Nonparametric Inference Under Biased Sampling
from a Finite Population

P.J. Bickel
University of California, Berkeley

V.N. Nair
AT&T Bell Laboratories, Murray Hill

and

P.C.C. Wang
University of Calgary

Nonparametric Inference Under Biased Sampling from a Finite Population

P. J. Bickel

University of California, Berkeley

V. N. Nair

AT&T Bell Laboratories, Murray Hill

and

P. C. C. Wang

University of Calgary

ABSTRACT

We consider a biased sampling model that has been found useful in incorporating size-biases inherent in many types of discovery data. The model postulates that the data are obtained from a finite population by selecting successively without replacement and with probability proportional to some measure of size. Unlike the *ppswor* scheme in survey sampling, we assume here that the size measure is a function of the unknown population values. In this article, we consider maximum likelihood estimation of the finite population parameters under this biased sampling model. We study the large sample behavior of the MLE's and derive a simple, asymptotically efficient approximation to the MLE. The approximate MLE is structurally similar to the Horvitz-Thompson estimator. We show that information about the order in the sample can be used to make inference even when the population size is unknown, which in fact can be estimated. Small sample behavior of the estimators are investigated through a limited simulation study, and the results are used to analyze oil and gas discovery data from the North Sea Basin.

Acknowledgements

Nonparametric Inference Under Biased Sampling from a Finite Population

P. J. Bickel

University of California, Berkeley

V. N. Nair

AT&T Bell Laboratories, Murray Hill

and

P. C. C. Wang

University of Calgary

1. Introduction

Let $\mathcal{U} = \{U_1, \ldots, U_N\}$ denote a finite population of N units, and let Y_j be a characteristic associated with U_j, $j = 1, \ldots, N$. Let $\mathcal{S}_n = (U_{i_1}, \ldots, U_{i_n})$ be an ordered sample of size n that is selected successively without replacement and with probability proportional to some measure of size $\{w_1, \ldots, w_N\}$. More specifically,

$$P\{(U_{i_1}, \ldots, U_{i_n})\} = \prod_{j=1}^{n} \frac{w_{i_j}}{\sum\limits_{i=1}^{N} w_i - \sum\limits_{k=1}^{j-1} w_{i_k}}, \tag{1.1}$$

where $w_j = w(Y_j)$, a positive function of the unknown population characteristic, and $w_{i_0} \equiv 0$. A special case of this model is the *ppswor* scheme in survey sampling where the w_j's are auxiliary attributes whose values are known for all the units in the finite population. We are interested in the general case where the selection probabilities can depend on the a priori unknown population characteristics (which are in fact the parameters of interest). In this case, the w_j's associated with the $(N-n)$ unobserved units will not be known with certainty even after the sample is observed.

The sampling design (1.1) where the selection probabilities depend on the parameters of interest is sometimes known as an informative design (Cassel, Sarndal and Wretman, 1977). It has been used by Cozzolino (1972) and Barouch and Kaufman (1976) as a

model for analyzing oil and gas discovery data. They considered the case $w(y) = y$ and suggested the *ppswor* successive sampling scheme as a mechanism for incorporating the size-bias inherent in the discovery process. Bloomfield *et al.* (1979) and Lee and Wang (1983) extended their work by taking $w(y) = y^\theta$. The parameter θ is called the coefficient of discoverability in the petroleum estimation literature. Expression (1.1) can also be interpreted as a (marginal) likelihood of suitably defined rank statistics under a proportional hazards model (Kalbfleisch and Prentice, 1980). This relationship will be considered in more detail elsewhere.

Throughout this paper, we shall ignore any information in the labels $U_1, ..., U_N$. We can then represent the ordered sample $\mathcal{S}_n$ by its attributes which we denote by $(y_1, ..., y_n)$. Let $z_1, ..., z_K$ denote the distinct values in the finite population with multiplicities $N_1, ..., N_K$. We are interested in estimating the parameters $N_1, ..., N_K$ from an observed, ordered sample of size n based on (1.1).

Statistical inference under the model (1.1) has been considered by various authors. Most of these results were developed under the assumption that a population characteristic such as the population size N or the population total $R = \sum_{k=1}^{K} N_k z_k$ is known. Barouch and Kaufman (1977), Barouch, Kaufman and Nelligan (1983), Lee and Wang (1983, 1985), and Nair and Wang (1989) developed parametric estimators of the distribution under a superpopulation framework where the finite population itself is assumed to be an *iid* sample from some underlying population. Gordon (1989) and Andreatta and Kaufman (1986) discussed inference procedures for the finite population itself and considered, among other things, Horvitz-Thompson type estimators. Gordon (1989) also provided an approach for estimating the population size by moment-type estimator using a split-sample technique. Andreatta and Kaufman (1986) considered the situation where the information

about the order in the sample is either unavailable or ignored. They discussed the use of Murthy's (1957) estimator assuming that R or some other population parameter such as N is known. See also Godambe and Rajashari (1989) for a related biased sampling problem with Bernoulli sampling.

The present paper deals with maximum likelihood estimation. We show that it is possible to do inference without assuming any knowledge of the population size or other population characteristics. We can in fact estimate these characteristics through maximum likelihood. It is the information in the *ordered* sample that allows us to do this unusual inference. Maximum likelihood estimation of the finite population was considered in Smith and Ward (1981) but they did not get any explicit results and did not study the properties of the MLE's.

In Section 2, we study MLE's of $\{N_1, ..., N_K\}$. In particular, we indicate how parameters such as N and R can be estimated. The asymptotic behavior of the MLE suggests a simple, asymptotically equivalent estimator that is easily computed. This estimator is structurally similar to the Horvitz-Thompson estimator, but it depends on the order in which the data in the sample $\mathscr{S}_n$ are observed. Consistency, asymptotic distribution and efficiency of the MLE are also discussed. Proofs are deferred to Section 6.

Section 3 compares the asymptotic and finite-sample behavior of the MLE's through a limited simulation study. In Section 4, the results are applied to the analysis of oil and gas discovery data from the North Sea Basin. Section 5 develops some large sample results of estimators under the *ppswor* model. These are extensions of the results in Holst (1973), Sen (1979), and Gordon (1989). Section 6 deals with some auxiliary results and proofs of the theorems in Section 2. For the asymptotic results, we assume that K, the number of distinct values in the finite population, is fixed as $N \rightarrow \infty$. The more general case where $K = K(N) \rightarrow \infty$ will be considered elsewhere. We also assume that the sampling proportion

$f_N = n/N$ satisfies $0 < \lim_{N \to \infty} f_N < 1$. If $f_N = 1$, all the N units of the finite population have been observed, so there is nothing to infer. If $f_N \to 0$ (n fixed and $N \to \infty$), sampling without replacement is equivalent to sampling with replacement, and the model in (1.1) simplifies to the familiar length-biased estimation problem considered in the literature (Cox, 1964; Patil and Rao, 1977; and Vardi, 1982).

In practice, the size-bias mechanism $w(y)$ is unknown and may have to be estimated from the data. One approach is to parameterize the shape of $w(y)$, for example by letting $w(y) = e^{\beta y}$, and to estimate all the unknown parameters simultaneously. However, for our results we will assume that $w(y)$ is a fixed, known, positive function.

2. Estimation

Let $z_1, \ldots, z_K$ be the distinct values of the finite population with multiplicities $N_1, \ldots, N_K$. In this section, we consider nonparametric maximum likelihood estimation of $\{N_1, \ldots, N_K\}$, and hence of the population size N.

If the z_k's are known, the problem of estimating $\{Y_1, \ldots, Y_N\}$ is equivalent to estimating $\{N_1, \ldots, N_K\}$. In general, of course, the z_k's will not be known a priori. However, as we shall see below, the nonparametric MLE gives nonzero mass only to the z_k's which are observed in the sample. Therefore, we can restrict attention to the z_k's which are the distinct values in the observed sample.

2.1 Notation

Throughout the paper, we will use the following notation. Let

$$I_j^n = I[U_j \in \mathcal{S}_n], \tag{2.1}$$

the indicator of the event that the jth unit is included in the sample $\mathcal{S}_n$ of size n. Let

$z_1, \ldots, z_K$ be the distinct values of Y_j in the observed sample, and let $I[Y_j = z_k] = 1$ if $Y_j = z_k$ and 0 otherwise. Then, $n_k = \sum\limits_{j=1}^{N} I[Y_j = z_k] I_j^n$ denotes the multiplicity of $z_k, k = 1, \ldots, K$. In addition, let $w_k = w(z_k)$,

$$\pi_k(t) = 1 - e^{-t w_k}, \qquad (2.2)$$

and

$$\bar{\pi}_k(t) = 1 - \pi_k(t) = e^{-t w_k}, \qquad (2.3)$$

for $k = 1, \ldots, K$. Let

$$f = n/N, \qquad \theta_k = N_k/N, \quad k = 1, \ldots, K, \qquad (2.4)$$

be, respectively, the sampling fraction and the cell probabilities. It will also be useful to consider the quantities

$$v_k = N_k/n, k = 1, \ldots, K. \qquad (2.5)$$

Note that

$$f^{-1} = \sum\limits_{k=1}^{K} v_k. \qquad (2.6)$$

2.2 Maximum Likelihood Estimation

Define

$$D(i) = \sum\limits_{j=0}^{i-1} w(y_j), \qquad i = 1, \ldots, n, \qquad (2.7)$$

with $w(y_0) = 0$. The likelihood of $N = (N_1, \ldots, N_K)$ is obtained from (1.1) as

$$e^{L(N)} = \prod\limits_{k=1}^{K} \frac{N_k!}{(N_k - n_k)!} \prod\limits_{i=1}^{n} \frac{w(y_i)}{\sum\limits_{r=1}^{K} N_r w_r - D(i)}. \qquad (2.8)$$

The combinatorial coefficients indicate the number of different ways in which the ordered sample could have been obtained. It is easily seen from (2.8) that the nonparametric MLE gives nonzero mass only to the z_k's for which $n_k > 0$.

