SOME CONSIDERATIONS ON THE MULTIPLE INVERSE SAMPLING METHOD

Kudo, Akio
Kyushu University

Yao, Jing-Shing
Kyushu University

KYUSHU UNIVERSITY

SOME CONSIDERATIONS ON THE MULTIPLE INVERSE SAMPLING METHOD

By

Akio KUDÔ and Jing-Shing YAO

§ 1. Introduction.

The multiple inverse sampling is, in this peper, understood as follows. Suppose we have a multinomial population with m exclusive events, E_1, E_2, $\cdots$, E_m, with the respective probabilities P_1, P_2, $\cdots$, P_m. We have m prefixed non-negative integers R_1, R_2, $\cdots$, R_m, and we continue sampling until $R_i \leq r_i$ ($i = 1, \cdots, m$) is satisfied, where r_i is the number of occurrences of E_i respectively.

Let k_i denote $r_i - R_i$ for $i = 1, 2, \cdots, m$. At the termination of the sampling, therefore, at least one of k_i's ($i = 1, 2, \cdots, m$) is zero, and the remains are all non-negative.

When $m = 2$ and $R_2 = 0$, the distribution of k_2 is called the negative binomial distribution, Feller [6] pp. 155–157, and when m is more than 2 and all R_i's are zero except for R_1, it may be legitimate to call the joint distribution of $k_2, \cdots, k_m$ the negative multinomial distribution.

This type of sampling seems to have some importance but does not seem to have been systematically treated so far.

The authors may mention the works of D. G. Chapman [2], M. H. De-Groot [3], C. E. Deiulefait [4], B Epstein and M. Sobel [5], D. J. Finney [7], M. A. Girshick, F. Mosteller, and L. J. Savage [8], J. B. S. Haldane [9], J. Nadler [12], M. C. K. Tweedie [15], which are all of the works pertinent to this type of sampling so far as the authors are aware of.

The most of the works seem to be concerned with the simple rather than multiple inverse sampling and the results are on the estimation of the parameters rather than the test of significance, or more generally the multiple decision problems. This paper is motivated by a sampling problem in a large scale survey now proceeding, and we have forcused our attention to the problems to meet our immediate needs. We have so far obtained the result mostly in the case $m = 2$.

In § 2 the distribution of $K = k_1 + k_2$, etc. and their means and variances will be obtained, which, of course, depend on the underlying probabilities

p_1 and p_2, and prefixed numbers R_1 and R_2. The behaviours of the mean and the variance is not very easy to analyse but some numerical considerations have been made and some interesting findings will be presented.

In § 3, we shall derive a uniformly most powerful test of the hypothesis that the underlying probability is a specified value. In § 4, a symmetric multiple decision problem will be discussed. § 5 is devoted to the derivation of the limit distribution when R_1 and R_2 are large. In § 6, there will be presented the outline of the survey by which this paper is motivated and the application of the uniformly most powerful test obtained in § 3.

The authors are grateful for the encouragements and criticism given by Professor T. Kitagawa while this paper is being worked out. The authors express their gratitude to Mr. N. Furukawa for his advices in surveying previous works of this line, and also assistance in the manuscript.

§ 2. Distribution, expectation and variance.

Let E_1 and E_2 be two exclusive events with the probabilities p and $q = 1 - p$ respectively. The probability that the sampling stated in the introduction terminates at the (R_1+R_2+k)-th. trial is equal to the probability that at the (R_1+R_2+k)-th. trial either (a) : E_1 occurs and this is the R_1-th occurence of E_1 or (b) : E_2 occurs and this is the R_2-th occurence of E_2. By considering that E_2 and E_1 occured exactly $R_2 + k$ and $R_1 + k$ times in each case respectively, the probability is found to be

$$(2.1) \qquad P_r(k) = \binom{R_1+R_2+k-1}{R_1-1} p^{R_1} q^{R_2+k} + \binom{R_1+R_2+k-1}{R_2-1} p^{R_1+k} q^{R_2},$$

$$k = 0, 1, 2, \cdots,$$

where the first term represents the event (a) and the second the event (b). By making use of the recurrence relation concerning the incomplete beta functions

$$(2.2) \qquad I_p(r, N+1) - I_p(r, N) = \frac{(N+r-1)!}{(r-1)!N!} \, p^r (1-p)^N,$$

$$(r, \ N \text{ positive integers})$$

the cummulative distribution function is given by

$$(2.3) \qquad F(k_0) = \sum_{k=0}^{k_0} P_r(k) = I_p(R_1, \ R_2, +k_0+1) - I_p(R_1 + k_0 + 1, \ R_2),$$

$$k_0 = 0, 1, 2, \cdots,$$

and also the probability that the sampling terminates with the occurence of E_1 is

$$(2.4) \qquad \sum_{k=0}^{\infty} \binom{R_1+R_2+k-1}{R_1-1} p^{R_1} q^{R_2+k} = \sum_{k=R_2}^{\infty} \binom{R_1+k-1}{R_1-1} p_{R_1} q_k = 1 - I_p(R_1, R_2),$$

from which follows that with E_2 is $I_p \ (R_1, \ R_2)$.

The probability that the sampling terminates at the $(R_1 + R_2 + k)$-th. trial with E_1 is the first term of (2.1), and that with E_2 is the second.

The corresponding sums are given by

$$(2.5) \quad F_2(k_0) = \sum_{k=0}^{k_0} \binom{R_1+R_2+k-1}{R_1-1} p^{R_1} q^{R_2+k} = I_p(R_1, R_2+k_0+1) - I_p(R_1, R_2),$$

and

$$(2.6) \quad F_1(k_0) = \sum_{k=0}^{k_0} \binom{R_1+R_2+k-1}{R_2-1} p^{R_1+k} q^{R_2} = I_q(R_2, R_1+k_0+1) - I_q(R_2, R_1),$$

and the distribution function (2.3) is indeed the sum of (2.5) and (2.6).

