BAYESIAN ESTIMATION OF POPULATION SIZE
WITH REMOVAL SAMPLING

by

Richard G. Cornell

FSU Statistics Report M179
ONR Technical Report No. 43

January 1970
Department of Statistics
Florida State University
Tallahassee, Florida
BAYESIAN ESTIMATION OF POPULATION SIZE WITH
REMOVAL SAMPLING

by

Richard G. Cornell
Florida State University

The posterior distribution of the number of items remaining in a
population after binomial removal sampling is derived assuming a diffuse
or negative binomial prior distribution on the number of items initially
in the population and known probability of removal per item. Applications
to animal sampling and epidemiology are described. Extensions to removal
sampling for successive sampling periods and for more than one sub-
population are also given.

1. INTRODUCTION

We consider the estimation of a population size \( n \) based on an
observation \( x \) on a binomial random variable with probability function

\[
f(x|n) = \frac{n!}{(n-x)!x!} p^x q^{n-x}, \quad x = 0, 1, \ldots, n; \quad 0 < p < 1,
\]

for \( p \) assumed known and \( q = 1 - p \). Our motivation for investigating
this problem arises from a study of removal sampling from finite populations
when it can be assumed that the probability, \( p \), of observing any item
during a single sampling period is known.

We assume that either a negative binomial or diffuse prior distribution
can be prescribed for \( n \). It is shown in Sections 2 and 3 that the
posterior distribution of the number of items not observed, namely
\( m = n - x \), is negative binomial for both types of prior distributions.
Moreover, the posterior mean with a diffuse prior distribution differs by
at most unity from the maximum likelihood estimate. A Poisson prior
distribution is also considered for n in Section 2 but is not suggested
for use because the resultant Poisson posterior distribution for m is
independent of the sample observation. The posterior distribution of
n given x follows immediately from that of m. Our results are generalized
in Section 4 to successive samples of the items not previously observed
when those items which are observed are removed from the population
before another sample is taken.

Our model is equivalent, when p is known, to that considered by
procedures for maximum likelihood estimation utilizing removal trapping
results from several trapping periods. Another estimation procedure
proposed by Hayne [4] is also studied by Zippin. Our model could also
be used in estimating the number of susceptibles remaining in a population
after an epidemic for which it is reasonable to specify a common probability
of becoming ill for each susceptible.

Feldman and Fox [3] have studied maximum likelihood and other
asymptotically equivalent estimation procedures for n in equation (1) for
independent samples from a population with that binomial probability
function. Although the removal sampling procedure described here does not
result in repeated sampling from a population of constant size n, the
maximum likelihood estimation presented by Feldman and Fox is appropriate
for our model for a single sampling period.

Section 5 contains an extension of our model to the situation where
the population sampled is just one of several subpopulations in a larger
population of interest. A Poisson distribution on n with mean \lambda is
introduced as part of the model and leads to a Poisson distribution for x
when \( \lambda \) is fixed. Bayesian estimation of \( \lambda \) is considered with either a diffuse or gamma prior distribution, both of which result in a gamma posterior distribution. The combination and comparison of results for more than one subpopulation are also discussed.

2. NEGATIVE BINOMIAL PRIOR DISTRIBUTION

Let us consider a negative binomial prior distribution for \( n \) with probability function \( g(n) \) such that

\[
g(n) = \frac{\Gamma(n+\alpha)[1/(1+\beta)]^n}{\Gamma(n+1)}, \quad n = 0, 1, \ldots; \quad \alpha, \beta > 0. \tag{2}
\]

This distribution has mean \( E(n) \) and variance \( V(n) \) given by

\[
E(n) = \frac{\alpha}{\beta}; \quad V(n) = E(n)[(1+\beta)/\beta]. \tag{3}
\]

This prior distribution has adequate flexibility for use in a variety of situations.

One way of deriving this negative binomial probability function is to assume that \( n \) has a Poisson distribution with a mean \( \lambda \) which is itself a gamma random variable with density function given by

\[
g(\lambda) = \lambda^{\alpha-1}e^{-\beta\lambda}, \quad \lambda > 0; \quad \alpha, \beta > 0, \tag{4}
\]

with mean and variance

\[
E(\lambda) = \frac{\alpha}{\beta}; \quad V(\lambda) = E(\lambda)/\beta. \tag{5}
\]

Then the unconditional distribution of \( n \) with \( \lambda \) not given is negative binomial as defined by (2). Note that \( E(\lambda) \) as given by (5) is the same as \( E(n) \) in (3). This characterization of the negative binomial distribution can be helpful in rationalizing its use as a prior distribution for \( n \) and
along with the moments in (3) and (5) can aid in the selection of a prior distribution. In particular, it helps us specify a diffuse prior distribution for $n$ in the next section.