To obtain the MLE $\{\hat{N}_k\}_{k=1}^K$, one can maximize the likelihood (2.8) numerically subject to the constraint $\hat{N}_k \geq n_k$, $k = 1, \ldots, K$. Note that the solution can occur on the boundary, i.e., $\hat{N}_k = n_k$ for some k. The computation of the MLE involves maximization over a K-dimensional space and can be difficult in practice. In the next section, we obtain a simple estimator that is asymptotically equivalent to the MLE. But first we study the large sample behavior of the likelihood equations. This is done within the following asymptotic framework. To avoid cumbersome notation, we will suppress the index N from the subscripts of f_N, z_{Nk}, ν_{Nk} and so on. *But one should remember that the various quantities considered depend on N.*

A1 As $N \to \infty$, the number of distinct values in the finite population, K, is fixed, and the distinct values $\{z_1, \ldots, z_K\}$ converge to some fixed, finite values.

A2 $\theta_k = N_k/N$ satisfies $0 < \lim_{N \to \infty} \theta_k < 1$ for $k = 1, \ldots, K$, and $f_N = n/N$ satisfies

$$0 < \lim_{N \to \infty} f_N < 1.$$

The more general situation where $K = K(N) \to \infty$ as $N \to \infty$ will be considered elsewhere.

As $N \to \infty$, the parameters of interest $N_k \to \infty$ for $k = 1, \ldots, K$. So we will work with the parameters $\{\theta_k\}_{k=1}^{K-1}$ and f defined by (2.4) or the equivalent set of normalized parameters $\{\nu_k\}_{k=1}^K$ defined by (2.5). The observed sample proportions will be denoted by

$$p_k = \frac{n_k}{n}, k = 1, \ldots, K. \tag{2.9}$$

Our basic approach is to replace factorials by gamma functions in the likelihood in (2.8). Our parameters can then be treated as continuous and we can differentiate the log-

likelihood and obtain the likelihood equations. We argue that MLE's (more precisely first local MLE's to be defined below) satisfy these equations to a high enough order of approximation. Then we deduce their asymptotic theory by the usual linearization process.

With some abuse of notation, we will let $L(v)$ denote the log-likelihood of v. From (2.8),

$$L(v) = \text{const} + \sum_{k=1}^{K} \left\{ \log \Gamma(n v_k + 1) - \log \Gamma(n(v_k - p_k) + 1) \right\}$$

$$- \sum_{i=1}^{n} \log \left[\sum_{r=1}^{K} v_r w_r - n^{-1} D(i) \right]. \tag{2.10}$$

We will use the notations v_0, f_0, etc. to distinguish the true population parameters from the argument of the likelihood function. Note that these values also vary with N since we are considering a sequence of finite populations for our asymptotic framework. But we will continue to suppress their dependence on N for notational convenience. The following result is known (see Rosén, 1972; Holst, 1973; and Gordon, 1989) and can be obtained as a consequence of Corollary 5.5. For $1 \leq k \leq K$,

$$| p_k - v_{k0} \pi_k(\lambda_0) | = o_p(1) , \tag{2.11}$$

where $\lambda_0 = \lambda(f_0)$, and $\lambda(u)$ satisfies

$$N^{-1} \sum_{k=1}^{K} N_{0k} e^{-\lambda(u) w_k} = \sum_{k=1}^{K} \theta_k \bar{\pi}_k(\lambda(u)) = 1 - u , \quad 0 < u < 1. \tag{2.12}$$

Note that $\lambda(u)$ is an increasing, convex function of u; so there is a one-to-one relationship between the sampling proportion f_0 and the parameter λ_0.

We see from (2.11) that, if λ_0 is known, one can estimate v_{k0} consistently by

$$\hat{v}_k = \frac{p_k}{\pi_k(\lambda_0)} = \frac{p_k}{1 - e^{-\lambda_0 w_k}}. \tag{2.13}$$

When we replace the unknown inclusion probabilities, $1 - e^{-\lambda_0 w_k}$, by their asymptotic approximations, this yields the Horvitz-Thompson estimator. We will see in the next section that the approximate maximum likelihood estimator also has this structure but now the data are used to estimate the unknown parameter λ_0.

The likelihood in (2.10) can, in general, have several local maxima. We shall consider the "first" local maximum defined as follows. Reparameterize v by (θ, f) given by (2.4), and (with some abuse of notation) write $L(\theta, f)$ for $L(v)$. For each fixed f, the θ_k's vary over a non-empty subset of the simplex, and $L(\theta, f)$ is continuous in θ. So, for a given f, $L(f) = \max_\theta \{L(\theta, f) : \theta_k \geq f p_k, k = 1, ..., K\}$ is achieved by some $\hat{\theta}(f)$. We call $L(f)$ a "profile log-likelihood" of f. Let $\mathcal{F}$ be the set of all local maximizers of $L(f)$ in $[0, 1]$. Note that for $f \in (0, 1)$, f is a local maximizer if $L'(f) = 0$, and $L''(f) < 0$. At the boundaries, $f = 0$ is a local maximizer if $L'(0^{plus}) < 0$, and similarly, $f = 1$ is a local maximizer if $L'(1^-) > 0$. We consider the "first local maximizer" of $L(f)$ given by $\hat{f} = \min\{f \in \mathcal{F}\}$. Since $L(f)$ is continuous in f, the set $\mathcal{F}$ is non-empty and $\hat{f}$ exists. Let $\hat{v}$ be the value of v that corresponds to $(\hat{\theta}(\hat{f}), \hat{f})$. We establish the consistency and asymptotic normality of $\hat{v}$ in Theorem 1 below. We consider the first local maximizer rather than the global maximizer because of technical constraints which require f to be bounded away from 1 in our proofs. It may well be that the global MLE is also consistent but we have not shown this.

Theorem 1:   Under A1−A2,

   i) $\hat{v}$ is consistent;

and   ii) $\mathcal{L}(n^{1/2}(\hat{v} - v_0)) \to s N(0, I^{-1}(v_0))$,

where

$$I^{-1}(v_0) = \|\delta_{jk}d_{k0} + \frac{w_j d_{j0} w_k d_{k0} f_0 \int_0^{f_0} [\lambda'(u)]^2 \, du}{\left(1 - f_0 \int_0^{f_0} [\lambda'(u)]^2 \, du \sum_{r=1}^{K} w_r^2 d_{r0}\right)}\|_{K \times K}, \qquad (2.14)$$

$$d_{k0} = \frac{v_{k0} \bar{\pi}_k(\lambda_0)}{\pi_k(\lambda_0)}, \qquad (2.15)$$

and from (2.12),

$$\lambda'(u) = \frac{d\lambda(u)}{du} = \left(\sum_{k=1}^{K} w_k \theta_k \bar{\pi}_k(\lambda(u))\right)^{-1}. \qquad (2.16)$$

Remark 1: Since v_0 depends on N, the statement (ii) above (and those in the sequel) should be formally interpreted as $\mathcal{L}\left(n^{\frac{1}{2}} I^{\frac{1}{2}}(v_0) [\hat{v} - v_0]\right) \to_s N(0, I)$.

2.3 Approximate MLE

The estimator $\hat{v}$ can be difficult to compute in practice. We obtain an approximation that is easy to compute and is asymptotically equivalent to $\hat{v}$ as follows. Differentiating (2.10) with respect to the v_k's, we get the "likelihood equations" as

$$\frac{\partial L}{\partial v_k} = n \left[\psi(nv_k+1) - \psi(n[v_k-p_k]+1)\right] - w_k \sum_{i=1}^{n} \left(\sum_{r=1}^{K} v_r w_r - n^{-1} D(i)\right)^{-1} \qquad (2.17)$$

for $k = 1, \ldots, K$, where $\psi(\cdot)$ denotes the digamma function. Recall that $\psi(x+1) = \log x + O(x^{-1})$ as $x \to \infty$. So,

$$n^{-1} \frac{\partial L}{\partial v_k} = \left(\log \frac{v_k}{v_k - p_k} - w_k \lambda_n(v;y)\right) + O_p(n^{-1}), \qquad (2.18)$$

where $\lambda_n(v;y)$ is the data-dependent function

$$\lambda_n(v;y) = n^{-1} \sum_{i=1}^n \left(\sum_{k=1}^K v_k w_k - n^{-1} D(i) \right)^{-1}. \tag{2.19}$$

If we ignore the second term on the right-hand side in (2.18) which is of smaller order than the first term and solve for v by setting the first term equal to zero, we obtain an estimator implicitly defined by

$$v_k = \frac{p_k}{\pi_k(\lambda_n(v;y))}, \quad k=1,...,K. \tag{2.20}$$

Note the structural similarity between this and the Horvitz-Thompson estimator in (2.13).

