The Ninth Korea and Japan Joint Conference of Statistics

Varying Probability Sampling and Maximum Entropy

IK Chan Kim[1]

ABSTRACT

Consider a random variable X, which must take on one of the values $x_1, x_2, \cdots, x_n$ with respective probabilities $p_1, p_2, \cdots p_n$. As $-\log p_i$ represents the surprise evoked if X takes on the value x_i, it follows that the expected amount of surprise we shall receive upon learning the value of X is given by

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \ .$$

The quantity $H(X)$ is known in information theory as the entropy of the random variable X.

As $H(X)$ represents the average amount of surprise one receives upon learning the value of X, it can be interpreted as representing the amount of uncertainty that exists as to the value of X.

We consider $H(X)$ as the randomness in sampling distribution.

We suggest the maximum entropy model in varying probability sampling, especially Probabilities Proportional to Size(PPS) and we present two methods for stepwise selection procedures of sampling unit, 'forword' and 'backward' in cases of n is fixed and n is non-fixed.

KEYWORDS: Independent Bernoulli trials, Maximum entropy, Sampling with varying probabilities, PPS, forward and backward procedure.

1. INTRODUCTION

Random sampling of n distinct units from a finite population of N units without replacement may be called varying probability sampling or weighed sampling when the probabilities associated with $\binom{N}{n}$ possible choices are not all equal.

The random sample may be denoted by X where $X = (X_1, X_2, \cdots, X_N)$ and the random variable X_i takes the values 1 or 0 according as the ith unit is in or out of the sample for $i = 1, 2, \cdots, N$.

1 Department of Computer Science and Statistics, Cheju National University, Cheju 690-756, Korea

Let $D^n = \{x = (x_1, x_2, \cdots, x_N)\ ;\ x_i = 0$ or 1, and $x_1 + x_2 + \cdots + x_N = n\}$

The random vector X takes values in D^n. We denote the pdf of a sampling scheme by $p(x)$ for $x \in D^n$, where $p(x) > 0$ and $\sum_{x \in D^n} p(x) = 1$. The associated probability that the sample includes the ith unit is

$$\pi_i = E(X_i) = \sum_{x \in D^n} x_i p(x) \tag{1}$$

The marginal probabilities π_i that the sample includes the ith population unit are prespecified, where

$$0 < \pi_i < 1 \text{ for } i = 1, 2, \cdots, N \quad \text{and} \quad \sum_{i=1}^{N} \pi_i = n. \tag{2}$$

We propose the particular family of sampling schemes.

Pick any vector of probabilities $P = (p_1, p_2, \cdots, p_N)$, where $0 < p_i < 1$ for $i = 1, \cdots, N$ and define $Z = (Z_1, Z_2, \cdots, Z_N)$ to be independent Bernoulli trials with $p_1, p_2, \cdots p_N$. Then define the sampling distribution of X to be the same as the conditional distribution of Z given $\sum_{i=1}^{N} Z_i = n$.

If the w_i are proportional to $p_i/(1-p_i)$, it means $p(x) \propto \prod_{i=1}^{N} w_i^{x_i}$, where vector of weights $w = (w_1, \cdots, w_N)$, $w_i > 0$.

We will refer to

$$p(x) = \prod_{i=1}^{N} w_i^{x_i} \Big/ \sum_{y \in D^n} \left(\prod_{i=1}^{N} w_i^{y_i} \right) \propto \exp\left(\sum_{i=1}^{N} \theta_i x_i \right), \ x \in D^n \tag{3}$$

as the maximum entropy model, where w or θ are determined by π through (1).

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

To compute w from π, we recast (1) in the form of a set of equations (4) below, and solve these iteratively as in (6). 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) $\sum_{j \in C} w_j R(k-1, C \setminus \{j\}) = kR(k,C)$

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 notatioh, we may rewrite (1) as

$$\pi_i = \frac{w_i R(n-1, \{i\}^c)}{R(n,S)} \qquad (i = 1,2,\cdots,N) \tag{4}$$

By a) $\sum \pi_i = \sum_i \frac{w_i R(n-1, \{i\}^C)}{R(n,S)} = n$, which is the result in (2).

Thus for fixed n, there are $N-1$ linearly independent relations among the N relations of (4). Without loss of generality, we assume that $\pi_1 \le \pi_2 \le \cdots \le \pi_N$ and let

$\pi_N = w_N$, then $\pi_N = \frac{w_N R(n-1, \{N\}^c)}{R(n,S)}$ by (4) i.e. $R(n,S) = R(n-1, \{N\}^c)$.

Hence we get

$$w_i = \frac{\pi_i R(n-1, \{N\}^c)}{R(n-1, \{i\}^c)} \qquad (i = 1,\cdots,N-1), \ w_N = \pi_N \tag{5}$$

Although a closed-form solution of (5) 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 (5);

$$w_i^{(t+1)} = \left. \frac{\pi_i R(n-1, \{N\}^c)}{R(n-1, \{i\}^c)} \right|_{w = w^{(t)}} ; \ (i = 1,\cdots,N-1), \quad w_N^{(t+1)} = w_N^{(t)} = \pi_N \tag{6}$$

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

3. SELECTION PROCEDURES

We discuss selection procedures for drawing a sample from the maximum entropy model. 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 population and take the remaining n units as the sample. Noticing that for any $x \in D^n$, $p(x) \infty \Pi_{i\in A^c} w_i^{-1}$, where $A=\{i;\, x_i=1\}$, it is obvious that for any 'forward' procedure, there is a corresponding 'backward' procedure, which selects a sample in the same way using w_i^{-1} instead of w_i, and vise versa.

