On second-order inclusion probabilities and variances among large randomized PPS samples

Paul Knottnerus

Discussion paper (201127)

Statistics Netherlands

The Hague/Heerlen, 2011

Explanation of symbols

.	= data not available
*	= provisional figure
**	= revised provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2010–2011	= 2010 to 2011 inclusive
2010/2011	= average of 2010 up to and including 2011
2010/'11	= crop year, financial year, school year etc. beginning in 2010 and ending in 2011
2008/'09–2010/'11	= crop year, financial year, etc. 2008/'09 to 2010/'11 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands
Grafimedia

Cover
TelDesign, Rotterdam

Information
Telephone +31 88 570 70 70
Telefax +31 70 337 59 94
Via contact form:
www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

60083201127 X-10

On second-order inclusion probabilities and variances among large randomized PPS samples

Paul Knottnerus

Summary: This paper presents and discusses some new results on the second-order inclusion probabilities of a systematic probability proportional to size sample drawn from a randomly ordered list, also called randomized PPS sampling. It is shown that some standard approximations of these second-order inclusion probabilities meant for relatively small sample sizes, need not be valid when the sample size n is of the same order as the population size N. In addition, it is shown that under a number of assumptions the variance formulas for rejective Poisson sampling can be applied to randomized PPS sampling designs when both n and N-n are large.

Keywords: Horvitz-Thompson estimator; Rejective Poisson sampling; Sampling autocorrelation coefficient; Second-order inclusion probability.

1. Introduction

When the study variable *y* is more or less proportional to a size variable *x*, a widely used estimator is the Horvitz-Thompson (HT) estimator in combination with a systematic probability proportional to size sample from a randomly ordered list. This procedure is often called the randomized PPS sampling design or the Goodman-Kish design.

In the literature there is some debate on the question of which type of approximation should be used for the second-order inclusion probabilities in this design, say π_{ijPPS}.

Brewer (2002, page 154) discusses two different types of approximations. One family of approximations is based on the well-known approximation of Hartley and Rao (1962). Another family of approximations is based on the approximation for rejective Poisson sampling of Hájek (1964). Brewer argues that the former type of approximation should be preferred. In contrast, Berger (1998, 2004 and 2005) advocates the use of the approximation of Hájek (1964) for this type of highly randomized sampling designs such as, for instance, the procedures proposed by Sampford (1967) or Chao (1982). In addition, Asok and Sukhatme (1976) provide a convenient approximation formula for the second-order inclusion probabilities in Sampford's procedure when $1<<n<<N$. Their approximation is slightly different from the one of Hartley and Rao for randomized PPS sampling.

The main aim of this paper is to show that under a number of assumptions Hájek's variance approximation is asymptotically valid for randomized PPS sampling particularly when the sample size *n* is of the same order as the population size *N*. Moreover, the variance approximation of Hartley and Rao (1962) need not be valid in such a situation.

The outline of the paper is as follows. Section 2 introduces some notation, and gives two standard expressions for the variance of the HT estimator and an alternative expression based on the sampling autocorrelation coefficient. Furthermore, section 2 discusses various approximations for the π_{ijPPS}. In section 3 approximations are derived for the sampling autocorrelation coefficients and variances in randomized PPS sampling and rejective Poisson sampling. Section 4 starts with a counterexample that some standard approximations for the π_{ijPPS} need not be valid when the underlying condition $n<<N$ is not satisfied. Furthermore, it is shown that under a number of assumptions two terms of a Taylor series expansion of the π_{ijPPS} are necessary and sufficient to obtain asymptotically valid sampling autocorrelation coefficients and variances in randomized PPS sampling when $0< f_0 < n/N < f_1 <1$. In addition, by reasons of symmetry, the indispensable conclusion appears to be that under mild assumptions Hájek's approximation for the π_{ij} in rejective Poisson sampling can also be used for approximating the π_{ijPPS}. In section 5 similar results are derived when n/N tends to unity or zero.

2. Notation and approximations for π_{ij} in randomized PPS sampling

Consider a population $U=\{1,...,N\}$ and let *s* be a sample of fixed size *n* drawn from *U* without replacement according to a given sampling design with first order inclusion probabilities π_i and second-order inclusion probabilities π_{ij} $(i,j=1,...,N)$. The HT estimator of the population total, $Y=\Sigma_{i\in U}Y_i$, is defined by $\hat{Y}_{HT}=\Sigma_{i\in s}Y_i/\pi_i$. Suppose there is a measure of relative size X_i (i.e., $X_i>0$ and $X=\Sigma_{i\in U}X_i=1$) such that all $X_i \le 1/n$. In fact, it is assumed here that units with $X_i>1/n$ are put together in a separate certainty-stratum. When the π_i are proportional to these size measures, $\pi_i=nX_i$. Defining $Z_i=Y_i/X_i$, we can write *Y* as a weighted mean of the Z_i, that is, $Y=\mu_z=\Sigma_{i\in U}X_iZ_i$. Likewise, we can write the HT estimator for *Y* in

randomized PPS sampling as $\hat{Y}_{HT} = \hat{Y}_{PPS} = \bar{z}_s$ where $\bar{z}_s$ stands for the sample mean of the Z_i.

The variance of the randomized PPS estimator is

$$\text{var}(\hat{Y}_{PPS}) = \frac{1}{n^2}\sum_{i\in U}\sum_{j\in U}(\pi_{ij} - \pi_i\pi_j)Z_iZ_j \tag{1}$$

$$= -\frac{1}{2n^2}\sum_{i\in U}\sum_{j\in U}(\pi_{ij} - \pi_i\pi_j)(Z_i - Z_j)^2 \tag{2}$$

with $\pi_{ii} = \pi_i$; recall that $Z_i = nY_i/\pi_i$. The former is attributed to Horvitz and Thompson (1952) and the latter is due to Sen (1953) and Yates and Grundy (1953). An alternative expression for the variance is

$$\text{var}(\hat{Y}_{PPS}) = \text{var}(\bar{z}_s) = \{1 + (n-1)\rho_z\}\frac{\sigma_z^2}{n}, \tag{3}$$

where $\sigma_z^2 = \Sigma_{i\in U}X_i(Z_i - \mu_z)^2$, and

$$\rho_z = \sum_{i\in U}\sum_{\substack{j\in U \\ j\neq i}}\frac{\pi_{ij}}{n(n-1)}\left(\frac{Z_i - \mu_z}{\sigma_z}\right)\left(\frac{Z_j - \mu_z}{\sigma_z}\right). \tag{4}$$

For a proof of (3), see Knottnerus (2003, page 103).

The sampling autocorrelation coefficient ρ_z in (4) is a generalization of the more familiar intraclass correlation coefficient ρ in systematic sampling with equal probabilities; see, for instance, Cochran (1977, pages 209 and 240) and Särndal et al. (1992, page 79). The phrase *sampling autocorrelation* is used because ρ_z refers to the correlation coefficient between two randomly chosen observations, say z_{s1} and z_{s2}, from *s*. Consequently, the value of ρ_z depends on the sampling design. In particular, when sampling with replacement, $\rho_z = 0$, while under simple random sampling without replacement (SRS), $\rho_z = -1/(N-1)$.