Since $\lambda_n(v;y)$ in (2.19) depends on the unknown parameters, the approximate MLE (2.20) is still not easily computable. However, the v_k's are easily obtained once $\lambda_n(v;y)$ is determined. So let

$$v_k(\lambda) = \frac{p_k}{\pi_k(\lambda)} = \frac{p_k}{1-e^{-\lambda w_k}}, \quad k=1,...,K. \tag{2.21}$$

Substitute $v_k(\lambda)$ in the log-likelihood (2.10), treat λ as an unknown parameter, and consider the "profile log-likelihood" of λ (which we denote as $L_{PR}(\lambda)$). The derivative of this likelihood with respect to λ is

$$\frac{dL_{PR}(\lambda)}{d\lambda} = -n \sum_{k=1}^K \frac{p_k w_k \bar{\pi}_k(\lambda)}{[\pi_k(\lambda)]^2} \tag{2.22}$$
$$\times \left[\psi(n v_k(\lambda) + 1) - \psi(n[v_k(\lambda) - p_k] + 1) - w_k Z_n(\lambda) \right],$$

where

$$Z_n(\lambda) = n^{-1} \sum_{i=1}^n \left(\sum_{k=1}^K v_k(\lambda) w_k - n^{-1} D(i) \right)^{-1} \tag{2.23}$$

$$= n^{-1} \sum_{i=1}^n \left(\sum_{k=1}^K \frac{p_k w_k}{\pi_k(\lambda)} - n^{-1} D(i) \right)^{-1}.$$

Recall that $\psi(x+1) = \log(x) + O(x^{-1})$ as $x \to \infty$. So the term in square brackets in (2.22) reduces to $w_k[\lambda - Z_n(\lambda)] + O_p(n^{-1})$. Hence, setting the "profile log-likelihood" equation (2.22) equal to zero and solving for λ is equivalent to solving the equation

$$Z_n(\lambda) = \lambda. \tag{2.24}$$

It is now natural to estimate v by $v(\hat{\lambda})$ where $\hat{\lambda} > 0$ solves (2.24) and $Z_n'(\hat{\lambda}) < 1$. $\lambda = 0$ is always a solution of (2.24), but this corresponds to $f = 0$; we are interested in a strictly positive, finite solution.

We have thus reduced the problem of computing the MLE's (a K-dimensional problem) to finding an appropriate root of (2.24) — a one-dimensional problem. The asymptotic equivalence between the $v(\hat{\lambda})$ and the MLE $\hat{v}$ is formally stated in the following theorem.

Theorem 2:   Under A1-A2,

         i) $v(\hat{\lambda})$ is consistent;

         ii) $v(\hat{\lambda}) = \hat{v} + o_p(n^{-\frac{1}{4}})$;

and hence   iii) $n^{\frac{1}{4}}(v(\hat{\lambda}) - v_0)$ also has $N(0, I^{-1}(v_0))$ as a limit in law.

Remark 2: The asymptotic variance-covariance matrix $I^{-1}(v_0)$ can be estimated by substituting consistent estimates for the unknown parameters in (2.14). The right hand term in (2.16) can be used for estimating $\lambda'(u)$.

Remark 3: When the underlying population is continuous, equation (2.24) can still be used to estimate λ. Thus, parameters such as the cumulative distribution function, N, and R can also be estimated from (2.24) and (2.21) for this problem. Properties of these estimators will be investigated elsewhere.

Remark 4: As $\lambda \to 0$, it can be seen from (2.12) that the sampling proportion $f_0 \to 0$. For this limiting case, we can see from (2.21) that $\hat{v}_k$ is proportional to p_k/w_k. (Note that only the probabilities θ_k are meaningful in this case.) These are indeed the MLE's for this special case: $\hat{\theta}_k = p_k/\alpha w_k$ where α is a proportionality constant (see Cox, 1969 or Vardi, 1982).

2.4 Efficiency

We define efficiency in this context as follows. Let $\bar{v}$ be a competing estimator to $v(\hat{\lambda})$ such that

$$\mathcal{L}_{v_n}(n^{\frac{1}{2}}(\bar{v} - v_n)) \to sN(0, \Sigma_{v_0}) \tag{2.25}$$

for all sequences of populations with $v = v_n$ such that $n^{\frac{1}{2}}(v_n - v_0)$ is bounded and $v_0 \in R^+ \times \cdots \times R^+$. Then,

Theorem 3:

$$\Sigma_{v_0} \geq I^{-1}(v_0)$$

where $\geq$ is as usual defined by $A \geq B$ iff $A - B$ is nonnegative definite.

Alternative equivalent formulations of efficiency are a Hájek convolution theorem when the limit in (2.25) is not necessarily Gaussian and an asymptotic minimax theorem where all competitors are permitted but optimality is measured by the minimax risk in $n^{-\frac{1}{2}}$ neighborhoods of v_0.

3. A Simulation Study

In this section, we examine the finite sample behavior of the MLE's through a small simulation study. The distinct values of the finite population were taken as the quantiles of a standard exponential distribution $z_k = -\ln\left(1 - \dfrac{k-1/2}{5}\right)$, $k = 1, \ldots, 5$, with

multiplicities $N_k = 20$, $k = 1, ..., 5$, and hence $N = 100$. We generated 1000 samples of size $n = 50$ each from this population with *ppswor* according to the size measure $w(y) = y$. From each sample, we computed the approximate MLE's of N_k and θ_k based on $v_k(\hat{\lambda})$ in Section 2.

[Figure 1 about here.]

Figure 1 shows the box plots of the (approximate) MLE's of N_k, $k = 1, ..., 5$. The estimates are reasonably centered around the true values of 20. The variability in the distributions decreases as one goes from left to right, corresponding to increasing values of z_k's. This is intuitively reasonable because of the selection bias which results in more units being observed from cells with large z_k-values. The heavier upper tails are also be expected since the lower tails are bounded by the constraint $\hat{N}_k \geq n_k$, the observed frequencies.

[Table 1 about here.]

Table 1 compares the asymptotic errors of the estimators with the observed standard errors from the simulation. We see that the asymptotic formulas provide reasonable approximations to the true standard errors in almost all the cases. It is still possible, however, that interval estimators based on asymptotic normality may not do well because the finite sample distributions are quite nonnormal. This appears to be the case for the data in Figure 1. A referee has suggested the use of transformations to enhance normality and the quality of interval estimation. While this is possible, it is unlikely that a particular transformation, like the log-transformation, will work well in all cases. Our own preference is to use the bootstrap technique to obtain interval estimates in these situations.

[Figure 2 about here.]

Figure 2 shows the box plots of $\hat{\theta}_k$, $k = 1, ..., 5$. These distributions are fairly symmetrically distributed around the true values of 0.2. The variabilities do not exhibit a definite pattern as they did with Figure 1. Table 1 also shows that, for the $\hat{\theta}_k$'s, there is little difference between the actual standard errors and those obtained from the asymptotic approximations. Thus, the asymptotic normal approximation for $\hat{\theta}_k$ (or something like logit ($\hat{\theta}_k$)) is likely to provide reasonable interval estimates for θ.

4. Application

We shall now illustrate our results by applying them to oil and gas discovery data from the North Sea Petroleum Basin. The data, given in Table 2, are the estimated recoverable reserves (in millions of barrels of oil and gas equivalent) from 99 reservoirs that were discovered during the period 1967-1976. At the time of the last indicated discovery, drilling had not advanced far enough to exhaust the play. So, it was of interest to estimate the remaining amount of petroleum resources in this reservoir.

[Table 2 about here.]

Smith and Ward (1981) and Andreatta and Kaufman (1986) have modeled the North Sea data by using the successive sampling scheme (1.1) to approximate the discovery process. Unfortunately, their original data are unavailable and the data in Table 2 appear to differ slightly from theirs; so we cannot make a comparison of our estimates with theirs which are based on other estimation schemes.

[Table 3 about here.]

Following the previous analyses, we grouped the data into seven classes and replaced the data by the class mid-points. The class sizes, mid-points, and observed frequencies are given in Table 3. We estimated the parameters through the approximate MLE's in Section 2 with the weight function $w(y) = y$. The estimates of the cell multiplicities, N_k, and the cell proportions, θ_k, together with their estimated standard errors are given in Table 3. The population size N is estimated to be 404 with a standard error of 144. Most of the $404 - 99 = 305$ unobserved reservoirs are in the smallest class-size. As expected, the units in the larger class sizes have been mostly observed. The total amount of available resources from this pool is estimated to be $\hat{R} = 46{,}942$ million barrels with a standard error of 8,259. Since 31,925 million barrels have already been discovered (based on the grouped data), the remaining undiscovered resources are approximately 15,907 million barrels. Table 3 also gives the estimated cell proportions and their standard errors.

[Figure 3 about here.]

Although there are 99 discovered pools in Table 3, we can actually estimate the population parameters based on each sequential sample of size i, $2 \leq i \leq 99$. Figure 3 shows how the estimates of N vary with time as more discoveries become available. There is initially an increase, from an estimate of 268 based on the first two discoveries, to around 2600. But after the first 30 discovered pools, the estimates become quite stable and vary only from about 300 to 500.

5. Related Asymptotic Theory

In this section, we develop some results on the asymptotic behavior of estimators under *ppswor* sampling. Our results are in part restatements and in part extensions of the results of Rosén (1972), Holst (1973), and Gordon (1989). While the conditions we give can be

considerably weakened, they are adequate for our purposes.

