1} a_i \to 0
\end{aligned}
$$

as $n \to \infty$. Similarly, we have $c_{2,n}/c^*_{2,n} \to 1$ as $n \to \infty$. Thus, as $n \to \infty$

$$\frac{c_{1,n}c_{2,n}}{c^*_{1,n}c^*_{2,n}} \to 1 \quad \text{or} \quad \frac{c^*_{1,n}c^*_{2,n}}{c_{1,n}c_{2,n}} \to 1. \tag{22}$$

Let $s_0' = \min\{s_1', s_2'\}$. Since for $s \geq \max\{s_1', s_2'\} \geq s_0'$

$$\begin{aligned}
&(s+n+1)^2 - (s_1'+n)(s_2'+n) \\
&= n(2s+2-s_1'-s_2') + (s+1)^2 - s_1's_2' \\
&\leq n(2s+2-2s_0') + (s+1)^2 - (s_0')^2 \\
&= 2n(s+1-s_0') + (s+1-s_0')(s+1+s_0') \\
&\leq 2n(s+1-s_0') + (s+1-s_0')2(s+1) \\
&= 2(s+1-s_0')(n+s+1),
\end{aligned}$$

we have that

$$1 - \frac{(s_1'+n)(s_2'+n)}{(s+n+1)^2} \leq \frac{2(s+1-s_0')}{s+n+1} \leq \frac{2(s+1-s_0')}{n}. \tag{23}$$

Also, since $s_2'$ is the posterior mode of $\pi_2(s_2 \mid n, s_2')$, it follows that $\prod_{i=s_2'}^{s_2} b_i \leq 1$ for all $s_2 \geq s_2'$. Thus, together with (20), (21), (22), and (23)

$$\begin{aligned}
&|f(n, s_1', s_2') - f(n+1, s_1', s_2')| \\
&= \left| \frac{1}{c_{1,n}c_{2,n}} \sum_{s=s'}^{\infty} (\prod_{i=s_1'}^{s} a_i)(\prod_{i=s_2'}^{s} b_i) - \frac{1}{c^*_{1,n}c^*_{2,n}} \sum_{s=s'}^{\infty} (\prod_{i=s_1'}^{s} a_i^*)(\prod_{i=s_2'}^{s} b_i) \right| \\
&\leq \left| \frac{1}{c_{1,n}c_{2,n}} \sum_{s=s'}^{\infty} (\prod_{i=s_1'}^{s} a_i)(\prod_{i=s_2'}^{s} b_i) - \frac{1}{c_{1,n}c_{2,n}} \sum_{s=s'}^{\infty} (\prod_{i=s_1'}^{s} a_i^*)(\prod_{i=s_2'}^{s} b_i^*) \right| \\
&\quad + \left| \frac{1}{c_{1,n}c_{2,n}} \sum_{s=s'}^{\infty} (\prod_{i=s_1'}^{s} a_i^*)(\prod_{i=s_2'}^{s} b_i^*) - \frac{1}{c^*_{1,n}c^*_{2,n}} \sum_{s=s'}^{\infty} (\prod_{i=s_1'}^{s} a_i^*)(\prod_{i=s_2'}^{s} b_i^*) \right| \\
&\leq \sum_{s=s'}^{\infty} (1 - \frac{(s_1'+n)(s_2'+n)}{(s+n+1)^2})(\prod_{i=s_1'}^{s} a_i)(\prod_{i=s_2'}^{s} b_i) \\
&\quad + \left| \frac{c^*_{1,n}c^*_{2,n}}{c_{1,n}c_{2,n}} - 1 \right| \frac{1}{c^*_{1,n}c^*_{2,n}} \sum_{s=s'}^{\infty} (\prod_{i=s_1'}^{s} a_i^*)(\prod_{i=s_2'}^{s} b_i^*) \\
&\leq \sum_{s=s'}^{\infty} \frac{2(s+1-s_0')}{s+n+1} (\prod_{i=s_1'}^{s} a_i) + \left| \frac{c^*_{1,n}c^*_{2,n}}{c_{1,n}c_{2,n}} - 1 \right| f(n+1, s_1', s_2') \\
&\to 0 \quad \text{as } n \to \infty,
\end{aligned}$$

since $f \leq 1$.

## BIBLIOGRAPHY


Andrews, J. H., Kinkel, L. L., Berbee, F. M., and Nordheim, E. V. (1987) Fungi, Leaves, and the Theory of Island Biogeography, *Microbial Ecology*, 14, 277-290.

Berger, J. O. (1985) *Statistical Decision Theory and Bayesian Analysis*, Springer–Verlag, New York.

Bunge, J. and Fitzpatrick, M. (1993) Estimating the Number of Species: A Review, *Journal of the American Statistical Association*, 88, 364-373.

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.

Chao, A. (1981) On Estimating The Probability of Discovering a New Species, *Annals of Statistics*, 6, 1339-1342.

Chao, A. (1984) Non-parametric Estimation of the Number of Classes in a Population, *Scandinavian Journal of Statistics*, 11, 265-270.

Chao, A. and Lee, S. (1992) Estimating the Number of Classes via Sample Coverage, *Journal of the American Statistical Association*, 87, 210-217.

Clayton, M. K. (1985) A Bayesian Nonparametric Sequential Test for the Mean of a Population, *Annals of Statistics*, 13, 1129-1139.

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.

Colwell, R. K. and Coddington, J. A. (1994) Estimating Terrestrial Biodiversity, *Philosophical Transactions of the Royal Society of London*, 345, 101-118.

Engen, S. (1975) *Stochastic Abundance Models*, Chapman and Hall, London.

Fisher, R. A., Corbet, A. S., and Williams, C. B. (1943) The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population, *Journal of Animal Ecology*, 12, 42-58.

Harris, B. (1968) Statistical Inferences in the Classical Occupancy Problem Unbiased Estimation of the Number of Classes, *Journal of the American Statistical Association*, 63, 837-847.

Heltshe, J. F. and Forrester, Nancy E. (1983) Estimating Species Richness using the Jackknife Procedure, *Biometrics*, 39, 1-11.

Hill, B. M. (1979) Posterior Moments of the Number of Species in a Finite Population and the Posterior Probability of Finding a New Species, *Journal of the American Statistical Association*, 74, 668-673.

Kinkel, L. L., Nordheim, E. V., and Andrews, J. H. (1992) Microbial Community Analysis in Incompletely or Destructively Sampled System, *Microbial Ecology*, 24, 227-242.

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.

Lewontin, R. C. and Prout, T. (1956) Estimation of the Number of Different Classes in a Population, *Biometrics*, 12, 211-223.

Pielou, E. C. (1975) *Ecological Diversity*, John Wiley and Sons, New York.

Pielou, E. C. (1979) *Biogeography*, John Wiley and Sons, 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.

Ray, S. N. (1965) Bounds on the Maximum Sample Size of a Bayes Sequential Procedure, *Annals of Mathematical Statistics*, 36, 859-878.

Sichel, H.S. (1982) Asymptotic Efficiencies of Three Methods of Estimations for the Inverse Gaussian-Poisson Distribution, *Biometrika*, 69, 467-472.

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.