J. Austral. Math. Soc. (*Series A*) **44** (1988), 197–213

SAMPLING WITHOUT REPLACEMENT: APPROXIMATION TO THE PROBABILITY DISTRIBUTION

J. N. DARROCH, M. JIRINA and T. P. SPEED

(Received 10 May 1985; revised 2 February 1987)

Communicated by T. C. Brown

Abstract

Let P be the probability distribution of a sample without replacement of size n from a finite population represented by the set $\mathbf{N} = \{1, 2, \ldots, N\}$. For each $r = 0, 1, \ldots$, an approximation P_r is described such that the uniform norm $\|P - P_r\|$ is of order $(n^2/N)^{r+1}$ if $n^2/N \to 0$. The approximation P_r is a linear combination of uniform probability product-measures concentrated on certain subspaces of the sample space $\mathbf{N}^n$.

1980 *Mathematics subject classification* (*Amer. Math. Soc.*): 60 F 05, 62 E 20.

1. Introduction

A finite population of size N will be represented by the set $\mathbf{N} = \{1, 2, \ldots, N\}$. A sample of size n will be denoted by $x = (x_1, \ldots, x_n)$ (with $x_i \in \mathbf{N}$). The set $X = \mathbf{N}^n$ is the set of all sample realizations associated with sampling with replacement; the corresponding probability measure will be denoted by P_0. The set $X_0 = \{x \in X$: all x_i different$\}$ is the set of all sample realizations associated with sampling without replacement; the corresponding probability measure will be denoted by P. Clearly $P_0(x) = 1/N^n$ for all $x \in X$ and $P(x) = 1/[N]_n$ if $x \in X_0$, $P(x) = 0$ if $x \notin X_0$; here we have used $[z]_n$ as a symbol for the downward factorial $z(z-1)\cdots(z-n+1)$. Finally, if Q is a signed measure on X, we shall denote by $\|Q\|$ its uniform norm, that is, $\|Q\| = \max_A |Q(A)|$.

It is clear that if N is much larger than n, the two probability measures P and P_0 should not differ much. D. Freedman proved in fact in [3] that

$$\|P - P_0\| = 1 - \prod_{j=1}^{n-1}\left(1 - \frac{j}{N}\right). \tag{1}$$

From this, simple inequalities like

$$1 - \exp\left(-\binom{n}{2}/N\right) < \|P - P_0\| < \binom{n}{2}/N$$

and an asymptotic relation

$$\|P - P_0\| \sim \frac{1}{2}\frac{n^2}{N} \quad \text{as } \frac{n^2}{N} \to 0 \tag{2}$$

follow.

The proof of (1) is so simple that it is worth repeating it here: For the signed measure $Q = P - P_0$ we have

$$Q(x) = \frac{1}{[N]_n} - \frac{1}{N^n} \quad \text{if } x \in X_0 \quad \text{and} \quad Q(x) = -\frac{1}{N^n} \quad \text{if } x \notin X_0,$$

so that the corresponding Hahn decomposition of X is

$$A^+ = \{x: Q(x) \geqslant 0\} = X_0, \qquad A^- = \{x: Q(x) \leqslant 0\} = X - X_0.$$

As $Q(X) = 0$, $\|Q\| = Q(A^+) = |Q(A^-)|$ and $Q(A^+) = 1 - [N]_n/N^n$. This proves (1).

We may consider P_0 as an approximation of P with the maximum error of order $O(\frac{n^2}{N})$. In the next section we shall derive an improved approximation P_1 and show that its maximum error is of order $O((\frac{n^2}{N})^2)$. In Section 3, an approximation with maximum error of order $O((\frac{n^2}{N})^{r+1})$ is presented for an arbitrary $r = 0, 1, 2, \ldots$. The proof concerning the general case may be found in Section 4. It utilizes certain results from the theory of partition lattices and applies to any $r = 0, 1, \ldots$. A direct elementary proof demonstrated for $r = 1$ in Section 2 becomes more and more complicated as r increases and would not be feasible as a general proof.

2. Second order approximation

Let Φ be the generating function of P and Φ_0 the generating function of P_0, that is, for any $v = (v_1, \ldots, v_n)$

$$\Phi(v) = \frac{1}{[N]_n} \sum_{x \in X_0} v_1^{x_1} \cdots v_n^{x_n}$$

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

and $\Phi_0(v) = \prod_{i=1}^{n} \varphi(v_i)$ with $\varphi(u) = (1/N)\sum_{y=1}^{N} u^y$. For each v, $\Phi(v)$ is the coefficient at $z_1 z_2 \cdots z_n$ in the power series expansion of

$$G(z) = \frac{N^n}{[N]_n} \prod_{y=1}^{N} \left(1 + \frac{1}{N}(z_1 v_1^y + \cdots + z_n v_n^y)\right).$$

The logarithm of the product $\prod_{y=1}^{N}$ can be written in the form

$$\sum_{y=1}^{N} \left[\frac{1}{N} \sum_i z_i v_i^y - \frac{1}{N^2} \sum_{i<j} z_i z_j (v_i v_j)^y\right] + O\left(\frac{1}{N^2}\right) + Z$$

$$= \sum_i z_i \varphi(v_i) - \frac{1}{N} \sum_{i<j} z_i z_j \varphi(v_i v_j) + O\left(\frac{1}{N^2}\right) + Z$$

where Z stands for terms containing at least one z_i in higher power. Further

$$\frac{N^n}{[N]_n} = \left[\prod_{j=1}^{n-1} \left(1 - \frac{j}{N}\right)\right]^{-1} = 1 + \binom{n}{2} \Big/ N + O\left(\frac{1}{N^2}\right).$$