Consider a successive sampling scheme *with* replacement and with probability of selection proportional to the size measures $\{w_j\}_{j=1}^N$. Let $X_j(m)$ be the number of times U_j is selected in such a *ppswr* scheme of size m, and let

$$e_j(m) = \min(X_j(m), 1), \tag{5.1}$$

the indicator of the event that U_j is selected. Further, let $\bar{w}_j = w_j / \sum_{r=1}^N w_r$, the normalized weights, and $e_j^c(m) = 1 - e_j(m)$. Evidently under the *ppswr* scheme,

$$E\, e_j^c(m) = (1 - \bar{w}_j)^m. \tag{5.2}$$

Let $N(t)$ be a homogeneous Poisson process independent of everything with intensity N, i.e. $E(N(t)) = Nt$. If we define $X_j(N(t))$ in the obvious manner, then it is easy to see that $X_j(N(t))$ are independent Poisson $(Nt\bar{w}_j)$. Therefore, $e_j^c(N(t))$, $1 \le j \le N$ are independent. Further, let $[x]$ denote the greatest integer function, and define

$$M_N(t) = \inf\{s : \sum_{j=1}^N e_j(N(s)) = [Nt]\}. \tag{5.3}$$

Then, evidently $\left\{e_j(M_N(n))\right\}_{j=1}^N$ have the same distribution as $\{I_j^n\}_{j=1}^N$ given by (2.1). This embedding of ppswor in a Poisson sampling scheme also leads immediately to Gordon's (1989) formula. See also Holst (1986). Let T_j be the waiting time to the first "event" in the Poisson process $X_j(N(s))$, $1 \le j \le N$. The T_j are independent exponential $(N\bar{w}_j)$ rv's. The probability of obtaining a particular sequence $U_{j_1}, ..., U_{j_n}$ of objects in the ppswor scheme with $n = N$ is simply

$$P[T_{j_1} < T_{j_2} < ... < T_{j_n}] = \prod_{i=1}^{n} \frac{\bar{w}_{j_i}}{1 - \sum_{r=1}^{i-1} \bar{w}_{j_r}},$$

and this is Gordon's (1989) formula.

Given a function $g_N : [0,1] \to R$, define a stochastic process on $[0,\infty)$ by

$$W_N(t) = N^{-\frac{1}{2}} \sum_{j=1}^{N} g_N(j/N)(e_j^c([Nt]) - (1 - \bar{w}_j)^{[Nt]}). \qquad (5.4)$$

It is convenient to also think of the w_j as being defined by

$$w_j = w_N(j/N)$$

where $w_N:[0,1] \to R^+$. We shall assume:

R1: The w_N are bounded and

$$w_N(t) \to w(t) > 0 \quad \text{on } [0,1]$$

so that if $\bar{w}_N(t) = w_N(t)/[N^{-1} \sum_{j=1}^{N} w_j]$, then

$$\bar{w}_N(t) \to w(t)/\int_0^1 w(s)ds = \bar{w}(t) .$$

Note that under A2 we can take $\bar{w}(t) = w_k$ on an interval of length θ_k, $1 \le k \le K$ where the θ_k's are defined in (2.4).

R2: The g_N are bounded and

$$g_N(t) \to g(t) \quad \text{on } [0,1] .$$

Define

$$W_N^p(t) = W_N(N(t)/N) . \qquad (5.5)$$

From (5.4), we can write $W_N^p(t)$ as

$$W_N^P(t) = N^{-\frac{1}{2}} \sum_{j=1}^{N} g_N(j/N) \left\{ [e_j^c(N(t)) - e^{-N\bar{w}_j}] - [(1 - \bar{w}_j)^{N(t)} - e^{-N\bar{w}_j}] \right\}.$$

Under our conditions,

$$N(t) \log (1 - \bar{w}_j) = -\bar{w}_j N(t) + O_p(N^{-1}),$$

and

$$(1 - \bar{w}_j)^{N(t)} - e^{-N\bar{w}_j} = - e^{-N\bar{w}_j} \bar{w}_j[N(t) - Nt] + O_p(N^{-1}).$$

All remainders which are written in O_p form here are functions of the form $R_{jN}(t)$ such that $\sup_{j,t} |R_{jN}(t)|$ has the prescribed order. Therefore,

$$W_N^P(t) = N^{-\frac{1}{2}} \left\{ \bar{g}_N(t)[N(t) - Nt] + \sum_{j=1}^{N} g_N(j/N)[e_j^c(N(t)) - e^{-N\bar{w}_j}] \right\} + O_p(N^{-\frac{1}{2}}),$$

where

$$\bar{g}_N(t) = \sum_{j=1}^{N} g_N(j/N) \bar{w}_j e^{-N\bar{w}_j}.$$

Proposition 5.1: The processes W_N^P converge weakly in $D[0,T]$ to a Gaussian process W with mean 0 and, if $s < t$,

$$\text{cov}(W(s), W(t)) = \lim_N N^{-1} \left(\sum_{j=1}^{N} g_N^2(j/N) e^{-N\bar{w}_j t}(1 - e^{-N\bar{w}_j s}) - s\bar{g}_N(s)\bar{g}_N(t) \right). \quad (5.6)$$

Proof: Finite dimensional convergence follows from Lindeberg-Feller. Tightness of $W_N^P(t)$ can be inferred from the proof of Proposition 5.3, which is a more general version of this result. See also Sen (1979).

To verify (5.6), since

$$\bar{g}_N(t) = - \operatorname{cov}\left(\sum_{j=1}^{N} g_N(j/N)e_j^c(N(t)), \ N(t)\right) / Nt,$$

note that

$$\operatorname{cov}(W_N^P(t), N(t)) = 0,$$

and

$$\operatorname{cov}(W_N^P(s), N(t) - N(s)) = 0.$$

Hence,

$$\operatorname{cov}(W_N^P(s), W_N^P(t)) = N^{-1} \sum_{j=1}^{N} g_N^2(j/N)\operatorname{cov}(e_j^c(N(s)), e_j^c(N(t)))$$

$$+ N^{-1} \operatorname{cov}(\sum_{j=1}^{N} g_N(j/N)e_j^c(N(t)), N(s))$$

which yields (5.6). □

By our assumptions, if

$$\bar{g}(s) = \int_0^1 g(u) \, \bar{w}(u)e^{-s\bar{w}(u)} \, du,$$

then, for $s < t$,

$$\operatorname{cov}(W(s), W(t)) = \int_0^1 g^2(u) \, e^{-t\bar{w}(u)}(1 - e^{-s\bar{w}(u)}) \, du - s \, \bar{g}(s)\bar{g}(t). \tag{5.7}$$

It is easy to see that the process W has continuous sample functions.

By Proposition 5.1

$$N^{-1} \sum_{j=1}^{N} e_j^c(N(t)) = N^{-1} \sum_{j=1}^{N} e^{-N t \bar{w}_j} + O_p(N^{-\frac{1}{2}})$$

uniformly on $[0, T]$. So, from (5.3),

$$M_N(t) = \tau_N(t) + o_p(1)$$

uniformly on $[0, T]$, where

$$N^{-1} \sum_{j=1}^{N} e^{-N\tau_N(t)\tilde{w}_j} = 1 - t. \tag{5.8}$$

Comparing (5.8) and (2.12), we see that $\tau_N(\cdot)$ is just a scaled version of $\lambda(\cdot)$. Under our assumptions,

$$\tau_N(t) = \tau(t) + o(1) \qquad \text{uniformly on } [0, T], \tag{5.9}$$

where

$$\int_0^1 e^{-\tau(t)\tilde{w}(x)} \, dx = 1 - t. \tag{5.10}$$

Since $(e_1(M_N(t)), \ldots, e_N(M_N(t)))$ are distributed as $\{I_j^{[Nt]}\}$, $j = 1, \ldots, N$, we obtain the basic result proved by Rosén (1972), Holst (1973), Sen (1979) and Gordon (1989) under different conditions. A more general version of this result is given in Proposition 5.3. In this section, we will let $\mathcal{L}$ denote weak convergence in $D[0, T]$.

Proposition 5.2: Let

$$V_N(t) = N^{-\frac{1}{4}} \sum_{j=1}^{N} g_N(j/N) \left(I_j^{[Nt]} - [1 - e^{-N\tau_N(t)\tilde{w}_j}] \right).$$

Then, under our assumptions, for all $T < \infty$,

$$V_N(\cdot) \xrightarrow{\mathcal{L}} W(\tau(\cdot)). \tag{5.11}$$

We actually need a slight extension of this result which is new. Let

$$h_N : [0, T] \times [0, 1] \rightarrow R,$$

and define

$$\tilde{W}_N(t) = N^{-\frac{1}{4}} \sum_{j=1}^{N} h_N(t, j/N) \left(e_j^c([Nt]) - (1 - \tilde{w}_j)^{[Nt]} \right),$$

and $\tilde{W}_N^p$, $\tilde{V}_N$ and $\bar{h}_N(t)$ analogously. Replace R2 by

R2′:

  a) $h_N(t,u) \to h(t,u)$

  b) $\sup_{N,t,u} |h_N(t,u)| \le M_1 < \infty$

  c) $\sup_{N,t,u} |\frac{\partial h_N}{\partial t}(t,u)| \le M_2 < \infty$

Proposition 5.3: Under $R1$, $R2′$ and $A1-A2$,

$$\tilde{W}_N^P \overset{\mathcal{L}}{\to} \tilde{W}, \tag{5.12}$$

$$\tilde{V}_N \overset{\mathcal{L}}{\to} \tilde{W}(\tau(\cdot)), \tag{5.13}$$

where, if $s \le t$,

$$\mathrm{cov}(\tilde{W}(s), \tilde{W}(t)) = \int_0^1 h(s,u)h(t,u)e^{-\tilde{w}(u)}(1 - e^{-s\tilde{w}(u)}) \, du - s\,\bar{h}(s)\bar{h}(t), \tag{5.14}$$

and

$$\bar{h}(s) = \int_0^1 h(s,u) \, \tilde{w}(u)e^{-s\tilde{w}(u)} \, du, \tag{5.15}$$

and τ is given by (5.10).

