Ann. Inst. Statist. Math.
28 (1976), Part A, 429-444

CERTAIN ESTIMATION PROBLEMS FOR MULTIVARIATE HYPERGEOMETRIC MODELS

K. G. JANARDAN

(Received Oct. 1, 1973; revised Feb. 25, 1974)

Abstract

Certain estimation problems associated with the multivariate hypergeometric models: the property of completeness, maximum likelihood estimates of the parameters of multivariate negative hypergeometric, multivariate negative inverse hypergeometric, Bayesian estimation of the parameters of multivariate hypergeometric and multivariate inverse hypergeometrics are discussed in this paper.

A two stage approach for generating the prior distribution, first by setting up a parametric super population and then choosing a prior distribution is followed. Posterior expectations and variances of certain functions of the parameters of the finite population are provided in cases of direct and inverse sampling procedures. It is shown that under extreme diffuseness of prior knowledge the posterior distribution of the finite population mean has an approximate mean $\bar{x}$ and variance $(N-n)S^2/Nn$, providing a Bayesian interpretation for the classical unbiased estimates in traditional sample survey theory.

1. Introduction

The subject of estimation including Bayes's estimation of the parameters of a finite population (viz. the parameters of multivariate hypergeometric (MH)) has received the attention of several authors including Wilks [13], Sarndal [11], Hartley and Rao [4] and Hoadley [5] among others. Janardan [6], and Hartley and Rao [4] have discussed the maximum likelihood (ML) estimates of the parameters of MH. Wilks [13] has given the unbiased estimates of mean and variance, in sampling without replacement from a finite population using a two-stage sampling procedure. Sarndal [11], [12] has discussed the Bayesian estimation of the parameters of MH with Bose-Einstein uniform and multivariate negative hypergeometric (MNH) type distributions as priors.

A number of people working on sampling inspection plans have also been interested in Bayes's estimation for the hypergeometric distribution, most notably Hald [3] who gives a number of interesting results including the use of Polya distribution as the prior distribution. His results are for the case $s \leqq 2$. Following Hald's approach in the univariate case, Hoadley [5] obtains the Bayesian estimates of the parameters of hypergeometric in the multivariate case. Sarndal [11] and Hoadley [5] have used in their derivations of prior distributions the concept of super population which has been widely used earlier in sample survey work. Hartley and Rao [4] are mainly concerned with finding unbiased minimum variance and ML estimates of certain functions of the population parameters but they do also have a section on Bayes' approach with MNH as a prior distribution.

Janardan and Patil [8] have shown that multivariate hypergeometric (MH), multivariate inverse hypergeometric (MIH), multivariate negative hypergeometric (MNIH), multivariate Polya (MP) and multivariate inverse Polya (MIP) distributions belong to a class of multivariate distributions called unified multivariate hypergeometric (UMH) class. The probability functions of these models are given in Table 1.

Mosimann [9], [10] has discussed the moment estimates of the parameters of MNH and MNIH. Other distributions viz. MIH, MP and MIP do not seem to have been considered in literature for the purposes of parametric estimation. In this paper we discuss certain estimation problems associated with MH models. We discuss briefly, the completeness of MIH family in Section 2, the ML estimates of the parameters of MNH and MNIH in Section 3, Bayesian estimates of the parameters of MIH with MNIH as a prior distribution and of the parameters of MH with UMH as a prior distribution in Section 4.

2. Completeness of MIH family

Hartley and Rao [4] have shown that MH family is complete. They give an elegant proof due to B. K. Kale to show that $\boldsymbol{x}$ is completely sufficient for N for MH. Here we ask an interesting question, that is, do the models which come under the UMH class enjoy this property of completeness? We answer this question partly by showing that MIH is complete. Yet, it remains to prove or disprove the completeness of the other families of distributions which come under UMH.

The probability mass function of the s-variable inverse hypergeometric distribution, given in Table 1, will be denoted by

$$p_s(x_1, \cdots, x_s; k, N_0, \cdots, N_s) ,$$

or, on occasion, more simply by

$$p_s(x_1, \cdots, x_s; k) .$$

Write $x_s = j$, $j = 0, \cdots, N_s$. One can show the following relation between the $(s-1)$-variable and s-variable probabilities:

$$\begin{aligned} (2.1)\qquad & p_s(x_1, \cdots, x_{s-1}, j; k, N_0, \cdots, N_s) \\ & = C_j(N_s) p_{s-1}(x_1, \cdots, x_{s-1}; k+j, N_0+N_s, N_1, \cdots, N_{s-1}) . \end{aligned}$$

Here the constant $C_j(N_s) > 0$ does not depend on $x_1, \cdots, x_{s-1}$,

$$(2.2)\qquad C_j(N_s) = \frac{k}{k+j} \frac{\binom{N_0}{k}\binom{N_s}{j}}{\binom{N_0+N_s}{k+j}} ,$$

$j = 0, \cdots, N_s$. This relation will be used in proving the following theorem.