Hence

$$G(z) = \left[1 + \binom{n}{2}\frac{1}{N} + O\left(\frac{1}{N^2}\right)\right] \prod_{i=1}^{n} (1 + z_i \varphi(v_i) + Z)$$

$$\cdot \prod_{i<j} \left(1 - \frac{1}{N} z_i z_j \varphi(v_i v_j) + O\left(\frac{1}{N^2}\right)\right) \cdot \left(1 + O\left(\frac{1}{N^2}\right) + Z\right).$$

Equating the coefficients at $z_1 z_2 \cdots z_n$ on both sides we get

$$\Phi(v) = \Phi_1(v) + O\left(\frac{1}{N^2}\right)$$

where

$$\Phi_1(v) = \left(1 + \binom{n}{2}\frac{1}{N}\right)\Phi_0(v) - \frac{1}{N}\Phi_0(v) \sum_{i<j} \frac{\varphi(v_i v_j)}{\varphi(v_i)\varphi(v_j)}.$$

The (signed) measure P_1 corresponding to this generating function can be described in the following way. For any subset $\{i, j\}$ of $\{1, 2, \ldots, n\}$ put $X_{\{i,j\}} = \{x = (x_1, \ldots, x_n) \in X\colon x_i = x_j\}$ and let $P_{\{i,j\}}$ be the uniform probability distribution concentrated on $X_{\{i,j\}}$, that is, $P_{\{i,j\}}(x) = \frac{1}{N^{n-1}}$ if $x \in X_{\{i,j\}}$, $P_{\{i,j\}}(x) \doteq 0$ otherwise. Then

$$P_1 = \left(1 + \binom{n}{2} \Big/ N\right) P_0 - \frac{1}{N} \sum_{i<j} P_{\{i,j\}} = P_0 + \frac{1}{N}\left[\binom{n}{2} P_0 - \sum_{i<j} P_{\{i,j\}}\right].$$

We have not specified the exact meaning of $O(\frac{1}{N^2})$ and, even if we had done so, it would not be clear what implication this would have for the difference $P - P_1$. However we may use the above mentioned Freedman method and prove directly

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

that

$$\|P - P_1\| \sim \frac{1}{4}\left(\frac{n^2}{N}\right)^2 \quad \text{as } \frac{n^2}{N} \to 0. \tag{3}$$

PROOF OF (3). Let X_1 be the set of all $x = (x_1, \ldots, x_n) \in X$ with exactly one tie among the x_i, that is, with the x_i assuming exactly $(n-1)$ distinct values. The signed measure $Q = P - P_1$ satisfies

$$Q(x) = \begin{cases} \dfrac{1}{[N]_n} - \left(1 + \binom{n}{2}\Big/N\right)\dfrac{1}{N^n} & \text{if } x \in X_0, \\ \dfrac{1}{N^n}\left[N^{n-1}\sum_{i<j} P_{\{i,j\}}(x) - 1 - \binom{n}{2}\Big/N\right] & \text{if } x \notin X_0. \end{cases}$$

Hence,

$$Q(x) = \frac{1}{N^n}\left[\prod_{j=1}^{n-1}\left(1 - \frac{j}{N}\right)\right]^{-1} - 1 - \binom{n}{2}\Big/N \geqslant 0 \quad \text{if } x \in X_0,$$

$$Q(x) \geqslant \frac{1}{N^n}\left[2 - 1 - \binom{n}{2}\Big/N\right] \geqslant 0 \quad \text{if } x \in X - (X_0 \cup X_1) \text{ and } \binom{n}{2}\Big/N \leqslant 1,$$

$$Q(x) = -\frac{1}{N^{n+1}}\binom{n}{2} \leqslant 0 \quad \text{if } x \in X_1.$$

This proves that $X_1 = A^-$ = the negative part of the Hahn decomposition for Q if $\binom{n}{2}/N \leqslant 1$ and we have again $Q(x) = 0$. Hence

$$\|Q\| = |Q(A^-)| = \binom{n}{2}^2 \frac{[N]_{n-1}}{N^{n+1}} \quad \text{if } \binom{n}{2}\Big/N \leqslant 1.$$

From this (3) follows easily.

3. The general case

Several results from [1] will be quoted later and, therefore, we shall use the notation of [1] whenever possible. For any natural number n, $\mathscr{P}(n)$ will denote the set of all partitions $\pi = \{A_1, \ldots, A_b\}$ of the set $\mathbf{n} = \{1, 2, \ldots, n\}$. The subsets A_i will be called blocks. For any $\pi \in \mathscr{P}(n)$, $b(\pi)$ will denote the number of blocks in π and $\mathscr{P}_i(n) = \{\pi \in \mathscr{P}(n)\colon b(\pi) = n - i\}$. We shall say that $\pi = \{A_1, A_2, \ldots, A_b\}$ is of type $k_1, k_2, \ldots, k_b$, if, for each i, $k_i = |A_i|$ = the number of elements in A_i. Occasionally, when describing the type of partition, we shall write $1^{a_1}2^{a_2}\cdots$ instead of repeating 1 a_1 times, 2 a_2 times, etc. For any partition

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

π of type $k_1, \dots, k_b$ we shall put

$$\nu(\pi) = \prod_{i=1}^{b} (k_i - 1)!.$$

For each $\pi = \{A_1, \dots, A_{b(\pi)}\} \in \mathscr{P}(n)$, X_π will denote the set of all $x = (x_1, \dots, x_n) \in X$ with the x_i tied at least on each A_j and P_π the uniform probability measure concentrated on X_π. X_π is a rectangular subset of X containing $N^{b(\pi)}$ elements and P_π is a product-type measure with $P_\pi(x) = 1/N^{b(\pi)}$ if $x \in X_\pi$. Note: $X_{\{i,j\}}$ and $P_{\{i,j\}}$ of Section 2 become X_π and P_π if π is of type $1^{n-2}2^1$ with $\{i, j\}$ as the only block with 2 elements.