Proof: (5.13) follows immediately from (5.12) and (5.11). Finite dimensional convergence of $\tilde{W}_N^P$ and (5.13) follow from the Lindeberg-Feller theorem as in Proposition 5.1. For tightness we note that the scaled Poisson processes

$$N^{-\frac{1}{2}}\bar{h}_N(t) \, [N(t) - Nt]$$

are evidently tight. If $t_1 < t_2 < t_3$ and $\Delta_{jN}(t) = e_j^c(N(t)) - e^{-N\tilde{w}_j t}$, then

$$E\left[(\sum_{j=1}^N V_{1j})^2(\sum_{j=1}^N V_{2j})^2\right] \le \left[2(\sum_{j=1}^N E(V_{1j}V_{2j}))^2 + (\sum_{j=1}^N EV_{1j}^2)(\sum_{j=1}^N EV_{2j}^2)\right], \tag{5.16}$$

where

$$V_{1j} = h_N(t_2, j/N)\Delta_{jN}(t_2) - h_N(t_1, j/N)\Delta_{jN}(t_1),$$

and

$$V_{2j} = h_N(t_3, j/N)\Delta_{jN}(t_3) - h_N(t_2, j/N)\Delta_{jN}(t_2).$$

Now

$$
\begin{aligned}
E(V_{1j}V_{2j}) =\; & h_N(t_2, j/N)h_N(t_3, j/N)\, e^{-N\tilde{w}_j t_3}(1 - e^{-N\tilde{w}_j t_2}) \qquad (5.17)\\
& - h_N^2(t_2, j/N)e^{-N\tilde{w}_j t_2}(1 - e^{-N\tilde{w}_j t_2})\\
& - h_N(t_1, j/N)h_N(t_3, j/N)\, e^{-N\tilde{w}_j t_3}(1 - e^{-N\tilde{w}_j t_1})\\
& + h_N(t_1, j/N)\, \tilde{g}_N(t_2, j/N)e^{-N\tilde{w}_j t_2}(1 - e^{-N\tilde{w}_j t_1}) \; .
\end{aligned}
$$

The right-hand side of (5.17) is bounded by $(M_1^2 N\tilde{w}_j + M_1 M_2)(t_2 - t_1)$. Similarly,

$$EV_{1j}^2 = O(t_2 - t_1), \qquad EV_{2j}^2 = O(t_3 - t_2), \qquad (5.18)$$

and from (5.16), (5.17), (5.18) we see that the conditions of Theorem 15.6 of Billingsley (1968) are satisfied. Thus, (5.11) and Proposition 5.3 follow. $\square$

Let $\{c_{ijN}\}$, $1 \le i \le n$, $1 \le j \le N$ be a double array representable as

$$c_{ijN} = m_{iN}\, h_N(i/N, j/N) \; .$$

Assume

R3:

a) $\displaystyle\sup_N N^{-1}\sum_{i=1}^{n} |m_{iN}| < \infty$

b) $\displaystyle M_N(s) = N^{-1}\sum_{i \le Ns} m_{iN} \to M(s)$ on $[0, f_0]$.

Proposition 5.4: Suppose A1-A2, R1, R2', R3 hold. Then,

$$N^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{N} c_{ijN}\left(I_j^{[i]} - [1 - e^{-N\tau_N(i/N)\tilde{w}_j}]\right) \xrightarrow{\mathscr{L}} N(0, \sigma^2), \qquad (5.19)$$

where

$$\sigma^2 = 2\int_0^1 \left\{ \int_0^{f_0} \int_0^t (h(\tau(s), u)\, h(\tau(t), u)\, e^{-\tau(t)\bar{w}(u)}(1 - e^{-\tau(s)\bar{w}(u)}) \right. \tag{5.20}$$

$$\left. - \tau(s)\, \bar{h}(\tau(s))\bar{h}(\tau(t)))\, dM(s)dM(t) \right\} du .$$

Proof: (5.19) follows from the fact that $\int_0^{f_0} \bar{V}_N(s)\, dM_N(s) \overset{\mathscr{L}}{\rightarrow} \int_0^{f_0} \bar{W}(\tau(s))\, dM(s)$. $\square$

As an immediate consequence of Propositions 5.3 and 5.4, we obtain

Corollary 5.5: Under the conditions of Proposition 5.4,

$$\sup_{0 < s < f_0} \left| N^{-1} \sum_{j=1}^N h_N(s,j/N)\, I_j^{[Ns]} - \int_0^1 h(\tau(s),u)(1 - e^{-\tau(s)\bar{w}(u)})\, du \right| \tag{5.21}$$

$$= O_p(N^{-\frac{1}{4}}),$$

and

$$N^{-2} \sum_{i=1}^n \sum_{j=1}^N c_{ijN} I_j^{[i]} = \int_0^1 \int_0^{f_0} h(\tau(s), u)(1 - e^{-\tau(s)\bar{w}(u)})dM(s)\, du \tag{5.22}$$

$$+ O_p(N^{-\frac{1}{4}}).$$

6. Proofs

We now complete the proofs of Theorems 1-3 in Section 2. But first we need some notation and preliminary results. Define

$$\pi_r^*(u) = \pi_r(\lambda(u)), \quad \bar{\pi}_r^*(u) = 1 - \pi_r^*(u),$$
$$\pi_{r0} = \pi_r^*(f_0), \; \bar{\pi}_{r0} = 1 - \pi_{r0}, \text{ and } p_{r0} = \nu_{r0}\pi_{r0}.$$

Lemma 6.1 establishes the properties of $\Lambda(\nu)$, an asymptotic approximation to the likelihood. We will need these results in our proof of the consistency of maximum likelihood and approximate maximum likelihood estimators. Let

$$\Lambda(v) = \sum_{k=1}^{K} \Big(v_k \ \log v_k - (v_k - p_k)\log (v_k - p_k)$$

$$- v_{k0} \log v_{k0} + (v_{k0} - p_{k0}) \log (v_{k0} - p_{k0}) \Big) \qquad (6.1)$$

$$- f_0^{-1} \int_0^{f_0} \ \log \left\{ \frac{\sum_{r=1}^{K} w_r(v_r - \pi_r^*(u)v_{r0})}{\sum_{r=1}^{K} w_r v_{r0} \overline{\pi}_r^*(u)} \right\} du.$$

Lemma 6.1:

$$\Lambda(v) \leq \Lambda(v_0) = 0 \qquad (6.2)$$

with $<$ unless $v = v_0$.

Proof: Maximize Λ subject to $\sum_{k=1}^{K} w_k v_k = c$, $v_k \geq p_{k0}$, $1 \leq k \leq K$. There is a unique solution of $\nabla(\Lambda - \alpha(c)\Sigma_k w_k v_k) = 0$ given by,

$$\log \frac{v_k}{v_k - p_{k0}} = \alpha w_k,$$

where $\alpha(c)$ is the unique solution of

$$\sum_{k=1}^{K} w_k \ p_{k0} \ (1 - e^{-\alpha w_k})^{-1} = c, \quad c \geq \sum_{k=1}^{K} w_k p_{k0}. \qquad (6.3)$$

We claim

$$v_k(c) = \frac{p_{k0}}{1 - e^{-\alpha(c)w_k}}$$

gives a local and hence a global maximum. Evidently, $v_k(c)$ maximizes Λ subject to the constraint iff it maximizes

$$\overline{\Lambda}(v) = \sum_{k=1}^{K} \{v_k \log v_k - (v_k - p_{k0}) \log (v_k - p_{k0}) - \alpha(c)w_k \ v_k\}.$$

But the Hessian of $\bar{\Lambda}$ is just $Diag(-p_{k0}\,[\nu_k(\nu_k-p_{k0})]^{-1})$ and hence $\nu_j(c)$ is a global maximizer. Now,

$$\frac{\partial}{\partial c}\Lambda(\nu(c)) = \sum_{k=1}^{K} \log\left(\frac{\nu_k}{\nu_k-p_{k0}}\right)\frac{\partial\nu_k}{\partial c} - f_0^{-1}\int_0^{f_0}\left(c - \sum_{k=1}^{K} w_k\pi_k^*(u)\nu_{k0}\right)^{-1} du$$

$$= -\left(\sum_{k=1}^{K} \alpha\,w_k^2\,p_{k0}\,e^{-\alpha w_k}(1-e^{-\alpha w_k})^{-2}\,\frac{\partial\alpha}{\partial c}\right.$$

$$\left. + f_0^{-1}\int_0^{f_0}(c - \sum_{k=1}^{K} w_k\,\pi_k^*(u)\,\nu_{k0})^{-1}\,du\right)$$

$$= -\alpha(c) + f_0^{-1}\int_0^{f_0}(c - \sum_{k=1}^{K} w_k\,\pi_k^*(u)\,\nu_{k0})^{-1}\,du,$$

which is increasing in c. Since $\frac{\partial\Lambda}{\partial c}(\nu(c)) = 0$ if $\alpha(c) = \lambda_0$, or equivalently $\nu(c) = \nu_0$, the lemma follows. $\square$

