The Annals of Statistics
1992, Vol. 20, No. 2, 1094–1109

FROM THE SPECIES PROBLEM TO A GENERAL COVERAGE PROBLEM VIA A NEW INTERPRETATION

BY SHAW-HWA LO

Columbia University and Harvard University

A basic interpretation is given which provides a new way of understanding the structure of the species problem and which leads to the popular Turing–Good–Robbins estimator. Through this interpretation we provide an explanation why the Turing–Good–Robbins estimators are always biased. An iterative procedure is suggested and applied to these estimators, which leads to new estimators whose biases are reduced. Using this basic construction we are able to generalize our discussion to a much broader class of coverage problems with the species problem as a special case. Three examples are studied in detail: the species problem, the problem of estimating the volume of a convex set and the missile-coverage problem. Furthermore, we derive general (new) estimators and study their properties by applying the interpretation to the framework of the general coverage problem. It is pointed out that, as in species problem, the general estimators derived from the interpretation are usually biased, we then apply our construction together with the iterative procedure to the previous three examples to produce new estimators whose biases are reduced. Finally, we extend our construction to the conditional cases.

1. Introduction. The problem of estimating the total probability of unseen species goes back at least to A. M. Turing according to Good (1953). To describe the problem comprehensively, we use the notation of Robbins (1956, 1968). Let $e_1, e_2, \ldots$ be the possible distinct species with probabilities $p_1, p_2, \ldots,$ of being selected in a single trial. In n independent trials suppose that n_r species appear r times, $r = 1, 2, \ldots$ and so $\sum_{r=1}^{\infty} r n_r = n$. We also use n_0 to denote the number of species which are not present in the sample. It is clear that $n_1, n_2, \ldots$ are observable but n_0 is not. In fact n_0 is infinite if there are infinitely many species. Let $X_i = j$ if and only if the ith trial results in outcome e_j. For $r \geq 0$, let $\psi_j(r; n) = 1$ if the number of species e_j appearing in the sample is r and 0 otherwise. The sum of the probabilities of all species that are each represented r times in the sample is $C_r = \sum_{j=1}^{\infty} p_j \psi_j(r; n)$. It is clear that C_r is a random quantity. To estimate C_r, Turing and Good [see Good (1953)] suggested the formula

$$\frac{(r+1) n_{r+1}}{n}. \tag{1.1}$$

Using a uniform prior, Good (1953) gave a derivation of these estimators from a Bayesian point of view. Since then several interpretations of these

Received April 1989; revised, August 1991.
[1]Research supported by ONR Grant N00014-86-K-0246 and by NSF Grant DMS-85-96024.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62G09.
Key words and phrases. Species problem, general coverage problem, new interpretation.

estimators have appeared in the literature. These include Robbins (1956, 1968) and Diaconis and Stein (1983) among others. It should be noted that Robbins (1968) constructed an unbiased estimator for C_0 which is very similar to (1.1). Here an estimator is called unbiased for estimating a random variable if E(estimate) = E(random variable). The problem continues to attract the attention of many researchers. To name a few: Starr (1979), Clayton and Frees (1987), Esty (1986), Bickel and Yahav (1986), Cohen and Sackrowitz (1990) and Banerjee and Sinha (1985). As an important application, the species problem is currently of great interest to researchers in automated speech identification [Bahl, Jelinek and Mercer (1983), Jelinek (1976) and Katz (1987) among others].

The objective here is to introduce a basic derivation of these estimators which leads to interesting applications other than the species problem. We consider the estimation of a random quantity $\theta_n(\xi_n)$ which can be expressed in the form $E\{H_n(\xi_n; X_{n+1})|\xi_n\}$, where $\xi_n = \xi_n(X_1, \dots, X_n)$ is the observed sample, X_{n+1} is an additional independent observation and $H_n(\xi_n; x)$ is a real-valued function of ξ_n and x, typically an indicator function. When H_n is an indicator function, we can interpret $\theta_n(\xi_n)$ as the conditional probability of some event regarding X_{n+1} given the sample ξ_n. In the species example, $C_r = \theta_n(\xi_n) = E\{I(X_{n+1} \in S_n(r))|\xi_n\} = P\{X_{n+1} \in S_n(r)|\xi_n\}$, where $H_n(\xi_n; x) = I(x \in S_n(r))$ and $S_n(r) = \{j; \psi_j(r; n) = 1\}$—a function of the observed sample ξ_n.

The key idea of this approach is to create information about the unknown random quantity $\theta_n(\xi_n)$ by deleting one observation from the sample at a time and comparing the deleted observation with the remaining $n - 1$ observations. The deletion method of constructing estimators can be made precise as follows: Let $\xi_{n,j}$ be the sample with the jth observation deleted. An estimator of $\theta_{n-1}(\xi_{n,j})$, that is conditionally unbiased given $\xi_{n,j}$, is $H_{n-1}(\xi_{n,j}; X_j)$; call this $T^*_{n-1,j}$. The final estimator is defined to be $T^{**} = \text{avg}(T^*_{n-1,j}) = (1/n)\sum_j T^*_{n-1,j}$. Since the final estimator directly estimates a probabilistic phenomenon involving $n - 1$ observations, it will be named $(n - 1)$-estimator hereafter.

If we apply this method to the question of estimating C_r in the species example, we end up with

$$T^*_{n-1,j} = I\big(X_j \in S_{n-1,j}(r)\big) \tag{1.2}$$

and

$$T^{**} = \text{avg}\big(T^*_{n-1,j}\big) = \frac{(r+1)n_{r+1}}{n}, \tag{1.3}$$

exactly the formula suggested by Turing and Good [Good (1953)], where $S_{n-1,j}(r) = \{i; \psi_{ij}(r; n-1) = 1\}$ and $\psi_{ij} = 1$ if and only if i appear exactly r times in the jth deleted sample $\xi_{n,j}$. Note that the relation $I(X_j \in S_{n-1,j}(r)) = 1$ iff $X_j \in S_n(r+1)$ is used in deriving (1.3). By taking expectations, we obtain $EC_r = P\{X_{n+1} \in S_n(r)\}$ and $E(T^{**}) = E((r+1)n_{r+1}/n) = P\{X_n \in S_{n-1}(r)\}$.