THEOREM 2.1. *The family of multivariate inverse hypergeometric distributions is complete.*

PROOF. Suppose for purposes of induction that the $(s-1)$-variable family is complete. Now consider the s-variable family and let g be a real-valued function of s arguments such that

$$(2.3)\qquad \sum_{\boldsymbol{x}_s} g(x_1, \cdots, x_s) p_s(x_1, \cdots, x_s; k) = 0 ,$$

for all $\boldsymbol{N}_s = (N_0, \cdots, N_s)$ with $\sum_0^s N_i = N$. The equation can be written

$$\sum_{x_s=0}^{N_s} \sum_{\boldsymbol{x}_{s-1}} g(x_1, \cdots, x_s) p_s(x_1, \cdots, x_s; k) = 0 ,$$

and using (2.1) becomes

$$(2.4)\qquad \begin{aligned} & \sum_{j=0}^{N_s} \sum_{\boldsymbol{x}_{s-1}} C_j(N_s) g(x_1, \cdots, x_{s-1}, j) p_{s-1}(x_1, \cdots, x_{s-1}; k+j) = 0 \\ \text{or} \\ & \sum_{j=0}^{N_s} C_j(N_s) \sum_{\boldsymbol{x}_{s-1}} g_j(x_1, \cdots, x_{s-1}) p_{s-1}(x_1, \cdots, x_{s-1}; k+j) = 0 . \end{aligned}$$

Here g_j is the real-valued function of $(s-1)$ arguments defined in the obvious manner so that

$$g_j(x_1, \cdots, x_{s-1}) = g(x_1, \cdots, x_{s-1}, j)$$

for all relevant $\boldsymbol{x}_{s-1}$. The summation in (2.4) can be written

$$(2.5)\qquad \sum_{j=0}^{N_s} C_j(N_s) \mathop{\mathrm{E}}_{\boldsymbol{x}_{s-1}} (g_j(\boldsymbol{x}_{s-1})) = 0 .$$

Now let $N_s=0$. Then

$$C_0(0) \underset{\boldsymbol{x}_{s-1}}{\mathrm{E}} (g_0(\boldsymbol{x}_{s-1}))=0 \quad \text{or} \quad \underset{\boldsymbol{x}_{s-1}}{\mathrm{E}} (g_0(\boldsymbol{x}_{s-1}))=0 \;. \tag{2.6}$$

This, and the completeness of the $(s-1)$-variable family, imply that

$$g_0(\boldsymbol{x}_{s-1})=0 \tag{2.7}$$

for all relevant $\boldsymbol{x}_{s-1}$ ($x_i \leqq N_i$, $i=0,\cdots,s-1$) and $\boldsymbol{N}_{s-1}$ such that $\sum_0^{s-1} N_i = N$. Next let $N_s=1$. Then

$$C_0(1) \underset{\boldsymbol{x}_{s-1}}{\mathrm{E}} (g_0(\boldsymbol{x}_{s-1}))+C_1(1) \underset{\boldsymbol{x}_{s-1}}{\mathrm{E}} (g_1(\boldsymbol{x}_{s-1}))=0$$

and using (2.6), we have then

$$\underset{\boldsymbol{x}_{s-1}}{\mathrm{E}} (g_1(\boldsymbol{x}_{s-1}))=0 \;.$$

Again, the completeness of the $(s-1)$-variable family implies

$$g_1(\boldsymbol{x}_{s-1})=0$$

for all relevant $\boldsymbol{x}_{s-1}$, $\boldsymbol{N}_{s-1}$. Continuing inductively on N_s one then has

$$g_j(\boldsymbol{x}_{s-1})=g(x_1,\cdots,x_{s-1},j)=0$$

for all j, and hence

$$g(\boldsymbol{x}_s)=0$$

for all relevant $\boldsymbol{x}_s$ and $\boldsymbol{N}_s$. The completeness of the $(s-1)$-variable family is seen to imply the completeness of the s-variable family.

To complete the proof of Theorem 2.1, we must show that the completeness of the family with $s=1$. One has

$$p_{N_1}(x)=\frac{\binom{-N-1}{-k-x}\binom{N_1}{x}}{\binom{-N_0-1}{-k}}$$

and we have to prove $\underset{N_1}{\mathrm{E}}(g(x)) \equiv 0 \Rightarrow g(x) \underset{x}{=} 0$.

$$\sum_x g(x)\frac{\binom{-N-1}{-k-x}\binom{N_1}{x}}{\binom{-N_0-1}{-x}} \equiv 0 \;.$$

Let $N_1=0$ then $g(0)\cdot 1+g(1)\cdot 0+\cdots+0\equiv 0$. Therefore $g(0)=0$. Let $N_1=1$ then $g(0)p_1(0)+g(1)p_1(1)=0$. Therefore $g(1)=0$. Continuing this way, we get $g(x)=0$ for all $x=0,1,2,\cdots,N_1$.

3. ML estimates of the parameters of MNH and MNIH

Since the ML estimates of the parameters are usually asymptotically more efficient than their moment estimates, we will discuss in this section the problem of obtaining ML estimates of the parameters of MNH and MNIH. Special cases of the results below are the results of Dubey [2] for beta-binomial and beta-Pascal distributions when $s=1$.

The logarithmic likelihood function for MNIH as defined in Table 1 is

$$\begin{aligned}(3.1)\qquad & \log L(N_0, N_1, \cdots, N_s) \\ &= \log \Gamma(k+x) + \log \Gamma(N) - \log \Gamma(k+x+N) \\ &\quad + \log \Gamma(k+N_0) - \log \Gamma(k) - \log \Gamma(N_0) \\ &\quad + \sum_{i=1}^{s} \log \Gamma(N_i + x_i) - \sum_{i=1}^{s} \log \Gamma(N_i) - \sum_{i=1}^{s} \log x_i! \,.\end{aligned}$$

This yields the $(s+1)$ likelihood equations as

$$(3.2)\qquad \frac{\partial \log L}{\partial N_0} = \Psi_{N_0}(N) - \Psi_{N_0}(k+x+N) + \Psi_{N_0}(k+N_0) - \Psi_{N_0}(N_0) = 0$$

and

$$(3.3)\qquad \frac{\partial \log L}{\partial N_i} = \Psi_{N_i}(N) - \Psi_{N_i}(k+x+N) + \Psi_{N_i}(N_i+x_i) - \Psi_{N_i}(N_i) = 0 \qquad \text{for } i=1, 2, \cdots, s$$

where $\Psi_v(u+v) = (d/dv) \log \Gamma(u+v)$ which is called the psi-function (Davis [1]). Also from Davis ([1], p. 282) we have

$$(3.4)\qquad \Psi(u) = -0.57722\cdots - \sum_{j=0}^{\infty} \left(\frac{1}{u+j} - \frac{1}{j+1} \right)$$

for positive real u. Using (3.4) we can write the (s+1) likelihood equations (3.2) and (3.3) as

$$(3.5)\qquad \sum_{j=0}^{k+x-1} \frac{1}{N+j} - \sum_{j=0}^{k-1} \frac{1}{N_0+j} = 0$$

and

$$(3.6)\qquad \sum_{j=0}^{k+x-1} \frac{1}{N+j} - \sum_{j=0}^{x_i-1} \frac{1}{N_i+j} = 0 \qquad \text{for } i=1, 2, \cdots, s\,.$$

We can solve the $(s+1)$ equations (3.5) and (3.6) iteratively by any suitable numerical methods. In this connection we may use the moment estimates of N_i ($i=0, 1, 2, \cdots, s$) obtained by Mosimann [10] as initial estimates or trial values.

The second order derivatives of the logarithmic likelihood function are:

$$(3.7)\quad \frac{\partial^2 \log L}{\partial N_0^2} = \Psi_{N_0^2}^{(1)}(N) - \Psi_{N_0^2}^{(1)}(k+x+N) + \Psi_{N_0^2}^{(1)}(k+N_0) - \Psi_{N_0^2}^{(1)}(N_0)$$

$$(3.8)\quad \frac{\partial^2 \log L}{\partial N_i^2} = \Psi_{N_i^2}^{(1)}(N) - \Psi_{N_i^2}^{(1)}(k+x+N) + \Psi_{N_i^2}^{(1)}(N_i+x_i) - \Psi_{N_i^2}^{(1)}(N_i) \qquad i=1, 2, \cdots, s$$

$$(3.9)\quad \frac{\partial^2 \log L}{\partial N_i \partial N_j} = \Psi_{N_i, N_j}^{(1)}(N) - \Psi_{N_i, N_j}^{(1)}(k+x+N) \qquad i \neq j = 1, 2, \cdots, s$$

where

$$\Psi_{N_i^2}^{(1)}(\cdot) = \frac{\partial}{\partial N_i} \Psi_{N_i}(\cdot) = \frac{\partial^2}{\partial N_i^2} \log \Gamma(\cdot) \qquad \text{for } i=0, 1, 2, \cdots, s$$

$$\Psi_{N_i, N_j}^{(1)}(\cdot) = \frac{\partial}{\partial N_i} \Psi_{N_j}(\cdot) = \frac{\partial^2}{\partial N_i \partial N_j} \log \Gamma(\cdot) \qquad \text{for } i \neq j = 1, 2, \cdots, s .$$

Using the formula for the polygamma (Davis [1], Vol. II, p. 111) functions

$$\Psi^{(m)}(u) = \frac{d^m}{du^m} \Psi(u) = (-1)^{m+1} m! \sum_{j=0}^{\infty} \frac{1}{(u+j)^{m+1}}$$

valid for $m > 0$ and positive real u, we can rewrite (3.7) to (3.9) as

$$(3.10)\quad \frac{\partial^2 \log L}{\partial N_0^2} = \left[\sum_{j=0}^{k-1} \frac{1}{(N_0+j)^2} - \sum_{j=0}^{k+x-1} \frac{1}{(N+j)^2} \right]$$

$$(3.11)\quad \frac{\partial^2 \log L}{\partial N_i^2} = \left[\sum_{j=0}^{x_i-1} \frac{1}{(N_i+j)^2} - \sum_{j=0}^{k+x-1} \frac{1}{(N+j)^2} \right] \qquad \text{for } i=1, 2, \cdots, s$$

$$(3.12)\quad \frac{\partial^2 \log L}{\partial N_i \partial N_j} = \sum_{j=0}^{k+x-1} \frac{1}{(N+j)^2} .$$

Expressions (3.10) to (3.12) could be used to obtain an estimate of the asymptotic variance-covariance matrix of ML estimates of N_i ($i=0, 1, 2, \cdots, s$). Exactly identical approach will enable one to obtain the ML estimates of the parameters of MNH and also an estimate of their variance-covariance matrix. The ML estimates of the parameters of multivariate Polya (MP) and Inverse Polya (MIP) distributions can be obtained on the same lines as above. The moment estimates of the parameter of MP and MIP have been obtained in Janardan [6].

4. Bayesian estimation of the parameters of MH and MIH

In this section, we will consider the Bayesian estimation of the parameters of MH using UMH as a prior distribution and then we will consider the Bayesian estimation of the parameters of MIH using MNIH as a prior distribution. We follow the two-stage approach (Sarndal [11], Hoadley [5]) for generating the prior distribution as MNIH, first by setting up a parametric superpopulation and then choosing a prior distribution for the parameters of the superpopulation.

Following an approach in sample survey theory let a finite population $\mathcal{P}$ of N distinguishable elements be identified by the tags $1, 2, \cdots, N$ and let $z=(z_1, z_2, \cdots, z_n)$ where z_i is the unknown value of some characteristic possessed by the ith individual of the population. Let

$$\mu=\frac{1}{N}\sum_{i=1}^{N} z_i \tag{4.1}$$

$$\mu'_r=\frac{1}{N}\sum_{i=1}^{N} z_i^r \tag{4.2}$$

and

$$\sigma^2=\frac{1}{N}\sum_{i=1}^{N} (z_i-\mu)^2 \tag{4.3}$$

be the population mean, the rth crude moment and the population variance. Here we obtain the Bayes's estimation of these functions of z.

Suppose that each z_i can only assume one of the finite set of numerical values $\{y_0, y_1, \cdots, y_s\}$ where $y_0<y_1<\cdots<y_s$ and s is assumed to be independent of the population size N. Now let N_j be the unknown numbers of the N population elements for which $z_i=y_i$, $j=0, 1, 2, \cdots, s$. Then the finite population $\mathcal{P}$ can be completely described (Hartley and Rao [4]) by $(s+1)$ nonnegative integer parameters N_i $(i=0, 1, \cdots, s)$ and (4.1) to (4.3) can be written as

$$\mu=\frac{1}{N}\sum_{j=0}^{s} y_j N_j \tag{4.4}$$

$$\mu'_r=\frac{1}{N}\sum_{j=0}^{s} y_j^r N_j \tag{4.5}$$

$$\sigma^2=\frac{1}{N}\sum_{j=0}^{s} y_j^2 N_j-\mu^2 . \tag{4.6}$$

If A is any subset of n elements taken from $\mathcal{P}$ of the N distinct population elements and if x_j denotes the number of z_i's $(i \in A)$ equal to y_j then the joint distribution of $\boldsymbol{x}=(x_1, x_2, \cdots, x_s)$ is clearly given by MH distribution as defined in Table 1.

If B denotes a subset of elements drawn without replacement from finite population $\mathscr{P}$ until we get k elements z_i's equal to y_0. If x_j denotes the number of those z_i's $(i \in B)$ equal to y_j then the joint distribution of $\boldsymbol{x}$ is clearly given by MIH distribution as defined in Table 1.

4.1. *Bayesian estimates of the parameters of MH*

Let the prior distribution be $\mathrm{p}(N)$, then the posterior distribution of N given $\boldsymbol{x}$ is given by the Bayes's theorem

$$\mathrm{p}(\boldsymbol{N}|\boldsymbol{x})=\frac{\mathrm{p}(\boldsymbol{N})\,\mathrm{p}(\boldsymbol{x}|\boldsymbol{N})}{\sum_{\boldsymbol{N}} \mathrm{p}(\boldsymbol{N})\,\mathrm{p}(\boldsymbol{x}|\boldsymbol{N})}\,. \tag{4.7}$$

THEOREM 4.1. *If* $\mathrm{p}(\boldsymbol{N})$ *is taken to be the UMH distribution given by*

$$\mathrm{p}(\boldsymbol{N})=\frac{\prod_{i=0}^{s}\binom{a_i}{N_i}}{\binom{a}{N}} \tag{4.8}$$

where $N_0+N_1+\cdots+N_s=N>0$, *and* $a_0+a_1+\cdots+a_s=a$ *and* $\mathrm{p}(\boldsymbol{x}|\boldsymbol{N})$ *is MH distribution, then the posterior distribution of* $\boldsymbol{N}$ *given* $\boldsymbol{x}$ *is*

$$\mathrm{p}(\boldsymbol{N}|\boldsymbol{x})=\frac{\prod_{i=0}^{s}\binom{a_i-x_i}{N_i-x_i}}{\binom{a-n}{N-n}}=\binom{N-n}{\boldsymbol{N}-\boldsymbol{x}}\frac{\Gamma(-a+n)}{\Gamma(-a+N)}\prod_{i=0}^{s}\frac{\Gamma(-a_i+N_i)}{\Gamma(-a_i+x_i)}\,. \tag{4.9}$$

PROOF. Straightforward on using (4.7).

We may note here that in (4.9) $\boldsymbol{N}-\boldsymbol{x}$ given $\boldsymbol{x}$ is a UMH with parameters $N-n$ and $a-n$, $\boldsymbol{a}-\boldsymbol{x}$, a_0-x_0.

Properties of the prior on $\boldsymbol{N}$. Before proceeding with the analysis of the aspects of the posterior distributions, it is useful for future use to list some properties of some characteristics (4.4) to (4.6) of the finite population. The prior expectation of the population moments are:

$$\mathrm{E}(\mu)=\sum_{j=0}^{s} y_j a_j/a \tag{4.10}$$

$$\mathrm{E}(\mu_r')=\sum_{j=0}^{s} y_j^r a_j/a\,. \tag{4.11}$$

The prior variance of the population mean and the rth crude moment are

(4.12) $$\mathrm{V}(\mu)=\frac{a-N}{N(a-1)}\left[\sum_{j=0}^{s} y_j^2 a_j/a-\left(\sum_{j=0}^{s} y_j a_j/a\right)^2\right]$$

(4.13) $$\mathrm{V}(\mu_r')=\frac{a-N}{N(a-1)}\left[\sum_{j=0}^{s} y_j^{2r} a_j/a-\left(\sum_{j=0}^{s} y_j^r a_j/a\right)^2\right]$$

and the prior expectation of the population variance is,

(4.14) $$\mathrm{E}(\sigma^2)=\frac{(N-1)a}{N(a-1)}\left[\sum_{j=0}^{s} y_j^2 a_j/a-\left(\sum_{j=0}^{s} y_j a_j/a\right)^2\right].$$

Properties of the posterior of N given $\boldsymbol{x}$. Writing $M_i=N_i-x_i$ and $M=N-n$ we can write down the posterior moments of UMH as

(4.15) $$\mathrm{E}(M_i\,|\,\boldsymbol{x})=M\frac{(a_i-x_i)}{(a-n)}$$

(4.16) $$\mathrm{E}(M_i^2\,|\,\boldsymbol{x})=M(M-1)\frac{(a_i-x_i)(a_i-x_i+1)}{(a-n)(a-n+1)}+M\frac{(a_i-x_i)}{(a-n)}$$

(4.17) $$\mathrm{E}(M_iM_j\,|\,\boldsymbol{x})=M(M-1)\frac{(a_i-x_i)(a_j-x_j)}{(a-n)(a-n+1)}$$

(4.18) $$\mathrm{V}(M_i\,|\,\boldsymbol{x})=M\frac{(a_i-x_i)(a-n-a_i+x_i)(a-N)}{(a-n)^2(a-n+1)}$$

(4.19) $$\mathrm{Cov}(M_i,\,M_j\,|\,\boldsymbol{x})=-M\frac{(a_i-x_i)(a_j-x_j)(a-N)}{(a-n)^2(a-n+1)}.$$

Posterior expectations and variances of the functions of the parameters of the finite population. Since

$$\mu=\frac{1}{N}\sum_{j=0}^{s} y_jN_j=\frac{1}{N}\sum_{j=0}^{s} y_i(M_j+x_j)=\frac{1}{N}\sum_{j=0}^{s} y_jM_j+\frac{1}{N}\sum_{j=0}^{s} y_jx_j\,.$$

Let $\bar{x}=\frac{1}{n}\sum_{j=0}^{s} y_jx_j$ be the sample mean, then

(4.20) $$\mu=\frac{1}{N}\sum_{j=0}^{s} y_jM_j+\frac{n}{N}\bar{x}$$

and using (4.15) and (4.10) we get the posterior expectation of μ given $\boldsymbol{x}$ as

(4.21) $$\mathrm{E}(\mu\,|\,\boldsymbol{x})=\frac{n(a-N)\bar{x}+(N-n)a\,\mathrm{E}(\mu)}{N(a-n)}$$

(4.22) $$=w\bar{x}+(1-w)\,\mathrm{E}(\mu)$$

where $w=n(a-N)/(N(a-n))$.

(4.22) shows that the posterior expectation is weighted average of the sample mean and the prior expectation. If a is small, which we could interpret as "diffuse" or "noninformative" prior, then

$$\mathrm{E}(\mu|\boldsymbol{x})\simeq\bar{x}\,. \tag{4.23}$$

Similarly we have the posterior expectation of the rth crude moment as

$$\mathrm{E}(\mu_r'|\boldsymbol{x})=\frac{(N-n)a\,\mathrm{E}(\mu_r')+n(a-n)m_r'}{N(a-n)}$$

where $m_r'=\frac{1}{n}\sum_{j=0}^{s}y_j^r x_j$ is the rth sample moment. If a is small, then

$$\mathrm{E}(\mu_r'|\boldsymbol{x})\simeq m_r'\,. \tag{4.24}$$

Since $\mu_r'=\frac{1}{N}\sum_{j=0}^{s}y_j^r N_j=\frac{1}{N}\sum_{j=0}^{s}y_j^r M_j+\frac{n}{N}m_r'$ and

$$\mathrm{V}(\mu_r'|\boldsymbol{x})=\frac{1}{N^2}\sum_{j=0}^{s}y_j^{2r}\,\mathrm{V}(M_j|\boldsymbol{x})+\frac{1}{N^2}\sum_{i\neq j}y_i^r y_j^r\,\mathrm{Cov}(M_i,M_j|\boldsymbol{x})\,.$$

Substituting from (4.18) and (4.19), we have

$$\mathrm{V}(\mu_r'|\boldsymbol{x})=\frac{M(a-N)}{N^2}\Bigg[\sum_{j=0}^{s}y_j^{2r}\frac{(a_j-x_j)(a-n-a_j+x_j)}{(a-n)^2(a-n+1)} -\sum_{i\neq j}y_i^r y_j^r\frac{(a_i-x_i)(a_j-x_j)}{(a-n)^2(a-n+1)}\Bigg] \tag{4.25}$$

and

$$\mathrm{V}(\mu|\boldsymbol{x})=\frac{(N-n)(a-N)}{N^2}\Bigg[\sum_{i=0}^{s}y_i^2\frac{(a_i-x_i)(a-n-a_i+x_i)}{(a-n)^2(a-n+1)} -\sum_{i\neq j}y_i y_j\frac{(a_i-x_i)(a_j-x_j)}{(a-n)^2(a-n+1)}\Bigg]\,. \tag{4.26}$$

Let the sample variance be

$$S^2=\frac{\sum_{j=0}^{s}y_j^2x_j-n\bar{x}^2}{(n-1)}\,.$$

Then we can express the posterior variance of μ given $\boldsymbol{x}$ in terms of S^2, $\bar{x}$, the posterior expectation and prior variance as

$$\mathrm{V}(\mu|\boldsymbol{x})=\Bigg[\frac{(N-n)(a-N)(n-1)S^2}{N^2(a-n)(a-n+1)}+\frac{a(a-1)(N-n)\,\mathrm{V}(\mu)}{N(a-n)(a-n+1)} +\frac{na(a-N)(N-n)(\bar{x}-\mathrm{E}(\mu))^2}{N^2(a-n)^2(a-n+1)}\Bigg]\,. \tag{4.27}$$

If a is chosen to be very small, that is, when we have no informative prior, then

$$\text{(4.28)} \qquad \mathrm{V}(\mu|\boldsymbol{x}) \simeq \frac{(N-n)}{N} \frac{(n-1)}{(n+1)} \frac{S^2}{n} \approx \frac{N-n}{N} \frac{S^2}{n} .$$

Thus under extreme diffuseness of prior knowledge the posterior distribution of the finite population mean μ has an approximate mean $\bar{x}$ and variance $((N-n)/N)\cdot(S^2/n)$. This provides a Bayesian interpretation for the classical unbiased estimates in traditional sample survey theory.

Since $\sigma^2 = \frac{1}{N} \sum_{j=0}^{s} y_j^2 N_j - \mu^2$ we get the posterior expectation of σ^2 as

$$\mathrm{E}(\sigma^2|\boldsymbol{x}) = \frac{1}{N}\Big[\sum_{j=0}^{s} y_j^2 \{x_j + \mathrm{E}(M_j|\boldsymbol{x}) - N\,\mathrm{V}(\mu|\boldsymbol{x}) - N\,\mathrm{E}^2(\mu|\boldsymbol{x})\} \Big]$$

and substituting from (4.15), (4.26) and (4.21) for $\mathrm{E}(M_j|\boldsymbol{x})$, $\mathrm{V}(\mu|\boldsymbol{x})$ and $\mathrm{E}(\mu|\boldsymbol{x})$ respectively we get

$$\text{(4.29)} \qquad \mathrm{E}(\sigma^2|\boldsymbol{x}) = \frac{1}{N}\Big[\frac{(a-N)(N(a-n)-n)(n-1)S^2}{N(a-n)(a-n+1)} + \frac{M(a-1)(N(a-n)-a)\,\mathrm{E}(\sigma^2)}{(N-1)(a-n)(a-n+1)} + \frac{nMa(a-n)(\bar{x}-\mathrm{E}(\mu))^2}{N(a-n)(a-n+1)} \Big] .$$

If 'a' very small, $\mathrm{E}(\sigma^2|\boldsymbol{x}) \simeq ((N+1)/N)\cdot((n-1)/(n+1))S^2$, where S^2 is the observed sample variance.

Inference about population and second sample. The posterior distribution of $\boldsymbol{N}$ given $\boldsymbol{x}$ obtained in (4.9) can also be written as

$$\text{(4.30)} \qquad \mathrm{p}(\boldsymbol{N}|\boldsymbol{x}) = \frac{\prod_{i=0}^{s} \binom{N_i - a_i - 1}{x_i - a_i - 1}}{\binom{N-a-1}{n-a-1}} .$$

If we take $M_i = N_i - x_i$ in (4.30), then we get the distribution of the second sample $\boldsymbol{M}$ given the first sample $\boldsymbol{x}$ (where $N-n$ is interpreted as a second sample after taking a sample of n from a finite population of N elements).

$$\text{(4.31)} \qquad \mathrm{p}(\boldsymbol{M}|\boldsymbol{x}) = \frac{\prod_{i=0}^{s} \binom{M_i + x_i - a_i - 1}{x_i - a_i - 1}}{\binom{M+n-a-1}{n-a-1}} .$$

Now the total proportion of the population that possesses the characteristics $y_0, y_1, \cdots, y_i$ is

$$\frac{H_i}{N}=\sum_{k=0}^{i}\frac{N_k}{N}.$$

Hence, the joint probability of H_i's is obtained from (4.30) as

$$\text{(4.32)}\qquad \mathrm{p}(\boldsymbol{H})=\frac{\prod_{i=0}^{s}\binom{H_i-H_{i-1}-a_i-1}{x_i-a_i-1}}{\binom{N-a-1}{n-a-1}}$$

where $\sum_{k=1}^{i} x_k \leqq H_i \leqq N-n+\sum_{k=1}^{i} x_k$, $i=1, 2, \cdots, s$ and $H_{-1}=0$, $H_s=N$.

The total proportion of the second sample that possesses the characteristic $y_0, y_1, \cdots, y_i$ is

$$\sum_{k=0}^{i}\frac{M_i}{M}=\frac{h_i}{M}.$$

Hence, the joint probability of h_i's is obtained from (4.31) as

$$\text{(4.33)}\qquad \mathrm{p}(\boldsymbol{h})=\frac{\prod_{i=0}^{s}\binom{h_i-h_{i-1}+x_i-a_i-1}{x_i-a_i-1}}{\binom{M+n-a-1}{n-a-1}}.$$

Asymptotic form of the posterior distribution of N given $\boldsymbol{x}$. Let $N_i \to \infty$ $(i=0, 1, 2, \cdots, s)$ and $N\to\infty$ such that

$$\mathrm{p}_i=\lim_{N\to\infty}\frac{N_i}{N}\qquad\text{and}\qquad (x_i-a_i)$$

remain fixed and positive, then clearly $0\leqq \mathrm{p}_i \leqq 1$ and $\sum_{i=1}^{s}\mathrm{p}_i\leqq 1$, we therefore obtain

$$\text{(4.34)}\qquad \lim_{N\to\infty} N^s\,\mathrm{p}(N|\boldsymbol{x})=\frac{(n-a-1)!}{\prod_{i=0}^{s}(x_i-a_i-1)!}\prod_{i=0}^{s}\mathrm{p}_i^{(x_i-a_i-1)}$$
$$=\frac{\Gamma(n-a)}{\prod_{i=0}^{s}\Gamma(x_i-a_i)}\prod_{i=0}^{s}\mathrm{p}_i^{(x_i-a_i-1)}$$

which is Dirichlet distribution with parameters (x_0-a_0) and $(\boldsymbol{x}-\boldsymbol{a})$. See Wilks [14].

Asymptotic form of $\mathrm{p}(\boldsymbol{H})$. Let $N_i\to\infty$ $(i=0, 1, 2, \cdots, \infty)$, $N\to\infty$ such that $d_i=\lim_{N\to\infty} H_i/N$ and (x_i-a_i) remain fixed and positive, then clearly

$$0 \leqq d_1 \leqq d_2 \leqq \cdots \leqq d_s \leqq 1$$

and

$$(4.35) \qquad \lim_{N\to\infty} N^s \, \mathrm{p}(\boldsymbol{H}) = \frac{(n-a-1)!}{\prod_{i=0}^{s}(x_i-a_i-1)!} \prod_{i=0}^{s} (d_i-d_{i-1})^{(x_i-a_i-1)}$$
$$= \frac{\Gamma(n-a)}{\prod_{i=0}^{s}\Gamma(x_i-a_i)} \prod_{i=0}^{s} (d_i-d_{i-1})^{x_i-a_i-1}$$

where $d_{-1}=0$, which is ordered Dirichlet distribution with parameters (x_0-a_0) and $(\boldsymbol{x}-\boldsymbol{a})$. See Wilks [14].

Asymptotic form of the posterior distribution of second sample given the first sample. Let $n\to\infty$ *and* let $\lim_{n\to\infty}(x_i-a_i-1)/(n-a-1)=\theta_i$, in (4.31), then

$$\lim_{n\to\infty} \mathrm{p}(\boldsymbol{M}|\boldsymbol{x}) = \lim_{n\to\infty} \left\{ \frac{M!}{\prod_{i=1}^{s} M_i!} \frac{\prod_{i=0}^{s}(x_i-a_i-1)^{(M_i)}}{(n-a-1)^{(M)}} \right\}$$
$$\approx \lim_{n\to\infty} \frac{M!}{\prod_{i=0}^{s} M_i!} \frac{\prod_{i=0}^{s}(x_i-a_i-1)^{M_i}}{(n-a-1)^{M}}$$
$$= \frac{M!}{\prod_{i=0}^{s} M_i!} \prod_{i=0}^{s} \theta_i^{M_i} .$$

4.2. *Bayesian estimates of the parameters of MIH*

Assume that the finite population of interest is a random sample obtained in inverse sampling from an infinite superpopulation for which p_j is the probability of category j. Then, conditional on p_j's, $\boldsymbol{N}$ has the NM distribution with parameters N_0+1 and $\boldsymbol{p}$. Now if the parameter vector $\boldsymbol{p}$ is assumed to follow Dirichlet distribution with parameters α_0 and $\boldsymbol{\alpha}$, then Mosimann [10] has shown that $\boldsymbol{N}$ follows MNIH with parameters N_0+1 and $\boldsymbol{\alpha}$. We show in Theorem 6.3 that $\boldsymbol{N}-\boldsymbol{x}$ given $\boldsymbol{x}$ is distributed as MNIH with parameters N_0-k+1 and $\boldsymbol{\alpha}+\boldsymbol{x}$.

THEOREM 4.2. *Let the conditional distribution of* $\boldsymbol{x}$ *given* $\boldsymbol{N}$ *be* MIH *with parameters* k *and* $\boldsymbol{N}$. *Let the prior distribution on* $\boldsymbol{N}$ *be* MNIH *with parameters* N_0+1 *and* $\boldsymbol{\alpha}$ *then* (i) *the unconditional distribution of* $\boldsymbol{x}$ *is* MNIH *with parameters* k *and* $\boldsymbol{\alpha}$ *and* (ii) *the posterior distribution of* $\boldsymbol{N}-\boldsymbol{x}$ *given* $\boldsymbol{x}$ *is* MNIH *with the pf*

$$p(\boldsymbol{N}-\boldsymbol{x}\,|\,\boldsymbol{x})=\frac{B(N-k-x+1,\,k+x+\alpha)}{B(\alpha_0+k,\,N_0-k+1)}\prod_{i=1}^{s}\frac{\Gamma(\alpha_i+N_i)}{\Gamma(\alpha_i+x_i)(N_i-x_i)!}\ .$$

PROOF. See Theorem 5.3 in Janardan [7].

The posterior moments of $\boldsymbol{M}$

(4.36) $$E(M_i\,|\,\boldsymbol{x})=(N_0-k+1)\frac{\alpha_i+x_i}{\alpha_0+k-1}\ ,$$

where $M_i=N_i-x_i,\ i=1,2,\cdots,s$

(4.37) $$V(M_i\,|\,\boldsymbol{x})=(N_0-k+1)\left(\frac{\alpha_i+x_i}{\alpha_0+k-1}\right)\left(1+\frac{\alpha_i+x_i}{\alpha_0+k-1}\right)\left(\frac{\alpha_0+N_0}{\alpha_0+k-2}\right)$$

(4.38) $$\mathrm{Cov}(M_i,\,M_j\,|\,\boldsymbol{x})=(N_0-k+1)\left(\frac{\alpha_i+x_i}{\alpha_0+k-1}\right)\left(\frac{\alpha_j+x_j}{\alpha_0+k-1}\right)\left(\frac{\alpha_0+N_0}{\alpha_0+k-2}\right)\ .$$

In this case also, one can write down expressions similar to (4.21) through (4.28) for the posterior expectations and variances of the functions of the parameters of the finite population utilizing expressions (4.36) to (4.38) for the posterior moments of $\boldsymbol{M}$, obtained in inverse sampling.

Acknowledgement

The author is grateful to the referee for his helpful comments on this paper, and thanks Dr. Mosimann for his helpful suggestions on the proof of Theorem 2.1.

SANGAMON STATE UNIVERSITY

Table 1

Name of the Distribution	Probability Function	Domain of Distributions Parameter Restrictions
Multivariate Hypergeometric	$\binom{N_0}{n-x}\prod_{i=1}^{s}\binom{N_i}{x_i}\Big/\binom{N}{n}$	x_i an integer, $0\leqq x_i\leqq\min(n, N_i)$ $\sum_1^s x_i\geqq n-\left(N-\sum_1^s N_i\right)$ $N_i=1, 2, \cdots, N$ for all i, $N_0=N-\sum_1^s N_i$
Multivariate Inverse Hypergeometric	$\binom{-N-1}{-k-x}\prod_{i=1}^{s}\binom{N_i}{x_i}\Big/\binom{-N_0-1}{-k}$	$x_i=0, 1, 2, \cdots, \infty$, $0<k<\infty$ $0<N_i$, $N<\infty$, $N_0=N-\sum_1^s N_i$
Multivariate Negative Hypergeometric	$\frac{n!\Gamma(N)}{\Gamma(N+n)}\prod_{i=0}^{s}\frac{\Gamma(N_i+x_i)}{\Gamma(N_i)x_i!}$	$x_i=0, 1, 2, \cdots, \infty$, $x_0=n-\sum_{i=1}^s x_i$ $0<N_i$, $N<\infty$, $N_0=N-\sum_1^s N_i$
Multivariate Negative Inverse Hypergeometric	$\frac{B(k+x, N)}{B(k, N_0)}\prod_{i=1}^{s}\frac{\Gamma(N_i+x_i)}{\Gamma(N_i)x_i!}$	$x_i=0, 1, 2, \cdots, \infty$, $0<k<\infty$ $0<N_i$, $N<\infty$, $N_0=N-\sum_1^s N_i$
Unified Multivariate Hypergeometric	$\binom{a_0}{n-x}\prod_{i=1}^{s}\binom{a_i}{x_i}\Big/\binom{a}{n}$	$x_i=0, 1, 2, \cdots$ for all i $x=\sum_1^s x_i$ $\quad a=\sum_{i=1}^s a_i$ $\quad a_i$ real

References

[1] Davis, H. T. (1963). *Tables of the Mathematical Functions*, Vols. I and II, The Principia University Press.

[2] Dubey, Satya D. (1966). Compound Pascal distributions, *Ann. Inst. Statist. Math.*, **18**, 357-365.

[3] Hald, A. (1960). The compound hypergeometric distribution and a system of single sampling inspection plans based on prior distributions and costs, *Technometrics*, **2**, 275-340.

[4] Hartley, H. O. and Rao, J. N. K. (1968). A new estimation theory for sample surveys, *Biometrika*, **55**, 547-557.

[5] Hoadley, Bruce (1969). The compound multinomial distribution and Bayesian analysis of categorical data from finite populations, *J. Amer. Statist. Ass.*, **64**, 216-229.

[6] Janardan, K. G. (1970). On a class of multivariate hypergeometric models, a Ph.D. thesis submitted to the graduate school of Penn State University, U.S.A.

[7] Janardan, K. G. (1973). Chance mechanisms for multivariate hypergeometric models, to appear in *Sankhya*.

[8] Janardan, K. G. and Patil, G. P. (1972). A unified approach for a class of multivariate hypergeometric models, *Sankhya*, **34** (4), 363-376.

[9] Mosimann, James E. (1962). On the compound multivariate distribution, the multivariate β-distribution and correlations among proportions, *Biometrika*, **49**, 65-82.

[10] Mosimann, James E. (1963). On the compound negative multinomial distribution and correlations among inversely sampled pollen counts, *Biometrika*, **50**, 47-54.

[11] Sarndal, Carl-Erik (1964). A unified derivation of some nonparametric distributions, *J. Amer. Statist. Ass.*, **59**, 1042-1053.

[12] Sarndal, Carl-Erik (1965). Derivation of a class of frequency distributions via Bayes' theorem, *J. R. Statist. Soc.* Ser. B, **27**, 290-300.

[13] Wilks, S. S. (1960). A two-stage scheme for sampling without replacement, *Bull. Int. Statist. Inst.*, **37** (2), 241-248.

[14] Wilks, S. S. (1962). *Mathematical Statistics*, John Wiley & Sons, Inc., New York.