Ann. Inst. Statist. Math.
Vol. 42, No. 4, 623–636 (1990)

ADMISSIBILITY OF ESTIMATORS OF THE PROBABILITY OF UNOBSERVED OUTCOMES*

ARTHUR COHEN AND HAROLD B. SACKROWITZ

Department of Statistics, Hill Center, Busch Campus, Rutgers University, New Brunswick, NJ 08903, U.S.A.

(Received April 12, 1988; revised March 20, 1989)

Abstract. The problem of estimating the probability of unobserved outcomes or, as it is sometimes called, the conditional probability of a new species, is studied. Good's estimator, which is essentially the same as Robbins' estimator, namely the number of singleton species observed divided by the sample size, is studied from a decision theory point of view. The results obtained are as follows: (1) When the total number of different species is assumed bounded by some known number, Good's and Robbins' estimators are inadmissible for squared error loss. (2) If the number of different species can be infinite, Good's and Robbins' estimators are admissible for squared error loss. (3) Whereas Robbins' estimator is a UMVUE for the *unconditional* probability of a new species obtained in one extra sample point, Robbins' estimator is not a uniformly minimum mean squared error unbiased estimator of the conditional probability of a new species. This answers a question raised by Robbins. (4) It is shown that for Robbins' model and squared error loss, there are admissible Bayes estimators which do not depend only on a minimal sufficient statistic. A discussion of interpretations and significance of the results is offered.

Key words and phrases: Probability of new species, Good's estimator, Robbins' estimator, admissibility, uniformly minimum variance unbiased, uniformly minimum mean squared error unbiased, sufficiency, completeness.

1. Introduction

The problem of estimating the probability of the unobserved outcomes of an experiment has been considered in a variety of forms by many researchers. Most recently Clayton and Frees (1987) study what they call the problem of estimating the unconditional probability of a new species

*Research supported by NSF Grant DMS-88-22622.

on the $(n+1)$-st trial. Clayton and Frees (1987) give a summary of the work done on this subject. One other recent reference is Esty (1986). The main references that relate to the results of this paper are Good (1953) and Robbins (1968). In order to summarize our results we must distinguish between Good's model and Robbins' model. Within Robbins' model we must also distinguish between the unconditional probability of discovering a new species on the $(n+1)$-st trial and the conditional probability of discovering a new species on the $(n+1)$-st trial.

We describe the problem formally as done in Clayton and Frees (1987). Consider a population composed of distinct species and use M_i to represent the i-th species $(i = 1, 2, \ldots)$. Let k be the total number of distinct species in the population. We will allow $k = \infty$. We assume that the species have no natural order and that the number of species may be countably infinite. Suppose that n independent drawings are made from the population, with replacement if the population is finite, and $Y_j = i$ when the j-th draw is from M_i. Let $X_i^n = \sum_{j=1}^{n} I(Y_j = i)$ be the number of representatives of the species M_i in n drawings, where $I(A)$ is the indicator function of the set A. Let

$$U_n = \sum_i p_i I(X_i^n = 0) = 1 - \sum_i p_i I(X_i^n \neq 0)\,, \tag{1.1}$$

where $p_i = \Pr\{Y_j = i\}$ for all j. The quantity U_n represents the probability of the unobserved outcomes of an experiment. We note that U_n is a random variable that depends on $\boldsymbol{X}^n = (X_1^n, X_2^n, \ldots)$ and $\boldsymbol{p} = (p_1, p_2, \ldots)$. Good's problem is to estimate U_n. Good's estimator is

$$G(\boldsymbol{X}^n) = n^{-1} \sum_i I(X_i^n = 1)\,. \tag{1.2}$$

The quantity U_n also represents the conditional probability of discovering a new species in one additional search given $\boldsymbol{X}^n$. Robbins also wishes to estimate U_n, but in Robbins' model an additional drawing is made and Robbins' estimator is

$$V(\boldsymbol{X}^{n+1}) = (n+1)^{-1} \sum_i I(X_i^{n+1} = 1)\,. \tag{1.3}$$

Robbins noted that V is an unbiased estimator for U_n in the sense that $EV = EU_n$ and raised the question as to whether V is a uniformly minimum mean squared error unbiased estimator (UMMSEUE) of U_n.

Starr (1979), in an effort to extend Robbins' work, changed the problem to that of estimating $\theta_n = EU_n$. Starr conjectured that V is a MVUE for θ_n and Clayton and Frees (1987) proved that conjecture.

Consider a squared error loss function. The problem is to estimate U_n.