For each $n = 1, 2, \dots$, the sequence $v_0(n), v_1(n), \dots$ is defined by its generating function $F_n(z) = [\prod_{j=1}^{n-1}(1 - jz)]^{-1}$. Obviously, all $v_i(n)$ are nonnegative (in fact strictly positive if $n \geqslant 2$) and they satisfy the recursion

$$v_{i+1}(n+1) = n v_i(n+1) + v_{i+1}(n).$$

They are also related to the Stirling numbers $S(n, i)$ of the second kind by

$$v_i(n) = S(n + i - 1, n - 1) \qquad (n \geqslant 1, i \geqslant 0). \tag{4}$$

(See, for example, [1], page 232.)

If we repeated the procedure used at the beginning of the previous section keeping this time all terms associated with $1/N, \dots, 1/N^r$ $(r = 0, 1, \dots)$, we would obtain a generating function of a (signed) measure

$$\begin{aligned} P_r &= \sum_{i=0}^{r} \left(-\frac{1}{N}\right)^i \left(\sum_{j=0}^{r-i} \frac{v_j(n)}{N^j} \right) \sum_{\pi \in \mathscr{P}_i(n)} \nu(\pi) P_\pi \\ &= \sum_{k=0}^{r} \frac{1}{N^k} \left(\sum_{i=0}^{k} v_{k-i}(n)(-1)^i \sum_{\pi \in \mathscr{P}_i(n)} \nu(\pi) P_\pi \right). \end{aligned} \tag{5}$$

The measures P_0 and P_1 defined by this formula agree with those of Sections 1 and 2. For $r = 2$ we get

$$P_2 = P_1 + \frac{1}{N^2} \left[\binom{n+1}{3} \frac{3n-2}{4} P_0 - \binom{n}{2} \sum_{\substack{\pi \text{ of type} \\ (1^{n-2}2^1)}} P_\pi + \sum_{\substack{\pi \text{ of type} \\ (1^{n-4}2^2)}} P_\pi + 2 \sum_{\substack{\pi \text{ of type} \\ (1^{n-3}3^1)}} P_\pi \right].$$

We accept P_r as the rth approximation to P and we shall prove the following theorem.

THEOREM.

$$\|P - P_r\| \simeq \frac{1}{2(r+1)!} \left(\frac{n^2}{N}\right)^{r+1} \quad \text{as } \frac{n^2}{N} \to 0.$$

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

The proof of the theorem is based on three lemmas. The following notation is used: For $j = 0, 1, \ldots, n-1$,

$$X_j = \{x = (x_1, \ldots, x_n) \in X: \text{ the } x_i \text{ assume exactly } n - j \text{ distinct values}\}$$

and A^+, A^- is the Hahn decomposition of X for the signed measure $P - P_r$.

LEMMA 1. *For each* $r = 0, 1, \ldots, P_r(X) = 1$.

LEMMA 2. *For sufficiently small* n^2/N,

$$A^+ = X_0 \cup X_2 \cup \cdots \cup X_r \quad \text{if } r \text{ is even}$$

and

$$A^- = X_1 \cup X_3 \cup \cdots \cup X_r \quad \text{if } r \text{ is odd}.$$

LEMMA 3. *For sufficiently small* n^2/N,

$$\|P - P_r\| = (P - P_r)(X_0) - P_r(X_2) - \cdots - P_r(X_r) \quad \text{if } r \text{ is even}$$

and

$$\|P - P_r\| = P_r(X_1) + P_r(X_3) + \cdots + P_r(X_r) \quad \text{if } r \text{ is odd}.$$

Lemmas 1 and 2 will be proved in Section 4. Lemma 3 is an easy consequence of Lemmas 1 and 2. Explicit formulae for $P_r(X_j)$, involving Stirling numbers, may be written down and we may then prove the main theorem using well-known asymptotic formulae for Stirling numbers. For more details, see the end of Section 4.

4. Proofs

For any π, σ from $\mathscr{P}(n)$ we shall write $\pi \leqslant \sigma$ if and only if π is a refinement of σ. With this partial ordering, $\mathscr{P}(n)$ is a lattice; its least element (= the finest partition) will be denoted by 0. The function ζ will be defined by

$$\zeta(\sigma, \pi) = \begin{cases} 1 & \text{if } \sigma \leqslant \pi, \\ 0 & \text{otherwise,} \end{cases}$$

and μ will denote the corresponding Möbius function ($\mu = \zeta^{-1}$, see [1], page 141). By a formula in [1], page 163,

$$\mu(0, \pi) = (-1)^{n - b(\pi)} \nu(\pi) \quad \text{for each } \pi \in \mathscr{P}(n). \tag{7}$$

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

The numbers

$$w_i(n) = \sum_{\pi \in \mathscr{P}_i(n)} \mu(0, \pi) \tag{8}$$

are called the ith level numbers of the first kind ([1], page 155) and they are related to the Stirling numbers s of the first kind by

$$w_i(n) = s(n, n - i) \tag{9}$$

(see [1], 4.20(iv), page 155). It follows from the definition of the Stirling numbers ([1], page 88) that the sequence $w_\cdot(n)$ has for each n the generating function $G_n(z) = \prod_{j=1}^{n-1}(1 - jz)$.

The sequences $v_\cdot(n)$ (defined in Section 3) and $w_\cdot(n)$ are related by

$$(v_\cdot(n) * w_\cdot(n))(k) = \delta_{0,k} \quad \text{for } k = 0, 1, \ldots \tag{10}$$

where $*$ denotes convolution and δ is the Kronecker symbol. This follows from $F_n(z)G_n(z) \equiv 1$.