Proof of Theorem 1: We can apply Corollary 5.5 by identifying, for $k = 1,\ldots,K$,

$$h_N^{(k)}(i/N,j/N) = \frac{I(Y_j = z_k)}{\nu_{k0}\bar{\pi}_{k0}} - \frac{w_k w_j}{\left(\sum_{r=1}^{K} w_r\,\nu_{r0}\,\pi_r^*(i/N)\right)^2}\,,\ 1\le i\le n,\ 1\le j\le N,$$

and

$$m_{iN} = 1$$

to conclude that

$$\sup_{i\le n}\left|\frac{D(i)}{n} - \sum_{k=1}^{K} w_k\,\nu_{k0}\,\pi_k^*(i/N)\right| = o_p(1)\,. \tag{6.4}$$

Also,

$$\log\Gamma(n\nu_k + 1) - \log\Gamma(n(\nu_k - p_k) + 1) = \log n + \sum_{j=0}^{n_k-1}\log(\nu_k - j/n),$$

$$= \log n + \int_{0}^{p_{k0}} \log(v_k - u)\,du + o_p(1)$$

uniformly in v_k, for $k = 1, \ldots, K$. The second equality follows from (2.11) since $p_k = p_{k0} + o_p(1)$. Let $\Lambda(v)$ be given by (6.1). It then follows that

$$n^{-1}(L(v) - L(v_0)) = \Lambda(v) + o_p(1) \tag{6.5}$$

uniformly for $\{v : 0 \leq f \leq 1 - \epsilon\}$ for every ϵ in $(0, 1)$. Recall (2.6) and the $1-1$ relationship between v and (θ, f). Uniform convergence here and in the sequel means that if $R_n(v)$ is the remainder in (6.5), then $\sup\{|R_n(v)| : v \text{ as specified}\} = o_p(1)$. By Lemma 6.1, $\Lambda \leq 0$ with equality iff $v = v_0$. (Note that v_0 still depends on N as does Λ.) By identifying $h_N^{(k)}(i/N, j/N)$ appropriately as before, it can be shown using Corollary 5.5 that, for $j = 1, 2$,

$$n^{-1}\frac{\partial^J L}{\partial v_{k_1} \cdots \partial v_{k_j}}(v) = \frac{\partial^J \Lambda}{\partial v_{k_1} \cdots \partial v_{k_j}}(v) + o_p(1) \tag{6.6}$$

uniformly for $0 \leq f \leq 1 - \epsilon$ for every ϵ in $(0, 1)$. From this and the fact that $\frac{\partial \Lambda}{\partial v_k}(v_0) = 0, k = 1, \ldots, K$, with probability $\to 1$, there is a unique maximum of L in a neighborhood of $v_0 = (\theta_0, f_0)$. Further, with probability $\to 1$ the first local maximum defined in Section 2 coincides with this value, thus proving consistency.

To establish asymptotic normality, from (6.6) and the consistency of $\hat{v}$, we see that

$$-n^{-1}\frac{\partial L}{\partial v_k}(v_0) = \sum_{j=1}^{K} \frac{\partial^2 \Lambda}{\partial v_j \partial v_k}(v_0)(\hat{v}_j - v_{j0}) + O_p(|\hat{v} - v_0|^2) + O_p(n^{-1}), \tag{6.7}$$

for $1 \leq k \leq K$. On the other hand, if we expand $n^{-1}\frac{\partial L}{\partial v_k}(v_0)$ around p_{k0} and the limits of $n^{-1}D(i)$ and apply Corollary 5.5 again, we obtain

$$n^{-1}\frac{\partial L}{\partial v_k}(v_0) = (p_k - p_{k0})/v_{k0}\bar{\pi}_{k0} \tag{6.8}$$

$$- w_k n^{-1}\sum_{i=1}^{n}\left(\left[\sum_{r=1}^{K}[w_r v_{r0}\bar{\pi}_r^*(i/N)]\right]^{-2} \times [n^{-1}D(i) - \sum_{r=1}^{K}w_r v_{r0}\pi_r^*(i/N)]\right)$$

$$+ O_p(n^{-1}).$$

We can now apply Proposition 5.4 to the right-hand side of (6.8), identifying $h_N^{(k)}(i/N, j/N)$ and m_{iN} as before to conclude that

$$n^{-\frac{1}{2}}\left(\frac{\partial L(v_0)}{\partial v_1}, \ldots, \frac{\partial L(v_0)}{\partial v_K}\right) \to N(0, I(v_0)). \tag{6.9}$$

Lemma 6.4 shows that the covariance matrix can be obtained explicitly as

$$I(v_0) = -\|\frac{\partial^2 \Lambda(v_0)}{\partial v_j \partial v_k}\|_{K \times K} .$$

Now

$$\frac{\partial^2 \Lambda(v_0)}{\partial v_j \partial v_k} = w_j w_k f_0 \int_0^{f_0}[\lambda'(u)]^2 \, du - \delta_{jk} d_{k0}, \tag{6.10}$$

where d_{k0} and $\lambda'(u)$ are defined in (2.15)-(2.16). By Theorem 8.3.3 of Graybill (1983), $I(v_0)$ is invertible iff

$$\int_0^{f_0}[\lambda'(u)]^2 \, du \neq \left(f_0\sum_{k=1}^{K}w_k^2 d_{k0}\right)^{-1} = \left(\sum_{k=1}^{K}w_k^2 \theta_k \bar{\pi}_k(\lambda_0)/\pi_k(\lambda_0)\right)^{-1}. \tag{6.11}$$

But (2.23) holds by Lemma 6.2 below, hence completing the proof. □

Lemma 6.2: Under A1-A2,

$$f_0\int_0^{t}(\lambda'(u))^2 \, du < \left(\sum_{k=1}^{K}w_k^2 \theta_k \bar{\pi}_k(\lambda(t))/\pi_k((\lambda(t))\right)^{-1} \quad \text{for all } t \in (0, 1).$$

Proof: Let $u(\lambda)$ be the inverse mapping of $u \to \lambda(u)$. Then, $\lambda'(u) = \left(\sum\limits_{k=1}^{K} \theta_k w_k \bar{\pi}_k(\lambda(u))\right)^{-1}$

implies $u'(\lambda) = \sum\limits_{k=1}^{K} \theta_k w_k \bar{\pi}_k(\lambda)$. Therefore,

$$\int_0^t (\lambda'(u))^2 \, du = \int_0^{\lambda(t)} (u'(\lambda))^{-1} \, d\lambda$$

$$= \int_0^{\lambda(t)} \left(\sum_{k=1}^{K} \theta_k w_k \bar{\pi}_k(\lambda)\right)^{-1} \, d\lambda.$$

So, we need to show that, for $\lambda \in (0, \infty)$,

$$A(\lambda) = \int_0^{\lambda} \left(\sum_{k=1}^{K} \theta_k w_k \bar{\pi}_k(s)\right)^{-1} \, ds < \left(\sum_{k=1}^{K} \theta_k w_k^2 \bar{\pi}_k(\lambda)/\pi_k(\lambda)\right)^{-1} \equiv B(\lambda).$$

Note that $A(0) = B(0) = 0$. Further,

$$A'(s) = \left(\sum_{k=1}^{K} \theta_k w_k \bar{\pi}_k(s)\right)^{-1},$$

and

$$B'(s) = \frac{\sum\limits_{k=1}^{K} \theta_k w_k^3 \bar{\pi}_k(s)/[\pi_k(s)]^2}{\left(\sum\limits_{k=1}^{K} \theta_k w_k^2 \bar{\pi}_k(s)/\pi_k(s)\right)^2}.$$

Define the measure μ_s by

$$d\mu_s(k) = \frac{\theta_k w_k \bar{\pi}_k(s)}{\sum\limits_{r=1}^{K} \theta_r w_r \bar{\pi}_r(s)}.$$

Then

$$\frac{B'(s)}{A'(s)} = \frac{E_\mu [W/(1-e^{-sW})]^2}{\left(E_\mu [W/(1-e^{-sW})]\right)^2} > 1.$$

Therefore, $B(\lambda) > A(\lambda)$ for all $\lambda \in (0, \infty)$, as required. □

We need the following lemma in the proof of Theorem 2.

Lemma 6.3: λ_0 is the only solution of $Z(\lambda) = \lambda > 0$.

Proof: Let

$$c(\lambda) = \sum_{k=1}^{K} w_k \, v_{k0} \, \frac{\pi_{k0}}{\pi_k(\lambda)}$$

Then, for $\lambda > 0$,

$$-\infty < c'(\lambda) = -\Sigma_{k=1}^{K} w_k^2 \, v_{k0} \, \frac{\pi_{k0}}{\pi_k^2(\lambda)} \, \bar{\pi}_k(\lambda) < 0$$

and in the notation of Lemma 6.2,

$$\frac{d}{d\lambda}\Lambda(v(c(\lambda))) = \frac{\partial \Lambda}{\partial c}(v(c(\lambda))) \, c'(\lambda) .$$

On the other hand, a direct calculation gives

$$\frac{\partial \Lambda(v(c(\lambda)))}{\partial v} = \lambda - Z(\lambda)$$

and the result follows from Lemma 6.2.0 □

Proof of Theorem 2:   To prove consistency, we see from Corollary (5.5) that

$$Z_n(\lambda) = Z(\lambda) + o_p(1) \tag{6.12}$$

uniformly for $0 \leq \lambda \leq M$ where

$$Z(\lambda) = \bar{f_0}^{-1} \int_0^{f_0} \left[\sum_{k=1}^K w_k v_{k0} \left[\frac{\pi_{k0}}{\pi_k(\lambda)} - \pi_k^*(u) \right] \right]^{-1} du .$$ (6.13)

Note that $Z(0) = 0$. By Lemma 6.3, $Z(\lambda) = \lambda$ has the unique positive root λ_0 at which point $Z'(\lambda_0) < 1$. Since $Z'_n(\lambda) \xrightarrow{P} Z'(\lambda)$ uniformly for $0 \le \lambda \le M$, we conclude that $\hat{\lambda}$ is consistent for λ_0. Since p_k tends in probability to p_{k0}, we have consistency of $v(\hat{\lambda})$.