Our results are as follows:

(1) If k is bounded, say $k \leq K$, K known, Good's estimator is inadmissible. Also Robbins' estimator is inadmissible. (The proof provides a nice illustration of the use of the class of "Bayes wide sense" procedures.)

(2) If $k = \infty$, Good's and Robbins' estimators are admissible.

(3) The estimator V is *not* a UMMSEUE of U_n. In fact no UMMSEUE exists.

(4) The estimator V is a UMVUE for θ_n immediately, by virtue of sufficiency and completeness.

(5) Robbins' model serves as an example of a situation where admissible Bayes estimators do *not* depend only on the minimal sufficient statistics.

As previously mentioned there are a large number of detached references on this problem. The purposes of this paper are to clarify the issues, identify what we feel are the important issues, and make a theoretical contribution to some of these issues.

Before proceeding we need to recognize a distinction between an indexed multinomial model and an unindexed multinomial model. An indexed multinomial model is the usual one where the classes are known and can be identified before any observations are made. The unindexed model is where we cannot identify the classes until observations are made. Good's model is of this latter type. In the unindexed model it makes sense to base all estimators on the frequencies of frequencies, i.e., the number of different singleton species observed, the number of doubleton species observed, the number of times we observed exactly 3 of a species, and so on. Clearly, Good's estimator is of this type. The rational and formal way to describe this requirement is through permutation invariance. In fact, permutation invariant estimators are reasonable, desirable and even compelling for this model. We note that the frequencies of the frequencies are equivalent to the ordered frequencies defined through $\boldsymbol{X}^n$ by $X^n_{(1)} \geq X^n_{(2)} \geq \cdots$ where $X^n_{(1)}$ is the largest observed frequency, $X^n_{(2)}$ is the second largest observed frequency, and so on. Of course, ties among the ordered values are possible and permissible. The ordered frequencies represent a maximal invariant statistic under the permutation group and all invariant estimators are a function of the maximal invariant statistic. See, for example, Lehmann (1986). Formally, a statistic $T(\boldsymbol{X}^n)$ is invariant under the transformation g if $T(\boldsymbol{X}^n) = T(g\boldsymbol{X}^n)$.

In our development below we assume the indexed multinomial model. However, we will indicate that nearly all our results are appropriate for the unindexed model as well. That is, we will indicate when and why a stated result will be true for the case where estimators are required to be permutation invariant. For example, result (1) that Good's estimator is inadmissible if k is bounded by a known integer, will be true even if the class of estimators is limited to permutation invariant estimators (i.e., those based only on frequencies of frequencies).

Section 2 contains the results pertaining to admissibility and to the remark above concerning Bayes estimators that do not depend only on sufficient statistics. Section 3 contains the other results. Section 4 contains a discussion on the significance of the results and offers interpretations as to which problem is meaningful.

2. Admissibility

First suppose $k \leq K$ (known) and consider Good's model and Good's estimator. Let $\xi(\boldsymbol{p})$ represent a prior distribution on $\boldsymbol{p}$. For ease of presentation only we let $n \geq 5$. We prove

THEOREM 2.1. *If k is bounded by known K, Good's estimator* (1.2) *is inadmissible. Furthermore if k is bounded by known K, Good's estimator is inadmissible within the class of permutation invariant estimators.*

PROOF. The proof is given in the Appendix.

Next we study Robbins' model. We first observe that the joint distribution of the observations is

$$\prod_{i=1}^{\infty} p_i^{x_i^n} \prod_{i=1}^{\infty} p_i^{I\{y_{n+1}=i\}} . \tag{2.1}$$

Whereas $(\boldsymbol{X}^n, Y^{n+1})$ is seen to be a sufficient statistic, $(\boldsymbol{X}^{n+1})$ is a minimal sufficient statistic. Write the distribution of $(\boldsymbol{X}^n, Y_{n+1})$ as

$$f(\boldsymbol{x}^n, y_{n+1}; \boldsymbol{p}) = \frac{n!}{\prod_{i=1}^{\infty} x_i^n!} \Pi p_i^{x_i^n} \cdot \Pi p_i^{I\{y_{n+1}=i\}} . \tag{2.2}$$

Notice that the expected risk of an estimator of $U_n(\boldsymbol{X}^n, \boldsymbol{p})$ when $\xi(\boldsymbol{p})$ is a prior distribution is

$$\int \sum_{\boldsymbol{x}^n, y_{n+1}} (t(\boldsymbol{x}^n, y_{n+1}) - U_n(\boldsymbol{x}^n, \boldsymbol{p}))^2 f(\boldsymbol{x}^n, y_{n+1}; \boldsymbol{p}) d\xi(\boldsymbol{p}) \tag{2.3}$$

and so to minimize (2.8) for each $(\boldsymbol{x}^n, y_{n+1})$ we minimize

$$\int (t(\boldsymbol{x}^n, y_{n+1}) - U_n(\boldsymbol{x}^n, \boldsymbol{p}))^2 f(\boldsymbol{x}^n, y_{n+1}; \boldsymbol{p}) d\xi(\boldsymbol{p}) . \tag{2.4}$$

If $\xi(\boldsymbol{p})$ is a prior such that $\int f(\boldsymbol{x}^n, y_{n+1}; \boldsymbol{p}) d\xi(\boldsymbol{p}) > 0$ for a given $(\boldsymbol{x}^n, y_{n+1})$ the

Bayes estimator is

$$\int U_n(\boldsymbol{x}^n,\boldsymbol{p})f(\boldsymbol{x}^n,y_{n+1};\boldsymbol{p})\,d\xi(\boldsymbol{p})\Big/\int f(\boldsymbol{x}^n,y_{n+1};\boldsymbol{p})\,d\xi(\boldsymbol{p})\,. \tag{2.5}$$

Observe that despite the fact that $f(\boldsymbol{x}^n, y_{n+1})$ can be expressed entirely as a function of $\boldsymbol{x}^{n+1}$, the Bayes estimator in (2.5) depends also on $\boldsymbol{x}^n$ through $U(\boldsymbol{x}^n,\boldsymbol{p})$. Since a realization of the vector $\boldsymbol{X}^{n+1}$ can arise from different $\boldsymbol{X}^n$ vectors, and these different $\boldsymbol{X}^n$ vectors lead to different U_n expressions, we have a situation where the Bayes estimator does not depend only on a minimal sufficient statistic.

Using the above definition of a Bayes estimator we can prove the following.

THEOREM 2.2. *If k is bounded by known K, Robbins' estimator* (1.3) *is inadmissible.*

PROOF. The method of proof is essentially the same as that used in proving Theorem 2.1. We omit the details.

THEOREM 2.3. *For $k=\infty$ Good's estimator and Robbins' estimator are admissible.*

PROOF. We prove the result for Good's estimator and remark that the proof for Robbins' estimator is similar. Suppose G is not admissible. Then there exists an estimator $\delta(\boldsymbol{X}^n)$ which is better. That is,

$$\begin{aligned}\sum_{\boldsymbol{x}^n}(G(\boldsymbol{x}^n) - U_n(\boldsymbol{x}^n,\boldsymbol{p}))^2 f(\boldsymbol{x}^n;\boldsymbol{p})\\ \geq \sum_{\boldsymbol{x}^n}(\delta(\boldsymbol{x}^n) - U_n(\boldsymbol{x}^n,\boldsymbol{p}))^2 f(\boldsymbol{x}^n;\boldsymbol{p})\end{aligned} \tag{2.6}$$

where $f(\boldsymbol{x}^n;\boldsymbol{p}) = \left(n!\Big/\prod_i x_i^n!\right)\prod_i^{\infty} p_i^{x_i^n}$. Since (2.6) must be true for all $\boldsymbol{p}$ our approach is to iteratively examine (2.6) for particular choices of $\boldsymbol{p}$. We will show that the validity of (2.6) for each particular $\boldsymbol{p}$ implies the equality of G and δ for certain sample points. Also as we consider all our $\boldsymbol{p}$ choices we will cover all sample points.

(i) (2.6) must be true for $p_1 = 1$, $p_i = 0$, $i \neq 1$. Here (2.6) reduces to $G^2(n,0,\dots) \geq \delta^2(n,0,\dots)$. To see this note that in this case $f(\boldsymbol{x}^n;\boldsymbol{p}) = 0$ for all $\boldsymbol{x}^n$ except $\boldsymbol{x}^n = (n,0,\dots)$. Since $G(n,0,\dots) = 0$, $\delta(n,0,\dots)$ must equal 0. Similarly, this can be done for all points which are permutations of $(n,0,\dots)$. Thus we may regard the summations in (2.6) to be over all $\boldsymbol{x}^n$ except $(n,0,\dots)$, $(0,n,0,\dots)$, $\dots$.

(ii) (2.6) must be true for $p_1 = 1 - 1/n$, $p_2 = \cdots = p_{r+1} = 1/nr$ for all $r = 1, 2, \ldots$ and also for the limit as $r \to \infty$. For $r \to \infty$, examination of the dominant terms (note that by (i) the point $(n, 0, \ldots)$ has been eliminated from the sums) and noting that $U_n(\boldsymbol{x}^n, \boldsymbol{p})$ converges to $1/n$ for these terms, reduces (2.6) to

$$\sum_{[\boldsymbol{x}^n: x_1^n = n-1]} \left(G(\boldsymbol{x}^n) - \frac{1}{n} \right)^2 f(\boldsymbol{x}^n; \boldsymbol{p}) \geq \sum_{[\boldsymbol{x}^n: x_1^n = n-1]} \left(\delta(\boldsymbol{x}^n) - \frac{1}{n} \right)^2 f(\boldsymbol{x}^n; \boldsymbol{p}) .$$

Since $G(\boldsymbol{x}^n) = 1/n$ for all $\boldsymbol{x}^n$ such that $x_1^n = n - 1$ so must $\delta(\boldsymbol{x}^n) = 1/n$ for such $\boldsymbol{x}^n$. Again this may be done for all permutations of $(n-1, 1, 0, \ldots)$ and these points may be removed from the summations in (2.6).

(iii) (2.6) must be true for $p_1 + p_2 = 1$. Now (2.6) reduces to

$$\sum_{y=2}^{n-2} (G(n-y, y, 0, \ldots) - 0)^2 f((n-y, y, 0, \ldots); \boldsymbol{p})$$
$$\geq \sum_{y=2}^{n-2} (\delta(n-y, y, 0, \ldots) - 0)^2 f((n-y, y, 0, \ldots); \boldsymbol{p}) .$$

Since $G(n-y, y, 0, \ldots) = 0$ for $2 \leq y \leq n-2$ so must $\delta(n-y, y, 0, \ldots) = 0$. Again this may be done for all permutations of $(n-y, y, 0, \ldots)$ and these points may be removed from the summations in (2.6).

(iv) Continue as indicated above. That is, the choices of $\boldsymbol{p}$ actually occur in sets. (i) above by itself would be set 1, (ii) and (iii) together comprise set 2. In general set k, $k = 1, \ldots, n$ consists of k parameter points used in the following order:

(1) $\quad p_1 = 1 - \dfrac{k-1}{n}, \quad p_2 = \cdots = p_{r+1} = \dfrac{k-1}{nr} \quad$ with $r \to \infty$,

(2) $\quad p_1 + p_2 = 1 - \dfrac{k-2}{n}, \quad p_3 = \cdots = p_{r+2} = \dfrac{k-2}{nr} \quad$ with $r \to \infty$,

$(k-1)$ $\quad p_1 + \cdots + p_{k-1} = 1 - \dfrac{1}{n}, \quad p_k = \cdots = p_{r+k-1} = \dfrac{1}{nr} \quad$ with $r \to \infty$,

(k) $\quad p_1 + \cdots + p_k = 1$.

It should be noted that (depending on n) there may be more parameter points than necessary and that some of the above mentioned parameters might not serve to reduce the summation. However if all the parameter points are used, in the order given, we have that $\delta(\boldsymbol{x}^n) = G(\boldsymbol{x}^n)$ for all $\boldsymbol{x}^n$.

Remark 2.1. The fact that Good's estimator is admissible when

$k = \infty$ surely implies its admissibility among permutation invariant estimators.

Remark 2.2. Let $Z(\boldsymbol{X}^n)$ be the number of coordinates of $\boldsymbol{X}^n$ that are one. Let $\delta^*(\boldsymbol{X}^n)$ denote any estimator which depends only on $Z(\boldsymbol{X}^n)$. In other words δ^* takes on the same value for every $\boldsymbol{X}^n$ with the same number of ones. Any such δ^*, provided $0 \le \delta^* \le 1$, and $\delta^* = 0$ whenever $Z = 0$ is admissible. The proof follows, step by step, the proof of Theorem 2.3 as long as the parameter points are chosen judiciously. For example, the parameter point corresponding to the sample point $(n-1, 1, 0, \ldots, 0)$ is $p_1 = 1 - \Delta, p_2 = \cdots = p_{r+1} = \Delta/nr$ if $\delta^* = \Delta$ for that sample point.

3. *V* is not a UMMSEUE

Consider Robbins' model for $k = \infty$. Let $S^{(n)} = \sum_{i=1}^{\infty} I\{X_i^n \neq 0\}$, which is the number of different species observed at time n. Let $\rho(\boldsymbol{X}^{n+1}) = \Pr(S^{(n)} = S^{(n+1)} | \boldsymbol{X}^{n+1})$. Note that this probability does not depend on $\boldsymbol{p}$. Let $I = 1$ if $S^{(n)} = S^{(n+1)}$ and $I = 0$ otherwise.

THEOREM 3.1. *Under Robbins' model with $k = \infty$, V is not a UMMSEUE.*

PROOF. For V defined in (1.3) and any integer $t \ge 2$ define

$$V_t^*(\boldsymbol{X}^n, \boldsymbol{X}^{n+1}) = V(\boldsymbol{X}^{n+1}) - (1/t)I + (1/t)\rho(\boldsymbol{X}^{n+1}) . \tag{3.1}$$

Note that

$$EV_t^*(\boldsymbol{X}^n, \boldsymbol{X}^{n+1}) = EE(V_t^*(\boldsymbol{X}^n, \boldsymbol{X}^{n+1}) | \boldsymbol{X}^{n+1}) = EV(\boldsymbol{X}^{n+1}) ,$$

which means that both V and V^* have expected values equal to EU_n so they are unbiased in the sense described in the introduction.

Now consider the parameter point $p_1 = p_2 = \cdots = p_t = 1/t$, $p_i = 0$ for $i \ge t+1$. We show that the MSE for V^* is less than the MSE of V at this point by showing that the conditional MSE of V at this point given $\boldsymbol{X}^{n+1}$ is less than the conditional MSE of V for every $\boldsymbol{X}^{n+1}$. That is, consider

$$E\{(V^* - U_n)^2 | \boldsymbol{X}^{n+1}\} = E\{(V + (\rho/t) - (1/t)I - U_n)^2 | \boldsymbol{X}^{n+1}\} . \tag{3.2}$$

The right side of (3.2) is

$$(3.3) \qquad \begin{aligned} &(V + (\rho - 1)/t - (t - S^{(n+1)})/t)^2 \rho \\ &\quad + (V + \rho/t - (t - S^{(n+1)} + 1)/t)^2 (1 - \rho) \\ &\qquad = [V + (\rho - 1 - t + S^{(n+1)})/t]^2 \\ &\qquad = (V - (1 - \rho)/t - (t - S^{(n+1)})/t)^2 . \end{aligned}$$

On the other hand,

$$(3.4) \qquad \begin{aligned} &E\{(V - U_n)^2 | X^{n+1}\} \\ &\quad = (V - (t - S^{(n+1)})/t)^2 \rho + (V - (t - S^{(n+1)} + 1)/t)^2 (1 - \rho) \\ &\quad = (V - (t - S^{(n+1)}/t))^2 \rho + (V - 1/t - (t - S^{(n+1)})/t)^2 (1 - \rho) . \end{aligned}$$

Use Jensen's inequality on the right-hand side of (3.4) to find that it is greater than

$$(3.5) \qquad (V - (1 - \rho)/t - (t - S^{n+1})/t)^2 ,$$

which is (3.3).

Note that Theorem 3.1 is true for any unbiased estimator that depends only on X^{n+1} and furthermore the proof did not require $k = \infty$.

COROLLARY 3.1. *Under Robbins' model, no UMMSEUE of U_n exists.*

PROOF. The proof follows from Theorems 2.3 and 3.1.

We remark that Theorem 3.1 and Corollary 3.1 would hold even if the class of estimators was restricted to permutation invariant estimators. This follows since V and V^* are both permutation invariant.

Clayton and Frees (1987) prove that V is a MVUE for θ_n. We remark that this follows from the facts that V is based on a minimal sufficient statistic X^{n+1} and X^{n+1} has a distribution which is complete. This latter fact follows from the definition of completeness and the fact that any multinomial distribution with b cells, $b = 1, 2, \ldots, n + 1$, is complete.

4. Interpretations and discussion

The comments in this section are to some extent subjective and reflect the opinions of the authors. The first set of comments refer to which problem is most meaningful and the second set of comments pertain to the significance of the results of this paper.

We regard Good's model and the problem of estimating the conditional probability of a new species as relevant, intuitive, and meaningful. Robbins also wishes to estimate U_n and we feel this is the relevant quantity of interest. However, Robbins' model is somewhat questionable since one is likely to still be interested in the conditional probability of observing a new species after $(n + 1)$ drawings. That is, the relevant quantity is U_{n+1}. Robbins' approach is interesting as a device to come up with an "unbiased" estimator, when none exists based on the sample of size n. This idea has potential use in other problems where one seeks unbiased or conditionally unbiased estimators.

We do not feel that $\theta_n = EU_n$, the quantity studied in Starr (1979) and Clayton and Frees (1987) is really the relevant quantity to study. It is not intuitive to us as to why one would be interested in estimating the unconditional probability of observing a new species on the $(n + 1)$-st trial in this particular setting. There is one justification for studying estimators of θ_n in that θ_n may be regarded as an approximation to U_n and so estimators that do well for θ_n may do well for U_n. However, for example, we feel the Monte Carlo study of Clayton and Frees should consider the MSE of the estimators with U_n in place of θ_n. Such would provide a meaningful comparison of their estimator and V.

The inadmissibility or admissibility of the Good-Robbins estimator depends on whether k is bounded by a known number or not. It seems that more often than not, in a practical problem one can find an upper bound for k. Often n will be substantially less than k and therefore less than any upper bound for k. The admissibility result when $k = \infty$ has some theoretical appeal. However, from Remark 2.2 we note that many estimators, even some strange ones are admissible by the same argument.

The proofs of inadmissibility and admissibility when k is bounded and $k = \infty$ respectively are based on Bayesian ideas and the role of estimating 0. When k is bounded by known K, an estimator which estimates 0 as often as the Good-Robbins estimator cannot be Bayes in the wide sense. Yet, when $k = \infty$, sequences of priors can be found that are more tolerant of estimating zero so often.

The inadmissibility results above and the fact that V is not an MMSEUE for U_n suggest that competitors to Good-Robbins estimator still need to be found and studied. This is easier said than done and at this point we cannot recommend any such competitors. We feel however that our results help clarify some of the issues on this difficult problem.

Appendix

PROOF OF THEOREM 2.1. For the case where $n \geq K$, Good's estimator is trivially inadmissible. The estimator which estimates by 0 when all K cells are occupied and estimates by $G(X^n)$ otherwise is better.

The case $n < K$, surely one which often occurs, requires proof. The proof will be done in three steps which we will designate as Lemmas A.1, A.2 and A.3. In Lemma A.1 we will prove that Good's estimator is inadmissible for a conditional problem. In Lemma A.2 we prove that inadmissibility of an estimator for the conditional problem of Lemma A.1 implies inadmissibility of the estimator for the unconditional problem. In Lemma A.3 we prove that inadmissibility of the estimator for the unconditional problem implies inadmissibility among the class of permutation invariant estimators.

We proceed. Let K be the least upper bound for k. Assume K is known in all arguments below we prove inadmissibility for $k = K$, which suffices. Define the orbits

$$F_1 = \{\boldsymbol{x}^n : \boldsymbol{x}^n = (n-1, 1, 0, \ldots, 0) \text{ and all its permutations}\},$$

$$F_2 = \{\boldsymbol{x}^n : \boldsymbol{x}^n = (n-2, 2, 0, \ldots, 0) \text{ and all its permutations}\}.$$

Now consider the conditional problem of estimating U_n given that the sample space is $F_1 \cup F_2$ and the parameter space is

$$\Omega = \left\{ \boldsymbol{p} = (p_1, p_2, \ldots, p_K) : p_i \geq 0, \sum_{i=1}^{k} p_i = 1, p_i \neq 1 \right\}. \tag{A.1}$$

LEMMA A.1. *For k bounded by known K, Good's estimator is inadmissible for the conditional problem.*

PROOF. For the conditional problem

$$P\{X_1^n = 0, \ldots, X_i^n = n-1, \ldots, X_j^n = 1, \ldots, X_K^n = 0\} = p_i^{n-1} p_j / D, \tag{A.2}$$

where $D = \sum_{1=i\neq j=1}^{K} p_i^{n-1} p_j + [(n-1)/2] \sum_{1=i\neq j=1}^{K} p_i^{n-2} p_j^2$. Note by the definition of Ω, $D > 0$ (since $p_i \neq 1$). Also

$$\begin{aligned} P\{X_1^n = 0, \ldots, X_i^n = n-2, \ldots, X_j^n = 2, \ldots, X_K^n = 0\} \\ = [(n-1)/2] p_i^{n-2} p_j^2 / D. \end{aligned} \tag{A.3}$$

For the sample point in brackets of (A.2) or (A.3) $U_n = \sum_{\substack{\nu=1 \\ \nu\neq i, \nu\neq j}}^{K} p_\nu$. We prove the theorem by showing that (1.2) cannot be Bayes in the wide sense for the conditional problem. Bayes in the wide sense is defined in Wald ((1950), p. 17). Since the estimators which are Bayes in the wide sense are an essentially complete class and the loss function is squared error (1.2) is inadmissible. The essentially complete class result appears in Wald ((1950), Theorem 3.17) and also in Le Cam (1955). The fact that the estimand U_n in this problem is a random variable does not effect the Wald or Le Cam results since for the squared error loss function the risk function of any estimator is lower semi-continuous in the space of decisions for each fixed $\boldsymbol{p}$.

To show (1.2) is not Bayes in the wide sense define the sets $B_\varepsilon = \{\boldsymbol{p}: p_i \geq 1 - \varepsilon \text{ for some } i\}$, $0 < \varepsilon \leq 1$. Consider any sequence of priors $\{\xi_n\}$ for which there exists an $\varepsilon > 0$ such that

$$\liminf_{m\to\infty} P_{\xi_m}(B_\varepsilon) < 1 .$$

For such a sequence $\{\xi_m\}$ there exists a subsequence of priors $\{\xi_{m'}\}$ such that

$$\lim_{m'\to\infty} P_{\xi_{m'}}(B_\varepsilon) < 1 .$$

Note that $P_{\xi_{m'}}(B_\varepsilon) < 1$ is equivalent to

$$P_{\xi_{m'}}(p_i \leq 1 - \varepsilon \text{ all } i = 1,\ldots,k) > 0 .$$

For such a prior distribution $\xi_{m'}$, the Bayes estimator is given by

$$\text{(A.4)}\qquad \delta_{m'}(0,\ldots,n-2,\ldots,2,\ldots 0) = \frac{\int\left(\sum_{\nu\neq i,j} p_\nu\right)[p_i^{n-2}p_j^2/D]\,d\xi_{m'}(\boldsymbol{p})}{\int[p_i^{n-2}p_j^2/D]\,d\xi_{m'}(\boldsymbol{p})}$$

and

$$\text{(A.5)}\qquad \delta_{m'}(0,\ldots,n-1,\ldots,1,\ldots 0) = \frac{\int\left(\sum_{\nu\neq i,j} p_\nu\right)[p_i^{n-1}p_j/D]\,d\xi_{m'}(\boldsymbol{p})}{\int[p_i^{n-1}p_j/D]\,d\xi_{m'}(\boldsymbol{p})}$$

whenever the denominators are positive (note that there may be (i, j) pairs for which the denominators in (A.4) and (A.5) are zero. Then the Bayes estimator can be anything. However the denominators cannot be 0 for all (i, j) pairs). Also $\lim_{m'\to\infty} P_{\xi_{m'}}(B_\varepsilon) < 1$ is equivalent to $\lim P_{\xi_{m'}}(p_i \leq 1 - \varepsilon$ all $i = 1,\ldots,k) > 0$ so that

$$\lim \max_{i\neq j} \int[p_i^{n-2}p_j^2/D]\,d\xi_{m'}(\boldsymbol{p}) > 0$$

and

$$\lim_m \max_{i \neq j} \int [p_i^{n-1} p_j / D] d\xi_{m'}(\boldsymbol{p}) > 0 .$$

Thus if the terms in (A.4) were, for all pairs i, j, to converge to 0, then the same would be true for at least one term of the form (A.5). Hence Good's estimator cannot be Bayes in the wide sense with respect to such sequence of priors.

Next consider any sequence of priors for which

$$\lim_{m \to \infty} P_{\xi_m}(B_\varepsilon) = 1$$

for every $\varepsilon > 0$. For such a sequence the (marginal) probability of the set of sample points

$$A = \{(n-1, 1, 0, \ldots, 0) \text{ and all its permutations}\}$$

converges to 1 and so U_n converges to 0 in probability. Thus Good's estimator has an expected risk which converges to $(1/n)^2$ while the estimator $\delta^*(\boldsymbol{X}^n) \equiv 0$ has expected risk which converges to 0. Thus $\delta^*(\boldsymbol{X}^n)$ is Bayes in the wide sense for such a sequence of priors and $G(\boldsymbol{X}'')$ cannot be Bayes in the wide sense. This completes the proof of Lemma A.1.