The following asymptotic formulae hold for each fixed $i = 0, 1, \ldots$.

$$v_i(n) \simeq \frac{n^{2i}}{2^i i!} \quad \text{as } n \to \infty \tag{11}$$

and

$$w_i(n) \simeq (-1)^i \frac{n^{2i}}{2^i i!} \quad \text{as } n \to \infty. \tag{12}$$

These relations follow from the well-known asymptotic formulae for Stirling numbers (see, for example, [2], page 293), but they may be also derived directly. The sequences $v_i(n)/n^{2i}$ and $w_i(n)/n^{2i}$ $(i = 0, 1, \ldots)$ have generating functions $F_n(z/n^2)$ and $G_n(z/n^2)$ respectively and these converge (as $n \to \infty$) to $e^{z/2}$ and $e^{-z/2}$ respectively.

For each $\pi \in \mathscr{P}(n)$, we shall put

$$w_i(n, \pi) = \sum_{\sigma \in \mathscr{P}_i(n)} \mu(0, \sigma)\zeta(\sigma, \pi)$$

and

$$R_j(\pi) = \sum_{i=0}^{j} w_i(n, \pi).$$

The numbers $w_i(n, \pi)$ are in fact the ith level numbers for the sublattice $\{\sigma \in \mathscr{P}(n): \sigma \leqslant \pi\}$ and this sublattice is isomorphic to the product lattice $\mathscr{P}(k_1) \times \cdots \times \mathscr{P}(k_b)$, if π is of type $k_1, \ldots, k_b$. Using this fact we can show through a simple calculation that

$$w_i(n, \pi) = s(k_1, \cdot) * \cdots * s(k_b, \cdot)(n - i). \tag{13}$$

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

The set $X = \mathbf{N}^n$ is the set of all maps x: $\mathbf{n} \to \mathbf{N}$ and, therefore, ker x (the kernel of the map x) is well defined (see, for example, [1], page 5). It is a partition of $\mathbf{n}$ such that x is constant on each block of ker x and assumes distinct values on distinct blocks. We shall write $b(x) = b(\ker x)$; $b(x)$ is the number of distinct values in the image of x. We shall also write $R_j(x) = R_j(\ker x)$ and use R_j as a symbol for the measure defined by $R_j(x)$.

The sets X_i and X_π of Section 3 and the probability measure P_π may be now re-defined as follows: $X_i = \{x \in X: b(x) = n - i\}$, $X_\pi = \{x \in X;\ \pi \leqslant \ker x\}$ and $P_\pi(x) = \frac{1}{N^{b(\pi)}}\zeta(\pi, \ker x)$.

Using this, we may write down another formula for P_r, namely

$$P_r = \frac{1}{N^n} \sum_{j=0}^{r} \alpha_j(n, N) R_{r-j} \tag{14}$$

where

$$\alpha_j(n, N) = \frac{v_j(n)}{N^j}.$$

PROOF OF LEMMA 1. Using second part of (5), (7), (8) and $P_\pi(X) = 1$ we get

$$P_r(X) = \sum_{k=0}^{r} \frac{1}{N^k}(v_\cdot(n) * w_\cdot(n))(k).$$

The rest follows from (10).

PROOF OF LEMMA 2. The proof will consist of several steps. The following additional symbols will be used: If $\pi \in \mathscr{P}(n)$ is of type $k_1, \ldots, k_b$, then $\tilde{b}(\pi)$ will denote the number of blocks with $k_i \geqslant 2$ (while $b = b(\pi) =$ the number of all blocks),

$$m(\pi) = \max_i \{k_1, \ldots, k_b\} \quad \text{and} \quad d(\pi) = \sum_{i=1}^{b} \binom{k_i}{2}.$$

They satisfy

$$m(\pi) - 1 \leqslant \sqrt{2d(\pi)} \quad \text{and} \quad \tilde{b}(\pi) \leqslant d(\pi). \tag{15}$$

(a) *Let $\pi \in \mathscr{P}(n)$ be of type $k_1, \ldots, k_b$. Then the sequence $\{(-1)^j R_j(\pi)\}$, $j = 0, 1, \ldots,$ has a generating function*

$$K(z) = \frac{1}{1+z} \prod_{i=1}^{b} G_{k_i}(-z) = \frac{1}{1+z} \prod_{i=1}^{b} \prod_{j=1}^{k_i - 1} (1 + jz).$$

PROOF OF (a). By (13), the sequence $w_\cdot(n, \pi)$ has the generating function

$$H(z) = z^n \prod_{i=1}^{b} \left[\frac{1}{z}\right]_{k_i} = \prod_{i=1}^{b} G_{k_i}(z).$$

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

The sequence $R_j(\pi) = \sum_{i=0}^{j} w_i(n, \pi)$ may be considered as a convolution of the sequence $1, 1, \ldots$ and $w_i(n, \pi)$. Hence, the sequence $R_j(\pi)$ has the generating function $(1 - z)^{-1}H(z)$ and the sequence $\{(-1)^j R_j(\pi)\}$ has the generating function $(1 + z)^{-1}H(-z)$.

(b) *If* $\pi \in \mathscr{P}_0(n)$, *then* $R_j(\pi) = 1$ *for all* $j \geqslant 0$.

If $\pi \in \mathscr{P}_l(n)$ *with* $l \geqslant 1$, *then*

$$(-1)^j R_j(\pi) \geqslant 1 \quad \text{for } 0 \leqslant j \leqslant l - 1,$$

$$R_j(\pi) = 0 \quad \text{for } j \geqslant l.$$

PROOF OF (b). The case $\pi \in \mathscr{P}_0(n)$ is trivial. If $l \geqslant 1$, then, in the formula for $K(z)$ in (a), at least one k_i is $\geqslant 2$, so that $K(z)$ is a polynomial of degree $(l - 1)$ and all its coefficients are strictly positive integers.

(c) *To each pair* A, B *of real numbers there exists a constant* $C_{A,B}$ (*independent of* n) *such that*

$$|R_j(\pi)| \leqslant C_{A,B}$$

for all j *and for all* $\pi \in \mathscr{P}(n)$, $n = 1, 2, \ldots$ *satisfying* $m(\pi) \leqslant A$, $\tilde{b}(\pi) \leqslant B$.