About the mean and the variance we have the following

Theorem 2.1 *The mean $E(K)$ and the variance $V(K)$ of the variable defined as above are given by*

$$(2.7) \quad E(K) = \left(\frac{R_1}{p} - \frac{R_2}{q}\right)[q - I_p(R_1, R_2)] + \frac{(R_1+R_2-1)!}{(R_1-1)!(R_2-1)!} p^{R_1-1} q^{R_2-1}$$

$$(2.8) \quad V(K) = (q-p)\left(\frac{R_1}{p} - \frac{R_2}{q}\right) E(K) + \frac{qR_1}{p^2} + pq\left(\frac{R_1}{p} - \frac{R_2}{q}\right)^2$$

$$+ \frac{(R_1+R_2-1)!}{(R_1-1)!(R_2-1)!} p^{R_1-1} q^{R_2-1}\left(\frac{1}{pq} - 3\right) + \left(\frac{pR_2}{q^2} - \frac{qR_1}{p^2}\right) I_p(R_1, R_2)$$

$$- [E(K)]^2.$$

Proof.

By definition we have

$$(2.9) \quad E(K) = \sum_{k=0}^{\infty}\left[\binom{R_1+R_2+k-1}{R_1-1} p^{R_1} q^{R_2+k} k + \binom{R_1+R_2+k-1}{R_2-1} p^{R_1+k} q^{R_2} k\right].$$

The first term is

$$(2.10) \quad \sum_{k=0}^{\infty} \frac{(R_1+R_2+k-1)!}{(R_1-1)!(R_2+k)!} p^{R_1} q^{R_2+k}[(R_2+k) - R_2]$$

$$= R_1 \frac{q}{p} \sum_{k=0}^{\infty} \frac{(R_1+R_2+k-1)!}{R_1!(R_2+k-1)!} p^{R_1+1} q^{R_2+k-1} - R_2 \sum_{k=R_2}^{\infty} \frac{(R_1+k-1)!}{(R_1-1)!k!} p^{R_1} p^k$$

$$= R_1 \frac{q}{p}[1 - I_p(R_1+1, R_2-1)] - R_2[1 - I_p(R_1, R_2)].$$

On the other hand as we have

$$(2.11) \quad I_p(R_1+1, R_2-1) = -\frac{(R_1+R_2-1)!}{R_1!(R_2-1)!} p^{R_1} q^{R_2-1} + I_p(R_1, R_2),$$

$$(R_1 > 0, \ R_2 > 1).$$

The first term is equal to

$$(2.12) \quad R_1 \frac{q}{p}\left[1 + \frac{(R_1+R_2-1)!}{R_1!(R_2-1)!} p^{R_1} q^{R_2-1} - I_p(R_1, R_2)\right] - R_2[1 - I_p(R_1, R_2)].$$

Similarily the second term is

$$(2.13) \quad R_2 \frac{p}{q}\left[1 + \frac{(R_1+R_2-1)!}{R_2!(R_1-1)!} p^{R_1-1} q^{R_2} - I_q(R_2, R_1)\right] - R_1[1 - I_q(R_2, R_1)].$$

Combining these two

(2.14) $E(K) = \left(R_1\dfrac{q}{p} + R_2\dfrac{p}{q}\right) + \dfrac{(R_1+R_2-1)!}{(R_1-1)!\,(R_2-1)!}\,p^{R_1-1}q^{R_2-1} - (R_1+R_2)$

$\qquad\qquad + I_p(R_1,\ R_2)\left(R_2 - R_1\dfrac{q}{p}\right) + I_q(R_2,\ R_1)\left(R_1 - R_2\dfrac{p}{q}\right).$

And after some simple calculations we have that (2.7) holds true.

In order to prove (2.8), we calculate the expectation of K^2.

(2.15) $E(K^2) = \sum\limits_{k=0}^{\infty}\left[\dbinom{R_1+R_2+k-1}{R_1-1}p^{R_1}q^{R_2+k}k^2 + \dbinom{R_1+R_2+k-1}{R_2-1}p^{R_1+k}q^{R_2}k^2\right].$

The first term in the summation expressed as the sum of four terms

(2.16) $\sum\limits_{k=0}^{\infty}\dbinom{R_1+R_2+k-1}{R_1-1}p^{R_1}q^{R_2+}k^2 = \sum\limits_{k=0}^{\infty}\dfrac{(R_1+R_2+k-1)!}{(R_1-1)!\,(k_2+k-2)!}\,p^{R_1}q^{R_2+k}$

$\qquad + \sum\limits_{k=0}^{\infty}\dfrac{(R_1+R_2+k-1)!}{(R_1-1)!\,(R_2+k-1)!}\,p^{R_1}q^{R_2+k} - R_2^2\sum\limits_{k=0}^{\infty}\dfrac{(R_1+R_2+k-1)!}{(R_1-1)!\,(R_2+k)!}p^{R_1}q^{R_2+k}$

$\qquad - 2R_2\sum\limits_{k=0}^{\infty}\dfrac{(R_1+R_2+k-1)!}{(R_1-1)!\,(R_2+k)!}\,p^{R_1}q^{R_2+k}k.$

The summations of each of four terms in (2.16) are as follows

(2.17) $R_1(R_1+1)\,\dfrac{q^2}{p_2}[1 - I_p(R_1+2,\ R_2-2)],$

(2.18) $R_1\dfrac{q}{p}\,[1 - I_p(R_1+1,\ R_2-1)],$

(2.19) $-R_2^2\,[1 - I_p(R_1,\ R_2)],$

(2.20) $-2R_2R_1\dfrac{q}{p}\,[1 - I_p(R_1+1,\ R_2-1)] + 2R_2^2[1 - I_p(R_1,\ R_2)].$

In view of the symmetricity between the first and the second terms with respect to p, q and R_1, R_2, we have

(2.21) $E(K^2) = R_1(R_1+1)\,\dfrac{q^2}{p^2}[1 - I_p(R_1+2,\ R_2-2)] + R_1\dfrac{q}{p}[1 - I_p(R_1+1,\ R_2-1)]$

$\qquad - R_2^2[1 - I_p(R_1,\ R_2)] - 2R_1R_2\dfrac{q}{p}[1 - I_p(R_1+1,\ R_2-1)] + 2R_2^2[1 - I_p(R_1,R_2)]$