With $f(x|n)$ as given by our model equation (1) and with $g(n)$ as given by (2), the posterior probability function for $n$ given the observed $x$, denoted by $h(n|x)$, is such that

$$h(n|x) = f(x|n) g(n) = \Gamma(n+\alpha)q^{n-x}[1/(1+\beta)]^n/(n-x)! , \ n=x, x+1, \ldots , \ (6)$$

where $q$ and $x$ are both known constants. Recalling that $m$, the number of items remaining after removing the $x$ observed items, is equal to $n-x$, we have from (6) that the posterior probability function for $m$ given $x$, denoted by $h(m|x)$, is such that

$$h(m|x) = \Gamma(m+x+\alpha)[q/(1+\beta)]^m/\Gamma(m+1), \ m=0,1,\ldots . \ (7)$$

Comparing (7) with (2) shows that the posterior distribution of $m$ is negative binomial with prior parameters $\alpha$ and $\beta$ replaced by $(x+\alpha)$ and $(1-q+\beta)/q$, respectively. From (3) it follows that the posterior mean and variance of $m$ given $x$ are

$$E(m|x) = q(x+\alpha)/(1-q+\beta); \ \ \ \ (8)$$

$$V(m|x) = E(m|x)(1+\beta)/(1-q+\beta).$$

If we assume a quadratic loss function, the Bayes estimate of $m$ after removing $x$ items is given by the right side of equation (8). Percentiles of the posterior negative binomial distribution of $m$ can be used to form an interval estimate of $m$. Bartko [1,2] gives ways to compute or approximate these percentiles from tables of other distributions.
formation of a compound Poisson prior distribution. Lindley [5, page 155] shows that it is more reasonable to assign a uniform prior weighting to \( \ln \lambda \) by setting both \( \alpha \) and \( \beta = 0 \) in (8) and restricting our attention to positive integers for \( n \). This makes

\[
g(n) = 1/n, \ n = 1, 2, \ldots;
\]

\[
h(m|x) = (m+x-1)! \ q^m / m!, \ m = 0, 1, \ldots .
\]

Again our posterior distribution for \( m \) is negative binomial, provided now that \( x > 1 \), with mean and variance

\[
E(m|x) = qx/(1-q);
\]

\[
V(m|x) = E(m|x)/(1-q).
\]  \hspace{1cm} (10)

Alternative Bayes estimates of \( m \) assuming quadratic loss with diffuse prior information are given by the right sides of equations (9) and (10). If we add \( x \) to the latter, we obtain the Bayes estimate

\[
\hat{n} = x + qx/(1-q) = x/(1-q).
\]  \hspace{1cm} (11)

The largest integer in \( x/(1-q) \) is the maximum likelihood estimate of \( n \) and differs by at most one from \( \hat{n} \) as given by (11) for \( \alpha = \beta = 0 \) and \( x > 1 \). Similarly, using \( \alpha = 1 \) and \( \beta = 0 \) and (9), we find that

\[
\hat{n} = x/(1-q) + q/(1-q) \quad \text{for} \ x > 0, \ \text{which exceeds the estimate given by (11) by} \ q/(1-q).
\]

In view of this comparison with maximum likelihood estimation, we suggest taking \( \alpha = \beta = 0 \) and \( \hat{n} \) as given by (11) when prior information on \( n \) is vague. The restriction to values of \( x > 1 \) is not of practical concern since we would want considerable sample information, that is, we would want \( x \) to be large relative to \( n \), when we have vague prior information and
since \( n \) would ordinarily be much larger than one in applications.

4. EXTENSION TO ADDITIONAL SAMPLING PERIODS

To extend the results of Sections 2 and 3 to several sampling periods, we introduce the subscript \( i = 1, 2, \ldots, k \), where \( k \) is the total number of sampling periods. Let the subscript zero refer to initial or prior conditions. Also let \( p_i \) be the known probability that an item present at the beginning of the \( i \)th sampling period is observed and removed during that sampling period, let \( q_i = 1 - p_i \) and let \( r_i = \Pi_{j=1}^{i} q_j \). Let \( m_i \) be the number of items in the population after the \( i \)th sampling period with \( m_o = n \). Let \( x_i \) be the number of items observed and removed during the \( i \)th period and let \( t_i = x_1 + x_2 + \ldots + x_i \).

With this notation we can now present the posterior distribution of \( m_i \) after each of the \( k \) sampling periods by replacing \( x \) and \( q \) in Sections 2 and 3 by \( t_i \) and \( r_i \), respectively, that is, by successively defining our total sampling period to consist of the first, the first two, and so forth up to all \( k \) of the individual sampling periods. The results are as follows:

Suppose the prior distribution of \( m_o \) is either negative binomial with probability function given by (2) with positive parameters \( \alpha_o \) and \( \beta_o \), respectively, or diffuse with \( \alpha_o = \beta_o = 0 \) replacing \( \alpha \) and \( \beta \) in (2). Then the posterior distribution of \( m_i \) is negative binomial with parameters \( \alpha = \alpha_i \) and \( \beta = \beta_i \) in (2), where

\[
\alpha_i = t_i + \alpha_o;
\]

\[
\beta_i = (1 - r_i + \beta_o)/r_i, \quad i = 1, 2, \ldots, k.
\]
From (3),

$$E(m_1|t_1) = \alpha_1/\beta_1;$$

$$V(m_1|t_1) = E(m_1|t_1)((1+\beta_1)/\beta_1).$$

The right side of equation (12) is the Bayes estimate of $m_1$ given $t_1$ assuming quadratic loss. After any period of sampling the corresponding negative binomial posterior distribution can be used to construct posterior probability intervals on $m_1$ or to make other inferences as discussed in Section 2.