PROOF OF (c). If $\pi \in \mathscr{P}_0(n)$, then $R_j(\pi) = 1$ for all j by (b).

If $\pi \in \mathscr{P}(n) \setminus \mathscr{P}_0(n)$, then by (b)

$$|R_j(\pi)| \leqslant \sum_{j=0}^{\infty} |R_j(\pi)| = K(1) \leqslant \frac{1}{2}\left[\prod_{j=1}^{A-1}(1 + j)\right]^B.$$

(d) *To each* $l = 0, 1, \ldots$ *there exists a constant* C_l (*independent of* n) *such that*

$$|R_j(\pi)| \leqslant C_l$$

for all j *and for all* $\pi \in \mathscr{P}_l(n)$, $n = 1, 2, \ldots$.

PROOF OF (d). If $\pi \in \mathscr{P}_l(n)$, then $m(\sigma) \leqslant 2l$, $\tilde{b}(\sigma) \leqslant l$ and (c) applies.

(e) *Let a sequence of partitions* $\pi_n \in \mathscr{P}(n)$ *be such that* $d(\pi_n) \to \infty$. *Then for each* $j = 0, 1, \ldots$

$$|R_j(\pi_n)| \simeq \frac{(d(\pi_n))^j}{j!} \quad \text{as } n \to \infty.$$

PROOF OF (e). Put $d_n = d(\pi_n)$, $b_n = b(\pi_n)$, $m_n = m(\pi_n)$, assume that π_n is of type $k_{1,n}, \ldots, k_{b_n,n}$ and denote by $L_n(z)$ the generating function of $|R_j(\pi_n)|/d_n^j$. As $d_n \to \infty$, $\pi_n \notin \mathscr{P}_0(n)$ for sufficiently large n, $|R_j(\pi_n)| = (-1)^j R_j(\pi_n)$ by (b)

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

and $L_n(z) = K(z/d_n)$. Using the inequality $y - y^2/2 \leqslant \ln(1+y) \leqslant y$ we get

$$\ln L_n(z) \leqslant \frac{z}{d_n}\left[\sum_{i=1}^{b_n}\sum_{j=1}^{k_{i,n}-1} j - 1\right] = z\frac{d_n - 1}{d_n}$$

and

$$\ln L_n(z) \geqslant z\frac{d_n - 1}{d_n} - \frac{1}{2}\left(\frac{z}{d_n}\right)^2 Z_n$$

where

$$0 \leqslant Z_n = \sum_{i=1}^{b_n}\sum_{j=1}^{k_{i,n}-1} j^2 - 1 \leqslant \sqrt{2d_n}\sum_{i=1}^{b_n}\sum_{j=1}^{k_{i,n}-1} j = \sqrt{2}\, d_n^{3/2}.$$

In the last step we used $j \leqslant m_n - 1$ for all j involved, and (15). Hence, $Z_n/d_n^2 \underset{n}{\to} 0$ and $L_n(z) \underset{n}{\to} e^z$.

(f) *Let j, r be two (fixed) integers such that $0 \leqslant j \leqslant r$. Put, for each n,*

$$M_n = \max_{\pi \in \bigcup_{r<l<n} \mathscr{P}_l(n)} \left|\frac{R_j(\pi)}{R_r(\pi)}\right|.$$

Then $\sup_n M_n < \infty$.

Proof of (f). By (b), $|R_r(\pi)| \geqslant 1$ for all $\pi \in \bigcup_{r<l<n} \mathscr{P}_l(n)$ so that M_n are well defined. Suppose $\sup_n M_n = \infty$. Then there exists a sequence $n_p \underset{p}{\to} \infty$ and $\pi_p \in \bigcup_{r<l<n_p} \mathscr{P}_l(n_p)$ such that

$$\left|\frac{R_j(\pi_p)}{R_r(\pi_p)}\right| \underset{p}{\to} \infty. \tag{16}$$

If $d = \sup_p d(\pi_p) < \infty$, then $\sup_p |R_j(\pi_p)| < \infty$ by (15) and (c). This together with $|R_j(\pi_p)| \geqslant 1$ contradicts (16).

If $d = \infty$, we may assume without loss of generality that $d(\pi_p) \underset{p}{\to} \infty$. Then, by (e), $|R_j(\pi_p)/R_r(\pi_p)| \underset{p}{\to} 0$ contradicting again (16).

In the next assertion, it is important that all inequalities hold uniformly with respect to all n, N. Therefore, we shall write $X_l^{(n,N)}$ instead of X_l to stress the dependence on n, N. We shall also use the fact that $R_j(x) = R_j(\ker x)$ does not depend on N and that $x \in X_l^{(n,N)}$ if and only if $\ker x \in \mathscr{P}_l(n)$.

(g) *Let $r \geqslant 0$ be fixed. Then there exists $\varepsilon > 0$ such that for all pairs n, N satisfying $n^2/N \leqslant \varepsilon$,*

$$(-1)^l (P - P_r)(x) \geqslant 0 \quad \text{if } 0 \leqslant l \leqslant r \quad \text{and} \quad x \in X_l^{(n,N)},$$

$$(-1)^{r-1}(P - P_r)(x) \geqslant 0 \quad \text{if } x \in \bigcup_{l>r} X_l^{(n,N)}.$$

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

PROOF OF (g). We shall use the formula (14) for P_r.