It is easy to see that (6.7) and (6.8) must also hold for $v(\hat{\lambda})$. The rest of the results can be proved as in Theorem 1. □

Proof of Theorem 3: By Hájek (1971) to show efficiency it is enough to:

1) Establish that the uniformity in convergence specified by (2.25) holds for $v(\hat{\lambda})$, as is evident from our proof of Theorem 2.

2) Show that if $v_n = v_0 + n^{-\frac{1}{2}} t_n$ where $t_n \to t$ then

$$\mathcal{L}_{v_0}(L(v_n) - L(v_0)) \to sN\left(-\frac{t'I(v_0)t}{2} , \quad t'I(v_0)t \right),$$ (6.14)

and

$$\mathcal{L}_{v_n}(L(v_n) - L(v_0)) \to sN\left(\frac{t'I(v_0)t}{2} , \quad t'I(v_0)t \right).$$ (6.15)

But (6.14) and (6.15) are proved in the following lemma which establishes local asymptotic normality in our models, a result needed both for efficiency and an important "information" identity. Hence, efficiency follows. □

Lemma 6.4: (Local asymptotic normality)

If $v_n = v_0 + t_n n^{-\frac{1}{2}}$ and $t_n \to t$, then

$$\mathcal{L}_{\nu_0}(L(\nu_n) - L(\nu_0)) \rightarrow sN\left(-\frac{t'I(\nu_0)t}{2}, \quad t'I(\nu_0)t\right), \tag{6.16}$$

$$\mathcal{L}_{\nu_n}(L(\nu_n) - L(\nu_0)) \rightarrow sN\left(\frac{t'I(\nu_0)t}{2}, \quad t'I(\nu_0)t\right), \tag{6.17}$$

and

$$\left\|\frac{\partial^2 \Lambda(\nu_0)}{\partial \nu_j \partial \nu_k}\right\| = -I(\nu_0). \tag{6.18}$$

Proof:

A study of Le Cam (1960) actually shows that if we can establish convergence to $sN(\mu, \sigma^2)$ with $\sigma^2 = t'I(\nu_0)t$ in (6.16) and tightness (or a fortiori convergence to a normal distribution) in (6.17), then (6.16) and (6.17) hold as given.

Note from Corollary 5.5 that, for all $j \geq 0$,

$$n^{-1}\frac{\partial^j L(\nu)}{\partial \nu_{k_1} \cdots \partial \nu_{k_j}} = \frac{\partial^j \Lambda(\nu)}{\partial \nu_{k_1} \cdots \partial \nu_{k_j}} + o_p(1), \tag{6.19}$$

uniformly for $0 \leq f \leq 1-\epsilon$ for every ϵ in $(0, 1)$. Therefore,

$$L(\nu_n) - L(\nu_0) = n^{-\frac{1}{2}}t'_n \nabla L(\nu_0) + \frac{t'_n}{2}\left\|\frac{1}{n}\frac{\partial^2 L(\nu_0)}{\partial \nu_j \partial \nu_k}\right\| t_n + O_p(n^{-\frac{1}{2}}).$$

From Theorem 1,

$$\mathcal{L}_{\nu_0}(n^{-\frac{1}{2}}t'_n \nabla L(\nu_0)) \rightarrow sN(0, t'I(\nu_0)t),$$

and by (6.19) again,

$$\frac{1}{n}\frac{\partial^2 L(\nu_0)}{\partial \nu_j \partial \nu_k} = \frac{\partial^2 \Lambda}{\partial \nu_j \partial \nu_k}(\nu_0) + o_p(1).$$

So,

$$\mathcal{L}_{\nu_0}(L(\nu_n) - L(\nu_0)) \rightarrow sN\left(\tfrac{1}{2}t' \ \|\frac{\partial^2 \Lambda(\nu_0)}{\partial \nu_j \partial \nu_k}\| \ t, \ t'I(\nu_0)t\right).$$

Similarly,

$$\mathcal{L}_{\nu_n}(L(\nu_n) - L(\nu_0)) \rightarrow sN\left(- \tfrac{1}{2}t' \ \|\frac{\partial^2 \Lambda(\nu_0)}{\partial \nu_j \partial \nu_k}\| \ t, \ t'I(\nu_0)t\right).$$

By Le Cam (1960), for all t

$$t' \ \|\frac{\partial^2 \Lambda(\nu_0)}{\partial \nu_j \partial \nu_k}\| \ t = - t'I(\nu_0) \ t,$$

and the lemma follows. □

Table 1: Comparison of Asymptotic and Actual Standard Errors

Estimator	Asymptotic	Actual
$\hat{v}_1$	.38	.44
$\hat{v}_2$	.27	.30
$\hat{v}_3$	.22	.27
$\hat{v}_4$	.16	.17
$\hat{v}_5$	.08	.09
$\hat{\theta}_1$	.12	.12
$\hat{\theta}_2$	.07	.07
$\hat{\theta}_3$	.05	.05
$\hat{\theta}_4$	.04	.04
$\hat{\theta}_5$	.07	.06
$\hat{N}$	49	55
$\hat{R}$	30.5	33

Table 2: North Sea Oil and Gas Data

Name or Location	Discovery Date	Recoverable Reserves	Name or Location	Discovery Date	Recoverable Reserves
Balder	7-67	100	W. Beryl	8-74	115
Cod	6-68	100	N. Cormorant	8-74	367
2/3-1	4-69	12	W. Heather	9-74	75
Montrose	9-69	150	15/23-A1,2,4B	10-74	225
Ekofisk	9-69	2100	211/13	11-74	250
Josephine	9-70	88	15/22-1	11-74 .	†
Forties	10-70	1800	Tartan	1-75	268
Tor	11-70	340	Hod	1-75	88
W. Ekofisk	12-70	375	3/11-1	1-75	†
Auk	2-71	63	Mabel	2-75	100
Frigg	6-71	1264	14/20-1,5	2-75	62
30/2-1	6-71	†	Brae	4-75	649
Brent	7-71	2252	211/27-3	4-75	258
Statfjord	2-74	3017	Crawford	4-75	350
Argyll	8-71	28	Tern	4-75	238
3/25A	12-71	310	2/10-1A,2,3	4-75	50
Lomond	5-72	220	W. Ninian	5-75	100
Bream	6-72	75	9/13-7	5-75	650
S.E. Tor	6-72	45	Gudrun	6-75	450
Albuskjell	8-72	714	21/2-1	6-75	50
Beryl	9-72	617	3/2-1A	6-75	242
S. Cormorant	9-72	140	Valhall	6-75	647
Edda	9-72	126	3/4-5, 3/9-1	7-75	400
Eldfisk	12-72	1186	16/7-2	8-75	†
Heimdal	12-72	285	16/21	8-75	45
Piper	1-73	679	Murchison	9-75	365
Maureen	2-73	148	211/26-4	9-75	175
Dunlin	6-73	582	211/18-9	9-75	50
Thistle	7-73	530	15/30-1,2	9-75	†
3/19-1	7-73	†	15/21	10-75	62
3/15-2,3	8-73	150	15/13-2	10-75	200
E. Frigg	8-73	40	Fulmar	11-75	485
Hutton	9-73	240	3/23-1	11-75	†
3/29-1,2	9-73	†	3/8-4	11-75	†
Brisling	10-73	62	15/29-2	2-76	†
3/4-1,2,3	10-73	225	23/25A	3-76	100
Alwyn	11-73	358	Ranger	3-76	75
Heather	12-73	150	Renee	4-76	162
N.W. Tor	12-73	†	211/16	5-76	†
Ninian	1-74	1117	9/19-2	5-76	†
Odin	3-74	178	3/28-1	5-76	†
3/4-4,5	3-74	200	N. Thistle	7-76	175
Flyndre	4-74	†	Thelma	7-76	375
N.E. Frigg	4-74	66	30/7	8-76	†
Sleipner	4-74	660	7/12-2	8-76	200
Claymore	5-74	438	Beatrice	9-76	162
Andrew	6-74	150	35/3	10-76	†
S.E. Frigg	6-74	7	33/9-7	11-76	225
Bruce	7-74	450			
Magnus	7-74	430			
Buchan	8-74	155			

Source: O'Carrol, F. M. and Smith, J. L. (1980)
Notes:
1. Data shown are the constructed "consensus" data in O'Carrol and Smith
2. Reserve volumes are in millions of barrel of oil and gas equivalent.
3. † represents small reservoirs for which estimates are not available. These have been included in the smallest class size.