Although the quantity to be estimated $\theta_n(\xi_n)$ is random, it is still reasonable to evaluate an estimation $T_n(\xi_n)$ by mean square error of the estimator averaged over the sampling distribution of Ξ; that is, $E(T_n(\Xi) - \theta_n(\Xi))^2$, where Ξ denotes the random variable corresponding to the sample ξ_n. As usual, the mean-square error can be broken into variance and squared bias, where now bias is defined to be $E[T_n - \theta_n]$. From this viewpoint the estimator $T^*_{n-1,j}$ is an unbiased estimate of $E(I(X_j \in S_{n-1,j}(r))|\xi_{n-1,j})$ and T^{**} is an unbiased estimate of $\operatorname{avg}(E(I(X_j \in S_{n-1,j}(r)|\xi_{n-1,j})))$, a probabilistic statement based on $n-1$ observations. As an estimator of C_r, T^{**} is biased. This bias is slight [$O(n^{-1})$; see Section 3.1 for details] because C_r changes little as n changes. In Section 3 we shall show how to reduce the bias which could be substantial in other problems for which this approach applies.

In Section 2 we shall introduce a general coverage problem and show how to apply the deletion method to this problem. A key issue is the bias of the estimator, as it is clear that the $(n-1)$-estimator is biased for $\theta_n(\xi_n)$; instead, it unbiasedly estimates $\theta^*_n(\xi_n) = \operatorname{avg}(\theta_{n-1}(\xi_{n,j}))$. The deletion construction will not be useful unless the bias $E(\theta_n(\Xi_n) - \theta^*_n(\Xi_n))$ is small. In Section 2 it is shown how to evaluate the magnitude of this bias; it will be shown in Section 3 how bias corrected versions of the $(n-1)$-estimators can be constructed.

Section 3 presents three special examples. The first example is a further discussion of the species problem. The second example concerns the problem of estimating the volume of an arbitrary convex set V in the Euclidean space R^k. The data in this problem consists of n independent random observations $X_1, \dots, X_n$ uniformly distributed over V. By applying our construction to this problem, we obtain our estimator which is the volume of V_n, the convex hull formed by the n points, multiplied (enlarged) by a factor $n(n - \#$ of vertices of $V_n)^{-1}$ [see (3.2.2)]. Figure 1 shows a case when $n = 25$, $k = 2$. (The connection between the interpretation and the problem of estimating the volume of a convex polyhedron was pointed to me by Diaconis in a conversa-

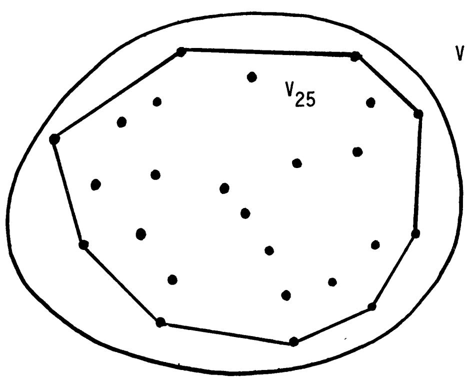

FIG. 1. $k = 2$, $n = 25$, $P(X_{26} \notin V_{25}|V_{25})$ *is estimated by* $(9/25)$, *where* 9 *is the number of vertices of* V_{25}. *The volume of* V *is estimated by* $(25/16)\mathrm{vol}(V_{25})$.

tion.) In the special case when $k = 2$, our estimator together with a recent result due to Groeneboom (1988) leads to interesting large sample results (Theorem 3.1). The last example deals with the following missile problem: Suppose n missiles are delivered and landing at a certain target area which is usually larger than the effective area caused by the explosion of a single missile. The typical questions we are interested in are: (i) if the $(n+1)$th missile is fired, what is the chance that this additional missile would involve an area which was not previously covered? (ii) How large is the newly covered area? (iii) How many more missiles are needed to cover 90% of the target area?

Section 4 returns to the species problem. The issue of concern here is to estimate a conditional parameter associated with the observations, such as the mean or median of those values associated with the unobserved species. The principles involved for a more general situation (other than the species problem) are also outlined.

2. A general coverage problem. In this section we shall introduce a framework where a general coverage problem can be defined. Let $(\Omega, \mathscr{F}, P)$ be a certain probability space, where Ω denotes a collection of a *certain subset of a fixed set* Δ in R^k, $k \geq 1$, $\mathscr{F}$ and P are an appropriate σ-field and a probability measure defined on $\mathscr{F}$, respectively. Let $X_1, \dots, X_n$ be n iid random elements defined on the usual product space $(\Omega^n, \mathscr{F}^n, P^n)$ such that each X_j is the jth coordinate variable; that is, $X_j(\omega_1, \dots, \omega_n) = \omega_j$ for all $1 \leq j \leq n$, where $(\omega_1, \dots, \omega_n) \in \Omega^n$. Typical sample outcomes of $X_1, \dots, X_n$ are n subsets of Δ. Let g be a measurable function from Ω to R^k. Some of the problems we are interested in are: Given a specified subset $S_n = S_n(\xi_n)$ of Δ, where $\xi_n = \xi_n(\mathbf{X}_n)$ denotes the observed data and $\mathbf{X}_n = (X_1, \dots, X_n)$, estimate, for a new independent observation X_{n+1}, (i) the probability that $g(X_{n+1}) \in S_n$ given S_n. Furthermore, if all elements in Ω are Lebesgue-measurable, we are interested in estimating (ii) the expected volume of $S_n \cap X_{n+1}$ given S_n and (iii) the expected volume of $S_{n+1} \cap S_n$ given S_n.

The key idea can best be described as a one-step backward procedure as follows. Let X_j be removed from the observed sample ξ_n and let $\xi_{n,j}$ denote this sample. Let $S_{n-1,j} = S(\xi_{n,j})$ be the specified subset of Δ based on this jth deleted sample $\xi_{n,j}$. Let

$$I\big(g(X_j) \in S_{n-1,j}\big) = \begin{cases} 1, & \text{if } g(X_j) \in S_{n-1,j}, \\ 0, & \text{otherwise.} \end{cases}$$

Instead of estimating the probability $P(g(X_{n+1}) \in S_n | S_n)$ in (i), the $(n-1)$-estimator estimates $\text{avg}[P(g(X_j) \in S_{n-1,j} | S_{n-1,j})]$ unbiasedly. The construction can be described as follows: Since the probability that $g(X_j) \in S_{n-1,j}$ can be estimated unbiasedly by $I(g(X_j) \in S_{n-1,j})$, our final $(n-1)$-estimator is thus $n^{-1}\Sigma_{j=1}^n I(g(X_j) \in S_{n-1,j})$.

Likewise, in (ii) and (iii), the $(n-1)$-estimators are $n^{-1}\Sigma_{j=1}^n \text{vol}[S_{n-1,j} \cap X_j]$ and $n^{-1}\Sigma_{j=1}^n \text{vol}[S_{n-1,j} \cap S_n]$, respectively.

If $S_n = S_n(\xi_n)$ does not depend on $\mathbf{X_n}$, it is easy to see that our construction will lead to a naive estimator which is the function of the empirical measure.