$\qquad + R_2(R_2+1)\,\dfrac{p^2}{q^2}[1 - I_q(R_2+2,\ R_1-2)] + R_2\dfrac{p}{q}[1 - I_q(R_2+1,\ R_1-1)]$

$\qquad - R_1^2[1 - I_p(R_2,R_1)] - 2R_1R_2\dfrac{p}{q}[1 - I_q(R_2+1,\ R_1-1)] + 2R_1^2[1 - I_q(R_2,R_1)]$

$\qquad = R_1(R_1+1)\dfrac{q^2}{p^2}\Big[1 + \dfrac{(R_1+R_2-1)!}{(R_1+1)!\,(R_2-2)!}p^{R_1+1}q^{R_2-2} + \dfrac{(R_1+R_2-1)!}{R_1!\,(R_2-1)!}p^{R_1}q^{R_2-1}$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad - I_p(R_1,\ R_2)\Big]$

$\qquad + (1-2R_2)\,R_1\dfrac{q}{p}\Big[1 + \dfrac{(R_1+R_2-1)!}{R_1!\,(R_2-1)!}p^{R_1}q^{R_2-1} - I_p(R_1,R_2)\Big]$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad + R_2^2[1 - I_p(R_1,\ R_2)]$

$$+ R_2 \, (R_2+1)\frac{p^2}{q^2}\Big[1 + \frac{(R_1+R_2-1)!}{(R_2+1)! \, (R_1-2)!}p^{R_1-2}q^{R_2+1} + \frac{(R_1+R_2-1)!}{R_2!(R_1-1)!}p^{R_1-1}q^{R_2}$$

$$- I_q(R_2, \, R_1)\Big]$$

$$+ (1-2R_1)\,R_2\frac{p}{q}\Big[1 + \frac{(R_1+R_2-1)!}{R_2!(R_1-1)!}p^{R_1-1}q^{R_2} - I_q(R_2,R_1)\Big] + R_1^2[1 - I_q(R_2,R_1)]$$

$$= q^2\left(\frac{R_1}{p} - \frac{R_2}{q}\right)^2 + \frac{qR_1}{p^2} + \frac{(R_1+R_2-1)!}{(R_1-1)!(R_2-1)!}p^{R_1-1}q^{R_2-1}\Big[(q-p)\left(\frac{R_1}{p} - \frac{R_2}{q}\right) + \frac{p^3+q^3}{pq}\Big]$$

$$+ \Big[(p-q)\left(\frac{R_1}{p} - \frac{R_2}{q^2}\right)^2 - \frac{qR_1}{p^2} + \frac{pR_2}{q^2}\Big]\, I_p(R_1, \, R_2)$$

$$= (q-p)\left(\frac{R_1}{p} - \frac{R_2}{q}\right)E(k) + \frac{q}{p^2}R_1 + pq\left(\frac{R_1}{p} - \frac{R_2}{q}\right)^2$$

$$+ \frac{(R_1+R_2-1)!}{(R_1-1!(R_2-1)!}p^{R_1-1}q^{R_2-1}\left(\frac{1}{pq} - 3\right) + \left(\frac{pR_2}{q^2} - \frac{qR_1}{p^2}\right)I_p(R_1, \, R_2). \quad \text{Q. E. D.}$$

It is of some interest to observe how the mean and the variance vary with p and R_1 *when* $R_1 + R_2$ is fixed. In order to investigate their behaviours we have tabulated these values for all possible combinations of $p = 0.01$, (0.01), 0.99 and R_1, when R_1 is not less than R_2 and the sum of them is 10 and 20. Because of the symmetricity in R_1, p and R_2, q this covers all possible combinations of p, R_1 and R when the sum is 10 and 20.

In tables 1 and 2, there are shown (I) the value of the mean when $p = R_1/(R_1+R_2)$, (II) the minimum of the values of the mean tabulated at $p =$

TABLE 1. $R_1 + R_2 = 10$

R_1	Mean			Variance		
	I	II	III	IV	V	VI
5	2.4609	2.4609	0.500	6.4047	6.4047	0.500
6	2.5082	2.4372	0.570	7.3235	6.2824	0.560
7	2.6683	2.3562	0.650	10.6831	5.9059	0.620
8	3.0199	2.2015	0.730	19.5527	5.2514	0.680
9	3.8742	1.9011	0.810	51.9631	4.1966	0.760

TABLE 2. $R_1 + R_2 = 20$

R_1	Mean			Variance		
	I	II	III	IV	V	VI
10	3.5239	3.5239	0.500	11.1058	11.1058	0.500
11	3.5411	3.5151	0.540	11.5002	11.0639	0.530
12	3.5941	3.4883	0.580	12.7384	10.8626	0.570
13	3.6880	3.4419	0.620	15.0048	10.5311	0.600
14	3.8328	3.3733	0.660	18.6861	10.0993	0.640
15	4.0466	3.2786	0.700	24.5563	9.4691	0.670
16	4.3640	3.1519	0.740	34.2702	8.7243	0.710
17	4.8566	2.9812	0.790	51.9229	7.8015	0.740
18	5.7036	2.7269	0.830	90.6675	6.6555	0.790
19	7.5471	2.3254	0.880	223.3360	5.1682	0.830

0.01, (0.01), 0.99 and (III) the value of p where the minimum in the sense above is attained. In columns (IV), (V) and (VI) there are shown the same for the variance. An interesting fact is when R_1 is fixed the minumum values are attained at the value of p between $1/2$ and $R_1/(R_1+R_2)$, and the difference between the values at $p=R_1/(R_1+R_2)$ and that of the minimum is not a small amount as the authors anticipated at first, and moreover as R_1 increases from $R_1 = R_2$ the value of the former increases whereas the latter decreases for which the authors have not so far found analytic proof.

§ 3. The uniformly most powerful test

In this section we shall consider a test of the null hypothesis $H_0: p=p_0$ against the alternative hypothesis $H_1: p=p_0+\Delta$ where $0<\Delta<q_0=1-p$ with the significance level α.

