KYBERNETIKA — VOLUME 31 (1995), NUMBER 3, PAGES 303–314

ON HÁJEK'S CONJECTURE IN STRATIFIED SAMPLING

ZUZANA PRÁŠKOVÁ

In the paper a conjecture by Jaroslav Hájek, concerning asymptotic relations between including probabilities in conditional Poisson sampling and parameters of Poisson sampling in the case that a finite population is divided into strata, is verified. An alternative proving method to that proposed by Hájek is used.

1. INTRODUCTION

Let S be a population of N units, $s \subset S$ a sample and P a probability distribution defined on the set of all subsets of S. Let $I_i(s)$ be the including indicator of the unit i, i.e. a random variable with value 1 if $s \ni i$ and 0 otherwise, let $\pi_i = EI_i$ denote the probability of inclusion of the unit i into the sample.

One of the most important probability sampling scheme is Poisson sampling with parameters $0 < p_i < 1$, $i = 1, \ldots, N$ defined for any $s \subset S$ by probabilities

$$P(s) = \prod_{i \in s} p_i \prod_{i \in S-s} (1 - p_i).$$

The indicators of inclusion in this case are independent random variables satisfying $P(I_i = 1) = p_i = 1 - P(I_i = 0)$ and for including probabilities the identity $\pi_i = p_i$ holds for $i = 1, \ldots, N$. The size of sample $K(s) = \sum_{i=1}^{N} I_i(s)$ is a random variable.

The main role of Poisson sampling is to help to define and analyze other sampling procedures. It si known that sampling methods as simple random sampling, stratified or not, rejective sampling, stratified or not, two-stage sampling and others may be described as conditional Poisson sampling.

If we define a sampling plan as conditional Poisson sampling, the problem arises how to evaluate probabilities of inclusion, because the parameters $p_1, \ldots, p_N$ may not yield exact values of $\pi_1, \ldots, \pi_N$. For example, rejective sampling of size n can be defined as conditional Poisson sampling under the condition that the sample size is fixed and equal to n. Then only asymptotic approximation of π_i by means of $p_1, \ldots, p_N$ is available (see Hájek [3], Chapter 7 for more detail).

The same problem has to be solved if the population S is divided into strata $S_1, \dots, S_m$. Hájek [3] dealt with conditional Poisson sampling given fixed strata sample sizes $K_h = size(s \cap S_h), h = 1, \dots, m$ and pronounced a conjecture in which an asymptotic relation between including probabilities π_i and parameters of Poisson sampling is formulated. The strata are assumed to be disjoint as well as overlapping. Hájek proposed to solve the problem by means of multivariate quadratic regression and by using normal approximation for the vector of regressors.

Here we use another approach based on the Bayes theorem and on refined multivariate local limit theorem.

2. HÁJEK'S CONJECTURE

Let the population S is divided into strata $S_1, \dots, S_m$, $S = \cup_{h=1}^m S_h$. Let $s \subset S$ be the sample and let K_h denote the sample size in S_h. Consider a conditional Poisson sampling with parameters $p_1, \dots, p_n$ given the condition that the sample sizes K_h are fixed and equal to $n_h = EK_h$, $1 \le h \le m$ where n_j are integers. Let I_i be including indicators in Poisson sampling.

Define $\pi_i = E(I_i | K_h = n_h, 1 \le h \le m)$, $1 \le i \le N$ and $\pi_{ij} = E(I_i I_j | K_h = n_h, 1 \le h \le m)$, $1 \le i \ne j \le N$. In Hájek [3] the following conjecture (Conjecture 14.1) is pronounced:

Hájek's Conjecture. For $1 \le i \le N$

$$\pi_i = p_i \Big\{ 1 - \frac{1}{2}(1 - p_i)\left[(1 - 2p_i)\, v_i' D^{-1} v_i - \sum_{j=1}^N (v_i' D^{-1} v_j)(v_j' D^{-1} v_j)\, p_j (1-p_j)(1-2p_j)\right] + o(d^{-1}) \Big\} \quad \text{as } d \to \infty, \tag{1}$$

where D^{-1} is the inverse (or the generalized inverse) of the matrix

$$D = \left[\sum_{j=1}^N v_{hj} v_{kj} p_j (1 - p_j) \right]_{h,k=1}^m , \tag{2}$$

d is the minimal nonzero eigenvalue of D and v_j is the vector $(v_{1j}, \dots, v_{mj})'$ with values

$$v_{hj} = \begin{cases} 1 & j \in S_h \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Furthermore, for $1 \le i \ne j \le N$

$$\pi_i \pi_j - \pi_{ij} \sim (1 - \lambda_i)(1 - \lambda_j)\, \pi_i \pi_j v_i' D_\lambda^{-1} v_j \text{ as } d \to \infty, \tag{4}$$

where

$$D_\lambda = \left[\sum_{j=1}^N v_{hj} v_{kj} \pi_j (1 - \lambda_j) \right]_{h,k=1}^m \tag{5}$$

and the coefficients $1-\lambda_j,\ 1\le j\le N$ are solutions of

$$(1-\lambda_j)\,[1-\pi_j(1-\lambda_j)\,v_j' D_\lambda^{-1} v_j] = 1-\pi_j. \tag{6}$$

Relation (4) is uniform in $1\le i<j\le N$.

