Journal of Statistical Planning and Inference 199 (2019) 60–88

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Determinantal sampling designs

V. Loonis [a], X. Mary [b,*]

[a] Insee, Division des Méthodes et des Référentiels Géographiques, Paris, France
[b] Modal'X, UPL, Univ Paris Nanterre, F92000 Nanterre, France

ARTICLE INFO

Article history:
Received 24 March 2017
Received in revised form 23 May 2018
Accepted 27 May 2018
Available online 5 June 2018

ABSTRACT

In this article, recent results about point processes are used in sampling theory. Precisely, we define and study a new class of sampling designs: determinantal sampling designs. The law of such designs is known, and there exists a simple selection algorithm. We compute exactly the variance of linear estimators constructed upon these designs by using the first and second order inclusion probabilities. Moreover, we obtain asymptotic and finite sample theorems. We construct explicitly fixed size determinantal sampling designs with given first order inclusion probabilities. We also address the search of optimal determinantal sampling designs.

1. Introduction

The goal of sampling theory is to acquire knowledge of a parameter of interest θ using only partial information. The parameter θ is a function of $\{y_k, k \in U\}$, usually the sum or the mean of the y_k's. This is done by means of a sampling design, through which a random subset $\{y_k, k \in \mathbb{S}\}$ is observed, and the construction of an estimator $\hat{\theta}$ of θ based on this random sample. The properties of the sampling design are thus of crucial importance to get "good" estimators. In practice, the following issues are fundamental: simplicity of the design (in terms of its definition, theory and/or drawing algorithm), knowledge of the first and, possibly, second order inclusion probabilities, control of the size of the sample, effective construction, in particular with prescribed unequal probabilities, statistical amenability (consistency, central limit theorem, …), low Mean Square Error (MSE)/Variance of specific estimators based on the design.

In this article, we introduce a new parametric family of sampling designs indexed by Hermitian contracting matrices, *determinantal sampling designs*, that addresses all these issues. Section 2 gives their definition and probabilistic properties. In particular, it is shown that for this family, inclusion probabilities are known for any order. Section 2 also provides a sampling algorithm. Section 3 studies the statistical properties of linear estimators of a total. It gives algebraic and geometric formulas for the MSE which provide necessary and sufficient conditions for obtaining a perfectly balanced determinantal sampling design. In addition, we give asymptotic theorems and concentration inequalities. Section 4 provides effective constructions of fixed size determinantal sampling designs with fixed first order inclusion probabilities. Optimization problems and algorithms are then discussed in Section 5, and applied on a real data set. While the use of determinantal processes allows to derive directly statistical results in the field of survey sampling from well known results in point process theory (Definition 2.1, Theorems 2.1, 3.3, 3.5, Algorithm 2.1), innovative results can nevertheless be found: Theorems 3.1, 3.2, 4.1, 4.4, 5.1, or Algorithms 5.1 to 5.5. We also make connections with other theories (frame theory or semidefinite optimization).

* Corresponding author.
E-mail addresses: vincent.loonis@insee.fr (V. Loonis), xavier.mary@parisnanterre.fr (X. Mary).

https://doi.org/10.1016/j.jspi.2018.05.005
0378-3758/

2. Definition and general properties

2.1. Definition

According to its definition, an *unordered sampling design without replacement* (simply called *sampling design* afterwards) is a *simple point process* on a finite set U, that is to say a probability on 2^U, set of parts of U (Borodin, 2009; Tillé, 2011).

Among simple point processes, the general structure and properties of *determinantal point processes* have attracted a lot of attention recently (Borodin, 2009; Hough et al., 2006, 2009; Lyons, 2003; Soshnikov, 2000). This is (in part) due to the ubiquity of determinantal point processes in probability theory. They appear for instance in the study of random structures such as uniform spanning trees, zeros of random polynomials and spectra of random matrices. In the case of a finite set U, determinantal point processes are defined through associated matrices called kernels. Many probabilistic properties of these processes depend on algebraic properties of their kernels, but most of the results concern Hermitian matrices only. For this reason, and though there exist many interesting examples of determinantal point processes associated to non-Hermitian matrices, we restrict our attention to the Hermitian case.