The density function of K is given in the following slightly modified form,

$$(3.1) \quad f(k,\ p)=\begin{cases} f_1(k,\ p) = \dfrac{(R_1+R_2+k-1)!}{(R_2-1)!\ (R_1+k)!}p^{R_1+k}q^{R_2}, & \begin{array}{l}\text{if the sampling termi-}\\ \text{nates with } E_2,\end{array} \\[4mm] f_2(k,\ p) = \dfrac{(R_1+R_2+k-1)!}{(R_1-1)!\ (R_2+k)!}p^{R_1}q^{R_2+k}, & \begin{array}{l}\text{if the sampling termi-}\\ \text{nates with } E_1.\end{array} \end{cases}$$

For significance level α there may or may not exists a non-negative integer k_0 such that

$$(3.2) \qquad \sum_{k=k_0}^{\infty} f_1(k_1, p_0) \gtrless \alpha \geq \sum_{k=k_0+1}^{\infty} f_1(k, p_0).$$

Our case 1 assumes the existence of such k_0. When such k_0 does not exists, there exists k_0 such that

$$(3.3) \qquad \sum_{k=0}^{\infty} f_1(k, p_0) + \sum_{k=0}^{k_0} f_2(k, p_0) \leq \alpha < \sum_{k=0}^{\infty} f_1(k, p_0) + \sum_{k=0}^{k_0+1} f_2(k, p_0).$$

Our case 2 assumes the existence of k_0 satisfying (3.3) but not (3.2).

In this paper we propose the test procedure expressed in terms of the following test function.

In case 1

$$(3.4) \quad \varphi(k) = \begin{cases} 1 & \begin{array}{l}\text{if the sampling terminates with}\\ E_2 \text{ and } k \geq k_0+1\end{array} \\[4mm] \dfrac{\alpha - \sum\limits_{k=k_0+1}^{\infty} f_1(k,\ p_0)}{f_1(k_0, p_0)} & \begin{array}{l}\text{if the sampling terminates with}\\ E_2 \text{ and } k=k_0\end{array} \\[6mm] 0 & \text{otherwise.} \end{cases}$$

In case 2

$$(3.5) \qquad \varphi(k) = \begin{cases} 1 & \text{if the sampling terminates with } E_2 \text{ or with } E_1 \text{ and } k \leq k_0 \\[2mm] \dfrac{\alpha - \sum_{k=0}^{\infty} f_1(k, p_0) - \sum_{k=0}^{k_0} f_2(k_1, p_0)}{f_2(k_0+1, p_0)} & \text{if the sampling terminates with } E_1 \text{ and } k = k_0+1 \\[2mm] 0 & \text{otherwise} \end{cases}$$

where $\varphi(k)$ represents the probability of rejecting the null hypothesis when k is observed. It is clear that the significance level of the test procedure expressed by this test function is α.

Theorem 3.1 *The test procedure expressed in terms of* (3.4) *and* (3.5) *is the uniformly most powerful test with the significance level* α *of* H_0; $p = p_0$ *against the alternative hypothesis* H_1: $p = p_0 + \varDelta$, $(0 < \varDelta < q_0)$.

Proof. Under the condtion that the sampling terminates with E_2, the region

$$(3.6) \qquad \left\{ k \; ; \; \frac{f(k, p_0 + \varDelta)}{f(k, p_0)} > \lambda, \; (\lambda > 0) \right\}$$

can be proved, after some simple calculations, equal to

$$(3.7) \qquad \left\{ k \; ; \; k > A(\varDelta, \lambda) = \frac{\log \lambda \left(\dfrac{p_0}{p_0 + \varDelta}\right)^{R_1} \left(\dfrac{q_0}{q_0 - \varDelta}\right)^{R_2}}{\log \left(1 + \dfrac{\varDelta}{p_0}\right)} \right\}$$

Similarly under the condition that the sampling terminates with E_1 the region (3.6) is proved to be equal to

$$(3.8) \qquad \left\{ k \; ; \; k < B(\varDelta, \lambda) = \frac{\log \lambda \left(\dfrac{p_0}{p_0 + \varDelta}\right)^{R_1} \left(\dfrac{q_0}{q_0 - \varDelta}\right)^{R_2}}{\log \left(1 - \dfrac{\varDelta}{q_0}\right)} \right\} .$$

It is to be noted that $A(\varDelta, \lambda)$ and $B(\varDelta, \lambda)$ are monotone increasing and decreasing functions of λ for any fixed $\varDelta$ both ranging in the real line, and moreover we have

$$(3.9) \qquad A(\varDelta, \lambda) \; B(\varDelta, \lambda) \leq 0.$$

In case 1, for any $\varDelta$ there exists λ such that $k_0 \leq A(\varDelta, \lambda) < k_0 + 1$ and for this λ, $B(\varDelta, \lambda) \leq 0$ and hence the region (3.8) is null. In this case the region where $\varphi(k)$ in (3.4) is equal to 1 is equal to the region (3.7). In case 2, for any $\varDelta$ there exists λ such that $k_0 < B(\varDelta, \lambda) < k_0 + 1$ and $A(\varDelta, \lambda) \leq 0$ hence the region (3.7) is the whole non-negative integers, and the region where $\varphi(k)$ in (3.5) is equal to 1 is equal to the sum of the regions (3.7) and (3.8). Therefore we can apply the Neyman-Pearson's Funda-

mental Lemma (c.f. for instance Lehman [11] p. 65) and arrive at the conclusion. Q.E.D.

The similar procedure can be derived for the case of the other type of a one sided alternative H_1; $p = p_0 - \Delta$ $(0 < \Delta < p_0)$, and this can also be proved to be the uniformly most powerful.

§ 4. The optimum symmetric three decision procedure.

In the previous section we derived the uniformly most powerfull one sided test. We can device two sided test in many different ways and it is impossible to find the uniformly most powerful one. However if we impose a kind of symmetricity requirement on the power function, it would be possible to find one. In this paper we approach the problem in a different way.

In this section we assume $R_1 = R_2$, and we consider the following three hypothesis

$$(4.1) \qquad \begin{aligned} H_0; & \quad p = 1/2, \\ H_1; & \quad p = 1/2 + \Delta, \\ H_2; & \quad p = 1/2 - \Delta. \end{aligned} \qquad (0 < \Delta < 1/2)$$

Let $D_i (i = 0, 1, 2)$ denote the decisions that H_i is true, and let $P(D_i; H_j)$ $(i, j = 0, 1, 2)$ denote the probapbilty of accepting the decision D_i when H_j is true.

These probabilities, of course, depends on the decision procedure employed and also on Δ when $j = 1, 2$. On the decision procedure we impose the requirements similar to those in the slippage problems (c.f. E. Paulson [13]) or more generally in the symmetric multiple decision problems, (c.f. A. Kudô [10]). We require

$$(4.2) \qquad P(D_0; H_0) = 1 - \alpha$$

where α is a suitably chosen small positive number such as 0.05 or 0.01, and for other two probabilities of correct decisions we require

$$(4.3) \qquad P(D_1; H_1) = P(D_2; H_2), \text{ for all } \Delta .$$

In this sampling procedure the observation consists in whether the sampling terminates with E_1 or E_2 and also the number K when the sampling terminates at the $(R_1 + R_2 + K)$-th trial. For the sake of brevity we say that K_1 is observed when the sampling terminates with E_2 and K_2 is observed when E_1. In order to express the decisian procedure we intorduce a vector valued function $\varphi(k) = (\varphi_0(k), \varphi_1(k), \varphi_2(k))$ satisfying the conditions $0 \le \varphi_i(k) \le 1$ $(i = 0, 1, 2)$ and $\sum_{j=0}^{2} \varphi_j(k) = 1$. This vector valued function expresses the decision procedure of the following. When the observation is k, we decide D_i with the probability $\varphi_i(k)$ by making use of some chance

mechanism.

At first we prove the following

Lemma 4.1 *For any* α $(0<\alpha<1)$ *there exists a decision procedure* $\bar{\varphi}_\alpha(k)=(\bar{\varphi}_0(k),\ \bar{\varphi}_1(k),\ \bar{\varphi}_2(k))$ *satisfying* (4.2) *and* (4.3).

Proof. The density function under H_0 is obtained by putting $p=1/2$ and $R_1=R_2$ in (3.1), and from this we can evaluate

(4.4) $$\gamma = P_r(k_1 = 0 \mid H_0) + P_r(k_2 = 0 \mid H_0).$$

In case $1 - \alpha \leq \gamma$ we define $\varphi_\alpha(k_i)$ as follows

(4.5)
$$
\begin{cases}
\bar{\varphi}_0(k_i) = \dfrac{1-\alpha}{\gamma} \\[2mm]
\bar{\varphi}_i(k_i) = 1 - \dfrac{1-\alpha}{\gamma} & \text{if } k_i = 0 \\[2mm]
\bar{\varphi}_j(k_i) = 0 \ (j \neq i)
\end{cases}
$$
$$
\begin{cases}
\bar{\varphi}_0(k_i) = 0 \\
\bar{\varphi}_i(k_i) = 1 & \text{if } k_i \geq 1. \\
\bar{\varphi}_j(k_i) = 0 \quad (j \neq i)
\end{cases}
$$
$(i=1,2)$

In the other case $1 - \alpha > \gamma$, there exists a positive integer r such that

(4.6) $$\beta \equiv P_r(k_1 \leq r-1 \mid H_0) + P_r(k_2 \leq r-1 \mid H_0) \leq 1-\alpha$$
$$< P_r(k_1 \leq r \mid H_0) + P_r(k_2 \leq r \mid H_0) \equiv \delta$$

and in this case we define the decision function $\bar{\varphi}_\alpha(k_i)$ as follows

(4.7)
$$
\begin{cases}
\bar{\varphi}_0(k_i) = 1 \\
\bar{\varphi}_i(k_i) = 0 & \text{if } k_i \leq r - 1 \\
\bar{\varphi}_j(k_i) = 0 \quad (j \neq i)
\end{cases}
$$
$$
\begin{cases}
\bar{\varphi}_0(k_i) = \dfrac{1-\alpha-\beta}{\delta-\beta} \\[2mm]
\bar{\varphi}_i(k_i) = 1 - \dfrac{1-\alpha-\beta}{\delta-\beta} & \text{if } k_i = r \\[2mm]
\bar{\varphi}_j(k_i) = 0 \quad (j \neq i)
\end{cases}
$$
$(i=1,2)$
$$
\begin{cases}
\bar{\varphi}_0(k_i) = 0 \\
\bar{\varphi}_i(k_i) = 1 & \text{if } k_i > r \\
\bar{\varphi}_j(k_i) = 0 \quad (j \neq i)
\end{cases}
$$

It is a straightforward matter to see that this decision function satisfies the condition (4.2) and (4.3).

Theorem 4.1 *The decision function* $\bar{\varphi}_\alpha(k)$ *defined by* (4.5) *or* (4.7) *sttains the maximum of* (4.3) *among all the decision procedures satisfying*

the equalities (4.2) *and* (4.3).

Proof.

Let us here consider the following summation

$$(4.8) \quad F(\varphi, c, \varDelta) = \sum_{k_1=0}^{\infty} [P_r(k_1|H_0)\varphi_0(k_1) + cP_r(k_1H_1, \varDelta)\varphi_1(k_1) + cP_r(k_1H_2, \varDelta)\varphi_2(k_1)]$$

$$+ \sum_{k_2=0}^{\infty} [P_r(k_2|H_0)\varphi_0(k_2) + cP_r(k_2H_1, \varDelta)\varphi_1(k_2) + cP_r(k_2|H_2, \varDelta)\varphi_2(k_2)]$$

$$= \sum_{k_1=0}^{\infty} \{P_r(k_1|H_0)[\varphi_0(k_1) + c\varphi_1(k_1)(1+2\varDelta)^{R_1+k_1}(1-2\varDelta)^{R_1} + c\varphi_2(k_1)(1+2\varDelta)^{R_1}(1-2\varDelta)^{R_1+k_1}]\}$$

$$+ \sum_{k_2=0}^{\infty} \{P_r(k_2|H_0)[\varphi_0(k_2) + c\varphi_1(k_2)(1+2\varDelta)^{R_1}(1-2\varDelta)^{R_1+k_2} + c\varphi_2(k_2)(1+2\varDelta)^{R_1+k_2}(1-2\varDelta)^{R_1}]\}$$

for any fixed $\varDelta$ and positive number c, this summation is easily found to be maximized by the decision function satisfying

$$(4.9) \quad \begin{cases} \varphi_0(k_i) = 1 & \text{if } k_i < B(\varDelta, c) \\ \varphi_i(k_i) = 1 & \text{if } k_i > max (0, B(\varDelta, c)). \qquad (i=1, 2) \\ \varphi_j(k_i) = 0 \quad (j \neq i) & \text{if } k_i = 0, 1, 2, \cdots \end{cases}$$

where

$$(4.10) \qquad B(\varDelta, c) = \frac{-\log c - R_1\log(1-4\varDelta^2)}{\log (1+2\varDelta)}$$

for any fixed $\varDelta$, this function $B(\varDelta, c)$ is continuously decreasing function varying from infinity to minus infinity. Therefore for any $\varDelta$ we can select c such that $F(\varphi, c, \varDelta)$ is maximised by the decision function $\bar{\varphi}_\alpha(k)$. We write such c as $c(\varDelta, \alpha)$.

Now let us consider a class of decision functions satisfying (4.2) and (4.3), and try to maximise $F(\varphi, c(\varDelta, \alpha), \varDelta)$ in this class of decision functions. In this case the sum of the first terms in the summations is a constant, $1-\alpha$, and moreover the sums of the second terms and that of the third are identical, because of the conditions (4.2) and (4.3). Maximizing $F(\varphi, c(\varDelta, \alpha), \varDelta)$ is nothing but the maximization of the sum of the second terms or the third. On the other hand this is maximized unconditionally by the decision function $\bar{\varphi}_\alpha(k)$, which leads us to the conclusion.

§ 5. The limits of the density and the distribution.

In case R_1 and R_2 are both less than 50, we can evaluate the density and the distribution function by making use of the available table of the incomplete beta-function [14]. If one of the R's is larger than 50, there does not seem to exists any table available.

We have made some investigations about the limiting distributions, and

we shall here describe briefly the results.

We shall first discuss the limiting density function in more general way. Instead of two we shall here consider n exclusive events E_i with the probabilities p_i $(i=1, 2, \cdots, n)$. In this case we should continue sampling if one of the events E_i has occured less than a prefixed number R_i $(i=1, 2, \cdots, n)$, and we should stop sampling as soon as all E_i's have occured not less than R_i times respectively. At the termination of the sampling E_i must have occured R_i+K_i times where all the K_i should be nonnegative and at least one of them be zero. When the sampling terminates with E_i, $K_i=0$ and the joint density function of the rest of K_j's is given by

(5.1) $P_r(k_1, \cdots, k_{i-1}, k_{i+1}, \cdots, k_n)$

$$= \frac{(R+k_1+\cdots+k_{i-1}+k_{i+1}+\cdots+k_n-1)!}{(R_i-1)! \prod\limits_{j \neq i}(R_j+k_j)!} p_1^{R_1+k_1} \cdots p_{i-1}^{R_{i-1}+k_{i-1}} p_i^{R_i} p_{i+1}^{R_{i+1}+k_{i+1}} \cdots p_n^{R_n+k_n}$$

where $R=\sum\limits_{i=1}^{n} R_i$.

Let $p_j=(R_j+m_i)/R$, where $-R_j<m_j<R-R_j (j=1, 2, \cdots, n)$ and $\sum\limits_{j=1}^{n} m_j=0$,

then the joint distribution can be written in the following form

(5.2) $P_r(k_1, \cdots, k_{i-1}, k_{i+1}, \cdots, k_n)$

$$= \frac{(R+k_1+\cdots+k_{i-1}+k_{i+1}+\cdots+k_n)! R_i(R_i+m_i)^{R_i} \prod\limits_{j \neq i}(R_j+m_j)^{R_j+k_j}}{R_i! \prod\limits_{j \neq i}(R_j+k_j)! (R+k_1+\cdots+k_{i-1}+k_{i+1}+\cdots+k_n) R^{R+k_1+\cdots+k_{i-1}+k_{i+1}+\cdots+k_n}} .$$

We consider the limit of (5.2) when all the $R_j's$ tend to infinity under the condition $m_j^2/R_j \to 0$ and $k^3/R_j^2 \to 0$ $(j=1, 2, \cdots, n)$, where $k=\sum\limits_{j=1}^{n} k_j$.

By making use of the Stirling formula, we have that (5.2) tends to

(5.3) $\dfrac{1}{(\sqrt{2\pi})^{n-1}} \cdot \dfrac{(R+k-k_i)^{R+k-k_1+1/2} R_i(R_i+m_i)^{R_i} \prod\limits_{j \neq i}(R_j+m_j)^{R_j+k_j}}{R_i^{R_i+1/2} \prod\limits_{j \neq i}(R_j+k_j)^{R_j+k_j+1/2}(R+k-k_i) R^{R+\kappa-\kappa_i}}$

$$= \frac{1}{(\sqrt{2\pi})^{n-1}} \cdot \left(\frac{R_i}{R\prod\limits_{j \neq i}R_j}\right)^{1/2} \cdot \frac{\left(1+\dfrac{k-k_i}{R}\right)^{R+k-k_i}\left(1+\dfrac{m_i}{R}\right)^{R_i}\prod\limits_{j \neq i}\left(1+\dfrac{m_j}{R}\right)^{R_j+k_j}}{\prod\limits_{j \neq i}\left(1+\dfrac{k_j}{R_j}\right)^{R_j+k_j}\prod\limits_{j \neq i}\left(1+\dfrac{k_j}{R_j}\right)^{1/2}\left(1+\dfrac{k-k_i}{R}\right)^{1/2}}$$

Taking the logarithm of $(1+k_j/R_j)^{k_j}$, $(1+(k-k_i)/R)^{k-k_i}$ and $(1+m_j/R)^{k_j}$ and applying the condition, we arrive at the limit

(5.4) $\dfrac{1}{(\sqrt{2\pi})^{n-1}}\left(\dfrac{R_i}{R\prod\limits_{j \neq i}R_j}\right)^{1/2}\exp\left[-\dfrac{(k-k_i)^2}{R}+\sum\limits_{j \neq i}\dfrac{m_jk_j}{R_j}-\sum\limits_{j \neq i}\dfrac{k_j^2}{R_j}\right]$

$$= \frac{1}{(\sqrt{2\pi})^{n-1}}\left(\frac{R_i}{R\prod\limits_{j \neq i}R_j}\right)^{1/2}\exp\left[-(k-a)'\Lambda(k-a)\right]$$

where

$$
\boldsymbol{k} = \begin{pmatrix} k_1 \\ k_2 \\ \vdots \\ k_{i-1} \\ k_{i+1} \\ \cdots \\ k_n \end{pmatrix}, \quad \boldsymbol{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_{i-1} \\ a_{i+1} \\ \vdots \\ a_n \end{pmatrix}, \quad a_j = \frac{m_j R_i - m_i R_j}{2 R_i} (j=1,\cdots, i-1, i+1,\cdots, n),
$$

$$
\Lambda = \begin{pmatrix} 1/R_1 & & & & \\ & \ddots & & & 0 \\ & & 1/R^{i-1} & & \\ & & & 1/R^{i+1} & \\ & 0 & & & \ddots \\ & & & & 1/R_n \end{pmatrix} - \frac{1}{R} \begin{pmatrix} 1 & 1 & \cdots\cdots & 1 \\ 1 & 1 & \cdots\cdots & 1 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots\cdots & 1 \end{pmatrix}.
$$

It is interesting to note that the limits tends to a multivariate normal density function. In view of the condition, however, it does not cover entire range of k_i 's. Indeed the integral of the entire range of k is not equal to 1.

Now let us turn to the limiting distribution function. In this case our treatment is not so general as in the case of limiting density function. Our result may be ennunciated in the following.

Theorem 5.1 *When R_1 and R_2 tend to infinity, the limit of the cummulative distribution function given by* (2.3) *tends to*

(5.5) $$\Phi\left[\frac{qR_1 - pR_2 + qk_0}{\sqrt{(R_1 + R_2 + k_0)pq}}\right] - \Phi\left[\frac{qR_1 - pR_2 - pk_0 - 1}{\sqrt{(R_1 + R_2 + k_0)pq}}\right]$$

for all k_0.

Specially when $p = R_1/(R_1 + R_2)$, (5.5) *becomes*

(5.6) $$\Phi\left[\frac{R_2 k_0}{\sqrt{(R_1 + R_2 + k_0)R_1 R_2}}\right] - \Phi\left[\frac{-(R_1 k_0 + R)}{\sqrt{(R_1 + R_2 + k_0)R_1 R_2}}\right].$$

where Φ is the standardized normal distribution function.

Proof. Now let us recall to the relation between the distribution function of the binomial distribntion and the incomplete beta function.

(5.7) $$I_{1-p}(n-k, k+1) = \sum_{x=0}^{k} {}_n c_x p^x (1-p)^{n-x}.$$

Because of the central limit theorem applied to binomial distribution (c. f. De Moivre Laplace Theorem in Feller [6] p. 172), we have

(5.8) $$I_{1-p}(n-k, k+1) \to \Phi\left(\frac{k-np}{\sqrt{npq}}\right)$$

as k and n-k tends to infinity, and applying this relation to (2.3) we obtain (5.5), from which (5.6) follows immediately.   Q. E. D.

The more important limiting functions are these of (2.5) and (2.6), as they are used in the uniformly most powerful one sided test. The similar calculations as in Theorem 5.2 yields.

Corollary 5.1 *When R_1 and R_2 tends to infinity. the functions $F_2(k_0)$ and $F_1(k_0)$ tend to*

(5. 9)
$$\varPhi\left[\frac{qR_1-pR_2-q}{\sqrt{(R_1+R_2-1)pq}}\right]-\varPhi\left[\frac{qR_1-pR_2-pk_0-1}{\sqrt{(R_1+R_2+k_0)pq}}\right]$$

and

(5. 10)
$$\varPhi\left[\frac{qR_1-pR_2+qk_0}{\sqrt{(R_1+R_2+k_0)pq}}\right]-\varPhi\left[\frac{qR_1-pR_2-q}{\sqrt{(R_1+R_2-1)pq}}\right]$$

respectively for all k_0.

Specially when $p/R_1=q/R_2$, (5. 9) *and* (5. 10) *become*

(5. 11)
$$\varPhi\left[-\frac{\sqrt{R_2}}{\sqrt{(R_1+R_2-1)R_1}}\right]-\varPhi\left[\frac{-(R_1k_0+R)}{\sqrt{(R_1+R_2+k_0)R_1R_2}}\right]$$

and

(5. 12)
$$\varPhi\left[\frac{R_2k_0}{\sqrt{(R_1+R_2+k_0)R_1R_2}}\right]-\varPhi\left[-\frac{\sqrt{R_2}}{\sqrt{(R_1+R_2-1)R_1}}\right]$$

respectively for all k_0.

§ 6. An example.

A large scale survey on consaguinity is now proceeding in Fukuoka City by a group headed dy Prof. K. Yamaoka of Kyushu University, medical school under the sponsorship of U. S. Public Health Service (Gm 09470–02).

The survey has so far taken the following steps.

1) Listing of all the 55,315 females born between 1912 and 1931 from the address registrations at city offices in 8 different regions. 2) All of them have been sent questionairs as to their marital status, number of children, and consanguinity, and 46,901 (84.78 %) of them have answered, and among them the percentage of those who have answered as consanguineous has been approximately 7.37 %. 3) The confirmation of the consanguinity by Koseki (c. f. Yanase [16]) is now proceeding. 4) The sampling of the control group from non-consanguineous marriages is now proceeding by the following way.