Remark. Assumption $d\to\infty$ in (1) was not pronounced by Hájek explicitly but it follows from a context.

Theorem 1. Under assumption that D is regular the Hájek's conjecture holds.

3. PROOFS

It can be easily seen that

$$\pi_i = P(I_i=1|K_1=n_1,\dots,K_m=n_m) = p_i\frac{P(K_1=n_1,\dots,K_m=n_m|I_i=1)}{P(K_1=n_1,\dots,K_m=n_m)}, \tag{7}$$

where P is the probability measure induced by the Poisson sampling. Recall that in Poisson sampling $EI_i=p_i$, $\operatorname{Var} I_i = p_i(1-p_i)$ and I_i are independent. For any $1\le h\le m$ we have

$$K_h = \sum_{j\in S_h} I_j = \sum_{j=1}^N I_j v_{hj}$$

$$EK_h = \sum_{j=1}^N p_j v_{hj}, \qquad \operatorname{Var}(K_1,\dots,K_m) = D,$$

where the matrix D is defined by (2). If $m=1$, the conditional Poisson sampling described above reduces to the rejective sampling and the Conjecture holds with $D=d=\sum_{j=1}^N p_j(1-p_j)$, $D_\lambda=\sum_{j=1}^N \pi_j(1-\lambda_j)$ (see Hájek [3], Theorem 7.3, Theorem 7.4 and approximation (7.28)).

If $m>1$ and the strata are disjoint, the sample sizes $K_1,\dots,K_m$ in Poisson sampling are independent and

$$\pi_i = p_i\prod_{h=1}^m \frac{P(K_h=n_h|I_i=1)}{P(K_h=n_h)}.$$

Thus, we can apply rejective sampling in each stratum independently. Proceeding as in the proof of Theorem 7.3 in Hájek [3] we get (1) with $D=\operatorname{diag}[d_1,\dots,d_m]$ where $d_h=\sum_{j=1}^N p_j(1-p_j)\,v_{hj},\ h=1,\dots,m$. Further, we have

$$\pi_{ij} = P(I_i=1,I_j=1|K_1=n_1,\dots,K_m=n_m) = p_ip_j\prod_{l=1}^m \frac{P(K_l=n_l|I_i=1,I_j=1)}{P(K_l=n_l)}.$$

If $i \in S_h$, $j \in S_k$, $S_h \neq S_k$, the independence of I_j yields

$$\pi_{ij} = p_i \frac{P(K_h = n_h | I_i = 1)}{P(K_h = n_h)} p_j \frac{P(K_k = n_k | I_j = 1)}{P(K_k = n_k)} = \pi_i \pi_j$$

and $\pi_{ij} - \pi_i \pi_j = 0$. In this case (4) holds as identity since $D_\lambda = \operatorname{diag}[d_1(\lambda), \ldots, d_m(\lambda)]$ with $d_h(\lambda) = \sum_{j=1}^N \pi_j (1 - \lambda_j) v_{hj}$ and $v_i' D_\lambda^{-1} v_j = 0$.

For $i, j \in S_h$ we have

$$\pi_{ij} = p_i p_j \frac{P(K_h = n_h | I_i = 1, I_j = 1)}{P(K_h = n_h)}$$

and we can apply rejective sampling to the stratum S_h. The approximation (4) holds with D_λ as above.

Now, let us turn to the general case. First we introduce some notation. Denote

$$V = N^{-1} D = \left[\frac{1}{N} \sum_{j=1}^N p_j (1 - p_j) v_{hj} v_{kj} \right]_{h,k=1}^N \tag{8}$$

and suppose that D (and V) is regular. Put

$$w_j = (w_{1j}, \ldots, w_{mj})' = V^{-1/2} v_j \tag{9}$$

with the vectors v_j defined by (3).

For $t = (t_1, \ldots, t_m)'$ let $\varphi(t)$ denote the characteristic function of the random vector sample sizes $(K_1, \ldots, K_m)'$, i.e.

$$\varphi(t) = E \exp\{i(t_1 K_1 + \ldots + t_m K_m)\} \tag{10}$$

and for $j = 1, \ldots, N$ denote

$$\kappa_{3j} = p_j (1 - p_j)(1 - 2p_j), \tag{11}$$

$$\kappa_{4j} = p_j (1 - p_j)(1 - 2p_j)(1 - 6p_j + 6p_j^2). \tag{12}$$

For $u = (u_1, \ldots, u_m)'$ put

$$P_1(iu) = \frac{i^3}{3!} \frac{1}{N} \sum_{j=1}^N \kappa_{3j} (u' w_j)^3, \tag{13}$$

$$P_2(iu) = \frac{i^4}{4!} \frac{1}{N} \sum_{j=1}^N \kappa_{4j} (u' w_j)^4 + \frac{1}{2} (P_1(iu))^2 \tag{14}$$

and for $x = (x_1, \ldots, x_m)'$ let $Q_1(x)$, $Q_2(x)$ be polynomials in x such that for $\nu = 1, 2$

$$(2\pi)^{-m} \int_{R^m} \exp\left\{-iu'x - \frac{1}{2} u'u\right\} P_\nu(iu) \, du = Q_\nu(x) f(x), \tag{15}$$

where $f(x) = (2\pi)^{-m/2}\exp\{-x'x/2\}$ is the normal density. Especially, we have

$$\begin{aligned} Q_1(x) &= -\frac{1}{6}\sum_{h=1}^{m}(3x_h - x_h^3)\frac{1}{N}\sum_{j=1}^{N}\kappa_{3j}w_{hj}^3 \\ &- \frac{1}{2}\sum_{1\le h\ne k\le m}(x_h - x_h x_k^2)\frac{1}{N}\sum_{j=1}^{N}\kappa_{3j}w_{hj}w_{kj}^2 \\ &+ \sum_{1\le h<k<l\le m} x_h x_k x_l \frac{1}{N}\sum_{j=1}^{N}\kappa_{3j}w_{hj}w_{kj}w_{lj}. \end{aligned} \tag{16}$$

Finally, denote

$$\mu = \sum_{j=1}^{N} p_j v_j. \tag{17}$$

Now, we are able to prove the following theorem.

Theorem 2. Suppose that D is regular and $d_1 N \le d \le d_2 N$ for positive constants $d_1,\ d_2$. Then for any integers $k_1,\dots,k_m$

$$\begin{aligned} &|D|^{\frac{1}{2}}P(K_1 = k_1,\dots,K_m = k_m) = \\ &= (2\pi)^{-\frac{m}{2}}\exp\left\{-\frac{1}{2}x'x\right\}\left[1 + N^{-\frac{1}{2}}Q_1(x) + N^{-1}Q_2(x)\right] + o(N^{-1}), \end{aligned} \tag{18}$$

where $|D|$ is the determinant of D and for $k = (k_1,\dots,k_m)'$ we put $x = (x_1,\dots,x_m)' = D^{-\frac{1}{2}}(k-\mu)$. The result holds uniformly in k.

Proof. According to inverse formula for lattice vectors (see e.g. Bhattacharya and Rao [1], Chap. 5)

$$\begin{aligned} P(K_1 = k_1,\dots,K_m = k_m) &= (2\pi)^{-m}\int_{-\pi}^{\pi}\dots\int_{-\pi}^{\pi} e^{-it'k}\varphi(t)\,dt \\ &= (2\pi)^{-m}|D|^{-\frac{1}{2}}\int_{\Omega}\exp(-iu'D^{-\frac{1}{2}}k)\,\varphi(D^{-\frac{1}{2}}u)\,du, \end{aligned} \tag{19}$$

where $\Omega = \{u \in R^m : D^{-\frac{1}{2}}u \in [-\pi,\ \pi]^m\}$ and φ is the characteristic function defined by (10).

Now, notice that

$$\begin{aligned} \varphi(t) &= E\exp\left\{i\sum_{h=1}^{m}t_h K_h\right\} = E\exp\left\{i\sum_{j=1}^{N}I_j\sum_{h=1}^{m}t_h v_{hj}\right\} \\ &= E\exp\left\{i\sum_{j=1}^{N}I_j\ t'v_j\right\} = \prod_{j=1}^{N}\left[1 - p_j + p_j e^{it'v_j}\right] = \prod_{j=1}^{N}\varphi_j(t'v_j), \end{aligned} \tag{20}$$

where $\varphi_j(\tau) = 1 - p_j + p_j \exp(i\tau)$ is the characteristic function of the zero-one distribution with cumulants

$$\begin{aligned} \kappa_{1j} &= p_j, \\ \kappa_{\nu j} &= p_j(1-p_j)\frac{d\kappa_{\nu j-1}}{dp_j}, \quad \nu \geq 2. \end{aligned}$$

Obviously, κ_{3j} and κ_{4j} defined by (11) and (12) are the third and the fourth cumulants of the random variable with the characteristic function φ_j.