Unless specifically stated, matrices will be complex matrices. For a complex number z, $\bar{z}$ is its conjugate and $|z| = \sqrt{z\bar{z}}$ its modulus. We introduce the following notation. For any square matrix K indexed by U and $s \subseteq U$, $K_{|s}$ denotes the submatrix of K whose rows and columns are indexed by s. We will also use the following convention: the determinant of the empty matrix is 1, as is a product over the empty set ($\prod_{k\in\emptyset}\alpha_k = 1$). From the definition of determinantal point processes we derive the following definition of *determinantal sampling designs*:

Definition 2.1 (*Determinantal Sampling Design*). A sampling design $\mathcal{P}$ on a finite set U is a determinantal sampling design if there exists a Hermitian matrix K indexed by U, called kernel, such that for all $s \in 2^U$, $\sum_{s'\supseteq s}\mathcal{P}(s') = \det(K_{|s})$. This sampling design is denoted by $DSD(K)$.

A random variable $\mathbb{S}$ with values in 2^U and law $DSD(K)$ is called a determinantal random sample (with kernel K). It satisfies, for all $s \in 2^U$,

$$pr(s \subseteq \mathbb{S}) = \det(K_{|s}).$$

We will also write $\mathbb{S} \sim DSD(K)$.

In the following we will always identify the finite population U of size N with $\{1, \ldots, N\}$. It follows from the definition that determinantal sampling designs are unordered and without replacement. Macchi (1975) and Soshnikov (2000) proved that a Hermitian matrix K defines a determinantal point process, and as a consequence a $DSD(K)$, iff (if and only if) K is a *contracting matrix*, that is a matrix whose eigenvalues are in [0, 1]. We will use the notation $0 \leq K \leq I_N$ (Loewner partial order) for a contracting matrix. It follows from this fundamental result that determinantal sampling designs form a parametric family of sampling designs, parametrized by contracting matrices.

Example 2.1 (*Poisson Sampling*). Consider a diagonal matrix K^{Π} with diagonal elements $K^{\Pi}_{kk} = \Pi_k$ with values in [0, 1]. The corresponding determinantal sampling design satisfies, for all $s \in 2^U$,

$$pr(s \subseteq \mathbb{S}) = \prod_{k\in s}\Pi_k.$$

The inclusion–exclusion principle implies that

$$pr(\mathbb{S} = s) = \prod_{k\in s}\Pi_k\prod_{k\notin s}(1-\Pi_k).$$

This is precisely the equation of the Poisson sampling design (with first order inclusion probabilities $pr(k \in \mathbb{S}) = \Pi_k$), which therefore belongs to the family of determinantal sampling designs.

Let K be a Hermitian projection matrix. Then $K = \overline{K}^T$ and $K^2 = K$, hence K is an orthogonal projection matrix. Therefore, we will make no distinction between projections and orthogonal projections. As the eigenvalues of K are 0 or 1, then K is a contracting matrix. We can thus associate to K a determinantal sampling design $DSD(K)$. We will see that $DSD(K)$ enjoys interesting statistical and computational properties. Such determinantal point processes are sometimes called determinantal projection processes (Hough et al., 2006) or elementary determinantal point processes (Kulesza and Taskar, 2011) in the literature. We will usually write the rank n projection matrix K as $K = V\overline{V}^T$, where V is the $(N\times n)$ matrix of an n orthonormal basis of the range of K. Among these sampling designs, we single out three particular cases.

Example 2.2 (*Projection*). Let J_N be the square matrix of size N with all terms equal to 1.

1. $DSD(\frac{1}{N}J_N)$ is the simple random sampling (SRS) of size 1.
2. $DSD(I_N - \frac{1}{N}J_N)$ is the SRS of size $N-1$.
3. If K is a diagonal projection matrix, $DSD(K)$ is a non-random sampling design. In particular, if $K = I_N$, then the design is a census.

Apart from the cases $n = N-1$ and $n = 1$, Kulesza (2012) proved that the SRS is not a determinantal sampling design.

2.2. Inclusion probabilities

The following formulas for the inclusion probabilities of order 1 and 2 follow from Definition 2.1. As usual in sampling theory (Särndal et al., 2003), we denote them by π_k and π_{kl}, and let $\pi = (\pi_1, \ldots, \pi_N)^T$ be the vector of first inclusion probabilities. In matrix formulation, for all $k, l \in U$, setting

$$\pi_k = pr(k \in \mathbb{S}) = K_{kk}, \tag{1}$$

$$\pi_{kl} = pr(k, l \in \mathbb{S}) = K_{kk}K_{ll} - \mid K_{kl}\mid^2 (k \neq l), \tag{2}$$

$$\Delta_{kl} = \begin{cases} \pi_{kl} - \pi_k\pi_l = - \mid K_{kl}\mid^2 (k \neq l), \\ \pi_k(1-\pi_k) = K_{kk}(1-K_{kk})(k = l). \end{cases} \tag{3}$$

it holds that

$$\Delta = \overline{(I_N - K)} * K = (I_N - K) * \overline{K}, \tag{4}$$

where $*$ is the Schur–Hadamard (entrywise) matrix product.

Proposition 2.1. *From (3) a determinantal sampling design satisfies the so-called Sen–Yates–Grundy conditions:*

$$\pi_{kl} \leq \pi_k\pi_l \,(k \neq l). \tag{5}$$

More generally, a determinantal sampling design has *negative associations* (Lyons, 2003). In particular, for disjoint subsets A and B it holds that

$$pr(A \cup B \subseteq \mathbb{S}) \leq pr(A \subseteq \mathbb{S})pr(B \subseteq \mathbb{S}).$$

It was shown recently that determinantal point processes actually enjoy the *strong Rayleigh property* (Borcea et al., 2009; Pemantle and Peres, 2014), a technical property stronger than negative association. This property can be defined in terms of the localization of the zeros of the generating function of the process. These two properties (negative association, strong Rayleigh property) proved very useful for the study of statistics of determinantal processes (Yuan et al., 2003; Brändén and Jonasson, 2012; Pemantle and Peres, 2014). Some results will be used in Section 3.

2.3. Sample size

Of major importance to statisticians is the sample size of the random sample. It is for instance very common in practice to work with fixed size samples, that is with samples whose size is non-random and given. The sample size of a determinantal random sample follows from Theorem 7 in Hough et al. (2006). For a set A, let $\sharp A$ denotes its cardinal and for a Hermitian matrix K of size N, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be the N values on the diagonal of any diagonalizing matrix of K in descending order (vector of eigenvalues with their multiplicities, in descending order).

Theorem 2.1 (*Sample Size*)**.** *Let $\mathbb{S} \sim DSD(K)$. Then the random variable $\sharp\mathbb{S}$ has the law of a sum of N independent Bernoulli variables $B_1, \ldots, B_N$ of parameters $\lambda_1, \ldots, \lambda_N$.*

Corollary 2.1 (*Sample Size (2)*)**.** *Let $\mathbb{S} \sim DSD(K)$. Then*

1. $E(\sharp\mathbb{S}) = tr(K)$.
2. $var(\sharp\mathbb{S}) = tr(K - K^2) = \sum_{i=1}^{N} \lambda_i(1-\lambda_i) = \sum_{k,l\in U} \Delta_{kl}$.
3. $pr(\mathbb{S} = \emptyset) = 0$ *iff* 1 *is an eigenvalue of* K.
4. *The sample size is less than or equal to the rank of* K.
5. $DSD(K)$ *is a fixed size determinantal sampling design iff* K *is a projection matrix, and the size equals the rank of* K.

Proof. It holds that $var(\sharp\mathbb{S}) = \sum_{k,l\in U} \Delta_{kl}$ (see Särndal et al., 2003). The other results follow directly from Theorem 2.1 and the spectral decomposition of Hermitian matrices. □

Example 2.3 (*Unitary Transform*)**.** Let $K \in \mathcal{M}_{N\times N}(\mathbb{C})$ be a contracting matrix and $\mathbb{S} \sim DSD(K)$. Let also $W \in \mathcal{M}_{N\times N}(\mathbb{C})$ be a unitary matrix ($W\overline{W}^T = I_N$). Then $K_W = WK\overline{W}^T$ is a Hermitian matrix with the same eigenvalues as K. It follows that $\mathbb{S}_W \sim DSD(K_W)$ exists, with $\sharp\mathbb{S}_W \underset{law}{=} \sharp\mathbb{S}$.

2.4. Additional properties

We give here some other general probabilistic results on determinantal sampling designs and their interpretation in terms of sampling theory. We refer to Lyons (2003) and Hough et al. (2006) for their probabilistic versions.

Proposition 2.2 (*Complementary Sample*). *Let $\mathbb{S} \sim DSD(K)$. The complementary random sample $\mathbb{S}^c$ is a determinantal random sample with kernel $I_N - K$.*

Proposition 2.3 (*Domain*). *Let $DSD(K)$ be a determinantal sampling design on U with kernel K, and let $A \subseteq U$ be a subpopulation (or domain). Then the restriction $DSD(K)_{|A}$ of $DSD(K)$ to A is a determinantal sampling design on A with kernel $K_{|A}$, the submatrix of K whose rows and columns are indexed by A:*

$$DSD(K)_{|A} = DSD(K_{|A}).$$

Proposition 2.4 (*Stratification*). *Let $\{U_1, \ldots, U_H\}$ be a partition of U into H strata. The sampling design $DSD(K)$ is stratified iff the matrix K admits a block diagonal decomposition relative to these strata, that is $k \in U_h$, $l \in U_{h'}$, $h \neq h'$ implies $K_{kl} = 0$.*

By using the inclusion–exclusion principle, Lyons (2003) shows that the probabilities of disjunction are also given by a determinant (Theorem 5.1 Equation (5.2) for fixed size designs and Equation (8.1) for random size designs).

2.5. Sampling algorithm

A general algorithm for simulating a determinantal sampling design is provided in Hough et al. (2006), including a proof of its validity in a very general setup. Other implementations of this algorithm can be found in Scardicchio et al. (2009) and Lavancier et al. (2015). We consider the latter since it is more suitable and efficient when N is large. Algorithm 2.1 samples from fixed size n determinantal sampling designs. Let K be a projection matrix.

Algorithm 2.1.

1. Find a (N, n) matrix V such that $K = V\overline{V}^T$. Let v_k^T be the kth line of V.
2. Sample one element k_n of U with probabilities $\Pi_k^n = \|v_k\|^2/n$, $k \in U$.
3. Set $e_1 = v_{k_n}/\|v_{k_n}\|$.
4. For $i = (n-1)$ to 1 do:
   (a) sample one k_i of U with probabilities $\Pi_k^i = \frac{1}{i}[\|v_k\|^2 - \sum_{j=1}^{j=n-i}|\overline{e_j}^T v_k|^2]$, $k \in U$,
   (b) set $w_i = v_{k_i} - \sum_{j=1}^{j=n-i}|\overline{e_j}^T v_{k_i}|e_j$ and $e_{n-i+1} = w_i/\|w_i\|$.
5. End for.
6. Return $\{k_1, \ldots, k_n\}$.

The resulting sample is a realization of $DSD(K)$.

Step 1 of Algorithm 2.1 can be computationally costly, for it involves the decomposition of the $N \times N$ matrix K in $V\overline{V}^T$ which may be time consuming for very large population size N. Therefore it is preferable to have a description of the matrix K directly in terms of V. This is the case for the matrices defined in Theorems 4.1, 4.4 and E.1 (or at Step 2 of Algorithm 2.2).

Algorithm 2.2 describes a procedure to sample from any determinantal sampling design, by expressing it as a mixture of fixed size sampling designs (Theorem 7 in Hough et al., 2006).

Let K be a contracting matrix.

Algorithm 2.2.

1. Find the rank one decomposition $K = \sum_{i=1}^N \lambda_i \phi_i \overline{\phi_i}^T$.
2. Simulate a vector b whose components are independent Bernoulli variables with parameter $\lambda_1, \ldots, \lambda_N$, the eigenvalues of K.
3. Construct the projection matrix $K_b = \sum_{i=1}^N b_i \phi_i \overline{\phi_i}^T$.
4. Sample from $DSD(K_b)$ by Algorithm 2.1.

The resulting sample is a realization of $DSD(K)$.

3. Estimation of a total

3.1. Linear estimators and their mean square error

Let $y = (y_1, \ldots, y_N)^T$ be a variable of interest on the population $U = \{1, \ldots, N\}$. Typical parameters to estimate are the total $t_y = \sum_{k\in U} y_k$ or the mean value $m_y = t_y/N$. Let $DSD(K)$ be a determinantal sampling design on U with kernel K.

An estimator of t_y based on $DSD(K)$ is called linear and homogeneous if there exist real weights $w_k, k \in U$ such that the estimator writes

$$\hat{t}_{yw} = \sum_{k \in \mathbb{S}} w_k y_k, \text{ with } \mathbb{S} \sim DSD(K).$$

The Mean Square Error (MSE) decomposes as:

$$\mathrm{MSE}(\hat{t}_{yw}) = \overbrace{\sum_{k\in U}\sum_{l\in U} w_k w_l y_k y_l \Delta_{kl}}^{\text{Variance}} + \left[\overbrace{\sum_{k\in U}(w_k\pi_k - 1)y_k}^{\text{Bias}}\right]^2 \tag{6}$$

$$= \sum_{k\in U} w_k w_l y_k y_l (K_{kk}(1-K_{kk})) - \sum_{k\in U}\sum_{l\neq k} w_k w_l y_k y_l |K_{kl}|^2 + \left[\sum_{k\in U}(w_k K_{kk} - 1)y_k\right]^2 \tag{7}$$

where Δ_{kl} is defined by Eq. (3). An unbiased estimator (for all variables y!) then exists only if $\pi_k > 0$ for all $k \in U$. In this case it should satisfy $w_k = \pi_k^{-1}$. The corresponding estimator,

$$\hat{t}_{yHT} = \sum_{k\in\mathbb{S}} \pi_k^{-1} y_k,$$

is known as the Horvitz–Thompson estimator (Horvitz and Thompson, 1952). In the sequel, we will not restrict our attention to this estimator only. Indeed, we construct estimators with $w_k \neq \pi_k^{-1}$ in Section 5. In particular, we prove (Corollary 5.1) that in presence of an auxiliary variable x (approximately) proportional to y and for sampling designs of fixed size n, it may be interesting to consider the linear and homogeneous estimator with vector of weights $w^{opt} = \left((nx_1)^{-1}t_x, \dots, (nx_N)^{-1}t_x\right)^T$.

If the sampling design is of fixed-size, the MSE becomes:

$$\mathrm{MSE}(\hat{t}_{yw}) = -\frac{1}{2}\sum_{k\in U}\sum_{\substack{l\in U\\ l\neq k}}(w_k y_k - w_l y_l)^2 \Delta_{kl} + \left[\sum_{k\in U}(w_k\pi_k - 1)y_k\right]^2 \tag{8}$$

$$= \frac{1}{2}\sum_{k\in U}\sum_{\substack{l\in U\\ l\neq k}}(w_k y_k - w_l y_l)^2 |K_{kl}|^2 + \left[\sum_{k\in U}(w_k\pi_k - 1)y_k\right]^2. \tag{9}$$

Thus, to achieve small variance for a fixed size sampling design, $|K_{kl}|^2$ has to be small when $(w_k y_k - w_l y_l)^2$ is large. Equivalently, for a given set of first order inclusion probabilities, π_{kl} has to be as close as possible to $\pi_k\pi_l$ (Eq. (3), Proposition 2.1). Therefore, one has to find a trade-off between fixed-size sampling and Poisson sampling. The next sections provide instances of such sampling designs (Theorem 3.1, Corollary 4.1, Theorem 4.4, Algorithm 5.1).

3.2. *Mean square error for determinantal sampling designs*

In the case of a determinantal sampling design, the MSE of the homogeneous linear estimator $\hat{t}_{yw} = \sum_{k\in\mathbb{S}} w_k y_k$ of the total t_y of a variable of interest y admits algebraic and geometric formulations. They enable us to provide necessary and sufficient conditions for a perfect estimation of the total of auxiliary variables.

We introduce the following notations. We let $w = (w_1, \dots, w_N)^T$ and $e = (1, \dots, 1)^T$ (e is of size N). For a vector x, x^{-1} is its Schur–Hadamard inverse, and D_x denotes the diagonal matrix with diagonal x, whereas for a matrix A, $diag(A)$ is the vector of diagonal elements. For any two matrices $A, B \in \mathcal{M}_N(\mathbb{C})$, $\langle A, B\rangle = tr(\overline{A}^T B) = \sum_{k,l}\overline{a_{k,l}}b_{k,l}$ denotes the canonical scalar product on $\mathcal{M}_N(\mathbb{C})$. The associated Frobenius norm is denoted by $|A|$.