As estimators of (i), (ii) and (iii), these $(n-1)$-estimators are all biased. In many applications the biases are slight because (i), (ii) and (ii) changes little as n increases. However, in the general framework, this property is not automatically guaranteed. As a result, just how well these $(n-1)$-estimators estimate (i), (ii) and (iii) depends upon the forms of S_n and S_{n-1}. Let X be randomly selected from $(\Omega, \mathscr{F}, P)$ with $E\mathrm{vol}(X)^2 < \infty$. Let g be a measurable function from $(\Omega, \mathscr{F}, P)$ to R^k such that $g(w) \in w$ for all $\omega \in \Omega$. We use $\Delta(A, B)$ to denote the set $(A \setminus B) \cup (B \setminus A)$, the symmetric difference of sets A and B. Notice this the notations $S_n = S_n(\xi_n(\mathbf{X_n})) = \mathbf{S}(\mathbf{X_n})$ are used throughout the paper. The following proposition tells us that the bias of using $(n-1)$-estimates to estimate (i), (ii) and (iii) depends on the closeness of S_{n-1} to S_n. Since the proof of this proposition is elementary, we state it without proof.

PROPOSITION 2.1. *Suppose that X is independent of*

$$S_{n-1} = S(X_1, X_2, \dots, X_{n-1}),\ S_n \quad \textit{and} \quad S_{n+1} = S(X_1, X_2, \dots, X_n, X_{n+1}),$$

and set

$$P\{X \cap \Delta(S_n, S_{n-1}) \neq \varnothing\} = \delta_n \geq 0 \quad \textit{for all } n \geq 1.$$

Then

$$|P\{g(X_{n+1}) \in S_n\} - P\{g(X_n) \in S_{n-1}\}| \leq \delta_n \tag{1}$$

and

$$E(\mathrm{vol}[S_n \cap X_{n+1}]) - E(\mathrm{vol}[S_{n-1} \cap X_n]) = O(\delta_n^{1/2}). \tag{2}$$

If we further assume $\mathrm{vol}(\Delta) < \infty$, *where Δ is a fixed set in R^k described in the beginning of this section, then* (2) *becomes*

$$E(\mathrm{vol}[S_n \cap X_{n+1}]) - E(\mathrm{vol}[S_{n-1} \cap X_n]) = O(\delta_n). \tag{2'}$$

To calculate the biases, we pretend that the additional observation X_{n+1} is taken. The (n)-estimates obtained by applying (i), (ii) and (iii) to the $(n+1)$ observations should be unbiased. The biases of $(n-1)$-estimates can be evaluated by comparing these $(n-1)$-estimates with (n)-estimates. For example, as in (i), the bias of $(n-1)$-estimate is

$$E\left\{\frac{1}{n}\sum_{j=1}^{n} I(g(X_j) \in S_{n-1,j}) - \frac{1}{n+1}\sum_{j=1}^{n+1} I(g(X_j) \in S_{n,j})\right\}, \tag{2.1}$$

where $S_{n,j} = S(\mathbf{X}_{n+1,j})$ and $\mathbf{X}_{n+1,j} = (X_1, X_2, \dots, X_{j-1}, X_{j+1}, \dots, X_n, X_{n+1})$, the jth-deleted sample of size n. The bias term (2.1) can be calculated once the knowledge of relationship between S_{n-1} and S_n is given and this is possible only if the nature of the problems is specifically given. In this case, as we shall see in the next section, some bias-reduced estimators are always available.

The key idea of construction these estimators is to create (n)-estimates pretending the additional observation X_{n+1} is taken. These (n)-estimates are not real since their forms depend on the unobservable X_{n+1} and the relationship between S_n and S_{n+1}. However, we can replace the quantities involving X_{n+1} by their conditional expectations given the data (notice that, at this stage, the relationship between S_n and S_{n+1} is needed in calculating the expected value of those quantities); since underlying distribution depends on unknown parameters, we will plug in our $(n-1)$-estimate obtained via the interpretation. The resulting estimators are usually bias-reduced as we shall see in Section 3.

REMARK 1. It is clear that one can produce new estimators by repeatedly applying this procedure to the current estimators recursively. The biases of new estimator are usually further reduced and the typical order of the biases after the kth iteration is $O(n^{-(k+1)})$.

REMARK 2. In general, there is no guarantee that the mean square errors of these new estimators will be smaller than that of the ones before the iteration. The issue of MSE is certainly important and deserves further investigation. However, Example 3.1 below shows that the MSE of the new estimator is indeed smaller than that of the $(n-1)$-estimate.

REMARK 3. The idea of this recursive procedure has similar flavor to the EM algorithm [see Dempster, Laird and Rubin (1977)] and to the concept of self-consistency due to Efron (1967). The major difference between our procedure and the EM is that the M-step of the EM is replaced by our backward procedure of constructing the required estimator here.

3. Examples.

3.1. *Species problems.* In this section we shall continue our discussion of species problems introduced in Section 1. Let Δ denote the set of all positive integers. Let $\Omega = \Delta$, $\mathscr{F} = 2^{\Delta}$ and $P\{X = i\} = p_i$ for $i \in \Delta$. The collection of unseen species can be expressed as $S_n = S(X_1, \dots, X_n) = \{j; j \notin \{X_1, \dots, X_n\}\} \subset \Delta$. Let g be the identity map from Δ to Ω, that is, $g(i) = i$. The problem of estimating the total probability of unseen species is thus equivalent to estimating the probability of $g(X_{n+1}) \in S_n$ given S_n. More precisely, $C_0 = P\{g(X_{n+1}) \in S_n | S_n\}$. According to the previous section, the $(n-1)$-estimates of C_0 and C_r are n_1/n and

$$\frac{1}{n}\sum_{j=1}^{n} I\big(X_j \in S_{n-1,j}(r)\big) = \frac{(r+1)n_{r+1}}{n}, \tag{3.1.1}$$

where $S_{n-1,j}(r) = \{i; \sum_{h \neq j}^{n} I_{X_h}(i) = r,\ i \in \Delta\}$.

THE BIASES OF $(n-1)$-ESTIMATES. From (2.1) and (3.1.1), the bias of the $(n-1)$-estimate in estimating $P\{X_{n+1} \in S_n | S_n\}$ is $E\{(n_1/n) - (n_1 + \delta)/$

$(n + 1)\}$, where

$$\delta = \begin{cases} 1, & \text{if } X_{n+1} \notin \{X_1, \dots, X_n\}, \\ 0, & \text{if } X_{n+1} \text{ occurred at least twice among } \{X_1, \dots, X_n\}, \\ -1, & \text{if } X_{n+1} \text{ occurred once among } \{X_1, \dots, X_n\}. \end{cases}$$

It follows trivially that |bias of $(n - 1)$-estimate| $\leq 2/(n + 1) = O(1/n)$. The knowledge between the relationship of S_{n-1} to X_n enables us to construct a better estimate of which the bias is of order $O(n^{-2})$ in contrast with the order of $O(n^{-1})$ provided by the previous $(n - 1)$-estimate. The construction can be described heuristically as follows.