(α) Assume first that $x \in X_0^{(n,N)}$. Then, by (b),

$$(17)\qquad (P - P_r)(x) = \frac{1}{[N]_n} - \frac{1}{N^n}\sum_{j=0}^{r} \alpha_j(n,N)$$

$$= \frac{1}{N^n}\left[F_n\left(\frac{1}{N}\right) - \sum_{j=0}^{r} \frac{v_j(n)}{N^j}\right] = \frac{1}{N^n}\sum_{j=r+1} \frac{v_j(n)}{N^j}$$

$$\geqslant 0 \quad \text{for all } n, N.$$

(β) Assume now that $1 \leqslant l \leqslant r$ and $x \in X_l^{(n,N)}$. By (b),

$$(-1)^l(P - P_r)(x) = (-1)^{l-1}P_r(x)$$

$$= \frac{\alpha_{r-l+1}(n,N)}{N^n}\left[(-1)^{l-1}R_{l-1}(x) + (-1)^{l-1}\sum_{j=0}^{l-2} \beta_j(n,N)R_j(x)\right]$$

where

$$0 \leqslant \beta_j(n,N) = \frac{\alpha_{r-j}(n,N)}{\alpha_{r-l+1}(n,N)} \leqslant c\left(\frac{n^2}{N}\right)^{l-1-j} \quad \text{for some } c > 0$$

and all n, N; the last inequality follows from (11). By (e), all $R_j(x)$ are uniformly bounded with respect to n and $x \in X_l^{(n,N)}$. Hence, for sufficiently small n^2/N, the last sum is negligible with respect to $(-1)^{l-1}R_{l-1}(x) \geqslant 1$.

(γ) If $x \in \bigcup_{l>r} X_l^{(n,N)}$, then $(-1)^r R_r(x) \geqslant 1$ by (b) and

$$(-1)^{r-1}(P - P_r)(x)$$

$$= (-1)^r P_r(x) = \frac{(-1)^r R_r(x)}{N^n}\left[1 + \sum_{j=0}^{r-1} \alpha_{r-j}(n,N)\frac{R_j(x)}{R_r(x)}\right].$$

By (11), $0 \leqslant \alpha_{r-j}(n,N) \leqslant c(n^2/N)^{r-j}$ for all n, N and some $c > 0$ and, by (f), $R_j(x)/R_r(x)$ are uniformly bounded with respect to all $x \in \bigcup_{l>r} X_l^{(n,N)}$ and all n, N. Hence, for sufficiently small n^2/N, the last sum is negligible with respect to 1.

Lemma 2 follows now from (g).

Lemma 3 is an easy consequence of Lemmas 1 and 2, however to deduce the main theorem from Lemma 3, we must find the asymptotic behaviour of $(P - P_r)(X_0)$ and of $P_r(X_l)$ for $l \geqslant 1$.

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

Exact formulae. By (17),

$$(P - P_r)(X_0) = \frac{[N]_n}{N^n} \sum_{j=r+1}^{\infty} \frac{v_j(n)}{N^j}. \tag{18}$$

Using a well-known property of Stirling numbers of the second kind, namely that $S(n,k)$ is the number of partitions in $\mathscr{P}_{n-k}(n)$ (see for example, [1], 2.66, page 70), we can see easily that

$$\sum_{x \in X_l} \zeta(\pi, \ker x) = S(n-i, n-l)[N]_{n-l} \quad \text{if } 0 \leqslant i \leqslant l \quad \text{and} \quad \pi \in \mathscr{P}_i(n).$$

Hence, for $0 \leqslant j \leqslant l$,

$$R_j(X_l) = \sum_{x \in X_l} \sum_{i=0}^{j} w_i(n, \ker x) = [N]_{n-l} \sum_{i=0}^{j} v_{l-i}(n-l+1) w_i(n) \tag{19}$$

and, by (14) and (b),

$$P_r(X_l) = \frac{1}{N^n} \sum_{j=0}^{l-1} \frac{v_{r-j}(n)}{N^{r-j}} R_j(X_l) \tag{20}$$

if $0 \leqslant l \leqslant r$.

Asymptotic formulae (for $n^2/N \to 0$). Using (11) and keeping in (18) the term of lowest order in n^2/N, we get

$$(P - P_r)(X_0) \simeq \frac{1}{2^{r+1}(r+1)!} \left(\frac{n^2}{N}\right)^{r+1}. \tag{21}$$

Applying (11) and (12) to (19) we get, for $0 \leqslant j \leqslant l$,

$$\frac{1}{N^n} R_j(X_l) \simeq \beta(j,l) \frac{1}{2^l} \left(\frac{n^2}{N}\right)^l$$

where $\beta(j,l) = \sum_{i=0}^{j} (-1)^i / i!(l-i)!$; in particular, $\beta(l-1, l) = (-1)^{l-1}/l!$ if $l \geqslant 1$. Hence, by (11) and (20),

$$P_r(X_l) \simeq \frac{(-1)^{l-1}}{2^{r+1} l!(r-l+1)!} \left(\frac{n^2}{N}\right)^{r+1} \quad \text{if } 1 \leqslant l \leqslant r. \tag{22}$$

To conclude the proof of the main theorem of Section 3, assume first that r is even. Then, by Lemma 3, (21) and (22)

$$\begin{aligned} \|P - P_r\| &\simeq \frac{1}{2^{r+1}} \left(\frac{n^2}{N}\right)^{r+1} \left[\frac{1}{(r+1)!} + \frac{1}{2!(r-1)!} + \cdots + \frac{1}{r!1!}\right] \\ &= \frac{1}{2(r+1)!} \left(\frac{n^2}{N}\right)^{r+1}. \end{aligned}$$

The proof for r odd would be similar.

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

5. An application of the approximate distributions

The approximations P_r defined by (5) might seem more complicated than the exact distribution P. The purpose of this section is to show that, in spite of this, they do sometimes provide reasonably simple formulae for evaluating certain probabilities while the exact distribution does not.