Thus, for $\|t\| < m^{-\frac{1}{2}}$ we get

$$\begin{aligned} \log \varphi(t) &= \sum_{j=1}^{N} \log \varphi_j(t'v_j) \\ &= i\sum_j p_j(t'v_j) - \frac{1}{2}\sum_j p_j(1-p_j)(t'v_j)^2 + \frac{i^3}{3!}\sum_j \kappa_{3j}(t'v_j)^3 + \frac{i^4}{4!}\sum_j \kappa_{4j}(t'v_j)^4 + R(t) \\ &= it'\mu - \frac{1}{2}t'Dt + \frac{i^3}{3!}\sum_j \kappa_{3j}(t'v_j)^3 + \frac{i^4}{4!}\sum_j \kappa_{4j}(t'v_j)^4 + R(t), \end{aligned}$$

where $|R(t)| \leq cN\|t\|^5$ for a positive constant c.

For $t = D^{-\frac{1}{2}}u$ we have $\|t\| \leq \|u\|\, d^{-\frac{1}{2}}$ and for $\|u\| \leq d^{\beta} m^{-\frac{1}{2}}$, $\beta < \frac{1}{2}$ we get

$$\begin{aligned} \varphi\left(D^{-\frac{1}{2}}u\right) &= \exp\left\{iu'D^{-\frac{1}{2}}\mu - \frac{1}{2}u'u\right\} \\ &\times \exp\left\{\frac{i^3}{3!}\sum_j \kappa_{3j}(u'D^{-\frac{1}{2}}v_j)^3 + \frac{i^4}{4!}\sum_j \kappa_{4j}(u'D^{-\frac{1}{2}}v_j)^4 + R(D^{-\frac{1}{2}}u)\right\}. \end{aligned}$$

If we insert $D^{-\frac{1}{2}}v_j = N^{-\frac{1}{2}}w_j$ (according to (8) and (9)) into the second exponent, use the Taylor expansion and order terms in the powers of $N^{-\frac{1}{2}}$, we get for $\|u\| < m^{-\frac{1}{2}}d^{\beta}$, $\beta < \frac{1}{6}$

$$\begin{aligned} \varphi\left(D^{-\frac{1}{2}}u\right) &= \exp\left\{iu'D^{-\frac{1}{2}}\mu - \frac{1}{2}u'u\right\} \\ &\times \left[1 + N^{-\frac{1}{2}}P_1(iu) + N^{-1}P_2(iu) + Z(u)\right], \end{aligned} \tag{21}$$

where $|Z(u)| \leq cN^{-\frac{3}{2}}Z_1(\|u\|)$ and $Z_1(\|u\|)$ is a polynomial in $\|u\|$ of degree at most 9. Denote

$$\begin{aligned} \Omega_1 &= \{u : \|u\| < m^{-\frac{1}{2}}d^{\beta},\ \beta < 1/6\}, \\ \Omega_2 &= \{u : \|u\| \leq \pi m^{-\frac{1}{2}}d^{\frac{1}{2}}\}. \end{aligned}$$

Further, notice that

$$|\varphi_j(\tau)| = [1 - 2p_j(1-p_j)(1-\cos\tau)]^{\frac{1}{2}} \leq \exp\{-p_j(1-p_j)(1-\cos\tau)\}.$$

Thus,

$$\left|\varphi(D^{-\frac{1}{2}}u)\right| \le \exp\left\{-\sum_{j=1}^{N} p_j(1-p_j)\,(1-\cos(u'D^{-\frac{1}{2}}v_j))\right\}.$$

In our next considerations we will use the inequality $1-\cos x \ge 2x^2\pi^{-2}$ valid for $|x| \le \pi$. Obviously, for $u \in \Omega_2$ we have $|u'D^{-\frac{1}{2}}v_j| \le ||u||\,||v_j||d^{-\frac{1}{2}} \le \pi$ and thus

$$\left|\varphi(D^{-\frac{1}{2}}u)\right| \le \exp\left\{-\frac{2}{\pi^2}u'D^{-\frac{1}{2}}\sum_{j=1}^{N} p_j(1-p_j)\,v_jv_j'D^{-\frac{1}{2}}u\right\} = \exp\left\{-\frac{2}{\pi^2}u'u\right\}. \tag{22}$$

Furthermore,

$$\begin{aligned}\sum\nolimits_j p_j(1-p_j)\,(1-\cos(u'D^{-\frac{1}{2}}v_j)) &= \sum\nolimits_j p_j(1-p_j) - Re\left[\exp\{iu'D^{-\frac{1}{2}}v_j\}\right]\\ &\ge \sum\nolimits_j p_j(1-p_j)\left[1-\left|\sum\nolimits_j q_j\exp\{iu'D^{-\frac{1}{2}}v_j\}\right|\right],\end{aligned}$$

where $q_j = p_j(1-p_j)\,(\sum(p_j(1-p_j))^{-1}$.

Hence, $\sum q_j\exp\{iu'D^{-\frac{1}{2}}v_j\}$ is the characteristic function of a vector Y taking values v_j with the probabilities $q_j, j=1,\ldots,N$, calculated at the point $D^{-\frac{1}{2}}u$. The lattice character of Y yields that there exists $\delta < 1$ such that for $u \in \Omega - \Omega_2$

$$\left|\sum_{j=1}^{N} q_j\exp\{iu'D^{-\frac{1}{2}}v_j\}\right| < \delta$$

(see (22.13) in Bhattacharya and Rao [1]) and thus

$$\left|\varphi(D^{-\frac{1}{2}}u)\right| \le \exp\left\{-(1-\delta)\sum p_j(1-p_j)\right\}.$$

Since $d||u||^2 \le u'Du \le m||u||^2\sum p_j(1-p_j)$ and since we have assumed that $d \ge d_1N$, we can conclude that there exist positive constants C and γ such that $\sum p_j(1-p_j) \ge CN$ and

$$\left|\varphi(D^{-\frac{1}{2}}u)\right| \le e^{-\gamma N} \quad \text{for } u \in \Omega - \Omega_2. \tag{23}$$

Now, combining (19), (15) and (21) we can write

$$\begin{aligned}&|D|^{\frac{1}{2}}P(K_1=k_1,\ldots,K_m=k_m) - f(x)\left[1+N^{-\frac{1}{2}}Q_1(x)+N^{-1}Q_2(x)\right]\\ =\;&(2\pi)^{-m}(I_1+I_2+I_3),\end{aligned}$$

where

$$\begin{aligned}I_1 &= \int_{\Omega_1}\exp\left\{-iu'x-\frac{1}{2}u'u\right\}Z(u)\,du,\\ I_2 &= \int_{\Omega-\Omega_1}\exp\left\{-iu'D^{-\frac{1}{2}}k\right\}\varphi\left(D^{-\frac{1}{2}}u\right)\,du,\\ I_3 &= -\int_{R^m-\Omega_1}\exp\left\{-iu'x-\frac{1}{2}u'u\right\}\left[1+N^{-\frac{1}{2}}P_1(iu)+N^{-1}P_2(iu)\right]\,du.\end{aligned}$$

Further, it follows from (21) that

$$|I_1| \le cN^{-\frac{3}{2}} \int_{R^m} \exp\{-u'u/2\}\, Z_1(\|u\|)\, \mathrm{d}u = o(N^{-1}).$$

Similarly, if we utilize the assumption $d \ge d_1 N$ and the fact that $P_1(iu), P_2(iu)$ are polynomials in u with bounded coefficients, we get

$$|I_3| = o(N^{-1}).$$

To estimate I_2, write

$$I_2 = \left(\int_{m^{-\frac{1}{2}} d^{\beta} \le \|u\| \le \pi m^{-\frac{1}{2}} d^{\frac{1}{2}}} + \int_{\Omega - \Omega_2} \right) \exp\{-iu' D^{-\frac{1}{2}} k\}\, \varphi(D^{-\frac{1}{2}} u)\, \mathrm{d}u = J_1 + J_2.$$

Obviously, (22) yields, with a constant M,

$$\begin{aligned} |J_1| &\le \int_{m^{-\frac{1}{2}} d^{\beta} \le \|u\| \le \pi m^{-\frac{1}{2}} d^{\frac{1}{2}}} \exp\left\{ -\frac{2}{\pi^2} u'u \right\} \mathrm{d}u \\ &\le \int_{\|u\| \ge M d^{\beta}} \exp\left\{ -\frac{2}{\pi^2} u'u \right\} \mathrm{d}u = o(N^{-1}). \end{aligned}$$

Finally, according to (23)

$$|J_2| \le \int_{\Omega - \Omega_2} \exp\{-\gamma N\}\, \mathrm{d}u \le \exp\{-\gamma N\} \int_{\Omega} \mathrm{d}u \le \mathrm{const}\, N^{\frac{m}{2}} \exp\{-\gamma N\} = o(N^{-1}).$$

Now, combining all these results we can conclude that (18) holds true. □

Corollary. If $n_1, \ldots, n_m$ are the strata sizes, then

$$P(K_1 = n_1, \ldots, K_m = n_m) = (2\pi)^{-\frac{m}{2}} |D|^{-\frac{1}{2}} [1 + N^{-1} Q_2(0) + o(N^{-1})]. \tag{24}$$

Proof. It follows easily from the fact that $n_h = \mu_h$, $h = 1, \ldots, m$ and x is the zero vector and further, from the fact that $Q_1(x)$ is the polynomial without an absolute term (see (16)). □