Let n'_1 denote the number of species appearing once in the sample $\{X_1, X_2, \dots, X_n, X_{n+1}\}$. Since n'_1 is not observed, we will fill in its expected value given the data; since this distribution depends on unknown parameters, we will plug in our $(n - 1)$-estimate. Let $\hat{n}'_1$ denote this estimate which is defined by

$$\hat{n}'_1 = n_1 + 1, \quad \text{with probability } \frac{n_1}{n},$$

$$\hat{n}'_1 = n_1, \quad \text{with probability } \left(1 - \frac{n_1}{n} - \frac{2n_2}{n}\right),$$

$$\hat{n}'_1 = n_1 - 1, \quad \text{with probability } \frac{2n_2}{n}.$$

The expected value of $\hat{n}'_1$ given $(n_1, n_2, \dots)$ is $E(\hat{n}'_1|(n_1, n_2, \dots,)) = n_1 + n_1 n^{-1} - 2n_2 n^{-1}$. The final estimate of estimating the total probability of unseen species in the sample is $(n + 1)^{-1}E(\hat{n}'_1|(n_1, \dots,)) = (n + 1)^{-1}(n_1 + n_1 n^{-1} - 2n_2 n^{-1})$. The fact that the bias of this estimate is of order $O(n^{-2})$ can be seen by noticing that

$$(3.1.2) \quad E\left(\frac{n_1}{n} - \frac{2n_2}{n} - \delta\right) = E\left(\frac{n_1}{n} - \frac{2n_2}{n} - \frac{n'_1}{n+1} + \frac{2n'_2}{n+1}\right),$$

where n'_2 is the number of species appearing twice among $\{X_1, X_2, \dots, X_n, X_{n+1}\}$. It is clear that $|n_1 - n'_1| \leq 1$ and $|2n_2 - 2n'_2| \leq 2$ with probability 1. It follows from (3.1.2) that the absolute bias of the final estimate is bounded by $(3/n(n + 1))$, which is of order (n^{-2}).

One can mimic the above idea to find an estimator which has smaller bias than that of (3.1.1) in estimating C_r. The bias-reduced estimator is $[(r + 1)n_{r+1} + (r + 1)((r + 1)n_{r+1} - (r + 2)n_{r+2})n^{-1}](n + 1)^{-1}$, which has smaller bias.

3.2. *Estimating the volume of a convex set in R^k.* The problem of estimating the volume of a certain convex set can be described as follows: Let V denote a certain unknown convex set with finite volume in R^k. The data in this problem consists of independent random samples $X_1, \dots, X_n$ uniformly distributed over V. The first question we want to ask is: Having observed

$X_1, \dots, X_n$, how do we estimate vol(V)? The joint likelihood of $X_1, \dots, X_n$ is

$$\begin{aligned} \text{Lik}(X_1, \dots, X_n | V) &= \left[\frac{1}{\text{vol}(V)}\right]^n \prod_{i=1}^{n} I_V(X_i) \\ &= \left[\frac{1}{\text{vol}(V)}\right]^n I(V_n \subset V), \end{aligned} \tag{3.2.1}$$

where $V_n = V_n(X_1, \dots, X_n)$ is the convex hull formed by $\{X_1, \dots, X_n\}$.

It is easy to see from (3.2.1) that V_n, the convex hull formed by $\{X_1, \dots, X_n\}$, is a sufficient statistic of vol(V) according to Neyman's factorization theorem. This suggests that a reasonable estimate of vol(V) should be a function of V_n, the sufficient statistic of vol(V). To construct an estimate of vol(V), we first consider the problem of estimating the conditional probability $P(X_{n+1} \in V_n | V_n)$. Let $\Omega = V = \Delta$ and let $\mathscr{F}$ be the usual Borel field on V. Let P be the probability measure uniformly distributed over V. Define $g(w) = w$, the identity map from V to V. If we define $S_n = S(X_1, \dots, X_n) = V_n(X_1, \dots, X_n)$, the $(n-1)$-estimate of $P(X_{n+1} \in V_n | V_n)$ is $n^{-1}\Sigma_{j=1}^{n} I(X_j \in V_{n-1,j})$, where $V_{n-1,j}$ is the convex hull formed by the jth-deleted sample. Since

$$P(X_{n+1} \in V_n | V_n) = \int_{V_n} \left(\frac{1}{\text{vol}(V)}\right) dw = \frac{\text{vol}(V_n)}{\text{vol}(V)},$$

it follows that the $(n-1)$-estimate of vol(V) is

$$\widehat{\text{vol}_{n-1}}(V) = \text{vol}(V_n) \left[\frac{1}{n} \sum_{j=1}^{n} I(X_j \in V_{n-1,j})\right], \tag{3.2.2}$$

and the $(n-1)$-estimate of $P(X_{n+1} \notin V_n | V_n)$ is

$$\frac{1}{n} \sum_{j=1}^{n} \left[1 - I(X_j \in V_{n-1,j})\right] = \frac{\#\text{ of vertices of } V_n}{n}. \tag{3.2.3}$$

Let vtx(U) denote the set of vertices of a convex polyhedron U in R^k. Applying the similar idea of Section 3.1 to the current situation, we end up with a bias-reduced estimate [of $P(X_{n+1} \notin V_n | V_n)$]

$$\frac{\#\{\text{vtx}(V_n)\} + (1/n)\Sigma_{j=1}^{n}\left[\#\{\text{vtx}(V_n)\} - \#\{\text{vtx}(V_{n-1,j})\}\right]}{n+1}, \tag{3.2.4}$$

where $\#\{\text{vtx}(U)\}$ = number of vtx(U) for a convex polyhedron U. The bias-reduced estimates of $P(X_{n+1} \in V_n | V_n)$ and vol(V) are thus

$$1 - \frac{\#\{\text{vtx}(V_n)\} + (1/n)\Sigma_{j=1}^{n}\left[\#\{\text{vtx}(V_n)\} - \#\{\text{vtx}(V_{n-1,j})\}\right]}{n+1}, \tag{3.2.5}$$

and

$$\text{vol}(V_n)\left\{1 - \frac{\#\{\text{vtx}(V_n)\} + (1/n)\Sigma_{j=1}^{n}\left[\#\{\text{vtx}(V_n)\} - \#\{\text{vtx}(V_{n-1,j})\}\right]}{n+1}\right\}^{-1}. \tag{3.2.6}$$