5. EXTENSION TO SEVERAL SUBPOPULATIONS

So far we have been concerned with inferences about the size of a single population which initially contains $n$ items. In some applications the population being sampled is just one of several subpopulations comprising a larger population. For instance, in animal sampling we might study just one unit area because of the difficulty of adequately trapping a larger area, but we might prefer to make inferences about a larger, homogeneous area in which the study unit is located.

Suppose then that we are using removal sampling in a single unit from a population of several such units. Let $n$ be the population of that unit and let $\lambda$ be the population density for all the units together. We assume that items are distributed over units in accord with a Poisson process with mean $\lambda$ which with (1) implies that the distribution of $x$ given $\lambda$, but not $n$, is also Poisson but with mean $p\lambda$. Thus the probability function for our model is now

$$f(x|\lambda) = (p\lambda)^x e^{-p\lambda}/x!, \ x = 0, 1, \ldots.$$
Next we take a gamma prior distribution for \( \lambda \) with density function given by (4). Then the posterior distribution of \( \lambda \) given \( x \) is also gamma with density function

\[
h(\lambda|x) \propto \lambda^{x+\alpha-1} e^{-(p+\beta)\lambda}, \quad \lambda > 0. \tag{13}
\]

Comparison with (4) shows that the prior parameters \( \alpha \) and \( \beta \) have been replaced in (13) by \( x + \alpha \) and \( p + \beta \), respectively. From (5), the posterior mean and variance are

\[
E(\lambda|x) = \frac{(x+\alpha)}{(p+\beta)}; \quad V(\lambda|x) = E(\lambda|x)/(p+\beta).
\]

The posterior mean is the Bayes estimate of \( \lambda \) assuming quadratic loss. As discussed in Section 3, we can specify a diffuse prior function for \( \lambda \) by setting \( \alpha = \beta = 0 \). The posterior distribution is again gamma.

These results can be extended to several sampling periods as in Section 4 by replacing \( \alpha, \beta, x \) and \( p \) in (13) by \( \alpha_0, \beta_o, t_i \) and \( 1-r_i \), respectively, for \( i = 1, 2, \ldots, k \).

Since \( \lambda \) has a posterior gamma distribution with parameters \( x + \alpha \) and \( p + \beta \), the corresponding distribution of \( 2(p+\beta)\lambda \) is chi-square with \( 2(x+\alpha) \) degrees of freedom. This enables computation of posterior probability intervals for \( \lambda \) using tables of chi-square percentage points. It also implies that if \( h \) unit subpopulations with common \( \lambda \) were sampled independently, \( 2\lambda \sum_j (p_j+\beta_j) \) would have a posterior chi-square distribution with \( 2 \sum_j (x_j+\alpha_j) \) degrees of freedom, where \( j = 1, 2, \ldots, b \) indexes the units sampled. Hence a combined Bayes estimate of \( \lambda \) would be \( \sum_j (x_j+\alpha_j)/(\sum_j (p_j+\beta_j)) \). Moreover, if independent samples were obtained from two subpopulations with possibly different densities \( \lambda_1 \) and \( \lambda_2 \),
then \((p_1 + \beta_1)(x_2 + a_2) \lambda_1 / (p_2 + \beta_2)(x_1 + a_1)\lambda_2\) would have a posterior F-distribution with \(2(x_1 + a_1)\) and \(2(x_2 + a_2)\) degrees of freedom, where the subscripts 1 and 2 on \(\alpha, \beta, p\) and \(x\) are now used to associate them with either \(\lambda_1\) or \(\lambda_2\), respectively. This F-ratio could be used to compute posterior probability intervals for \(\lambda_1 / \lambda_2\), and, in particular, to decide if it is reasonable to infer that \(\lambda_1 = \lambda_2\).

A more complete discussion of Bayesian estimation of the means of Poisson distributions, including the results presented in this section because of their relevance to our topic, is given by Lindley [5, pages 153-7].

ACKNOWLEDGMENTS

This research has been done under the Office of Naval Research Contract NONR 988(08). Partial support for the author during the research was received from Biometry Training Grant 5 T01 GM 913 from the National Institute of General Medical Sciences.
REFERENCES


The posterior distribution of the number of items remaining in a population after binomial removal sampling is derived assuming a diffuse or negative binomial prior distribution on the number of items initially in the population and known probability of removal per item. Applications to animal sampling and epidemiology are described. Extensions to removal sampling for successive sampling periods and for more than one subpopulation are also given.