Table 3: Analysis of North Sea Data

Class Interval	z_k	n_k	$\hat{N}_k$ ($\pm$ s.e.)	$\hat{\theta}_k$ ($\pm$ s.e.)
[0, 50)	25	26	262 ($\pm$ 110)	.64 ($\pm$ .06)
[50, 100)	75	15	56 ($\pm$ 22.5)	.14 ($\pm$ .03)
[100, 200)	150	17	36 ($\pm$ 12.3)	.09 ($\pm$ .02)
[200, 400)	300	20	28 ($\pm$ 6.5)	.07 ($\pm$ .01)
[400, 800)	600	14	15 ($\pm$ 1.8)	.04 ($\pm$ .01)
[800, 1600)	1200	3	3 ($\pm$ .15)	.01 ($\pm$ .00)
[1600, 3200)	2400	4	4 ($\pm$ .01)	.01 ($\pm$ .00)
		99	404 ($\pm$ 144)	

Figure Legends

1. Figure 1: **Box** plots of the approximate MLE's, $\hat{N}_k$, $k = 1, ..., 5$ from the simulation study. Note the distributions are all reasonably centered around the true values of 20, the upper tails are heavier as to be expected, and that the variability decreases from the left to the right.

2. Figure 2: Box plots of the estimated proportions, $\hat{\theta}_k$, $k = 1, ..., 5$ from the simulation study. The distributions are centered around the true values of 0.2 and are reasonably symmetric.

3. Figure 3: Plot of the estimates of the population sizes for the North Sea Data based on increasing sample sizes. Note that the estimates settle down around sample size 40 and vary thereafter only in the region of 300-600. The final estimate of 404 is based on a sample of size 99.

REFERENCES

1. Andreatta, G. and Kaufman, G.M. (1986), " Estimation of finite population properties when sampling is without replacement and proportional to magnitude", *J. of Amer. Statist. Assoc.*, 81, 657-666.

2. Barouch, E. and Kaufman, G. M. (1976), "Probabilistic Modelling of Oil and Gas Discovery", in *Energy: Mathematics and Models*, ed. F. Roberts, Philadelphia: SIAM, 248-260.

3. Barouch, E. and Kaufman, G. M. (1977), "Estimation of Undiscovered Oil and Gas", *Proceedings of Symposia in Applied Mathematics*, 21, 77-91.

4. Barouch, E., Kaufman, G. M. and Nelligan, J. (1983), "Estimation of Parameters of Oil and Gas Discovery Process Models Using the Expectation-Maximization Algorithm", in *Energy Modelling and Simulation*, eds. A. S. Kydes et al., North-Holland, Amsterdam, 109-117.

5. Bloomfield, P., Deffeyes, K. S., Watson, G. S., Benjamini, Y. and Stine, R. A. (1979), "Volume and Area of Oil Fields and Their Impact on Order of Discovery", Resource Estimation and Validation Project, Departments of Statistics and Geology, Princeton University.

6. Cassel, C. M., Sarndal, C. E., and Wretman, J. H. (1977), *Foundations of Inference in Survey Sampling*, John Wiley, New York.

7. Cox, D. R. (1969), "Some Sampling Problems in Technology", in *New Developments in Survey Sampling*, eds. N. L. Johnson and H. Smith, Jr., Wiley-Interscience, New York, 506-527.

8. Cozzolino, J. M., (1972), "Sequential search for an unknown number of objects of nonuniform size," *Oper. Research*, V.20, 293-308.

9. Godambe, V. P. and Rajashari, M. B. (1989), "Optimal Estimation for Weighted Distributions: Semi-parametric Models," in *Statistical Data Analysis and Inference, edited by Y.*

Dodge, North-Holland, New York.

10. *Gordon, L. (1989), "Estimation for large successive samples with unknown inclusion probabilities", Advances in Applied Mathematics, (to appear).*

11. Graybill, F. (1983), *Matrices with applications in statistics, Wadsworth,* Belmont.

12. Hájek, J. (1971), "Limiting properties of likelihood and inference," V. P. Godambe and D. A. Spott eds. Holt, Rinehart and Winston, Toronto 142-162.

13. Holst, L. (1973), "Some limit theorems with applications in sampling theory," *Ann. Statist.,* 1, 644-658.

14. Holst, L. (1986), "On birthday, collectors, and the classical urn problems," *I. S. Review,* 54, 15-28.

15. Kalbfleisch, J. D. and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data,* John Wiley, New York.

16. Le Cam, L. (1960), Locally asymptotically normal families of distributions *Univ. of California Publications in Statistics.*

17. Lee, P. J. and Wang, P. C. C. (1983), "Probabilistic Formulation of a Method for the Evaluation of Petroleum Resources", *Jour. Math. Geol.,* 15, 163-181.

18. Lee, P. J. and Wang, P. C. C. (1985), "Prediction of Oil or Gas Pool. Sizes When Discovery Record is Available", *Jour. Math. Geol.,* 17, 95-113.

19. O'Carrol, F. M. and Smith, J. L. (1980), "Probabilistic methods for estimating undiscovered petroleum resources," in *Advances in the economics of energy and resources,* (ed. J. Moroney), V.3, JA1 Press, Inc., Greenwich, Conn. 31-63.

20. Nair, V. N. and Wang, P. C. C. (1989), "Maximum likelihood estimation under a successive sampling discovery model," *Technometrics* (to appear).

21. Patil, G. P. and Rao, C. R. (1977), "The Weighted Distributions: A Survey of Their Applications", in *Application of Statistics,* ed. P. R. Krishnaiah, North-Holland, Amsterdam,

383-405.

22. Rosén, B. (1972), "Asymptotic theory for successive sampling with varying probabilities without replacement I and II " *Ann. Math. Statist.*, 43, 373-397 and 748-776.

23. Sen, P. K. (1979), "Invariance principles for the coupon collector's problem: A martingale approach," *Ann. Statist.*, 7, 372-380.

24. Smith, J. L. and Ward, G. L. (1981), "Maximum Likelihood Estimates of the Size Distribution of North Sea Oil Fields", *Jour. Math. Geol.*, 13, 399-413.

25. Vardi, Y. (1982), "Nonparametric Estimation in the Presence of Length Bias", *Ann. Statistics*, 10, 616-620.

Figure Legends

1. Figure 1: Box plots of the approximate MLE's, $\hat{N}_k$, $k = 1, ..., 5$ from the simulation study. Note the distributions are all reasonably centered around the true values of 20, the upper tails are heavier as to be expected, and that the variability decreases from the left to the right.

2. Figure 2: Box plots of the estimated proportions, $\hat{\theta}_k$, $k = 1, ..., 5$ from the simulation study. The distributions are centered around the true values of 0.2 and are reasonably symmetric.

3. Figure 3: Plot of the estimates of the population sizes for the North Sea Data based on increasing sample sizes. Note that the estimates settle down around sample size 40 and vary thereafter only in the region of 300-600. The final estimate of 404 is based on a sample of size 99.
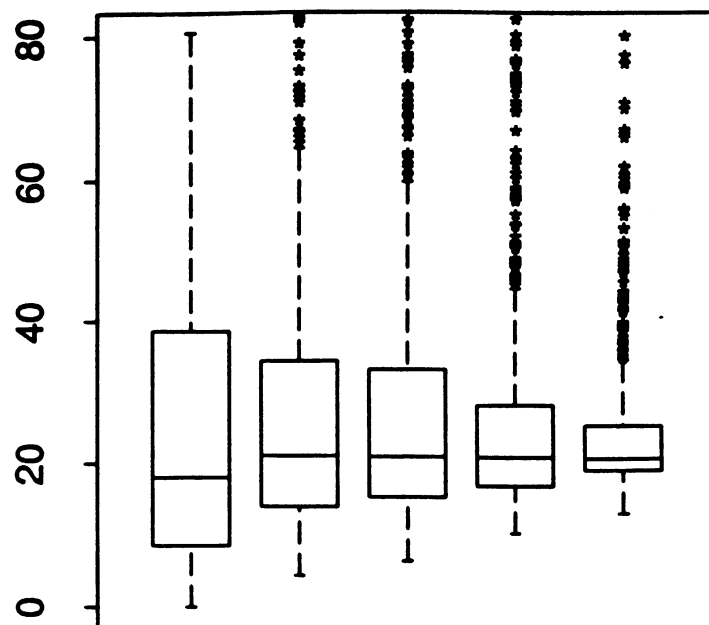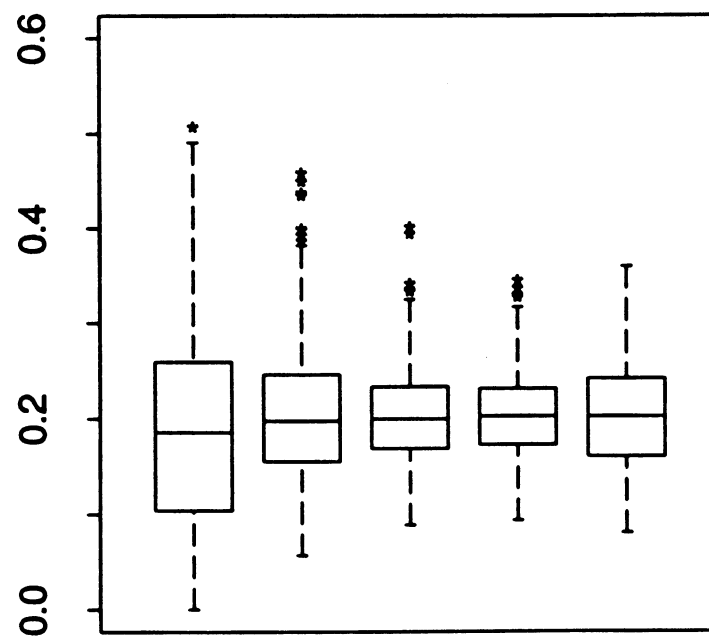
Figure 1:

Figure 2:

Figure 3: