J. of Basic Sciences, Cheju Nat. Univ.
9(1), 1~15, 1996

기초과학연구 제주대학교
9(1), 1~15, 1996

Unequal probability Sampling and Maximum Entropy

Ik-Chan Kim
Department of Computer Science and Statistics,
Cheju National University, Cheju 690-756, Korea

Abstract

Attention is drawn to a method of sampling a finite population of N units with unequal probabilities and without replacement. The method was proposed by Stern & Cover (1989) as a model for lotteries. The method can be characterized as maximizing entropy given coverage probabilities π_i, or equivalently as having the probability of a selected sample proportional to the product of a set of 'weights' w_i. We show the essential uniqueness of w_i given the π_i and describe practical, geometrically convergent algorithms for computing the w_i from the π_i. We present two methods for stepwise selection of sampling units, 'forward' and 'backward'.

Inclusion probabilities of any order can be written explictly in closed form. Second-order inclusion probabilities π_{ij} satisfy the condition $0 < \pi_{ij} < \pi_i\pi_j$ and shown the several properties of π_i and π_{ij}, which guarantees Yates & Grundy's variance estimator to be unbiased, definable for all samples and always nonnegative for any sample size.

1. Introduction

Random sampling of n distinct units from a finite population of N units without replacement may be called unequal probability sampling or weighed sampling when the probabilities associated with $\binom{N}{n}$ possible choices are

---

This paper was announced in Branch of Survey Sampling of the Korean Statistical Society, on 16. Mar. 1996.

not all equal. A problem is to define a particular weighted sampling scheme subject to prespecification of the marginal probabilities π_i that the sample includes the ith population unit, where

$$0 < \pi_i < 1 \quad (i=1,2,\cdots,N), \quad \sum_{i=1}^{N} \pi_i = n \tag{1}$$

The random sample may be denoted by X where $X=(X_1, X_2, \cdots, X_N)$ and

$$X_i = \begin{cases} 1 : i\text{th pop. unit is in the sample} \\ 0 : i\text{th pop. unit is not in the sample} \end{cases}$$

Let

$$D^n = \{x=(x_1, \cdots, x_n) : x_i = 0 \ \ or \ \ 1 \ \text{ and } \ x_1 + \cdots + x_n = n\}$$

Then random vector X takes values in D^n and denoted the probability density function of a sampling scheme by $p(x)$ for any vector $x \in D^n$, where $p(x) > 0$ and $\sum p(x) = 1$.

The associated probability that the sample includes ith units is

$$\pi_i = E(X_i) = \sum_{x \in D^n} x_i p(x), \tag{2}$$

when the π_i satisfy (1).

The particular family of sampling schemes that we propose can be defined in any of three ways that we show to be equivalent.

Method 1. Pick any vector of weights, $w=(w_1, \cdots, w_N)$, where $w_i > 0$ for $i=1,\cdots,N$, and define

$$p(x) \propto \prod_{i=1}^{N} w_i^{x_i} \tag{3}$$

It is obvious that rescaling the w_i by a positive constant multiplies determines the same $p(x)$ but that modulo rescaling the $p(x)$ corresponding to distinct u are distinct. It is less obvious that the coverage probabilities determined by (2) are in one to one correspondence with the weights w_i modulo scaling, but we show that this correspondence is a direct consequence of standard exponential family theory. If we put $w_i = e^{\theta_i}$ then

$$p(x) \propto \exp\left(\sum_{i=1}^{N} \theta_i x_i\right) \quad ; \quad \theta=(\theta_1, \cdots, \theta_N)$$

which makes exponential family connection .

Method 2. Pick any vector of coverage probabilities $\pi=(\pi_1,\cdots,\pi_N)$, where π_i satisfy (1) , and choose $p(x)$ subject to the constraints (2) to maximize the entropy $-\sum p(x)\log p(x)$.

If a weight vector w or θ can be determined such that $p(x)$ defined by Method 1 matches the π given for Method 2, then this Method 1 choice is the unique maximum entropy scheme proposed as Method 2.

1972 : Dorroch & Ratcliff [3] : A more general form of this result is proved

1989 : Stern & Cover [12] : This model was first proposed for determined optimal lottery strategies.

1990 : Joe [10] : further generalized

Method 3. Pick any vector of probabilities $p=(p_1,\cdots,p_N)$, where $0<p_i<1$ for $i=1,\cdots,N$ and define $Z=(Z_1,\cdots,Z_N)$ to be independent Bernoulli trials with probabilites $p_1,\cdots,p_N$. Then define the sampling distribution of X to be the conditional distribution of Z given $\sum Z_i = n$.

It is evident that Method 3 gives the same sampling scheme as Method 1 if and only if the w_i are proportional to $p_i/(1-p_i)$.

Thus the base model for all three models described above is

$$p(x) = \prod_{i=1}^{N} w_i^{x_i} \Big/ \sum_{y\in D^n}\left(\prod_{i=1}^{N} w_i^{y_i}\right) \propto \exp\left(\sum_{i=1}^{N}\theta_i x_i\right) \quad x\in D^n \tag{4}$$

where w, or equivalently θ, may be deternined by π through (2). We refer to (4) as the maximum entropy model.

2. Weights and coverage probabilities

The relation between w and π is a special case of that between natural and mean-value parameterizations for an exponential family. The following result can be proved by using Theorem 3.6 of Brown (1986, p74) [1]

Th 1. For any vector π satisfying (1), there exists a vector w for the maximum entropy model subject to the constraint (2), and w is unique up to rescaling.

To compute w from π, we recast (2) in the form of a set of equations (5) below, and solve these iteratively as in (7). Throughout the paper we use the following notation : $S=\{1,2,\cdots,N\}$, capital lettess such as A, B or C for subset of S, $A^C=S\setminus A$ for the complement of A in S, and $|A|$ for the number of elements of A. And we define

$$R(k,C)=\sum_{B\subset C,|B|=k}\left(\prod_{i\in B}w_i\right)$$

for any nonempty set $C\subset S$ and $1\le k\le|C|$, $R(0,C)=1$ and $R(k,C)=0$ for any $k>|C|$.

The following Proposition 1 follows immediately from the definiton.

Propositon 1. For any nonempty set $C\subset S$ and $1\le k\le|C|$:

a) $\sum_{j\in C}w_jR(k-1,C\setminus\{j\})=kR(k,C)$

b) $\sum_{j\in C}R(k,C\setminus\{j\})=(|C|-k)\cdot R(k,C)$

c) $\sum_{i=0}^{k}R(i,C)R(k-i,C^c)=R(k,C)$

Using this notation, we may rewrite (2) as

$$\pi_i=\frac{w_iR(n-1,\{i\}^c)}{R(n,S)}\quad,\quad(i=1,2,\cdots,N)\qquad(5)$$

By a) $\sum\pi_i=\sum_i\frac{w_iR(n-1,\{i\}^C)}{R(n,S)}=n$, which is the result in (1).

Thus for fixed n, there are $N-1$ linearly independent relations among the N relations of (5). Without loss of generality, we assume that $\pi_1\le\pi_2\le\cdots\le\pi_N$ and let $\pi_N=w_N$, then $\pi_N=\frac{w_NR(n-1,\{N\}^c)}{R(n,S)}$ by (5) i.e. $R(n,S)=R(n-1,\{N\}^c)$.

Hence we get

$$w_i=\frac{\pi_iR(n-1,\{N\}^c)}{R(n-1,\{i\}^c)}\quad,\quad(i=1,\cdots,N-1),\ w_N=\pi_N\qquad(6)$$

Although a closed-form solution of (6) seems impossible, the equations can be solved as a fixed-point problem by using an iterative procedure. Specially, the following updating scheme provides a solution of (6):

$$w_i^{(t+1)} = \left. \frac{\pi_i R(n-1, \{N\}^c)}{R(n-1, \{i\}^c)} \right|_{w=w^{(t)}}; \quad (i=1,\cdots,N-1), \quad w_N^{(t+1)} = w_N^{(t)} = \pi_N \tag{7}$$

where $w^{(t)} = (w_1^{(t)}, w_2^{(t)}, \cdots, w_N^{(t)})$.

Th 2. Define $W=\{w : 0 < w_i \le \pi_i,\ i=1,\cdots,N-1 ;\ w_N=\pi_N\}$, then

a) The set of equations in (6) has a unique solution, $w^* \in W$.

b) Starting from $w^{(0)}=\pi$, the sequence (7) of vectors $w^{(t)}$ $(t=1,2,\cdots)$ converges monotonically and geometrically to w^* with a rate bounded by π_N.

This procedure in (7) takes far fewer iterations to converge than Deming-stephan(1940) [4].
Also this algorithm is more direct and much faster than the one proposed by Sterm & Cover (1989) [12] which uses a generalized iterative scaling algorithm of Darroch & Ratcliff (1972) [3].

Lemma 1. For any $i,j \in S$, the following properties hold :

a) $\pi_i = \pi_j \Leftrightarrow w_i = w_j$

b) $\pi_i > \pi_j \Leftrightarrow w_i > w_j$

c) $\pi_i > \pi_j \Leftrightarrow \dfrac{w_i}{w_j} > \dfrac{\pi_i}{\pi_j}$

d) If $c_1 \le \pi_k < c_2$, for all π_k and some $c_1, c_2 \in (0,1) \Rightarrow w_i/w_j \to \pi_i/\pi_j$ as $N/n \to \infty$.

Proof

Take the ratio of the ith equation and the jth equation of (5):

$$\frac{\pi_i}{\pi_j} = \frac{w_i R(n-1, \{i\}^c)}{w_j R(n-1, \{j\}^c)} = \frac{w_i w_j R(n-2, \{i,j\}^c) + w_i R(n-1, \{i,j\}^c)}{w_j w_i R(n-2, \{i,j\}^c) + w_j R(n-1, \{i,j\}^c)} \tag{8}$$

Then a) and b) can be direatly deduced from (8).
Since the right-hand side of (8) $\le w_i/w_j$ if and only if $w_i \ge w_j$, c) is also strightforward by the condition $\pi_i \ge \pi_j$ and a) and b). By manipulating the equation in (8) for π_i instead of π_j and using c), we can show

$$\frac{\pi_i}{\pi_1} \le \frac{w_i}{w_1} \le \left\{\frac{\pi_i}{\pi_1}\left(\frac{N}{n}-1\right)\right\} \bigg/ \left(\frac{N}{n}-\frac{\pi_i}{\pi_1}\right) \tag{9}$$

As $N/n \to \infty$, the right-hand side of (9) approches π_i/π_1. Thus for any pair $i \neq j$, we have $w_i/w_j \to \pi_i/\pi_j$ as $N/n \to \infty$.

3. Draw-by-draw selection procedures

We discuss procedures for drawing a sample from the maximum entropy model.

We consider draw-by draw selection procedures, on acount of the large number $\binom{N}{n}$, of choices where we draw one unit at a time until n units are obtained.

We call a selection procedure 'forward' if it selects n units from the population as the sample, or 'backward' if it removes $N-n$ units from the propulation and takes the remaining n units as the sample. Then for any $x \in D^n$,

$$p(x) \propto \prod_{i \in A_x} w_i \propto \prod_{i \in A_x} w_i / \prod_{i \in S} w_i = \prod_{i \in A_x^c} w_i^{-1}$$

where $A_x = \{i\, ;\, x_i = 1\}$, it is obvious that for any 'forward' procedure, there is corresponding 'backward' procedure, which relects unsampled units in the same way using w_i^{-1} instead of w_i. We also distinguish among procedures by whether or not a procedure requires n fixed in advance. In the context of sample surveys, the π_i are usually prespecified. Thus the sample size n and the w_i are fixed in advance.

However, it is possible in some applications that the w_i are prespecified and different sample sizes are to be experimented with. In this case a selection procedure that does not depend on n is desired.

The output of a draw-by draw procedure is represented by $A_0, A_1, \cdots, A_n$ where $A_0 = \phi$ and $A_k \subset S$ denotes the set of selected indices after k draws. The following are two 'forward' procedures, one for fixed n and the other for nonfixed n.

The 'backward' version of these procedures can be defined accordingly.

Procedure 1. (forward, n fixed). At the kth draw $(k=1,\cdots,n)$, a unit $j\in A_{k-1}^c$ is selected with probability

$$P_1(j,A_{k-1}^c) = \frac{w_j R(n-k, A_{k-1}^c \setminus \{j\})}{(n-k+1)R(n-k+1, A_{k-1}^c)}$$

Using the relation $w_i \propto p_i/(1-p_i)$, the function R can be written as

$$R(k,C) = P_r(\sum_{i\in C} Z_i = k) \prod_{i\in C}(1+w_i)$$

for any nonempty set $C\subset S$ and $0\le k\le |C|$.

Thus P_1 has the interpretation

$$P_1(j,A_{k-1}^c) = \frac{1}{n-k+1} P_r(Z_j=1 \mid \sum_{t\in A_{k-1}^c} Z_t = n-k+1)$$

where $Z_1,\cdots,Z_N$ are independent Bernoulli trials as defined in Method 3 in Section 1.

It is easy to see by (a) in Proposition 1 that $P_1(\cdot, A_{k-1}^c)$ is a probability density on A_{k-1}^c. To see that a random sample of n units selected by Procedure 1 is a sample from the maximum entropy model, we first compute the probability of choosing an ordered set of indices $i_1,\cdots,i_n$ using Procedure 1, where in this case $A_k = \{i_1,\cdots,i_n\}$ for $k=1,\cdots,n$:

$$\prod_k = 1^n P_1(i_k, A_{k-1}^c) = \prod_{k=1}^{n} \frac{w_{i_k} R(n-k, A_k^c)}{(n-k+1)\cdot R(n-k+1, A_{k-1}^c)}$$

$$= \frac{1}{n!}\left(\prod_{k=1}^{n} w_{i_k}\right)\cdot \frac{R(0,A_n^c)}{R(n,S)} = \frac{1}{n!} P_r(X_t=1, t\in A_n)$$

Since there are $n!$ different ways of the ordering the indices $i_1,\cdots,i_n$, the probability of obtaining the units $i_1,\cdots,i_n$ without regard to order is exactly $P_r(X_t=1,\ t\in A_n)$.

Suppose that $\pi_{i_1\cdots i_k}$ is the kth-order inclusion probability for the units $i_1,\cdots,i_k$ to be in a sample of size n from the maximum entropy model. Then a property of Procedure 1 is that

$$P_r(A_k=\{i_1,\cdots,i_k\}) \propto \prod_{l=1}^{k} w_{i_l} \frac{R(n-k,\{i_1,\cdots,i_k\}^c)}{R(n,S)} = \pi_{i_1\cdots i_k} \tag{10}$$

Although P_1 can be calculated directly from R functions, the computation can be much simplified by noticing that $P_1(j,A_0^c)=\pi_j/n$ and using the following formula recursively for the consecutive draws.

Lemma 2. For any $1\le k\le n-1$ and $j\in A_k^c$,

$$P_1(j,A_k^c) = \frac{w_{i_k}P_1(j,A_{k-1}^c)-w_jP_1(i_k,A_{k-1}^c)}{(n-k)(w_{i_k}-w_j)P_1(i_k,A_{k-1}^c)} \tag{11}$$

Procedure 1 can be realized using the following algorithm, which requires $O(nN)$ operations.

1) For $j=1,2,\cdots,N$, calculates $P_1(j,S)$, which is given by π_j/n. Then draw unit i_1 according to the probability $P_1(i_1,S)$

2) If $n>1$, then $A_0\leftarrow\phi$, $A_1\leftarrow\{i_1\}$, $k\leftarrow2$, go to 3) ; otherwise stop.

3) For all $j\in A_{k-1}^c$, calculate $P_1(j,A_{k-1}^c)$ from $P_1(j,A_{k-2}^c)$ and $P_1(i_{k-1},A_{k-2}^c)$ using (11). Then draw unit i_k according to the probability $P_1(i_k,A_{k-1}^c)$.

4) If $k<n$, then $A_k\leftarrow A_{k-1}\cup\{i_k\}$, $k\leftarrow k+1$, go to 3 ; otherwise stop.

Procedure 2 (forward, n nonfixed). At the kth draw ($k=1,\cdots,n$), a unit $j\in A_{k-1}^c$ is selected with probability

$$P_2(j,A_{k-1}^c) = \sum_{i=0}^{k-1} \frac{w_jR(k-i-1,A_{k-1}^c\setminus\{j\})\cdot R(i,A_{k-1})}{(k-i)R(k,S)} \tag{12}$$

By a) and c) in Proposition 1,

$$\sum_{j\in A_{k-1}^c} P_2(j,A_{k-1}^c) = \sum_{i=0}^{k-1}\frac{R(i,A_{k-1})}{(k-i)R(k,S)}\left\{\sum_{j\in A_{k-1}^c} w_jR(k-i-1,A_{k-1}^c\setminus\{j\})\right\}$$

$$= \sum_{i=0}^{k-1}\frac{R(i,A_{k-1})}{(k-i)R(k,S)}(k-i)R(k-i,A_{k-1}^c) = 1$$

Thus $P_2(\cdot,A_{k-1}^c)$ is a probability density on A_{k-1}^c. Now we show by induction that a random sample seledcted by Procedure 2 is a sample from the maximum entropy model.

Let r_k be a random index of the unit selected at the kth draw. Assuming

$$P_r(A_{k-1}=A)=R(k-1,A)/R(k-1,S)$$

for any $A\subset S$ with $|A|=k-1$. which is true for $k=2$ by the definition of P_2. we show that $P_r(A_k=B)=R(k,B)/R(k,S)$ for any $B\subset S$ with $|B|=K$. By b) and c) in Proposition 1.

$$\begin{aligned}
P_r(A_k=B) &= \sum_{j\in B} P_r(A_{k-1}=B\setminus\{j\})P_r(r_k=j|A_{k-1}=B\setminus\{j\}) \\
&= \sum_{j\in B}\frac{R(k-1,B\setminus\{j\})}{R(k-1,S)}\left\{\sum_{i=0}^{k-1}\frac{w_j R(k-i-1,B^c)R(i,B\setminus\{i\})}{(k-i)R(k,S)}\right\} \\
&= \frac{R(k,B)}{R(k,S)}\sum_{i=0}^{k-1}\frac{R(k-i-1,B^c)}{(k-i)R(k-1,S)}\{\sum_{j\in B}R(i,B\setminus\{j\})\} \\
&= \frac{R(k,B)}{R(k,S)}\sum_{i=0}^{k-1}\frac{R(k-i-1,B^c)}{(k-i)R(k-1,S)}(k-i)R(i,B) \\
&= \frac{R(k,B)}{R(k,S)}
\end{aligned}$$

By induction. the probability of obtaining a set A_n is

$$R(n,A_n)/R(n,s)=P_r(X_t=1,t\in A_n)$$

It is evident from the proof above that using Procedure 2

$$P_r(A_k=\{i_1,\cdots,i_k\}) \propto \prod_{l=1}^{k} w_{i_l}$$

Thus Procedure 2 does not depend on n. The two procedures have different uses. By using (12). Procedure 1 requires less operations than Procedure 2. but cannot be used when n is nonfixed. Procedure 2 is useful for doing rotations in survey sampling.

The preference between forword and backward procedures depends on the scale of n. Forward procedures are preferred when $n\leq N/2$ while backward procedures are preferred when $n>N/2$.

When the w_i all are equal. both procedures reduce to simple random sampling without replacement.

4. Application to Survey Sampling

The goal of weighted sampling is typically to estimate the population total $Y=\sum y_i$ for a finite population of N units.

An associated estimator of Y :

$$\hat{Y}=\sum_{i=1}^{N}\frac{y_i}{\pi_i}X_i$$

where $X=(X_1,X_2,\cdots X_N)$: a random sample of n units from the population.

The variance of $\hat{Y}$ is :

$$V(\hat{Y})=\sum_{i=1}^{N}\frac{1-\pi_i}{\pi_i}y_i^2+2\sum_{1\le i\le j\le N}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}y_iy_j \tag{13}$$

where π_{ij} is the second-order inclusion probability for both the units i,j to be in the sample.

When n is fixed, an alternative expression is :

$$V(\hat{Y})=\sum_{1\le i\le j\le N}(\pi_i\pi_j-\pi_{ij})\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2. \tag{14}$$

If $\pi_{ij}\neq o$ for all pairs $i\neq j$ and n is fixed,

the estimator of $V(\hat{Y})$:

$$v(\hat{Y})=\sum_{1\le i\le j\le N}\frac{\pi_i\pi_j-\pi_{ij}}{\pi_{ij}}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2X_iX_j \tag{15}$$

property 1. The inclusion probabilities of any order are uniquely determined by the π_i and can be expressed in closed form. i.e.