LEMMA A.2. *If Good's estimator is inadmissible for the conditional problem, it is inadmissible for the unconditional problem.*

PROOF. Let $T(\boldsymbol{X}^n)$ be the estimator that beats $G(\boldsymbol{X}^n)$ for the conditional problem. Then we have

$$\text{(A.6)} \qquad \sum_{\boldsymbol{x}^n \in F_1 \cup F_2} (T(\boldsymbol{x}^n) - U_n)^2 f(\boldsymbol{x}^n; \boldsymbol{p}) / D \geq \sum_{\boldsymbol{x}^n \in F_1 \cup F_2} (G(\boldsymbol{x}^n) - U_n)^2 f(\boldsymbol{x}^n; \boldsymbol{p}) / D ,$$

for all $\boldsymbol{p} \in \Omega$, with strict inequality for some $\boldsymbol{p} \in \Omega$. Since $D > 0$ for all $\boldsymbol{p} \in \Omega$, multiply both sides in (A.6) by D and add

$$\sum_{\boldsymbol{x}^n \in (F_1 \cup F_2)^c} (G(\boldsymbol{x}^n) - U_n)^2 f(\boldsymbol{x}^n; \boldsymbol{p})$$

to both sides of (A.6). The resulting right-hand side is the risk function of $G(\boldsymbol{X}^n)$ while the left-hand side becomes the risk of an estimator, $T^*(\boldsymbol{X}^n)$ say, where

$$T^*(X^n) = \begin{cases} G(X^n) & \text{for } x^n \in (F_1 \cup F_2)^c , \\ T(X^n) & \text{for } x^n \in F_1 \cup F_2 , \end{cases}$$

and T^* is better than G for the unconditional problem. This completes the proof of Lemma A.2.

LEMMA A.3. *For k bounded by known K, Good's estimator is inadmissible among the class of permutation invariant estimators.*

PROOF. We use the invariance theory developed in Ferguson (1967), Chapter 4 with one modification. We use Definition 1 of Ferguson, p. 144 which defines the notion of a family of distributions invariant under the group G. Instead of Ferguson's Definition 2 on page 145 we define a decision problem to be invariant under the group G as follows: Let G be a group with element g. Let $\bar{g} \in \bar{G}$ be the group induced on the parameter space. Let $\tilde{g}$ be the identity transformation action on the action space. The decision problem is invariant under G if

(a) the family of distributions is invariant under G.

(b) $L(\boldsymbol{p}, a, \boldsymbol{X}^n) = L(\bar{g}\boldsymbol{p}, \tilde{g}a, g\boldsymbol{X}^n)$.

In other words the loss function, which for our problem depends on $\boldsymbol{p}$, a and $\boldsymbol{X}^n$, is invariant. With this definition of invariance of a decision problem, Ferguson's Lemma 4.2.1, Theorem 4.2.1 and Theorem 4.3.2 are easily established for our problem. The group operating on the sample space and parameter space is the permutation group which is finite (since K is known) and the group operating on the action space is merely the identity. We note $U_n(\boldsymbol{X}^n, \boldsymbol{p})$ is invariant and so the loss function $(t - U_n)^2$ is invariant as is the entire problem.

From Lemma A.2, Good's estimator is inadmissible. Therefore there exists an estimator, say T^* which is better. If T^* is invariant the proof is complete. If T^* is not invariant we use the proof of Theorem 2, Ferguson ((1967), p. 156) to construct an invariant estimator which is better than $G(\boldsymbol{X}^n)$. This completes the proof of Lemma A.3 and the proof of Theorem 2.1.

REFERENCES

Clayton, M. K. and Frees, W. (1987). Nonparametric estimation of the probability of discovering a new species, *J. Amer. Statist. Assoc.*, **82**, 305–311.

Esty, W. W. (1986). The efficiency of Good's nonparametric coverage estimator, *Ann. Statist.*, **14**, 1257–1260.

Ferguson, T. S. (1967). *Mathematical Statistics, A Decision Theoretic Approach*, Academic Press, New York.

Good, I. (1953). On the population frequencies of species and the estimation of population parameters, *Biometrika*, **40**, 237–264.

LeCam, L. (1955). An extension of Wald's theory or statistical decision functions, *Ann. Math. Statist.*, **26**, 69–81.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, Second Edition, Wiley, New York.

Robbins, H. (1968). Estimating the total probability of the unobserved outcome of an experiment, *Ann. Math. Statist.*, **39**, 256–257.

Starr, N. (1979). Linear estimation of the probability of discovering a new species, *Ann. Statist.*, **1**, 644–652.

Wald, A. (1950). *Statistical Decision Functions*, Wiley, New York.