It is not difficult to check that the biases of estimates (3.2.4) and (3.2.6) are of smaller order [$O(n^{-2})$, in fact] than those of $(n-1)$-estimates provided by (3.2.3) and (3.2.2). Since the arguments to verify this fact are very similar to those given in Section 3.1, we omit it. The problem of estimating the volume of newly covered area if an additional observation X_{n+1} is taken can thus be estimated by the $(n-1)$-estimate $n^{-1}\Sigma_{j=1}^{n}\operatorname{vol}[V_n \setminus V_{n-1,j}] = (\operatorname{vol}(\Delta_n)/n)$ (say), where $\operatorname{vol}(\Delta_n) = \Sigma_{j=1}^{n}\operatorname{vol}[V_n \setminus V_{n-1,j}]$. A bias-reduced estimate, using the similar idea again, can be expressed as

$$\frac{\operatorname{vol}(\Delta_n) + (1/n)\Sigma_{j=1}^{n}\left[\operatorname{vol}(\Delta_n) - \operatorname{vol}(\Delta_{n-1,j})\right]}{n+1}, \tag{3.2.7}$$

where $\operatorname{vol}(\Delta_{n-1,j}) = \Sigma_{j' \neq j}\operatorname{vol}[V_{n-1,j} \setminus V_{n-2,jj'}]$ and $V_{n-2,jj'}$, the convex hull formed by $\{X_i\}_{i=1}^{n}$ with X_j, $X_{j'}$ deleted.

EXAMPLE 3.1. Suppose $X_1, \dots, X_n$ are iid from $U(\theta_1, \theta_2)$, with unknown parameter θ_1 and θ_2. The volume (length, in fact) of the current convex set is $\theta_2 - \theta_1$. Let $X_{(1)} < \cdots < X_{(n)}$ be the ordered values of $\{X_i\}_{i=1}^{n}$. It follows from previous discussion that the $(n-1)$-estimate of $P(X_{n+1} \in (X_{(1)}, X_{(n)}) | (X_{(1)}, X_{(n)}))$ is

$$\frac{1}{n}\sum_{j=1}^{n} I(X_j \in V_{n-1,j}) = \frac{n-2}{n}. \tag{3.2.8}$$

In fact, from (i) of Section 2 this $(n-1)$-estimate is an unbiased estimate of $P(X_n \in (X_{(1)}, X_{(n-1)}))$ based on $n-1$ observations $\{X_i\}_{i=1}^{n-1}$. The bias-reduced estimates of $P(X_{n+1} \in (X_{(1)}, X_{(n)}) | (X_{(1)}, X_{(n)}))$ and $\theta_2 - \theta_1$ are [from (3.2.5) and (3.2.6)] thus

$$1 - \frac{2}{n+1} = \frac{n-1}{n+1} \quad \text{and} \quad \left(X_{(n)} - X_{(1)}\right)\frac{n+1}{n-1}$$

and both become unbiased.

If an additional observation X_{n+1} is taken, the length of newly covered area can be estimated by the $(n-1)$-estimate: $\tilde{l}_n = [(X_{(n)} - X_{(n-1)}) + (X_{(2)} - X_{(1)})]/n$. A bias-reduced estimate is

$$\hat{l}_n = \tilde{l}_n - \frac{2}{n(n+1)}\left\{\left[X_{(n-1)} - X_{(n-2)} + X_{(3)} - X_{(2)}\right]\right\}.$$

In this example, one can show that the bias-reduced estimator is indeed a better estimator. This can be seen, for example, from the fact that the MSE of the $(n-1)$-estimator in (3.2.8) can be expressed as

$$E\left[\frac{n-1}{n+1} - P\left\{X_{n+1} \in (X_{(1)}, X_{(n)}) \middle| (X_{(1)}, X_{(n)})\right\}\right]^2 + \frac{4}{n^2(n+1)^2},$$

which shows that the net gain of the bias-reduced estimator is $(4/n^2(n+1)^2)$, in terms of MSE. Simple calculations show that the MSE of the bias-reduced estimator is

$$\frac{2(n-1)}{(n+1)^2(n+2)}.$$

The relative decrease (gain) in MSE is thus $4[n^2(n+1)^2]^{-1}/[2(n-1)(n+1)^{-2}(n+2)^{-1} + 4n^2(n+1)^2] \sim (2/n^2)$.

It is heuristically clear that the volume of V_n would tend to the volume of V as n goes to infinity. It is desired to find the rate (and distribution, if possible) of how fast the volume of V_n tends to that of V as n becomes large. In R^2 we can go further in this problem and derive the limiting distributions of our $(n-1)$-estimators. Let N_n be the number of vertices of V_n. If V is a convex polygon in R^2 with r edges, it was shown in Proposition 1 of Rényi and Sulanke (1963) that

$$EN_n = \tfrac{2}{3}r(\log n + C) + D + o(1), \tag{3.2.9}$$

where C is Euler constant and D is another constant depending on V. It was also shown in the same paper that $EN_n \sim n^{1/3}$ if V has a smooth boundary in R^2. Since then much work has been done in this direction: Efron (1965), Geffroy (1959, 1961), Raynaud (1970), Eddy and Gale (1981), Buchta (1984) and Schneider (1987) among others. When V is a unit disk, Efron (1965) gave an exact formula for EN_n, from which an asymptotic expression of EN_n can be derived and expressed as

$$EN_n = 2\pi C_1 n^{1/3} + o(n^{1/6}), \tag{3.2.10}$$

where C_1 is a constant between 0 and 1. The following theorem presents the limiting distributions of our $(n-1)$-estimators.

THEOREM 3.1. (i) *If V is a convex polygon with r, $r \geq 3$ vertices, then as $n \to \infty$,*

$$\frac{n[(N_n/n) - P(X_n \notin V_{n-1})]}{\sqrt{(10/27)r\log n}} \to_{\mathscr{L}} N(0,1), \tag{3.2.11}$$

$$\frac{n[(N_n/n)\operatorname{vol}(V_n) - E[\operatorname{vol}(V \setminus V_{n-1})]]}{\sqrt{(10/27)r\log n}\operatorname{vol}(V)} \to_{\mathscr{L}} N(0,1). \tag{3.2.12}$$

(ii) *If V is the unit disk in the plane, then as $n \to \infty$, we have*

$$P(X_n \notin V_{n-1}) \approx (n^{-2/3})$$

and

$$n^{5/6}\left[\frac{N_n}{n} - P(X_n \notin V_{n-1})\right]\Big/\sqrt{2\pi C_2} \to_{\mathscr{L}} N(0,1), \tag{3.2.13}$$

$$n^{5/6}\left[\frac{N_n}{n}\operatorname{vol}(V_n) - E[\operatorname{vol}(V \setminus V_n)]\right]\Big/\sqrt{2\pi C_2}\operatorname{vol}(V) \to_{\mathscr{L}} N(0,1), \tag{3.2.14}$$

where C_1, C_2 are two positive constants between 0 and 1.

PROOF. It was shown in Groeneboom (1988) that if V is a convex polygon with r vertices, then

$$(3.2.15)\qquad n\left(\frac{N_n}{n}-\frac{2}{3}r\frac{\log n}{n}\right)\Big/\sqrt{\frac{10}{27}r\log n}\to_{\mathscr{L}} N(0,1);$$

if V is the unit disk on the plane, then

$$(3.2.16)\qquad n\left(\frac{N_n}{n}-2\pi C_1 n^{-2/3}\right)\Big/\sqrt{2\pi C_2 n^{1/3}}\to_{\mathscr{L}} N(0,1).$$

It follows from (3.2.9) and (3.2.10) that

$$E\left(\frac{N_n}{n}\right)=P(X_n\notin V_{n-1})$$

$$=\begin{cases}\frac{2}{3}r(\log n)/n+o\left((\log n)^{1/2}/n\right), & \\ & \text{if } V \text{ is a polygon with } r \text{ vertices},\\ 2\pi C_1 n^{-2/3}+o(n^{-5/6}), & \text{if } V \text{ is the unit disk.}\end{cases}$$

The results (3.2.11) and (3.2.13) are obtained by replacing the asymptotic means by their true means. To prove (3.2.12), we first notice that $\mathrm{vol}(V_n)-\mathrm{vol}(V)=O_p((\log n)n^{-1})$ since $(N_n/n)=O_p((\log n)n^{-1})$. The result (3.2.12) then follows from the facts that $E(\mathrm{vol}(V\setminus V_{n-1}))=\mathrm{vol}(V)P(X_n\notin V_{n-1})$ and $(N_n/n)[\mathrm{vol}(V_n)-\mathrm{vol}(V)]=o_p(n^{-1})$. The result (3.2.14) can be proved similarly. □

REMARK. In the case that V is a general convex set with smooth boundary, the results in (2) still hold, but with C_2 replaced by $C_2'=C_2(\pi/\mathrm{vol}(V))^{1/3}\int_{\partial V}k(s)^{1/3}\,ds/2\pi$, where ∂V is the boundary of V, and $k(s)$ is the curvature function of arc length. For details, see Rényi and Sulanke (1963) and Groeneboom (1988).

Some implications deserve further discussion here. From (3.2.3), the probability that the new observation X_{n+1} will fall outside the convex hull formed by the sample $\{X_1,\dots,X_n\}$ is determined by the knowledge about the number of vertices of the convex hull. This result [i.e., (3.2.3)] holds for any distribution on R^k and any $k\geq 1$. However, to estimate the volume of the convex set, the uniform distribution is used to create the relation like (3.2.2). We do not have a general theorem like Theorem 3.1 in R^k when $k\geq 3$, simply because a more general version of (3.2.16) is not available at the moment. However, from an applied point of view, we can always estimate the volume of a convex figure by (3.2.2), and the vertices of V_n will provide us with information about $V\setminus V_n$. It seems to this author that almost all relevant information about $V\setminus V_n$ is within the set of vertices of V_n. This point will be further justified in Section 4 in terms of species problem. From previous discussion, it is found that

$$\frac{\mathrm{vol}(V\setminus V_n)}{\mathrm{vol}(V)}\approx O(n^{-1})$$

or $O(n^{-2/3})$ depend on whether V is a smooth convex set in R^1 or in R^2. A question of interest is: What if $k \geq 3$?

3.3. *The missile problem.* n missiles are delivered and landing at a certain target area which is usually much larger than the effective area caused by the explosion of a single missile. The effective area here can be referred to as a covered area in the present terminology. A similar bombing problem was first considered by Robbins in the early 1940s during World War II while he was in the Navy [see Robbins (1985), pages 8–10]. A few years later the bombing problem finally lead to his important articles published in the *Annals of Mathematical Statistics* in (1944) and (1945). The same problem was treated by Bronowski and Neyman (1945). Unlike their works which solved the problem probabilistically, here we are mainly interested in the problem from a statistical viewpoint.

Let Δ denote the target area where the missiles would fall. Assuming that the locations of landing for all missiles are independent of each other and follow a certain unknown distribution G over Δ, let $Y_1, \dots, Y_n$ denote these n landing points. Associated with each Y_i there is a covered area $B(Y_i, r_i)$ which is the intersection of Δ and a disk with center Y_i and radius r_i. Each r_i may depend upon Y_i, but r_i and r_j are assumed independent for different i, j. If we let $X_i = B(Y_i, r_i)$ and $g(X_i) = Y_i$ for all $1 \leq i \leq n$, it is clear that the current model is within the framework of our general coverage problem described in Section 2. The chance that the $(n+1)$th missile would land at the uncovered area can be written as $P(g(X_{n+1}) \notin S_n | S_n)$, where $S_n = S(X_1, \dots, X_n) = \bigcup_{i=1}^n \{X_i\}$. From Section 2, the $(n-1)$-estimate is $n^{-1}\Sigma_{j=1}^n[1 - I(Y_j \in S_{n-1,j})]$, where $S_{n-1,j} = \bigcup_{i \neq j}\{X_i\}$.

Let us define $n_1(S_n) = \#$ of $\{Y_j;\, Y_j \notin S_{n-1,j}\}$ for brevity, and the previous $(n-1)$-estimate can thus be written as $(n_1(S_n)/n)$. Applying the similar idea of Section 3.1 to the current case, we come up with a bias-reduced estimate

$$\frac{n_1(S_n) + (1/n)\Sigma_{j=1}^n\left[n_1(S_n) - n_1(S_{n-1,j})\right]}{n+1}, \tag{3.3.1}$$

where

$$n_1(S_{n-1,j}) = \# \text{ of } \{Y_j;\, Y_j \notin S_{n-1,ji}\}, \qquad S_{n-1,ji} = \bigcup_{h \neq j,\, h \neq i} \{X_h\}.$$

To estimate the size of the newly covered area by the $(n+1)$th missile, it is easy to deduce from (ii) in Section 2 that the $(n-1)$-estimate is $(1/n)\Sigma_{j=1}^n \operatorname{vol}[X_j \setminus S_{n-1,j}] = (v_1(S_n)/n)$ (say), where

$$v_1(S_n) = \sum_{j=1}^n \operatorname{vol}\left[X_j \setminus S_{n-1,j}\right].$$

Similarly, one can deduce a bias-reduced estimate which is

$$\frac{v_1(S_n) + (1/n)\Sigma_{j=1}^n\left[v_1(S_n) - v_1(S_{n-1,j})\right]}{n+1}, \tag{3.3.2}$$

where $v_1(S_{n-1,j}) = \Sigma_{i \neq j}^n \operatorname{vol}[X_i \setminus S_{n-2,ji}]$ and $S_{n-2,ji} = \bigcup_{h \neq i,\, h \neq j}\{X_h\}$.

4. Extending the $(n-1)$-construction to the conditional cases. In this section we shall extend our $(n-1)$-construction to the conditional situation when data consist of n pairs $\{(X_i, Y_i), 1 \le i \le n\}$; that is to say, associated with each X_i a real Y_i (or a real vector) is observed. The issue of concern here is to estimate the conditional parameter associated with the observations. The following questions that arose from motor accidents may explain our concern. Suppose that n motor accidents were experienced by $m(m < n)$ drivers last year in a city. The insurance company is interested in (i) the age distribution among those drivers who were accident-free last year, (ii) the average damage loss made by the drivers who had two or more accidents in a year.

Let $S_n(\xi_n)$ be defined as in Section 2. We are interested in estimating a random parameter which can be expressed as $\theta = \theta(P_Y(\cdot|S_n))$—a smooth function of $P_Y(\cdot|S_n)$, where $P_Y(\cdot|S_n)$ is a conditional probability of Y_{n+1} given X_{n+1} is in $S_n(\xi_n)$. This conditional probability can be written precisely as

$$P_Y(E|S_n) = \frac{P\{Y_{n+1} \in E \text{ and } X_{n+1} \in S_n(\xi_n)\}}{P\{X_{n+1} \in S_n(\xi_n)\}} \tag{4.1}$$

for every Borel set E. To estimate this probability, we estimate the numerator and denominator separately, and this leads to an estimator $\hat{P}_Y(\cdot|S_n)$ which is defined as

$$\hat{P}_Y(E|S_n) = \frac{\sum_{j=1}^{n} I(Y_j \in E \text{ and } X_j \in S_{n,j}(\xi_{n,j}))}{\sum_{j=1}^{n} I(X_j \in S_{n,j}(\xi_{n,j}))}. \tag{4.2}$$

To estimate $\theta(P_Y(\cdot|S_n))$, we simply use $\theta(\hat{P}_Y(\cdot|S_n))$. In the species problem, suppose we want to estimate the conditional mean of Y among all unobserved outcomes, that is, $\theta = \int y\, dP_Y(y|S_n)$, where $S_n = \{j;\ j \notin \{X_1, \dots, X_n\}\}$, $P_Y(E|S_n) = \sum_{y_j \in E} p_j \psi_j(0; n) / \sum_{j=1}^{\infty} p_j \psi_j(0; n)$ and E is any Borel set in R (or in R^k if y is a vector in R^k). The conditional distribution of $P(E|S_n)$ can thus be written as $F(y|S_n) = P_Y((-\infty, y]|S_n)$ if $\{y_i\}$ are real-valued. Applying (4.2) to this example, we end up with an estimator

$$\hat{\theta}_n = \int y\, d\hat{P}_Y(y|S_n) = \sum_{j=1}^{n} I(X_j \in S_{n-1,j}) Y_j / n_1, \tag{4.3}$$

simply the sample mean of the corresponding observations which occur only once in the sample. Applying a similar idea to the conditional median of $\{y_j;\ j \notin \{X_1, \dots, X_n\}\}$, we obtain our estimator which is the sample median of Y_i of which the corresponding X_i occurs only once in the sample. In general, just how good $\hat{\theta}$ is as an estimator of θ depends upon the magnitude of $P\{X_{n+1} \in S_n(\xi_n)\}$, which is estimated by $(1/n)\sum_{j=1}^{n} I(X_j \in S_{n-1,j})$. The following two propositions provide some partial support of these estimators in terms of species problem. Let $F_n(y) = F(y|S_n)$ and $\hat{F}_n(y) = \hat{F}(y|S_n) = \#$ of $\{e_i;\ e_i$ appears once in $\{X_1, \dots, X_n\}$ and $y_i \le y\}/n_1$.

PROPOSITION 4.1. *If* $EY^2 < \infty$, $n^{1/2}(\sum_i p_i(1-p_i)^{n-1}) \to \infty$ *and* $\theta_n = \int y\,dP(y|S_n)$ *stay bounded in probability, then* $\hat{\theta}_n - \theta_n \to 0$ *in probability as* $n \to \infty$.

PROPOSITION 4.2. *If*

$$n^{1/2}\left(\sum_i p_i(1-p_i)^{n-1}\right) \to \infty \quad \text{as } n \to \infty,$$

then $\sup_y|\hat{F}_n(y) - F_n(y)| \to 0$ *in probability.*

The condition $n^{1/2}(\sum_i p_i(1-p_i)^{n-1}) \to \infty$ which appears in both propositions is equivalent to the condition $n^{-1/2}En_1 \to \infty$, which simply says that the unobserved probability can not be too small [smaller than $O(n^{-1/2})$] in order to ensure the validity of the estimator. The proofs of the propositions depend on the following three lemmas, of which the proofs are omitted.

LEMMA 1. *If* $k \geq 2$ *and* $E|Y| < \infty$, *then*

$$\sum_i p_i^k(1-p_i)^n = O(n^{-(k-1)}) \quad \text{and} \quad \sum_i p_i^k(1-p_i)^n y_i = O(n^{-(k-1)}).$$

LEMMA 2. *If* $n^{1/2}(\sum_i p_i(1-p_i)^n) \to \infty$ *and* $EY^2 < \infty$, *then*

$$\frac{(n_1/n) - \sum_i p_i(1-p_i)^n}{\sum_i p_i(1-p_i)^n} = o_p(1),$$

$$\frac{\sum_i p_i\psi_i(0;n) - \sum_i p_i(1-p_i)^n}{\sum_i p_i(1-p_i)^n} = o_p(1).$$

LEMMA 3. *If* $n^{1/2}(\sum_i p_i(1-p_i)^n) \to \infty$ *and* $EY^2 < \infty$, *then*

$$E\left[\frac{1}{n}\sum_{j=1}^{n} I(X_j \in S_{n-1,j})(Y_j) - \sum_i p_i\psi_i(0;n)y_i\right]^2 = O(n^{-1}).$$

PROOF OF PROPOSITION 4.1. Rewrite $\int y\,d\hat{P}(y|S_n) - \int y\,dP(y|S_n)$ as

$$D_n = \frac{b_n + \delta_n}{a_n + \mathscr{E}_n} - \frac{b_n}{a_n},$$

where $a_n = \sum_i p_i\psi_i(0;n)$, $b_n = \sum_i p_i\psi_i(0;n)y_i$, $\delta_n = (1/n)\sum_i \Psi_{i,n}y_i - b_n$, $\mathscr{E}_n = (n_1/n) - a_n$ and

$$\Psi_{i,n} = \begin{cases} 1, & \text{if } e_i \text{ appears exactly once in the sample,} \\ 0, & \text{otherwise.} \end{cases}$$