The main purpose of our study is to determine the effect of consanguinity on fertility and the mortality and morbidity of the children. As the concomittant variations, to be taken into account, caused by age and regional difference are expected to exist, we decided to sample the controls from each one of the age groups of 8 city offices.

The non-consanguinuous group has been already classified according to 8 regions, but classification and the enumeration of these 8 groups again according to the marital status and the age are almost prohibitive by our available means because of the large scale of the survey.

The only means to us seems to be the multiple inverse sampling in each of the 8 regions. As both the consanguinities and the controls are to be examined by direct interview, which means a considerable amount of cost and task, we decided to draw only the same number of controls in each

TABLE 3.

Years (born)	No. of consanguinity (R_i)	Sample size at the termination	Redundancy
1912	8	13	5
'13	11	25	14
'14	14	14	0
'15	12	26	14
'16	14	26	12
'17	12	24	12
'18	15	29	14
'19	16	19	3
'20	12	26	14
'21	17	24	7
'22	21	32	11
'23	14	26	12
'24	13	32	19
'25	15	28	13
'26	20	24	4
Total	214	368	—
'27	16	35	19
'28	19	27	8
'29	19	35	16
'30	12	38	26
'31	9	35	26
Total	75	170	—

one of the age groups in each region.

For the purpose of sampling, a program for a high speed computor was written to generate random numbers less than a perfixed value, which, in this case, 3964, the total number of non-consanguinity in this region.

Table 3 shows how the sampling was terminated in the region called KATAKASU, one of the 8 regions, and this seems to suggest the decrease in the consanguinity rate in recent years.

It is commonly believed that the consanguinity decreased after the World War II, namely after 1945. We seek for the evidence of it by means of statistical tests. But we have not developed a test procedure adequate to our purpose.

A drastic change might have occured between those who are born before and after 1926, who became age of twenty at the end of the World War II. For the purpose of the test, we classify into two according to this criterion. Ordinary χ^2 test of 2×2 table is in this case

TABLE 4.

	before 1926	after 1927	total
consanguinity	214	75	289
non-consanguinity	368	170	538
Total:	582	245	827

whose χ^2 value is 2.87 which is not significant.

For the illustration purpose, we sampled the controls by the double inverse sampling method according to the above two categories, where the sampling ended with the first category, and the redundancy in the 2nd category was 26. In this case the 2×2 table is

TABLE 5.

	before 1926	after 1927	total
consanguinity	214	75	289
non-consanguinity	214	101	315
Total:	428	176	604

whose χ^2 value is 2.723 which is again not significant. On the other hand, however, applying the test procedure described in § 3, taking $R_1 = 214$, $R_2 =$

75, $p_0 = 214/289 = 0.740$, $H_0 :$ $p = p_0$, $H_1 ;$ $p < p_0$, we obtain, from corollary 5.1, $Pr(k_2 > 26) = 0.047$, which shows the significance at the 5% level.

This fact illustrates the advantage of this test procedure in the double inverse sampling over the ordinary χ^2 test, which has been already proved by Themorem 3.1. It is interesting to observe that the χ^2 test based on Table 4 is not signifcant even though the total sample size is 827 much larger than 604 of table 5.

These figures are only provisional, subject to alteration by confirmation by Koseki and direct interview. They are borrowed here merely for illustrative purpose.

As this example indicates, our results ought to be extended to the case of inverse sampling from multinomial population and there has to be developed a test procedure for testing the equality of proportions against a kind of an ordered type of alternative (c.f. for instance D. J. Bartholomew [1]). The authors wish to develope these in a near future.

References

[1] BARTHOLOMEW, D. J.: *A test of homogeneity of means under restricted alternatives.* J. R. S. S. Ser. B Vol. **124** (1961), 239–281.

[2] CHAPMAN, D. G.: *Inverse, multiple and sequential sample.* Biometrics, Vol. **8** (1952), 286–306.

[3] DEGROOT, M. H.: *Unbiased sequential estimation for binomial populations,* Ann. Math. Stat., Vol. **30** (1959), 80–101.

[4] DIEULEFAIT, C. E.: *Note on a method of sampling,* Ann. Math. Stat., Vol. **13** (1942), 94–97.

[5] EPSTEIN, B. and SOBEL, M.: *Life testing,* J. Amer. Stat. Assc., Vol. **48** (1953), 486–502.

[6] FELLER, W.: *An introduction to probability theory and its application,* Vol. I, 2nd, ed., John Wiley (1957).

[7] FINNEY, D. J.: *On a method of estimating frequencies,* Biometrika, Vol. **36** (1949), 233–234.

[8] GIRSHICK, M. A., MOSTELLER, F. and SAVAGE, L. J.: *Unbiased estimates for certain binomial sampling problems with applications,* Ann. Math. Stat., Vol. **17** (1946), 13-23.

[9] HALDANE, J. B. S.: *On a method of estimating frequencies,* Biometrika, Vol. **33** (1945), 222 –225.

[10] KUDÔ, A.: *The symmetric multiple decision problems,* Memoirs of the Faculty of Science, Kyushu Univ. Ser. A, Vol. XIV (1960), 179–206.

[11] LEHMANN, E. L.: *Testing statistical hypotheses,* John Wiley, (1959).

[12] NADLER, J.: *Inverse binomial sampling plans when an exponential distribution is sampled with censoring,* Ann. Math. Stat., Vol. **31** (1960), 1201–1204.

[13] PAULSON, E.: *An optimum solution to the k-sample slippage problem for the normal distribution,* Ann. Math. Stat., Vol. **23** (1952), 610–616.

[14] PEARSON, E. S.: *Tables for incomplete beta functions.* Biometrika Office, London.

[15] TWEEDIE, M. C. K.: *Inverse statistical variates,* Nature. Vol. **155** (1945), 453.

[16] YANASE, T.: *The use of the Japanese family regrister for genetic studies,* Proceedings of the seminar on the use of vital and health statistics for genetic and radiation studies, sponsored by the United Nations and the World health organization, held in Geneva, 5–9 Sept., (1960), 119–133.