Let $A_k (k=1,\cdots,n)$ denote the collection of the indices of the selected units after k draws without regard for orders and $A_0=\phi$.

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

$$P_1(j, A_{k-1}^c)=\frac{1}{n-k+1} P(Z_j=1 \mid \sum_{t\in A_{k-1}^c} Z_t = n-k+1) \tag{7}$$

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

$$R(k, C) = P(\sum_{t\in C} Z_t = k) \prod_{i\in C}(1+w_i)$$

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

Thus P_1 can be rewritten as

$$P_1(j, A_{k-1}^c)=\frac{w_j \cdot R(n-k, A_{k-1}^c \setminus \{j\})}{(n-k+1)R(n-k+1, A_{k-1}^c)} \tag{8}$$

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

$$P_1(j, A_k^c)=\frac{w_{i_k} P_1(j, A_{k-1}^c)-w_j P_1(i_k, A_{k-1}^c)}{(n-k)(w_{i_k}-w_j)P_1(i_k, A_{k-1}^c)}. \tag{9}$$

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

1. For $j=1,2,\cdots,N$, calculate $P_1(j,S)$, 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 $A_0 \leftarrow \phi$, $A_1 \leftarrow \{i_1\}$, $k \leftarrow 2$, goto 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 (5). Then draw unit i_k according to the probability $P_1(i_k, A_{k-1}^c)$.
4. If $k<n$, then $A_k \leftarrow A_{k-1}\cup\{i_k\}$, $k\leftarrow k+1$, goto 3; otherwise stop.

Let $B_m (m=1,2,\cdots,N-n)$ denote the collection of the indices of the units dropped after m draws without regard for orders. And let $B_0=\phi$.

Procedure 2 (backward, n fixed) At the mth draw ($m=1,2,\cdots,N-n$), a unit $j\in B_{m-1}^{c}$ is selected with probability

$$P_2(j, B_{m-1}^c)=\frac{1}{N-n-m+1}P(Z_j=0\mid \sum_{t\in B_{m-1}^c} Z_t=n) \tag{10}$$

Procedure 3 (forward, n non-fixed) At the kth draw, a unit $j\in A_{k-1}^{c}$ is selected with probability

$$P_3(j, A_{k-1}^c)=\sum_{i=0}^{k-1}\frac{1}{k-i}P(Z_j=1,\ \sum_{t\in A_{i-1}^c} Z_t=k-i\mid \sum_{t\in S}Z_t=k) \tag{11}$$

Procedure 4 (backward, n non-fixed) At the mth draw ($m=1,2,\cdots,N-n$), a unit $j\in B_{m-1}^{c}$ is selected with probability

$$P_4(j, B_{m-1}^c)=\sum_{i=0}^{N-m}\frac{1}{N-m-i+1}P(Z_j=0,\ \sum_{t\in B_{m-1}^c} Z_t=i\mid \sum_{t\in S}Z_t=N-m) \tag{12}$$

The four procedures have different uses. Procedure 1 and 2 require less operations than Procedure 3 and 4, but can not be used when n is non-fixed. Procedure 3 and 4 are useful for doing rotations in survey sampling. The preference between forward and backward procedures depend on the scale on n.
Apparently forward are preferred when $n\le N/2$ while backward are preferred when $n > N/2$. When the w_i are all equal, all the four procedures reduce to the simple random sampling without replacement.

4. Application to Survey Sampling

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

$$\pi_{ij}=w_iw_jR(n-2,\{i,j\}^c)/R(n,S). \tag{13}$$

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

$$\pi_{i_1,\cdots,i_k}=(\prod_{l=1}^{k}w_{i_l})\cdot\frac{R(n-k,\{i_1,\cdots,i_k\}^c)}{R(n,S)} \tag{14}$$

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

property 3. $\sum_{i=1}^{N}\pi_i=n,\quad \sum_{j\ne i}^{N}\pi_{ij}=(n-1)\pi_i,\quad \sum_{i}^{N}\sum_{j>i}^{N}\pi_{ij}=\frac{1}{2}n(n-1).$

Example 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 1. PPS ($N=4$, $n=2$)

Sample(S)	U_iU_j	pps wor	unit	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	$\sum\pi_i=2$
			w_i	0.14727	0.30939	0.49508	0.715	

Then π_i are 0.235, 0.441, 0.609 and 0.715 respectively. The corresponding w_i found from (5) 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 2 by (13).

Table 2. 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 $\pi_{12} = p_1 \cdot p_{2|1} + p_2 : p_{1|2} = 0.022 + 0.025 = 0.047$ in tables 1 by calculator is same as $\pi_{12} = 0.04785$ in table 2 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 2, which are property 3.

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

REFERENCES

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

(2) Hanif, M. and Brewer, K.R.W. (1980) Sampling with unequal probabilities without replacement : A review, Intern. Statist. Review. 48, 317-335.

(3) Hartley, H.O. and Rao, J.N.K. (1962). Sampling with unequal probabilities and without replacement. Ann. Math. Statist., 33, 350-374.

(4) Kim, I.C. (1996). PPS to Maximize Entropy and Selection Procedures, 1996 Proceedings of the Autumn Conference, Korean Statistical Society 157-160.

(5) Stern, H and Cover, T.M. (1989). Maximum entropy and the lottery, Jour. of American Statist. Asso., 84, 980-985.