Further, we can write D_n as

$$D_n = \frac{a_n\delta_n - b_n\mathscr{E}_n}{(a_n + \mathscr{E}_n)a_n} = \frac{\delta_n}{a_n + \mathscr{E}_n} - \frac{b_n\mathscr{E}_n}{(a_n + \mathscr{E}_n)a_n}. \tag{4.4}$$

By Lemma 2, we have $n^{1/2}(n_1/n) \to \infty$ in probability. Since $\delta_n = O_p(n^{-1/2})$ by Lemma 3 and $\mathscr{E}_n = o_p(a_n)$ by Lemma 2, the proposition follows from the fact that

$$\frac{\delta_n}{a_n + \mathscr{E}_n} = o_p(1) \quad \text{and} \quad \frac{\mathscr{E}_n}{a_n + \mathscr{E}_n}\frac{b_n}{a_n} = o_p(1)$$

since $b_n a_n^{-1} = O_p(1)$ by assumption. □

PROOF OF PROPOSITION 4.2. It is easy to see $\hat{F}_n(y)$ can be written as $[\Sigma_i \Psi_{i,n} I(y_i \le y)][\Sigma_i \Psi_{i,n}]^{-1}$. From this, one can check that

$$E\left(\frac{1}{n}\sum_i \Psi_{i,n} I(y_i \le y)\right) = \sum_i p_i(1 - p_i)^{n-1} I(y_i \le y)$$

and

$$E\left(\sum_{y_i \le y} p_i \psi_i(0; n)\right) = \sum_i p_i(1 - p_i)^{n-1} I(y_i \le y).$$

From Lemma 1, it is easy to see

$$\sup_y \left|\sum_i p_i(1 - p_i)^{n-1} I(y_i \le y) - \sum_i p_i(1 - p_i)^n I(y_i \le y)\right| = O(n^{-1}).$$

Furthermore, with a similar argument as in Lemma 3, one can show that

$$nE\left[\frac{1}{n}\sum_i \Psi_{i,n} I(y_i \le y) - \sum_i p_i \psi_i(0; n) I(y_i \le y)\right]^2 < M < \infty$$

for some positive M, independent of y. The proposition is thus proved. □

Acknowledgments. I wish to thank Professors Herman Chernoff, Frederick Mosteller, Arthur Dempster, Herbert Robbins, Donald Rubin, B. H. Juang and Persi Diaconis for the constructive comments they made during this study. I owe a special thanks to Arthur Cohen for his very helpful suggestions. The detailed comments made by the Associate Editor and a referee improved the presentation of the paper.

REFERENCES

BAHL, L. R., JELINEK, F. and MERCER, R. C. (1983). Maximum likelihood approach to continuous speech recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **5** 179–190.

BANERJEE, P. K. and SINHA, B. K. (1985). Optimal and adaptive strategies in discovering new species. *Sequential Anal.* **4** 111–122.

BICKEL, P. J. and YAHAV, J. A. (1986). On estimating the total probability of the unobserved outcomes of an experiment. In *Adaptive Statistical Procedures and Related Topics* (J. Van Ryzin, ed.) 332–337. IMS, Hayward, Calif.

BRONOWSKI, J. and NEYMAN, J. (1945). The variance of the measure of a two-dimensional random set. *Ann. Math. Statist.* **16** 330–344.

BUCHTA, C. (1984). Stochastiche approximation konvexer polygone. *Z. Wahrsch. Verw. Gebiete* **67** 283–304.

CLAYTON, M. and FREES, E. (1987). Nonparametric estimation of the probability of discovering a new species. *J. Amer. Statist. Assoc.* **82** 305–311.

COHEN, A. and SACKROWITZ, H. (1990). Admissibility of estimators of the probability of unobserved outcomes. *Ann. Inst. Statist. Math.* **42** 623–636.

DEMPSTER, A., LAIRD, N. and RUBIN, D. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

DIACONIS, P. and STEIN, C. (1983). Decision theory. Lecture notes, Stanford Univ.

EDDY, W. F. and GALE, J. D. (1981). The convex hull of a spherically symmetric sample. *Adv. in Appl. Probab.* **13** 751–1260.

EFRON, B. (1965). The convex hull of a random set of points. *Biometrika* **52** 331–334.

EFRON, B. (1967). The two sample problem with censored data. In *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **4** 831–853. Univ. California Press, Berkeley.

EFRON, B. and THISTED, R. (1976). Estimating the number of unseen species: How many words did Shakespeare know? *Biometrika* **63** 435–447.

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

GEFFROY, J. (1959). Contribution à la théorie des valeurs extrémes. *Publ. Inst. Statist. Univ. Paris* **8** 123–185.

GEFFROY, J. (1961). Localization asymptotique du polyèdre d'appui dún échantillon Laplacien à k dimensions. *Publ. Inst. Statist. Univ. Paris* **10** 212–228.

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

GOOD, I. J. and TOULMIN, G. (1956). The number of new species and the increase in population coverage, when a sample is increased. *Biometrika* **43** 45–63.

GROENEBOOM, P. (1988). Limit theorems for convex hulls. *Probab. Theory Related Fields* **79** 329–368.

JELINEK, F. (1976). Continuous recognition by statistical method. *Proceedings of the IEEE* **64** 532–556.

KATZ, S. M. (1987). Estimation of probability from sparse data for the language model component of speech recognizer. *IEEE Trans. Acoust. Speech Signal Process.* **35** 400–401.

RAYNAUD, H. (1970). Sur l'enveloppe convexe des nuages de points aléatoires dans R^n. *J. Appl. Probab.* **7** 35–48.

RÉNYI, A. and SULANKE, R. (1963). Über die konevxe Hülle von n zufällig gewählten punkten. *Z. Wahrsch. Verw. Gebiete* **2** 75–84.

ROBBINS, H. (1944). On the measure of a random set. *Ann. Math. Statist.* **15** 70–74.

ROBBINS, H. (1945). On the measure of a random set, II. *Ann. Math. Statist.* **16** 342–347.

ROBBINS, H. (1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 137–163. Univ. California Press, Berkeley.

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

ROBBINS, H. (1985). *Herbert Robbins Selected Papers* (T. L. Lai and D. Siegmund, eds.). Springer, New York.

SCHNEIDER, R. (1987). Random approximation of convex sets. Math. Inst., Albertludwigs Univ.

STARR, N. (1979). Linear estimation of the probability of discovering a new species. *Ann. Statist.* **7** 644–652.

DEPARTMENT OF STATISTICS
MATH. 618, BOX 10 MATHEMATICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027