Theorem 3. Under the assumptions of Theorem 2 it holds

$$P(K_1 = n_1, \ldots, K_m = n_m | I_r = 1) = (2\pi)^{-\frac{m}{2}} |D|^{-\frac{1}{2}} \left\{ 1 - \frac{1}{2}(1 - p_r)(1 - 2p_r) v_r' D^{-1} v_r \right.$$
$$\left. + \frac{1}{2}(1 - p_r) \sum_{j=1}^{N} \kappa_{3j} (v_r' D^{-1} v_j)(v_j' D^{-1} v_j) + N^{-1} Q(0) + o(N^{-1}) \right\}. \tag{25}$$

Proof. If we denote $n = (n_1, \ldots, n_m)'$ and utilize the independence of indicators $I_j's$ we can write

$$\begin{aligned} &P(K_1 = n_1, \ldots, K_m = n_m | I_r = 1) = P(\tilde{K}_1 = n_1 - v_{1r}, \ldots, \tilde{K}_m = n_m - v_{mr}) \\ &= (2\pi)^{-m} \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \exp\{-it'(n - v_r)\}\, \tilde{\varphi}(t)\, \mathrm{d}t, \end{aligned} \tag{26}$$

where

$$(\tilde{K}_1, \ldots, \tilde{K}_m)' = \left(\sum_{j \neq r} I_j v_{1j}, \ldots, \sum_{j \neq r} I_j v_{mj} \right)'$$

and

$$\tilde{\varphi}(t) = \prod_{j \neq r} \varphi_j(t' v_j) = \varphi(t) / \varphi_r(t' v_r)$$

is the characteristic function of the vector $(\sum_{j \neq r} I_j v_{1j}, \ldots, \sum_{j \neq r} I_j v_{mj})'$.

Now, proceeding as in the proof of Theorem 2, we obtain that

$$\begin{aligned} & P(\tilde{K}_1 = n_1 - v_{1r}, \ldots, \tilde{K}_m = n_m - v_{mr}) \\ &= (2\pi)^{-\frac{m}{2}} |\tilde{D}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} x' x \right\} \left[1 + N^{-\frac{1}{2}} \tilde{Q}_1(x) + N^{-1} \tilde{Q}_2(x) + o(N^{-1}) \right], \end{aligned} \tag{27}$$

where

$$\begin{aligned} \tilde{D} &= \text{Var}\,(\tilde{K}_1, \ldots, \tilde{K}_m) = D - p_r(1 - p_r)\, v_r v_r', \\ x &= (p_r - 1)\, \tilde{D}^{-\frac{1}{2}} v_r, \end{aligned}$$

and polynomials $\tilde{Q}_1, \tilde{Q}_2$ are obtained from (15) with $P_\nu(iu)$, $\nu = 1, 2$ replaced by

$$\begin{aligned} \tilde{P}_1(iu) &= \frac{i^3}{3!} \frac{1}{N} \left[\sum_{j=1}^{N} \kappa_{3j} (u' \tilde{w}_j)^3 - \kappa_{3r} (u' \tilde{w}_r)^3 \right] \\ \tilde{P}_2(iu) &= \frac{i^4}{4!} \frac{1}{N} \left[\sum_{j=1}^{N} \kappa_{4j} (u' \tilde{w}_j)^4 - \kappa_{4r} (u' \tilde{w}_j)^4 \right] + \frac{1}{2} (\tilde{P}_1(iu))^2 \end{aligned}$$

for $\tilde{w}_j = \tilde{V}^{-\frac{1}{2}} v_j$, where $\tilde{V} = N^{-1} \tilde{D}$.

Now, let us introduce the following convention: For matrices A, B of the same type, the notation $A = B + o(\cdot)$ means that the asymptotic relation holds for each element of A, B, respectively.

According to Rao ([4], pp. 54–55),

$$\tilde{D}^{-1} = D^{-1} + p_r(1 - p_r)\, D^{-1} v_r v_r' D^{-1} \left[1 - p_r(1 - p_r)\, v_r' D^{-1} v_r \right]^{-1}.$$

For m fixed and N sufficiently large we get from here that $\tilde{D}^{-1} = D^{-1} + o(N^{-1})$.

Moreover,

$$|\tilde{D}| = |D| (1 + p_r(1 - p_r) v_r' \tilde{D}^{-1} v_r)^{-1}$$

and thus

$$\begin{aligned} |\tilde{D}|^{-\frac{1}{2}} &= |D|^{-\frac{1}{2}} \left(1 + p_r(1 - p_r)\, v_r' D^{-1} v_r + o(N^{-1}) \right)^{\frac{1}{2}} \\ &= |D|^{-\frac{1}{2}} \left(1 + \frac{1}{2} p_r (1 - p_r)\, v_r' D^{-1} v_r + o(N^{-1}) \right). \end{aligned} \tag{28}$$

Furthermore, notice that

$$\tilde{D} = D^{\frac{1}{2}}[E - p_r(1-p_r)\, D^{-\frac{1}{2}} v_r v_r' D^{-\frac{1}{2}}]\, D^{\frac{1}{2}} = D^{\frac{1}{2}}[E - B]\, D^{\frac{1}{2}},$$

where E is the identity matrix and $B = p_r(1-p_r)\, D^{-\frac{1}{2}} v_r v_r' D^{-\frac{1}{2}}$ is regular and symmetric. After some algebra we have $\tilde{D}^{-\frac{1}{2}} = D^{-\frac{1}{2}}(E - B)^{-\frac{1}{2}}$ and if we use the Taylor expansion for $(E-B)^{-\frac{1}{2}}$ (see i.e. Gantmacher [2], Chapt. V.) we get

$$\tilde{D}^{-\frac{1}{2}} = D^{-\frac{1}{2}} + o(N^{-1}). \tag{29}$$

Finally, we obtain

$$x = (p_r - 1) D^{-\frac{1}{2}} v_r + o(N^{-1}). \tag{30}$$

If we insert (28) and (30) into (27) and apply the Taylor expansion to the exponential function we get

$$\begin{aligned}
&P(\tilde{K}_1 = n_1 - v_{1r}, \ldots, \tilde{K}_m = n_m - v_{rm}) = \\
= \;& (2\pi)^{-\frac{m}{2}} |D|^{-\frac{1}{2}} \left[1 + \frac{1}{2} p_r(1-p_r)\, v_r' D^{-1} v_r + o(N^{-1})\right] \\
\times \;& \left[1 - \frac{1}{2}(p_r - 1)^2\, v_r' D^{-1} v_r + o(N^{-1})\right] \left[1 + N^{-\frac{1}{2}} \tilde{Q}_1(x) + N^{-1} \tilde{Q}_2(x) + o(N^{-1})\right] \\
= \;& (2\pi)^{-\frac{m}{2}} |D|^{-\frac{1}{2}} \Big[1 - \frac{1}{2}(1-p_r)(1-2p_r)\, v_r' D^{-1} v_r + N^{-\frac{1}{2}} \tilde{Q}_1(x) + N^{-1} \tilde{Q}_2(x) + \\
& + o(N^{-1})\Big]\,.
\end{aligned} \tag{31}$$

Now, let us turn to the terms $\tilde{Q}_1(x)$ and $\tilde{Q}_2(x)$.

According to (30) and (9) we can write $x = (p_r - 1)\, N^{-\frac{1}{2}} w_r + o(N^{-1})$ and according to (29) $\tilde{w}_j = w_j + o(N^{-\frac{1}{2}})$, $j = 1, \ldots, N$. Then the very careful calculation of $\tilde{Q}_1(x)$ (compare (16)) gives

$$\begin{aligned}
&N^{-\frac{1}{2}} \tilde{Q}_1(x) = \\
&= -\frac{1}{2N}(p_r - 1) \left[\sum_{k=1}^{m} w_{kr} \frac{1}{N} \sum_{j=1}^{N} \kappa_{3j} w_{kj}^3 + \sum_{1 \le k \ne l \le m} w_{lr} \frac{1}{N} \sum_{j=1}^{N} \kappa_{3j} w_{lj} w_{kj}^2\right] + o(N^{-1}) \\
&= -\frac{1}{2N}(p_r - 1) \frac{1}{N} \sum_{j=1}^{N} \kappa_{3j} \left[\sum_{k=1}^{m} w_{kr} w_{kj} \sum_{l=1}^{m} w_{lj}^2\right] + o(N^{-1}) \\
&= -\frac{1}{2N}(p_r - 1) \frac{1}{N} \sum_{j=1}^{N} \kappa_{3j} (w_r' w_j)(w_j' w_j) + o(N^{-1}) \\
&= -\frac{1}{2}(p_r - 1) \sum_{j=1}^{N} \kappa_{3j} (v_r' D^{-1} v_j)(v_j' D^{-1} v_j) + o(N^{-1}).
\end{aligned} \tag{32}$$

Inverting $\tilde{P}_2(iu)$ by (15) we obtain after some tedious calculations that $\tilde{Q}_2(x)$ is a polynomial in $x = (x_1, \ldots, x_m)$ of order 6 with the absolute term $Q_2(0) + O(N^{-1})$.