$$\pi_{ij}=w_iw_jR(n-2,\{i,j\}^c)/R(n,S). \tag{16}$$

In general, the kth-order ($1\le k\le n$) inclusion probability for the units $i_1,\cdots,i_k$ to be in the sample is

$$\pi_{i_1,\cdots,i_k}=(\prod_{l=1}^{k}w_{i_l})\cdot\frac{R(n-k,\{i_1,\cdots,i_k\}^c)}{R(n,S)} \tag{17}$$

property 2. For the maximum entropy model, $0 < \pi_{ij} < \pi_i \pi_j$, for any pair $i \neq j$.

property 3. $\pi_i \propto y_j \Leftrightarrow V(\hat{Y})=0 \Leftrightarrow v(\hat{Y})=0$ for all possible samples.

property 4. $\sum_{i=1}^{N} \pi_i = n,\quad \sum_{j \neq i}^{N} \pi_{ij} = (n-1)\pi_i,\quad \sum_{i}^{N}\sum_{j>i}^{N} \pi_{ij} = \frac{1}{2} n(n-1).$

Example 1.

Let N=100. A vector π^* is generated uniformly from the simplex.

$$\{\pi=(\pi_1,\cdots,\pi_{100}) : 0<\pi_i<1, i=1,\cdots,100 : \sum \pi_i = 50\}$$

i) w^* correponding to π^* $\Rightarrow$ is found from (6) via (7)

ii) This particular $w^*$$\Rightarrow$ is used as the weights for 5 different maximum entropy models with sample size n= 2, 5, 15, 30, 50 respectively

iii) The coverage probabilities π for each of five models $\Rightarrow$ are obtained by (5)

iv) Each of five π's $\Rightarrow$ is converted to a w by (7) then w should be the same as w^* up to a scalar.

v) We use max $|w_i^{(t)}/\tilde{w}_i - 1| < 0.01$ as the stopping rule, where $w^{(t)}$ is the value of w at step t and $\tilde{w} = w^* \pi_N / w_N^*$ is the fixed point.

Table 1. Number of iterations for computing w from π and the range of w and π for sample size $n=2,5,15,30,50$

n	Number of iterations	$\pi_{100}(w_{100})$	π_1	w_1	w_1/π_1
2	3	0.2634	5.350×10^{-5}	4.651×10^{-5}	0.8736
5	4	0.4576	1.489×10^{-4}	9.026×10^{-5}	0.5938
15	10	0.7730	6.201×10^{-4}	1.547×10^{-4}	0.2301
30	20	0.9165	2.020×10^{-3}	1.813×10^{-4}	0.08851
50	16	0.9903	7.627×10^{-3}	1.947×10^{-4}	0.02553

Example 2 Let consider PPS (prob. proportional to size).

$$S=\{U_i \mid i=1,2,\cdots,N\}$$

$$\rightarrow P(U_i) = \frac{M_i}{M} \equiv p_i$$

where $M_i=|U_i| \Rightarrow$ size of U_i unit and $M=\sum_{i=1}^{N} M_i$.

We put $N=4$, $n=2$ and $p_i=\frac{M_i}{M}$ is :

Unit	U_i	1	2	3	4
relative size	p_i	0.1	0.2	0.3	0.4

Table 2. PPS ($N=4$, $n=2$)

Sample(S)	U_iU_j	pps wor		1	2	3	4	
1	1,2	$p_1 \cdot p_{2\|1}$		0.022	0.022			
2	1,3	$p_1 \cdot p_{3\|1}$		0.034		0.034		
3	1,4	$p_1 \cdot p_{4\|1}$		0.044			0.044	
4	2,1	$p_2 \cdot p_{1\|2}$		0.025	0.025			
5	2,3	$p_2 \cdot p_{3\|2}$			0.075	0.075		
6	2,4	$p_2 \cdot p_{4\|2}$			0.100		0.100	
7	3,1	$p_3 \cdot p_{1\|3}$		0.043		0.043		
8	3,2	$p_3 \cdot p_{2\|3}$			0.086	0.086		
9	3,4	$p_3 \cdot p_{4\|3}$				0.171	0.171	
10	4,1	$p_4 \cdot p_{1\|4}$		0.067			0.067	
11	4,2	$p_4 \cdot p_{2\|4}$			0.133		0.133	
12	4,3	$p_4 \cdot p_{3\|4}$				0.200	0.200	
			π_i	0.235	0.441	0.609	0.715	$\sum \pi_i=2$

Then π_i are 0.235, 0.441, 0.609 and 0.715 respectively. The corresponding w_i found from (6) by maximum entropy model are 0.14727, 0.30939, 0.49508 and 0.715, respectively.

We use max $|w^{(t)}/w^{(t-1)}-1| < 0.001$ where $w^{(t)}$ is the value of w at step t and it needs 9 steps iterations to compute w from π.

The second-order inclusion properties π_{ij} are given in table 3 by (16).

Table 3. Second-order inclusion properties π_{ij}

i	j=2	j=3	j=4
1	0.04785	0.07658	0.11058
2		0.16087	0.23233
3			0.37178

We can certify that $\pi_{12}=p_1 \cdot p_{2|1}+p_2 \cdot p_{1|2}=0.022+0.025=0.047$ in tables 2 by calculator is same as $\pi_{12}=0.04785$ in table 3 by computer program for $i=1, j=2$ as an example.

Also we can find that $\sum_{i=1}^{4}\sum_{j>i}^{4}\pi_{ij}=1$ and $\sum_{i\neq j}^{N}\pi_{ij}=\pi_{12}+\pi_{13}+\pi_{14}=0.23491=\pi_1$ $=(n-1)\pi_i$ for $i=1$ in table 3. which are property 4.

It means that the distribution of maximum entropy model can be applied usefully for the sampling distribution of large sample sizes.

Appendix

(Program for Example 2)

```
proc iml;
reset print;
p={ 0.235, 0.441, 0.609, 0.715 };
w=p;
I={ 0 1 1 1, 1 0 1 1, 1 1 0 1, 1 1 1 0 };
k=J(4, 1, 1);

q=w[1: 3, ];
s=q[+, ];
d=w*diag(s);
n=w`*i;
w=d/n`;
oldw=w;
count=1;
print count, w;

do until (abs(m)<0.001);
oldw=w;
```

```
q=w[1:3, ];
s=q[+, ];
d=p*diag(s);
n=w`*i;
w=d/n`;
a=w/oldw-k;
a=a[1:3, ];
m=a[<>, ];
count=count+1;
print count, w;
end;
quit;
```

References

[1] Brown, L.D. (1986). Fundamentals of Statistical Exponential Families (with Applicati-ons in statistical Decision Theory). Hayward, CA : Institute of Mathematical Statistics

[2] Chaudhuri, A & Vos, J.W.E.(1988) Unified Theory and Strategies of Survey Sampling. New York : Elsevier Science.

[3] Darroch, J.N & Ratcliff, D. (1972) Generalized iterative scaling for log- linear models. Ann. Math. Statist. 43, 1470-80.

[4] Deming, W.E & Stephan, F.F(1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals ard known. Ann, Math. Statist. 11, 427-44.

[5] Fellegi, I.P. (1963). Sampling with varying probabilities without replacement : Rotating and non- rotating samples. Am. Statist. Assoc. J. 183-201.

[6] Hanif, M & Brewer, K.R.W. (1980) Sampling with unequal probabilities without replacement : A review Int. Statist. Rev. 48, 317-35.

[7] Hansen, M.H & Hurwitz, W.N. (1943) On the theory of sampling from a finite population. Ann. Math. Statist. 14, 332-62.

[8] Hartley, H. O & Rao, J.N.K. (1962). Sampling with unequal probabilities and without replacement. Ann. Math. Statist. 33, 350-74.

[9] Horvitz, D.G. & Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. J. Am. Statist. Assoc. 47, 663-85.

[10] Joe, H. (1990). A winning strategy for lotto games? Can. J. Statist. 18,

1233-44.
[11] Smith. A.F.M & Roberts. G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. J.R. Statist. Soc. B55. 3-23.
[12] Stern. H & Cover. T.M. (1989). Maximum entropy and lottery. J. Am. Statist. Assoc. 84. 980-85.
[13] Yates. F. & Grundy. P.M. (1953). Selection without replacement from within strata with probability proportional to size. J.R. Statist. Soc. B15. 253-61.
[14] 박홍래. (1989). 통계조사론. 영지문화사.