As an example consider the distribution of the sum of n random variables, which have been randomly sampled from a population of N random variables. Such a distribution often underlies frequency data. For instance, a cohort study of the etiology of a certain disease might begin by selecting a random sample of size n from a population of N 25-year old people and end twenty years later by recording the number of people in the sample who contracted the disease. This number is the sum of n Bernoulli variables randomly selected from a population of N Bernoulli variables.

In general let the jth individual bear a random variable Y_j. The random variables $Y_1, \dots, Y_N$ are assumed independent although not identically distributed; the probability distribution of Y_j will be denoted by Q_j. The sampling consists of first selecting at random n individuals without replacement and then reading the realizations of the corresponding n random variables $Y_{x_1}, \dots, Y_{x_n}$. Finally, let $S = Y_{\xi_1} + \cdots + Y_{\xi_n}$ where $\xi = (\xi_1, \dots, \xi_n)$ is the random vector the realizations of which are the labels $(x_1, \dots, x_n)$ of the n selected individuals. If we denote by $P^{(S)}$ the probability distribution of S based on the exact distribution P of ξ, we have

$$P^{(S)} = \sum_x (Q_{x_1} * \cdots * Q_{x_n}) P(x)$$

where $*$ denotes convolution. Similarly, if we denote by $P_r^{(S)}$ the approximate distribution of S based on the rth approximation P_r, we have

$$P_r^{(S)} = \sum_x (Q_{x_1} * \cdots * Q_{x_n}) P_r(x).$$

From the inequality

$$\| P^{(S)} - P_r^{(S)} \| \leqslant \| P - P_r \| \tag{23}$$

and from the main theorem it follows that $\|P^{(S)} - P_r^{(S)}\|$ is asymptotically $\leqslant 1/(2(r+1)!)(n^2/N)^{r+1}$. To prove (23), put $M = P - P_r$, $h(x) = Q_{x_1} * \cdots * Q_{x_n}$ and denote by (A^+, A^-) the Hahn decomposition of X with respect to M. Then for any B and $A = h^{-1}(B)$,

$$\begin{aligned} \left| \left(P^{(S)} - P_r^{(S)} \right)(B) \right| &= \left| \sum_{x \in A \cap A^+} h(x) M(x) + \sum_{x \in A \cap A^-} h(x) M(x) \right| \\ &\leqslant \max\{ M(A^+), |M(A^-)| \} = \| P - P_r \|. \end{aligned}$$

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

There is no simple formula for $P^{(S)}$. However for $r = 0, 1, 2$ we get

$$P_0^{(S)} = \bar{Q}^{*n}, \qquad P_1^{(S)} = P_0^{(S)} + \binom{n}{2} \Big/ N \left[\bar{Q}^n - \bar{Q}^{*(n-2)} * \bar{Q}_2\right],$$

$$P_2^{(S)} = P_1^{(S)} + \binom{n}{2} \Big/ N^2 \left[\frac{(n+1)(3n-2)}{12} \bar{Q}^{*n} - \binom{n}{2} \bar{Q}^{*(n-2)} * \bar{Q}_2 + \frac{2}{3}(n-2) \bar{Q}^{*(n-3)} * \bar{Q}_3 + \frac{1}{2}\binom{n-2}{2} \bar{Q}^{*(n-4)} * \bar{Q}_2 * \bar{Q}_2\right],$$

where

$$\bar{Q} = \frac{1}{N} \sum_{j=1}^{N} Q_j, \quad \bar{Q}_2 = \frac{1}{N} \sum_{j=1}^{N} Q_j * Q_j,$$

$$\bar{Q}_3 = \frac{1}{N} \sum_{j=1}^{N} Q_j * Q_j * Q_j, \quad \bar{Q}^{*n} = \bar{Q} * \cdots * \bar{Q} \quad (n\text{-times}).$$

The 0th approximation $P_0^{(S)}$ is a proper probability distribution. This is not necessarily true for $P_r^{(S)}$ with $r \geqslant 1$, as some very small exact probabilities might become negative, but the error of replacing these negative probabilities by 0s cannot exceed the overall error of the approximation and is therefore negligible. A similar comment applies to the complementary events, the "probabilities" of which become slightly greater than 1; it is always true that $P_r^{(S)}(R) = 1$.

If the random variables Y_j have finite expectations, say m_j, then it is easy to see that the exact probability distribution $P^{(S)}$ has the expectation $n\bar{m}$, where $\bar{m} = (1/N)\Sigma_{j=1}^{N} m_j$. Exactly the same holds for each $P_r^{(S)}$. For $r = 0, 1, 2$, this may be proved in the following way. The probability distributions $\bar{Q}$, $\bar{Q}_2$, $\bar{Q}_3$ have the expectations $\bar{m}$, $2\bar{m}$, $3\bar{m}$ respectively. Hence $P_0^{(S)}$ has the expectation $n\bar{m}$, $P_1^{(S)}$ has the expectation $n\bar{m} + \binom{n}{2}/N^2[n\bar{m} - n\bar{m}] = n\bar{m}$ and $P_2^{(S)}$ has the expectation

$$n\bar{m}\left\{1 + \binom{n}{2} \Big/ N^2 \left[(n+1)(3n-2)/2 - \binom{n}{2} + \frac{2}{3}(n-2) + \frac{1}{2}\binom{n-2}{2}\right]\right\} = n\bar{m}.$$

More involved techniques would be necessary to prove the same fact for $r > 2$.