Thus, we can see that for x given by (30), $N^{-1}\tilde{Q}_2(x) = N^{-1}Q_2(0) + o(N^{-1})$, which together with (31) and (32) completes the proof of (25). □

Proof of Theorem 1. When we use (7), (24) and (25) and the Taylor expansion we get

$$\begin{aligned} \pi_i &= p_i\left\{1 - \frac{1}{2}(1-p_i)\left[(1-2p_i)\,v_i'D^{-1}v_i\right.\right. \\ &- \left.\left.\sum\nolimits_{j=1}^{N} \kappa_{3j}(v_i'D^{-1}v_j)\,(v_j'D^{-1}v_j)\right] + N^{-1}Q_2(0) + o(N^{-1})\right\} \\ &\times \ [1 + N^{-1}Q_2(0) + o(N^{-1})]^{-1} \\ &= p_i\left\{1 - \frac{1}{2}(1-p_i)\left[(1-2p_i)\,v_i'D^{-1}v_i\right.\right. \\ &- \left.\left.\sum\nolimits_{j=1}^{N} \kappa_{3j}(v_i'D^{-1}v_j)\,(v_j'D^{-1}v_j)\right] + o(N^{-1})\right\} \end{aligned}$$

which is (1) with $d = O(N)$.

The same approach can be used for the general rate of convergence of $d \to \infty$ if we replace (21) by an expansion in powers of $d^{-\frac{1}{2}}$ and notice that instead of (23) we can use the inequality

$$|\varphi(D^{-\frac{1}{2}}u)| \le e^{-\rho d}$$

for a constant $\rho > 0$ and $u \in \Omega - \Omega_2$. The difference is in technicalities, only.

The proof of (4) runs in a similar way. It holds

$$\begin{aligned} \pi_{ij} &= P(I_i = 1, I_j = 1 | K_1 = n_1, \ldots, K_m = n_m) \\ &= p_i p_j \frac{P(K_1 = n_1, \ldots, K_m = n_m | I_i = 1, I_j = 1)}{P(K_1 = n_1, \ldots, K_m = n_m)}. \end{aligned} \tag{33}$$

Now, proceeding as in the proof of Theorem 3 we can establish an asymptotic expansion of the nominator of (33) with the remainder of order $o(N^{-1})$ (respectively, of order d^{-1}.) Combining it with the expansion (24) we obtain

$$\begin{aligned} \pi_{ij} = p_i p_j &\left\{1 - \frac{1}{2}\left[(1-p_i)(1-2p_i)\,v_i'D^{-1}v_i - (1-p_j)(1-2p_j)\,v_j'D^{-1}v_j\right]\right. \\ &-\frac{1}{2}\left[(p_i-1)\sum\nolimits_{l=1}^{N}\kappa_{3l}(v_i'D^{-1}v_l)\,(v_l'D^{-1}v_l) + (p_j-1)\sum\nolimits_{l=1}^{N}\kappa_{3j}(v_j'D^{-1}v_l)\,(v_l'D^{-1}v_l)\right] \\ &\left.-\frac{1}{2}\left[2(p_i-1)(p_j-1)\,v_i'D^{-1}v_j\right] + o(N^{-1})\right\}. \end{aligned}$$

Combining this result with (1), we get after some computations

$$\pi_{ij} = \pi_i \pi_j \left[1 - (1-p_i)(1-p_j)\,v_i'D^{-1}v_j + o(N^{-1})\right].$$

Since (1) implies $\pi_i = p_i[1 + O(N^{-1})]$ we can also write

$$\pi_i\pi_j - \pi_{ij} = \pi_i\pi_j(1-\pi_i)(1-\pi_j)\, v_i' D^{-1} v_j + o(N^{-1}).$$

Thus, we can conclude, that the relation (4) holds asymptotically. Moreover, Hájek's approximation

$$\pi_i\pi_j - \pi_{ij} \sim (1-\lambda_i)(1-\lambda_j)\pi_i\pi_j v_i' D_\lambda^{-1} v_j,$$

where D_λ is given by (5) and the coefficients $1-\lambda_j$ are solutions of (6), is tight, i. e. it satisfies the condition

$$\sum_{j\neq i}(\pi_i\pi_j - \pi_{ij})\, v_j' = \pi_i(1-\pi_i)\, v_i'.$$

Example 14.5 in Hájek [3] shows that the relation (4) can hold with very good accuracy even for small size of the population ($N = 20,\ N = 24$).

(Received February 17, 1995.)

REFERENCES

[1] R. N. Bhattacharya and Ranga R. Rao: Normal Approximation and Asymptotic Expansions. Nauka, Moscow 1982 (Russian translation).
[2] F. R. Gantmacher: Theory of Matrices. Nauka, Moscow 1966 (In Russian).
[3] J. Hájek: Sampling from a Finite Population. Marcel Dekker, Inc., New York 1981.
[4] C. R. Rao: Linear Statistical Inference and Its Application. Academia, Prague 1978 (Czech translation).

Doc. RNDr. Zuzana Prášková, CSc., Matematicko-fyzikální fakulta Univerzity Karlovy (Faculty of Mathematics and Physics – Charles University), Sokolovská 83, 186 00 Praha 8. Czech Republic.