Although exact expressions for the π_{ijPPS} $(i \neq j)$ in randomized PPS sampling are available, these calculations might be cumbersome when *N* is large. For an exact expression, see Connor (1966) and for a modification Hidiroglou and Gray (1980). A well-known approximation of π_{ijPPS} proposed by Hartley and Rao (1962) is

$$\begin{aligned}\pi_{ijHR} = n(n-1)X_iX_j\{1 + X_i + X_j - \mu_x + 2(X_i^2 + X_j^2 + X_iX_j) \\ - 3\mu_x(X_i + X_j - \mu_x - 2\Sigma_{i\in U}X_i^3)\},\end{aligned} \tag{5}$$

where $\mu_x = \Sigma_{i\in U} X_i^2$ (recall $\mu_z = \Sigma_{i\in U} X_i Z_i$). According to Thompson and Wu (2008), approximation (5) can be used when $n/N=o(1)$ as $N \to \infty$. Note that $\pi_{ijHR}/n(n-1)$ does not depend on n. Hence, the corresponding approximation of ρ_z doesn't depend on n (recall we have assumed that every $X_i \le 1/n$).

Brewer and Donadio (2003) examine approximations of the form

$$\pi_{ijBD} = \pi_i \pi_j (c_i + c_j)/2 . \tag{6}$$

Elaborating on (5), the authors propose $c_{iHR} = (n-1)/n(1+\mu_x - 2X_i)$. Choosing a somewhat different expression $c_{iK} = (n-1)/n\gamma(1-2X_i)$, Knottnerus (2003 and 2011) arrives for $X_i, X_j < 1/2$ at

$$\pi_{ijK} = n(n-1)\frac{X_i X_j}{\gamma}\left(\frac{1/2}{1-2X_i} + \frac{1/2}{1-2X_j}\right), \tag{7}$$

where

$$\gamma = \frac{1}{2} + \frac{1}{2}\sum_{i\in U}\frac{X_i}{1-2X_i}.$$

These π_{ijK} have been shown to satisfy the second-order restriction $\Sigma_{j\ne i}\pi_{ij} = (n-1)\pi_i$. Furthermore, (7) is exact for SRS sampling while the π_{ijK} coincide with the π_{ijBDu} from the special designs proposed by Brewer (1963) and Durbin (1967) for PPS samples with n=2. For a proof that, after dropping $O(1/nN)$ terms, c_{iK} is identical with c_{iHR} under mild conditions as $N \to \infty$, see Knottnerus (2011).

Another interesting approximation for the π_{ijPPS} stems from the related rejective Poisson (RP) sampling design

$$\pi_{ijRP} = \pi_i \pi_j \left\{1 - \frac{(1-\pi_i)(1-\pi_j)}{d}\right\}, \tag{8}$$

provided that $d = \Sigma_{i\in U}\pi_i(1-\pi_i) = n(1-n\mu_x) \to \infty$. Note that $d \to \infty$ implies $n \to \infty$ because $d < n$ and likewise, $N - n \to \infty$ because of the symmetry of $d = \Sigma_{i\in U}\pi_i(1-\pi_i)$. The formal derivation of (8) given by Hájek (1964) is somewhat cumbersome. A more intuitive derivation, to the author's best knowledge not mentioned elsewhere in the literature, is as follows. A simple approximation for π_{ij}

is $\pi_i\pi_j$. Define its error e_{ij} by $e_{ij} = \pi_{ij} - \pi_i\pi_j$. Noting that $e_{ij} = 0$ for $\pi_i = 0$ and $\pi_i = 1$, a quite natural and symmetric approximation for the error e_{ij} is

$$e_{ij} = \beta\{\pi_i(1-\pi_i)\pi_j(1-\pi_j)\}^\alpha . \tag{9}$$

An approximation for the constants α and β can be obtained by examining the following equality of sums

$$\begin{aligned}\beta\{\pi_i(1-\pi_i)\}^\alpha \sum_{j\in U(j\neq i)} \{\pi_j(1-\pi_j)\}^\alpha &= \sum_{j\in U(j\neq i)} (\pi_{ij} - \pi_i\pi_j) \\ &= (n-1)\pi_i - \pi_i(n-\pi_i) \\ &= -\pi_i(1-\pi_i)\end{aligned} \tag{10}$$

and, equivalently,

$$\beta\{\pi_j(1-\pi_j)\}^\alpha \sum_{i\in U(i\neq j)} \{\pi_i(1-\pi_i)\}^\alpha = -\pi_j(1-\pi_j) . \tag{11}$$

Assuming that $\Sigma_U\{\pi_i(1-\pi_i)\}^\alpha \to \infty$ $(0 < \alpha < C < \infty)$ as $d \to \infty$ and dividing (11) by (10), we get $\alpha \sim 1$ and consequently, $\beta \sim -1/d$. Throughout this paper the notation $A \sim B$ is used to indicate that $A/B \to 1$ as $d \to \infty$. Substituting $\alpha = 1$ and $\beta = -1/d$ into (9) yields Hájek's result (8).

The above heuristic derivation applies to any PPS sampling design where there is no detectable pattern or ordering in the selected sample of fixed size *n* without replacement, provided that $d \to \infty$. Such designs are also called high-entropy sampling designs; see, among others, Brewer and Donadio (2003) and Tillé and Haziza (2010).

3. Approximations of ρ_{zPPS}, ρ_{zRP}, $\mathrm{var}(\hat{Y}_{PPS})$ and $\mathrm{var}(\hat{Y}_{RP})$

Let $\bar{X}$ denote the population mean of $X_1, \cdots, X_N$ and define V_x^2 and σ_x^2 by

$$V_x^2 = \sum_{i\in U} (X_i - \bar{X})^2 / N$$

and

$$\sigma_x^2 = \sum_{i\in U} X_i (X_i - \mu_x)^2,$$

respectively. Suppose that there are positive constants C and c such that $V_x / \overline{X} < C$, $\sigma_x / \mu_x < C$ and $X_i + c < 1/2$. Furthermore, suppose that $(Z_i - Y)/\sigma_z = O(1)$ as $N \to \infty$. Then it can be shown that (7) as well as (5) and (6) lead to the following approximation ρ_{z12} for ρ_z in randomized PPS sampling, say ρ_{zPPS}. That is, when N is large and $n << N$

$$\rho_{zPPS} = \rho_{z12}[1 + O(\frac{1}{N})] + O(\frac{1}{N^2}), \tag{12}$$

where

$$\rho_{z12} = -\frac{\sum_{i \in U} X_i^2 (Z_i - Y)^2}{\sum_{i \in U} X_i (Z_i - Y)^2}.$$

For a proof of (12), see Knottnerus (2011). Substituting ρ_{z12} into (3), we get

$$\begin{aligned} \text{var}(\hat{Y}_{PPS}) &= \frac{\sigma_z^2}{n} - \frac{n-1}{n} \sum_{i \in U} X_i^2 (Z_i - Y)^2 \\ &= \frac{1}{n} \sum_{i \in U} X_i [1 - (n-1) X_i] (Z_i - Y)^2, \end{aligned} \tag{13}$$

which is also given by Hartley and Rao (1962). It is noteworthy that approximation ρ_{z12} also follows directly from substituting the simple approximation $\pi_{ijAP} = n(n-1) X_i X_j$ into (4); for the proof of a similar result, see the proof of Theorem 1 below. In contrast, direct use of π_{ijAP} in (1) or (2) for the SRS case with $X_i = X_j = 1/N$ may lead to errors of more than 100% for some specific populations; see Knottnerus (2003, pages 274-6).

Following Hájek (1964, page 1520), substitution of (8) into (2) yields

$$\text{var}(\hat{Y}_{RP}) = \frac{1}{n} \sum_{i \in U} X_i (1 - nX_i)(Z_i - Y^*)^2, \tag{14}$$

where $Y^* = \Sigma_{i \in U} \alpha_i Z_i$ with $\alpha_i = \pi_i (1 - \pi_i)/d$. The main difference between (13) and (14) is that Y is replaced by Y^*. In addition, Hájek proposes the following variance estimator

$$\hat{\text{var}}(\hat{Y}_{RP}) = \frac{1}{n(n-1)} \sum_{i \in s} (1 - nX_i)(Z_i - \hat{Y}^*)^2,$$

where $\hat{Y}^* = \Sigma_{i\in s}\hat{\alpha}_i Z_i$ with $\hat{\alpha}_i = (1-\pi_i)/\Sigma_{i\in s}(1-\pi_i)$. In order to get more insight into ρ_{zRP} corresponding to (8), define the correlation coefficient r_{xz} between the X_i and the Z_i by $r_{xz} = \Sigma_U X_i x_i z_i / \sigma_x \sigma_z$ where x_i and z_i stand for $(X_i - \mu_x)$ and $(Z_i - \mu_z)$, respectively. The following theorem gives an approximation for ρ_{zRP}.

Theorem 1. Suppose that there are positive constants *C* and *c* such that $V_x / \overline{X} < C$, $\sigma_x / \mu_x < C$ and $-C/N < \rho_{z12}, \rho_{zRP} < -c/N$. Furthermore, suppose that $z_i / \sigma_z = O(1)$ as $N \to \infty$. Then it holds for RP sampling that

$$\rho_{zRP} = (\rho_{z12} - \frac{n^2 r_{xz}^2 \sigma_x^2}{d})[1 + o(1)] \tag{15}$$

as $d \to \infty$.

Proof. Carrying out the multiplications on the right-hand side of (8) and neglecting symmetric terms, we obtain four mutually different terms, that is, with mutually different contributions to ρ_{zRP}. The contribution of the main term $\pi_i \pi_j$ in π_{ijRP} to ρ_{zRP} is according to (4)

$$\frac{1}{n(n-1)} \sum_{i\in U} \sum_{\substack{j\in U \\ j\neq i}} \frac{n^2 X_i X_j z_i z_j}{\sigma_z^2} = -\frac{n}{n-1} \sum_{i\in U} \frac{X_i^2 z_i^2}{\sigma_z^2} = \frac{n}{n-1} \rho_{z12},$$

where we used that $\Sigma_U X_j z_j = 0$. Consequently, the contribution of the term $\pi_i \pi_j / d$ to ρ_{zRP} is $o(1/N)$ provided $d \to \infty$. Also, the contribution of the term $\pi_i^2 \pi_j / d$ to ρ_{zRP} is $o(1/N)$ because under the assumptions of the theorem this contribution can be written as

$$\frac{n^2}{n-1} \sum_{i\in U} \sum_{\substack{j\in U \\ j\neq i}} \frac{X_i^2 X_j z_i z_j}{d\sigma_z^2} \sim -\frac{n}{d} \sum_{i\in U} \frac{X_i^3 z_i^2}{\sigma_z^2} = O(\frac{n}{dN^2}) = o(\frac{1}{N}),$$

where we again used that $\Sigma_U X_j z_j = 0$ and $\overline{X} = 1/N$ so that

$$\mu_x = \sum_{i\in U} X_i^2 = N(V_x^2 + \overline{X}^2) = O(\frac{1}{N})$$

and

$$\sum_{i\in U} X_i^3 = \sigma_x^2 + \mu_x^2 = O(\frac{1}{N^2}).$$

Hence, substituting (8) into (4) and omitting the irrelevant $o(1/N)$ contributions of $\pi_i\pi_j/d$ and $\pi_i^2\pi_j/d$, we get

$$\begin{aligned}\rho_{zRP} &\sim \frac{n}{n-1}\{\sum_{i\in U}\sum_{\substack{j\in U\\ j\neq i}} X_iX_j(1-\frac{n^2X_iX_j}{d})\frac{z_iz_j}{\sigma_z^2}\}\\ &= \frac{n}{n-1}\{\rho_{z12}-\frac{n^2}{d}\sum_{i\in U}X_i^2z_i(\sum_{j\in U}X_jx_jz_j-X_i^2z_i)/\sigma_z^2\}\\ &= \frac{n}{n-1}\{\rho_{z12}-\frac{n^2}{d}\sum_{i\in U}X_i^2z_i(r_{xz}\sigma_x\sigma_z-X_i^2z_i)/\sigma_z^2\}\\ &= \rho_{z12}+O(\frac{1}{nN})-\frac{n^2r_{xz}^2\sigma_x^2}{d}+O(\frac{n}{dN^2})\end{aligned}$$

as $d\to\infty$. Note that in the last line use is made of $z_i/\sigma_z=O(1)$, $nX_i=O(1)$ and $\Sigma_U X_i^3=O(\mu_x^2)=O(N^{-2})$. This concludes the proof.

In summary, apart from the main term $\pi_i\pi_j$ in (8) the only other term that may give an asymptotically relevant contribution to ρ_{zRP} is $-\pi_i^2\pi_j^2/d$. This contribution is $-n^2r_{xz}^2\sigma_x^2/d$ which is of order n^2/dN^2. Noting that $d=n(1-n\mu_x)$ and $n\mu_x=O(n/N)$, it is seen that when $n/N=o(1)$, this contribution is $O(n/N^2)=o(1/N)$ as $d\to\infty$ and hence, it can be ignored provided $\rho_{zRP}<-c/N$. In other words, under the assumptions of Theorem 1, Hájek's variance estimator can be applied to randomized PPS sampling when $n/N=o(1)$. To the author's best knowledge this result and its proof are not mentioned elsewhere in the literature. In addition, Brewer and Donadio (2003, page 191) also give a model-assisted check of the usefulness of their variance formulas derived from π_{ijBD} in (6) without the limiting assumption $n/N=o(1)$. For the models considered by the authors there holds $r_{xz}=o(1)$ so that approximation (15) amounts to $\rho_{zRP}\sim\rho_{z12}$.

4. $\mathrm{Var}(\hat{Y}_{PPS})$ and a Taylor series expansion of π_{ijPPS} when $f_0N<n<f_1N$

In this section we consider a p^{th}-order Taylor series expansion of $\pi_{ijPPS}/\pi_i\pi_j$. It is shown that under the assumptions of Theorem 1 and some additional assumptions only two terms of this Taylor series expansion are asymptotically relevant with respect to $\mathrm{var}(\hat{Y}_{PPS})$ and ρ_{zPPS} as $d\to\infty$ and $f_0N<n<f_1N$ $(0<f_0<f_1<1)$.

First, in order to give some more insight into the difference between (13) and (14), consider the following counterexample that the variance in (13) need not be valid when $n/N \neq o(1)$. Let U be a population consisting of two groups U_1 and U_2 with means $\overline{Y}_1$ and $\overline{Y}_2$, respectively. Both group sizes are $N/2$. Let s be a randomized PPS sample of size $n=3N/4$ from the whole population U. Let the X_i be such that

$$\pi_i = nX_i = \begin{cases} 1 & \text{if} \quad i \in U_1 \\ 0.5 & \text{if} \quad i \in U_2. \end{cases}$$

Obviously, group 1 doesn't contribute to the variance of $\hat{Y}_{PPS}$. The selected elements in s from U_2 constitute an ordinary SRS sample of size $N/4$. Hence, in this case the correct variance formula for $\hat{Y}_{PPS}$ is

$$\text{var}(\hat{Y}_{PPS}) = \left(\frac{N}{2}\right)^2 (1-\frac{1}{2}) \frac{S_{y2}^2}{N/4} = \frac{NS_{y2}^2}{2},$$

where

$$S_{y2}^2 = \frac{2}{N-2} \sum_{i \in U_2} (Y_i - \overline{Y}_2)^2.$$

Apart from an asymptotically negligible factor $(N-2)/N$, (14) gives the correct outcome. However, approximation (13) gives now an entirely different outcome unless $\overline{Y}_2 = 2\overline{Y}/3$ (recall that $Y^* = 3N\overline{Y}_2/2 = Y$ when $\overline{Y}_2 = 2\overline{Y}/3$). Consequently, also approximations (5)-(7) for the π_{ijPPS} need not be valid when both n and N are very large; see Knottnerus (2011).

From (14) and (15) it follows that for the above example ρ_{zPPS} can be approximated quite well by

$$\rho_{exmpl}^{PPS} = \rho_{exmpl}^{RP} = -\sum_{i \in U} \frac{X_i^2 z_i^2}{\sigma_z^2} - \frac{n^2 r_{xz}^2 \sigma_x^2}{d}, \tag{16}$$

irrespective of the values of Y_{ki} in group k ($k = 1,2$). Now the following natural questions arise (*i*) to what extent are (14) and (16) applicable to other randomized PPS samples and (*ii*) to what extent can (8) be seen as an appropriate approximation for π_{ijPPS} as $d \to \infty$ and $f_0 N < n < f_1 N$.

In order to shed some more light on these issues, consider an arbitrary randomized PPS sample of size n ($f_0 N < n < f_1 N$). Suppose without loss of generality that n depends on N [$n=n(N)$] and define R_N by $R_N = N\rho_{zPPS}$. Furthermore, suppose that

$R_N \to R \quad (R<0), \quad N^{h-1}\Sigma_U X_i^h \to m^{(h)} \quad (0<c<m^{(h)}<C<\infty) \quad$ as $\quad d \to \infty$ $(h=2,...,H),$ and that there exists a p such that for $d \to \infty$ the quantity $\pi_{ijPPS}/\pi_i\pi_j$ can be approximated appropriately by a p^{th}-order Taylor series expansion of $\pi_{ijPPS}/\pi_i\pi_j$ as function of π_i and π_j around the origin. Then π_{ijPPS} itself can be approximated by

$$\pi_{ijPPS} = \sum_{k,l\geq 1; k+l\leq p+1} a_N^{kl}\pi_i^k\pi_j^l. \tag{17}$$

In fact, (17) can be seen as a further generalization of approximations π_{ijHR} and π_{ijRP} from (5) and (8), respectively. Note that a_N^{kl} is independent of the (arbitrary) rank numbers i and j. Furthermore, we make the additional assumptions that $f_0 N < n < f_1 N$ and that for an arbitrary data set $X_1,\cdots,X_N$, $a_{N,arb}^{kl}$ is of the same order as $a_{N,exmpl}^{kl}$ from the above example as $n,d \to \infty$; for a justification of the latter assumption, see Appendix A. The assumptions so far are in line with the various approximations for π_{ij} mentioned in section 2.

For $n > f_0 N$ useful inequalities for a further analysis are

$$1 \leq N^{h-1}\Sigma_U X_i^h < C < \infty, \tag{18}$$

$(2 \leq h \leq H)$ irrespective of the data set $X_1,\cdots,X_N$. The first inequality in (18) follows from minimizing $\Sigma_U X_i^h$ $(h \geq 2)$ subject to $\Sigma_U X_i = 1$. This yields $X_i = 1/N$ and hence, $\Sigma_U X_i^h \geq N^{-h+1}$; for a similar constrained minimization problem, see Knottnerus (2003, page 166). The second inequality in (18) follows from the assumption $n > f_0 N$ and hence, $X_i \leq 1/n < 1/f_0 N$ so that $X_i^h < C/N^h$ with $C > 1/f_0^H$. Also suppose that $z_i/\sigma_z = O(1)$. The reason for assuming $n < f_1 N$ is that the above counterexample in its present form is not appropriate when $n/N \to 1$ as $d \to \infty$; for further details, see section 5.

According to (4) and (17), we get the following approximation for R_N

$$R_N = \frac{N}{n(n-1)} \sum_{i\in U} \sum_{\substack{j\in U \\ j\neq i}} \Big(\sum_{k,l\geq 1; k+l\leq p+1} a_N^{kl}\pi_i^k\pi_j^l \Big) \frac{z_i z_j}{\sigma_z^2},$$

which can alternatively be written as

$$R_N = \sum_{k,l \geq 1; k+l \leq p+1} R_N^{kl},$$

where

$$R_N^{kl} = \frac{N a_N^{kl}}{n(n-1)} \sum_{i \in U} \sum_{\substack{j \in U \\ j \neq i}} p_i^k p_j^l \frac{z_i z_j}{\sigma_z^2}.$$

Denote $N a_N^{kl} n^{k+l-1}/(n-1)$ by B_N^{kl} and define $r_{xz}^{(h)}$ as the correlation coefficient between the X_i^h and the Z_i $(h \geq 1)$. That is,

$$r_{xz}^{(h)} = \sum_{i \in U} X_i \frac{(X_i^h - \mu(x^h)) z_i}{\sigma(x^h)\sigma_z} = \sum_{i \in U} \frac{X_i^{h+1} z_i}{\sigma(x^h)\sigma_z},$$

where $\mu(x) = \mu_x$ and $\sigma(x) = \sigma_x$. In addition, define $r_{xz}^{(0)} = 0$. In order to trace the impact of R_N^{kl} on $R_N = N r_{zPPS}$ for $d \to \infty$, rewrite R_N^{kl} for $k,l \geq 1$ as

$$R_N^{kl} = B_N^{kl} \sum_{i \in U} \sum_{\substack{j \in U \\ j \neq i}} X_i^k X_j^l \frac{z_i z_j}{\sigma_z^2} = B_N^{kl} \sum_{i \in U} X_i^k \frac{z_i}{\sigma_z} \{ r_{xz}^{(l-1)} \sigma(x^{l-1}) - X_i^l \frac{z_i}{\sigma_z} \}$$

$$= B_N^{kl} \{ r_{xz}^{(k-1)} \sigma(x^{k-1}) r_{xz}^{(l-1)} \sigma(x^{l-1}) - \sum_{i \in U} X_i^{k+l} \frac{z_i^2}{\sigma_z^2} \} = R_{N1}^{kl} + R_{N2}^{kl},$$

where

$$R_{N1}^{kl} = \begin{cases} B_N^{kl} r_{xz}^{(k-1)} \sigma(x^{k-1}) r_{xz}^{(l-1)} \sigma(x^{l-1}) & \text{if} \quad k,l \geq 2 \\ 0 \quad \text{otherwise} \end{cases} \tag{19}$$

and

$$R_{N2}^{kl} = - B_N^{kl} \sum_{i \in U} X_i^{k+l} \frac{z_i^2}{\sigma_z^2} = O(a_N^{kl}). \tag{20}$$

In (20) we used (18) and the assumption that $z_i / \sigma_z = O(1)$.

Now suppose that the Z_i in the above example are such that $\left| r_{xz}^{(h)} \right| > c > 0$ (h=1, …, p). Let $R_{N1,expl}^{kl}$ denote the corresponding parameter in that example. Then it follows from the above example that, according to (16), only $r_{xz}^{(1)}$ $(= r_{xz})$ is asymptotically relevant. In other words, from the above example it emerges that the a_N^{kl} should be such that in spite of the nonzero correlations $r_{xz}^{(h)}$ we can set $R_{N1,expl}^{kl} = 0$ for each k and l unless k=l=2. Consider now an arbitrary data set, say

$X_{i,arb}$ (i = 1,…,N) for a randomized PPS sample with $f_0N < n < f_1N$. Noting that there are positive constants C_0 and c_0 such that $\sigma(x^h_{arb}) < C_0 / N^h$ when $X_{i,arb} \le 1/n < 1/f_0N$ and $c_0 / N^h < \sigma(x^h_{exmpl})$ for the above example, it is seen that $\sigma(x^h_{arb}) / \sigma(x^h_{exmpl}) < C_0 / c_0$. Since, by assumption, $a^{kl}_{N,arb} = O(a^{kl}_{N,exmpl})$, we get according to (19) for an arbitrary randomized PPS sample

$$\begin{aligned} R^{kl}_{N1,arb} &= O(B^{kl}_{N,arb}\sigma(x^{k-1}_{arb})\sigma(x^{l-1}_{arb})) \\ &= O(B^{kl}_{N,exmpl}\sigma(x^{k-1}_{exmpl})\sigma(x^{l-1}_{exmpl})) \\ &= O(R^{kl}_{N1,exmpl}) = o(1), \end{aligned}$$

unless k=l=2. Hence, we can set $R^{kl}_{N1,arb} = 0$ as $d \to \infty$ for any randomized PPS sample ($f_0N < n < f_1N$) irrespective of the correlations $r^{(h)}_{xz}$ unless k=l=2.

Next, we look at the role of the R^{kl}_{N2} when $f_0N < n < f_1N$. Choosing the Y_i in the above example such that $z_i^2 / \sigma_z^2 = 1$, it is seen from (16) that all R^{kl}_{N2} are asymptotically irrelevant in that example unless k=l=1. Moreover, for arbitrary data X_i we have for $k, l \ge 2$

$$\frac{R^{kl}_{N2,arb}}{R^{kl}_{N1,exmpl}} = O(\frac{\Sigma_U X^{k+l}_{i,arb} z_i^2 / \sigma_z^2}{\sigma^{k-1}_{x,exmpl}\sigma^{l-1}_{x,exmpl}}) = O(\frac{1}{N}),$$

where we used $z_i = O(\sigma_z)$ and (18). So in analogy with R^{kl}_{N1} we may set $R^{kl}_{N2} = 0$ for any other data set without affecting the results when according to (20) $z_i / \sigma_z = O(1)$ unless k=l=1.

It emerges that we can approximate R_N by

$$R_N \sim R^{11}_{N2} + R^{22}_{N1} = -Na^{11}_N \sum_{i\in U} X_i^2 \frac{z_i^2}{\sigma_z^2} + Na^{22}_N n^2 r^2_{xz}\sigma^2_x, \qquad (21)$$

where use is made of (20) and (19) for kl=11 and kl=22, respectively. In order to trace a^{11}_N and a^{22}_N, note that according to (16) we have for the above example

$$R_{N,exmpl} = R^{11}_{N2,exmpl} + R^{22}_{N1,exmpl} = -\sum_{i\in U} \frac{NX_i^2 z_i^2}{\sigma_z^2} - \frac{Nn^2 r^2_{xz}\sigma^2_x}{d}. \qquad (22)$$

Comparing the coefficients of the utmost right-hand sides of (21) and (22), it is seen that under the above assumptions $a^{11}_N = 1 + o(1)$ and $a^{22}_N = -d^{-1}(1 + o(1))$ as $d \to \infty$.

In summary, within the class of approximations for π_{ijPPS} that can be described by (17) with the order of the c_{N0}^{kl} being independent of the specific data X_i, the approximation

$$\pi_{ij23} = \pi_i \pi_j - \frac{\pi_i^2 \pi_j^2}{d} \tag{23}$$

leads under the assumptions of Theorem 1 to the appropriate ρ_{zPPS} for any data set, that is, $\rho_{zPPS} = \rho_{zRP}(1+o(1))$ as $d \to \infty$ and $f_0 N < n < f_1 N$; ρ_{zRP} is given by (15).

Comment. It may seem somewhat counterintuitive that for an arbitrary, randomized PPS sample fairly general conclusions can be drawn from such a specific counterexample as above. However, it should be noted that although in the above example the X_i can only assume the values 1/*n* or 1/2*n*, the quantities $\sigma(x^h)$ and $\Sigma_U X_i^h$ in the example are under mild conditions of the same order and magnitude as for other randomized PPS samples. From that point of view the above counterexample is sufficiently general to draw rather general conclusions with respect to a_N^{11}, a_N^{22}, the other a_N^{kl} and ρ_{zPPS}.

Finally, we look more closely at the π_{ijPPS}. Because a_N^{11} and a_N^{22} are the only asymptotically relevant coefficients in the Taylor series of π_{ijPPS} for calculating ρ_{zPPS} and the variance, an appropriate approximation of π_{ijPPS} should at least consist of the two components on the right-hand side of (23). As we have seen, use of (23) leads to asymptotically correct results for R_N, ρ_{zPPS} and $\mathrm{var}(\hat{Y}_{PPS})$. That is, all three results have relative $o(1)$ errors. However, the error approximation

$$e_{ij23} = \pi_{ij23} - \pi_i \pi_j = -\pi_i^2 \pi_j^2 / d$$

need not be very accurate. In order to further improve $e_{ij23} = -\pi_i^2 \pi_j^2 / d$, define a_i by $a_i = 1$ if unit *i* is selected in *s* and $a_i = 0$ otherwise. Define b_i by $b_i = 1 - a_i$. In fact, b_i indicates whether unit *i* is selected in $U \setminus s$ or not. Define $\tau_i = E(b_i)$ and $\tau_{ijPPS} = E(b_i b_j)$; recall that $\tau_i = 1 - \pi_i$ and $\tau_{ijPPS} = 1 - \pi_i - \pi_j + \pi_{ijPPS}$. Applying (23) to τ_{ijPPS} and making use of $e_{ij} = \mathrm{cov}(a_i, a_j) = \mathrm{cov}(b_i, b_j)$, we get

$$e_{ij24} = \pi_{ij24} - \pi_i \pi_j = \tau_{ij24} - \tau_i \tau_j = -\frac{\tau_i^2 \tau_j^2}{d}. \tag{24}$$

Obviously, the approximation of e_{ijPPS} should have a double symmetry: (*i*) in π_i and π_j and (*ii*) in π_i and τ_i. Noting that the double symmetric form $\pi_i\pi_j\tau_i\tau_j/d$ includes the indispensable terms $\pi_i^2\pi_j^2/d$ and $\tau_i^2\tau_j^2/d$, we get for $f_0N < n < f_1N$ under the assumptions of Theorem 1 and the additional assumption (17) the following approximation for e_{ijPPS}

$$e_{ij25} = -\frac{\pi_i\pi_j\tau_i\tau_j}{d}, \tag{25}$$

which is identical with expression (8) derived by Hájek for RP sampling. That is, within the class of π_{ijPPS} that can be described by (17) the results found so far can be written in the following general form

$$\begin{aligned}\pi_{ijPPS} - \pi_i\pi_j &= -\pi_i\pi_j\frac{A - B(\pi_i+\pi_j)+\pi_i\pi_j}{d}\\ &= -\frac{\pi_i\pi_j\tau_i\tau_j}{d} + \frac{A-1}{d}\pi_i\pi_j + \frac{(B-1)(\pi_i+\pi_j)\pi_i\pi_j}{d},\end{aligned}$$

where *A* and *B* should be negligible *o*(*d*) coefficients (*cf*. the proof of Theorem 1). By reasons of symmetry, we get *A*=*B*=1 which yields Hajek's result. Moreover, for the specific case of RP sampling Hájek (1964, page 1511) showed that $e_{ijRP} = e_{ij25}(1+o(1))$ as $d \to \infty$. For e_{ijPPS} this property still remains to be proved or disproved.

5. Taylor series expansion of π_{ijPPS} when *n*/*N* tends to unity or zero

In this section we briefly examine the form of the π_{ijPPS} when the sampling fraction f_N $(=n/N)$ of a randomized PPS sample tends to unity as $d \to \infty$. Suppose that, for instance, $n = N - \sqrt{N}$ or, equivalently, $f_N = 1 - 1/\sqrt{N}$. Furthermore, let the above counterexample be modified as follows. For group 1 we still assume $n_1 = N_1 = N/2$ but for group 2 we take $n_2 = (f_N - 0.5)N$. This means that

$$\begin{aligned}X_{1i} &= \frac{1}{N-\sqrt{N}} = \frac{1}{N}(1+\frac{1}{\sqrt{N}})[1+O(\frac{1}{N})];\\ X_{2i} &= \frac{1-2/\sqrt{N}}{N-\sqrt{N}} = \frac{1}{N}(1-\frac{1}{\sqrt{N}})[1+O(\frac{1}{N})].\end{aligned}$$

Obviously, only (8) again gives the asymptotically correct variance for $\hat{Y}_{PPS}$. For the X_i thus defined we have

$$\mu_x = \sum_{i\in U} X_i^2 = \frac{N}{2}\frac{1+(1-2/\sqrt{N})^2}{(N-\sqrt{N})^2} = \frac{1}{N}[1+O(\frac{1}{N})], \tag{26}$$
$$\sigma_x^2 = \sum_{i\in U} X_i(X_i - \frac{1}{N})^2 - (\mu_x - \frac{1}{N})^2 = \frac{1}{N^3}(1+o(1)),$$

and

$$\begin{aligned} d &= n(1-n\mu_x) \\ &= (N-\sqrt{N})\left(1-\frac{N}{2}\frac{2-4/\sqrt{N}+4/N}{N-\sqrt{N}}\right) \\ &= N-\sqrt{N}-N+2\sqrt{N}-2 = \sqrt{N}-2 \end{aligned}$$

as $d \to \infty$. For simplicity's sake but without loss of generality, it is assumed that in the present counterexample the Y_i are such that $z_i^2/\sigma_z^2 = 1$ and $\left|r_{xz}^{(h)}\right| > c > 0$ $(h = 1,..., p)$.

Under the assumption $z_i^2/\sigma_z^2 = 1$, it follows from the definition of ρ_{z12} and (26) that for the present example

$$\rho_{z12,exmpl} = -\mu_x = -\frac{1}{N}[1+O(\frac{1}{N})].$$

An essential difference with the previous counterexample is that for $d \to \infty$ and $n = N - \sqrt{N}$ we now have

$$1+(n-1)\rho_{z12,exmpl} = \frac{1}{\sqrt{N}} + O(\frac{1}{N}) << 1. \tag{27}$$

Consequently, an $O(1/\sqrt{N})$ contribution of some R_N^{kl} to R_N need not be irrelevant as was the case in section 4. Result (27) also holds for $1+(n-1)\rho_{zRP,exmpl}$ because in analogy with the proof of Theorem 1 ρ_{zRP} can be written as the sum of ρ_{z12} and three additional terms of order $1/N\sqrt{N}$. That is,

$$\rho_{zRP} = \rho_{z12} - \frac{\rho_{z12}}{d} + \frac{2R_N^{12}}{N} + \frac{R_N^{22}}{N}, \tag{28}$$

where $R_N^{kl} = R_{N1}^{kl} + R_{N2}^{kl}$ $(kl = 12, 22)$. R_{N1}^{kl} is given by (19) and R_{N2}^{kl} by (20) where, according to (8), B_N^{12} now stands for $Nn^2/d(n-1)$ and B_N^{22} for $-Nn^3/d(n-1)$. In order to show that the last three terms in (28) are $O(1/N\sqrt{N})$, substitute (19) and (20) into (28). Assuming that $n \approx n-1$, this yields

$$\rho_{zRP} = \rho_{z12} - \frac{\rho_{z12}}{d} - \frac{n}{d}(2\sum_{i\in U} X_i^3 + nr_{xz}^2\sigma_x^2 - n\sum_{i\in U} X_i^4). \tag{29}$$

Unlike in section 4 in (29) the last four terms are each separately of the same order $1/N\sqrt{N}$. However, just as in section 4 it can be shown that the combination of terms 2, 3 and 5 on the right-hand side of (29) gives a negligible $o(1/\sqrt{N})$ contribution to $1+(n-1)\rho_{zRP}$ for the above example. That is,

$$\rho_{zRP} = \rho_{z12,exmpl} - \frac{N^2 r_{xz}^2 \sigma_x^2}{d}[1+o(1)]. \tag{30}$$

To prove (30), note that it follows from (18) and $X_i \le 1/n$ that

$$\frac{1}{N^{h-1}} \le \sum_{i\in U} X_i^h \le \frac{N}{(N-\sqrt{N})^h} = \frac{1}{N^{h-1}}(1+O(\frac{1}{\sqrt{N}})). \tag{31}$$

Using (31) for *h*=3 and *h*=4, and the fact that $N\rho_{z12,exmpl} \sim -1$, it is seen that terms 2, 3 and 5 in (29) are together $o(1/Nd)$ so that their joint contribution to $1+(n-1)\rho_{zRP}$ is only $o(1/\sqrt{N})$. Hence, these three terms can be ignored and it follows from (27) and (30) that similar to section 4 only a_N^{11} and a_N^{22} are relevant in a Taylor series expansion of π_{ijPPS} for obtaining the appropriate ρ_{zPPS} and variance for an arbitrary randomized PPS sample when $n = N - \sqrt{N}$. For proofs that for arbitrary data (X_i, Z_i) $N\rho_{z12,arb} \sim -1$ and $\sigma(x_{arb}^h) = O(\sigma(x_{exmpl}^h)) = O(1/N^{h+0.5})$, see the end of this section. Moreover, it follows from (27) that R_{N1}^{kl} $[\max(k,l) > 2]$ is negligible only when $R_{N1}^{kl} = o(1/\sqrt{N})$ provided $r_{xz}^{(h)} \ne 0$. Noting that $\sigma(x_{arb}^h) = O(\sigma(x_{exmpl}^h)) = O(1/N^{h+0.5})$ as stated before, it is seen from (19) that we now should have $a_N^{kl} = o(1/\sqrt{N})$ $[\max(k,l) > 2]$. In addition, because $a_N^{kl} = o(1/\sqrt{N})$, we have by (20) and (31), $R_{N2}^{kl} = o(1/\sqrt{N})$ and hence, R_{N2}^{kl} is negligible as well when $\max(k,l) > 2$. More generally, similar results can be derived for PPS sampling when $n = N - N^{\alpha}$ $(0 < \alpha < 1)$.

Although it emerges that only a_N^{11} and a_N^{22} are relevant, we can draw some more general conclusions from (29) with respect to the Taylor series expansion of π_{ijPPS}. Using (4), (17) and the fact that $a_N^{22} \sim -1/d$, we get similar to (29) for ρ_{zPPS}

$$\rho_{zPPS} = \rho_{z12} + (a_N^{11} - 1)\rho_{z12} - 2a_N^{12} n \sum_{i \in U} X_i^3 - \frac{n^2 r_{xz}^2 \sigma_x^2}{d} + \frac{n^2}{d} \sum_{i \in U} X_i^4 . \tag{32}$$

It follows from the counterexample that in analogy with (29) the combination of terms 2, 3 and 5 on the right-hand side of (32) should be negligible, that is, $o(1/\sqrt{N})$. Again using (31) and the fact that $\rho_{z12,exmpl} \sim -1/N$, it is seen from (32) that terms 2, 3 and 5 together are only negligible if a_N^{11} and a_N^{12} satisfy the following approximation

$$-(a_N^{11} - 1) - 2a_N^{12} \sim -1/d. \tag{33}$$

As we have seen in RP sampling, a particular solution of (33) is $a_N^{11} = (d-1)/d$ and $a_N^{12} = 1/d$. The general approximate solution satisfying (33) is of the form

$$\begin{aligned} a_N^{11} &\sim 1 - \frac{1+D}{d}[1 + o(1)]; \\ a_N^{12} &= a_N^{21} \sim \frac{1 + D/2}{d}, \end{aligned} \tag{34}$$

where *D* is an arbitrary 'constant'. Substituting (34) into (17) with $a_N^{22} \sim -1/d$ gives

$$\begin{aligned} \pi_{ijPPS} - \pi_i \pi_j &= -\pi_i \pi_j \frac{(1+D) - (\pi_i + \pi_j)(1 + D/2) + \pi_i \pi_j}{d} \\ &= -(1+D)\frac{\pi_i \pi_j \tau_i \tau_j}{d} - \frac{D\pi_i \pi_j (\pi_i + \pi_j)}{2d} + \frac{D\pi_i^2 \pi_j^2}{d}. \end{aligned} \tag{35}$$

As we saw previously, the right-hand side of (35) should be asymptotically symmetric in π_i and τ_i. Hence, $D = o(1)$. In summary, the a_N^{kl} thus obtained for $d \to \infty$ are

$$\begin{aligned} a_N^{11} &= 1 + \varepsilon_1 - \frac{1+\varepsilon_2}{d}; \\ a_N^{12} &= a_N^{21} = \frac{1+\varepsilon_3}{d}; \\ a_N^{22} &= -\frac{1+\varepsilon_4}{d}, \end{aligned} \tag{36}$$

where $\varepsilon_k = o(1)$ (k=1,…,4). This is in line with Hájek's formula for π_{ijRP} in (8) and (25). From the latter version it is not difficult to see that by reasons of symmetry, (36) is still valid for approximating $\tau_{ij} - \tau_i\tau_j = \pi_{ij} - \pi_i\pi_j$ when $n = \sqrt{N}$ and more generally, when $n = N^{\alpha}$ $(0 < \alpha < 1)$.

We conclude this section with proofs that for an arbitrary data set (X_i, Z_i) $N\rho_{z12,arb} \sim -1$ and $\sigma(x^h_{arb}) = O(\sigma(x^h_{exmpl})) = O(1/N^{h+0.5})$ when $n = N - \sqrt{N}$. Write $X_{i,arb}$ as

$$X_{i,arb} = \frac{1}{N}(1 + \frac{\delta_i}{\sqrt{N}}).$$

Since $\overline{X}_{arb} = N^{-1}$, it holds that $\Sigma_U \delta_i = 0$. In addition, we assume that $\Sigma_U \delta_i^2 = O(N)$; a justification of this assumption is given below. Then $\mu(x_{arb})$ can be written as

$$\mu(x_{arb}) = \sum_{i\in U} X^2_{i,arb} = \frac{1}{N} + \frac{\Sigma_U \delta_i^2}{N^3} = \frac{1}{N} + O(\frac{1}{N^2}). \tag{37}$$

Furthermore, by (18),

$$\begin{aligned}\frac{1}{N^{h-1}} \le \sum_{i\in U} X^h_{i,arb} &= \sum_{i\in U} \frac{1}{N^h}(1 + \frac{\delta_i}{\sqrt{N}})^h \\ &= \frac{1}{N^{h-1}} + \frac{O(\Sigma_U \delta_i^2)}{N^{h+1}} = \frac{1}{N^{h-1}}(1 + O(\frac{1}{N})),\end{aligned}$$

where we used the binomial theorem for the utmost right-hand side of the first line. Hence,

$$\sigma^2(x^h_{arb}) = \sum_{i\in U} X^{2h+1}_{i,arb} - \left(\sum_{i\in U} X^{h+1}_{i,arb}\right)^2 = O(\frac{1}{N^{2h+1}}). \tag{38}$$

Furthermore, noting that $(\mu(x^h_{exmpl}) - N^{-h}) = O(1/N^{h+1})$, it is not difficult to see from the definition of the above $X_{i,exmpl}$ that $\sigma^2(x^h_{exmpl}) = h^2[1 + o(1)]/N^{2h+1}$ so that $\sigma(x^h_{arb}) = O(\sigma(x^h_{exmpl})) = O(1/N^{h+0.5})$ for h=1, …, H.

In order to show that $N\rho_{z12} \sim -1$ for any data set (X_i, Z_i) provided $z_i = O(\sigma_z)$, define the variable V_i for each unit i by $V_i = z_i^2 / \sigma_z^2$ so that $\mu_v = \Sigma_U X_i V_i = 1$ and $\sigma_v^2 = \Sigma_U X_i v_i^2$, where $v_i = V_i - 1$. Assuming $z_i = O(\sigma_z)$, it holds that $\sigma_v^2 = O(1)$. In addition, define the correlation coefficient r_{xv} by $r_{xv} = \Sigma_U X_i x_i v_i / \sigma_x \sigma_v$. Then for an arbitrary data set (X_i, Z_i) ρ_{z12} can be rewritten as

$$\rho_{z12} = -\sum_{i \in U} X_i^2 \frac{z_i^2}{\sigma_z^2} = -\mu_x - \sum_{i \in U} X_i (X_i - \mu_x) V_i$$
$$= -\mu_x - \sigma_x \sigma_v r_{xv} = -\frac{1}{N} + O(\frac{1}{N^2}) + O(\frac{1}{N\sqrt{N}}),$$

where in the last line we used (37) and (38) with h=1. Hence, $N\rho_{z12} \sim -1$ and result (30) with ρ_{z12} instead of $\rho_{z12,exmpl}$ can be proved along the same lines as before.