In the particular case mentioned above when the Y_j have Bernoulli distributions with parameters π_j, that is, $\mathscr{P}r(Y_j = 1) = \pi_j$, $\mathscr{P}r(Y_j = 0) = 1 - \pi_j$, the random variable S has a discrete distribution. If we denote by φ_0, φ_1, φ_2 the

generating functions for the distributions of S under the approximation $P_0^{(S)}$, $P_1^{(S)}$, $P_2^{(S)}$ respectively, we obtain from the previous formulae, after suitable re-arrangements

$$\varphi_0(z) = \psi(z)^n, \qquad \varphi_1(z) = \varphi_0(z) - \frac{1}{N}\binom{n}{2}\mu_2\psi(z)^{n-2}(1-z)^2,$$

$$\varphi_2(z) = \varphi_1(z) - \frac{1}{N^2}\binom{n}{2}\left\{\mu_2\psi(z)^{n-2}(1-z)^2 + \frac{2}{3}(n-2)\mu_3\psi(z)^{n-3}(1-z)^3 - \frac{1}{2}\binom{n-2}{2}\mu_2^2\psi(z)^{n-4}(1-z)^4\right\},$$

where $\psi(z) = 1 - \bar{\pi} + \bar{\pi}z$, $\bar{\pi} = (1/N)\Sigma_j \pi_j$, $\mu_2 = (1/N)\Sigma_j(\pi_j - \bar{\pi})^2$, $\mu_3 = (1/N)\Sigma_j(\pi_j - \bar{\pi})^3$. This may be rewritten as

$$P_0^{(S)}(k) = B_{n,\bar{\pi}}(k),$$

$$P_1^{(S)}(k) = P_0(k) - \frac{1}{N}\binom{n}{2}\mu_2\sum_{j=0}^{2}(-1)^j\binom{2}{j}B_{n-2,\bar{\pi}}(k-j),$$

$$P_2^{(S)}(k) = P_1^{(S)}(k) - \frac{1}{N^2}\binom{n}{2}\left\{\mu_2\sum_{j=0}^{2}(-1)^j\binom{2}{j}B_{n-2,\bar{\pi}}(k-j) + \frac{2}{3}(n-2)\mu_3\sum_{j=0}^{3}(-1)^j\binom{3}{j}B_{n-3,\bar{\pi}}(k-j) - \frac{1}{2}\binom{n-2}{2}\mu_2^2\sum_{j=0}^{4}(-1)^j\binom{4}{j}B_{n-4,\bar{\pi}}(k-j)\right\},$$

where $B_{n,\pi}$ denotes the binomial probability function with parameters n, π.

The distribution $P^{(S)}$ arises from two sources of stochastic variation, random sampling in the first place and Bernoulli variation in the second. When the second source is eliminated and each π_j is 0 or 1, $P^{(S)}(k)$ is the hypergeometric probability

$$P^{(S)}(k) = \frac{\binom{N_1}{k}\binom{N_0}{n-k}}{\binom{N}{n}}$$

where $N_1 = \sum_{j=1}^{N}\pi_j$, $N_0 = N - N_1$. The approximations to $P^{(S)}$ cease to have any practical value but have some theoretical interest. The moments are now $\bar{\pi} = N_1/N$, $\mu_2 = N_0N_1/N^2$, $\mu_3 = N_0N_1(N_0 - N_1)/N^3$. The approximation of $P^{(S)}(k)$ by $P_0^{(S)}(k)$ is the classical approximation of the hypergeometric probability by

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

the binomial probability, and the formulae for $P_1^{(S)}$, $P_2^{(S)}$, $P_3^{(S)}$ show how this approximation can be improved by the inclusion of further terms which are themselves binomial probabilities.

We conclude with two numerical examples. In the first example, $P^{(S)}$ is the hypergeometric distribution with $N = 30$, $N_1 = 15$ and $n = 5$. In the second example, we consider the more general model mentioned above with 30 Bernoulli random variables Y_j with parameters

$$(\pi_1, \pi_2, \ldots, \pi_{30}) = (0.1, 0.2, \ldots, 0.15, 0.85, 0.86, \ldots, 0.99).$$

The bottom line in the tables gives the numbers $d_r = \|P^{(S)} - P_r^{(S)}\|$.

EXAMPLE 1

x	$P^{(S)}(x)$	$P_0^{(S)}(x)$	$P_1^{(S)}(x)$	$P_2^{(S)}(x)$
0	0.0211	0.0313	0.0208	0.0203
1	0.1437	0.1562	0.1459	0.1432
2	0.3352	0.3125	0.3333	0.3365
3	0.3352	0.3125	0.3333	0.3365
4	0.1437	0.1562	0.1459	0.1432
5	0.0211	0.0313	0.0208	0.0203
		d_0	d_1	d_2
		0.0454	0.0044	0.0026

EXAMPLE 2

x	$P^{(S)}(x)$	$P_0^{(S)}(x)$	$P_1^{(S)}(x)$	$P_2^{(S)}(x)$
0	0.0239	0.0313	0.0238	0.0233
1	0.1477	0.1562	0.1488	0.1473
2	0.3285	0.3125	0.3274	0.3294
3	0.3285	0.3125	0.3274	0.3294
4	0.1477	0.1562	0.1488	0.1473
5	0.0239	0.0313	0.0238	0.0233
		d_0	d_1	d_2
		0.0320	0.0024	0.0020

https://doi.org/10.1017/S1446788700029785 Published online by Cambridge University Press

The 0th approximation $P_0^{(S)}$ is the same in both examples; the higher approximations are much closer to the respective exact distributions $P^{(S)}$. In the first example, calculating the exact distribution poses no problem. On the other hand, calculating the exact distribution $P^{(S)}$ in the second example required considerable computer time while calculating the corresponding approximations is as easy as in the first example.

References

[1] M. Aigner, *Combinational Theory* (Springer-Verlag, 1979).
[2] L. Comtet, *Advanced Combinatorix* (Reidel, 1974).
[3] D. Freedman, 'A remark on the difference between sampling with or without replacement', *J. Amer. Stat. Assoc.* **72** (1977), 681.

Department of Mathematics
The Flinders University of
South Australia
Bedford Park, S. A. 5042
Australia

Division of Mathematics and Statistics
C.S.I.R.O.
Yarralumla
Canberra, A.C.T. 2600
Australia