A last remark deals with the justification of the above assumption that $\Sigma_U \delta_i^2 = O(N)$. Suppose that $\Sigma_U \delta_i^2 = N^{1+\varepsilon}$ $(0 < \varepsilon \leq 0.5)$. Then the order of σ_x^2 would be $1/N^{3-\varepsilon}$; see the derivation of (38). Hence, the (negative) contribution of the 2nd term on the right-hand side of (30) to $\{1 + (n-1)\rho_{zRP}\}$ would become $O(N^3 r_{xz}^2 \sigma_x^2 / d) = O(1/N^{0.5-\varepsilon})$ while $\{1 + (n-1)\rho_{z12}\}$ is of the relatively smaller order $1/\sqrt{N}$; see (27) and (30). This may lead to a negative value for $\{1 + (n-1)\rho_{zRP}\}$ which contradicts the nonnegativity of variances. In addition, cases with $\varepsilon > 0.5$ would lead to $\mu_x > 1/n$ and, consequently, to $d < 0$.

Appendix A. Justification of $a_N^{kl} = O(a_{N,exmpl}^{kl})$

In this appendix we give a justification of the assumption that for an arbitrary data set $a_{N,arb}^{kl} = O(a_{N,exmpl}^{kl})$. Suppose that $N^{h-1} \Sigma_U X_i^h \to m_h$ $(0 < c < m_h < C < \infty)$ as $n, d \to \infty$ $(h = 2, ..., H)$; due to the inequalities in (18) this is a mild assumption. In analogy with the characteristic function for a probability distribution, we assume that the configuration of the X_i can be represented reasonably well by the moments of the finite population of the X_i or, equivalently, by the (rescaled) population totals of the powers of the X_i. So under the assumption that the a_N^{kl} in (17) may depend

on the configuration of the X_i, we can write a_N^{kl} quite generally as a function of these moments

$$\begin{aligned} a_N^{kl} &= g_N^{kl}(N\Sigma_U X_i^2, \cdots, N^{H-1}\Sigma_U X_i^H) \\ &\sim g_N^{kl}(m_2, \cdots, m_H) \\ &= \{c_{N1}^{kl} + c_{N2}^{kl}(m_2 - 1) + \cdots + c_{NH}^{kl}(m_H - 1)\}(1 + O(1)), \end{aligned} \tag{39}$$

where H is a sufficiently large integer. In the last line we used a first-order Taylor approximation of $g_N^{kl}(.)$ around the values $m_{h,SRS} = 1$ $(h = 2,..., H)$ from SRS sampling. By construction, the order of c_{Nh}^{kl} solely depends on the behaviour of $g_N^{kl}(.)$ at (1,…,1) irrespective of the X_i (h=1,…,H). Furthermore, it is assumed that certain regularity conditions are satisfied so that $g_N^{kl}(.)/c_{N1}^{kl} = O(1)$; this excludes exponential forms as $a_N^{kl} \sim N^{m_2 - 2}$ but forms as, for instance, $a_N^{kl} \sim m_2 m_3^2 / N$ or $a_N^{kl} \sim 1/(1 - nm_2 / N)$ are permitted. Under these assumptions, $a_{N,arb}^{kl} = O(c_{N1}^{kl})$. Noting that the sample in the example given in section 4 consists of two SRS samples from groups of size N/2, it is seen from $\pi_{ij,exmpl}$ with $i, j \in U_2$ that $a_{N,exmpl}^{kl} = a_{N/2,SRS}^{kl}$ is of the same order and magnitude as $a_{N,SRS}^{kl} = c_{N1}^{kl}$. Therefore, $a_{N,arb}^{kl} = O(a_{N,exmpl}^{kl})$.

A more formal argument that it suffices to only take the moments of the finite population of the X_i into account in (39) follows from the method of Hartley and Rao (1962, pages 357-360) for deriving an expression for π_{ijPPS}. To find the coefficients in their approximation of π_{ijPPS}, they apply Edgeworth expansions to their standardized variates T_v. Apart from π_i, π_j, N and n, these Edgeworth expansions solely depend on the corresponding cumulants k_l $(l = 1,2,...)$. As pointed out by the authors, these k_l can be expressed in terms of the standardized cumulants of the X_i which in turn can be expressed in terms of the moments of the X_i. This explains the form of (39).

References

Asok, C. and Sukhatme, B.V. (1976). On Sampford's procedure of unequal probability sampling without replacement, *Journal of the American Statistical Association*, 71, 912-918.

Berger, Y. (1998). Rate of convergence for asymptotic variance for the Horvitz-Thompson estimator, *Journal of Statistical Planning and Inference*, 74, 149-168.

Berger, Y. (2004). A simple variance estimator for unequal probability sampling without replacement, *Journal of Applied Statistics*, 31, 305-315.

Berger, Y. (2005). Variance estimation with Chao's sampling scheme, *Journal of Statistical Planning and Inference*, 127, 253-277.

Brewer, K.R.W. (1963). A model of systematic sampling with unequal probabilities, *Australian Journal of Statistics*, 5, 5-13.

Brewer, K.R.W. (2002). Combined Survey Sampling Inference: Weighing Basu's Elephants, Arnold, London.

Brewer, K.R.W. and Donadio, M.E. (2003). The high entropy variance of the Horvitz-Thompson estimator, *Survey Methodology*, 29, 189-196.

Chao, M.T. (1982). A general purpose unequal probability sampling plan, *Biometrika*, 69, 653-656.

Cochran, W.G. (1977). *Sampling Techniques*, John Wiley & Sons, New York.

Connor, W.S. (1966). An exact formula for the probability that two specified sampling units will occur in a sample drawn with unequal probabilities and without replacement, *Journal of the American Statistical Association*, 61, 384-390.

Durbin, J. (1967). Design of multi-stage surveys for the estimation of sampling errors, *Applied Statistics*, 16, 152-164.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population, *Annals of Mathematical Statistics*, 35, 1491-1523.

Hartley, H.O. and Rao, J.N.K. (1962). Sampling with unequal probabilities and without replacement, *Annals of Mathematical Statistics*, 33, 350-374.

Hidiroglou, M.A. and Gray, G.B. (1980). Construction of joint probability of selection for systematic P.P.S. sampling, *Applied Statistics*, 29, 107-112.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association*, 47, 663-685.

Knottnerus, P. (2003). *Sample Survey Theory: Some Pythagorean Perspectives*, Springer-Verlag, New York.

Knottnerus, P. (2011). On the efficiency of randomized PPS sampling, *Survey Methodology*, 37, 95-102.

Sampford, M.R. (1967). On sampling without replacement with unequal probabilities of selection, *Biometrika*, 54, 499-513.

Särndal, C.E., Swensson, B. and Wretman, J.H. (1992). *Model Assisted Survey Sampling*, Springer-Verlag, New York.

Sen, A.R. (1953). On the estimate of the variance in sampling with varying probabilities, *Journal of the Indian Society of Agricultural Statistics*, 5, 119-127.

Thompson, M.E. and Wu, C. (2008). Simulation-based randomized systematic PPS sampling under substitution of units, *Survey Methodology*, 34, 3-10.

Tillé, Y. and Haziza, D. (2010). An interesting property of the entropy of some sampling designs, *Survey Methodology*, 36, 229-231.

Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size, *Journal of the Royal Statistical Society*, B, 15, 253-261.