We also define $z = w * y$ (Schur–Hadamard product) and diagonal matrices $Z = D_{w*y}$, $Z^{1/2} = D_{\sqrt{w*y}}$ where the square root is taken in the complex sense for negative entries of $w * y$. Finally, we pose $\langle\langle A, B\rangle\rangle = \langle \overline{Z^{1/2}}^T A Z^{1/2}, Z^{1/2} B \overline{Z^{1/2}}^T\rangle$. Note that $Z^{1/2} = (Z^{1/2})^T$ and $\overline{Z} = Z$, two equalities that we will use thoroughly in the rest of this section.

Proposition 3.1 (*Algebraic and Geometric Forms of the* MSE)**.** *Let $\mathbb{S} \sim DSD(K)$. The* MSE *of $\hat{t}_{yw}$ satisfies*

$$\mathrm{MSE}(\hat{t}_{yw}) = (w*y)^T((I_N - K)*\overline{K})(w*y) + [e^T(K*I_N)(w*y) - e^T y]^2 \tag{10}$$

$$= \langle\langle I_N - K, K\rangle\rangle + [\langle D_y, KD_w - I_N\rangle]^2 \tag{11}$$

and, in the case of the Horvitz–Thompson estimator,

$$\mathrm{MSE}(\hat{t}_{yHT}) = var(\hat{t}_{yHT}) = (diag(K)^{-1} * y)^T((I_N - K) * \overline{K})(diag(K)^{-1} * y) \tag{12}$$

$$= \langle\langle I_N - K, K\rangle\rangle. \tag{13}$$

Proof. These formulas follow from the classical equality $tr(AB) = tr(BA)$ and the following equality relating the trace and the Schur–Hadamard product (Horn and Johnson, 1991): for any two vectors x, y and any two matrices A, B it holds that $\bar{x}^T A * By = tr(\overline{D_x} A D_y B^T)$. □

Recently, Deville (2012) raised the following question. For a given vector y, when can we estimate perfectly the total y, using a sampling design with fixed first order inclusion probabilities (and an homogeneous linear estimator)? Using the previous equations, we provide necessary and sufficient conditions within determinantal sampling designs.

Theorem 3.1 (*Perfect Estimation*)**.** *Assume y takes only non-zero values. Let $\mathbb{S} \sim DSD(K)$ and w be a vector of weights with non-zero values. Let $\alpha_1, \ldots, \alpha_q$, be the distinct values of $w_k y_k$, $k = 1, \ldots, N$, and A_j, $j = 1, \ldots, q$ be the associated sets of indexes k such that $w_k y_k = \alpha_j$.*

Then the following statements are equivalent:

1. *The total t_y is perfectly estimated (MSE $= 0$) by $\hat{t}_{yw}$.*
2. *K is a projection that commutes with $Z = D_{w*y}$, and $\sum_{k\in U} w_k K_{kk} y_k = t_y$.*
3. *$DSD(K)$ is a stratified determinantal sampling design with strata A_j, $j = 1, \ldots, q$, of fixed size within each stratum, and $\sum_{k\in U} w_k K_{kk} y_k = t_y$.*

In particular, the total t_y is perfectly estimated by $\hat{t}_{yHT}$ iff K is a projection with positive diagonal that commutes with $Z = D_{\pi^{-1}*y}$ iff $DSD(K)$ is a stratified determinantal sampling design of fixed size within each stratum, and with $\pi_k^{-1} y_k$ constant on each stratum.

Proof.

1 ⇒ 2 By Moutard–Fejer's Theorem (De Klerk, 2006 Appendix A), it holds that for any two semidefinite matrices A and B, $tr(AB) \geq 0$ with equality iff $AB = 0$. Assume $\mathrm{MSE}(\hat{t}_{yw}) = 0$. Then $tr\left(\overline{Z^{1/2}}^T (I_N - K) Z^{1/2} Z^{1/2} K \overline{Z^{1/2}}^T\right) = 0$. As $\overline{Z^{1/2}}^T (I_N - K) Z^{1/2}$ and $Z^{1/2} K \overline{Z^{1/2}}^T$ are semidefinite, then $\overline{Z^{1/2}}^T (I_N - K) Z K \overline{Z^{1/2}}^T = 0$. Multiplying on the left and on the right by $\overline{Z^{-1/2}}$ yields $ZK = KZK$ and taking the conjugate transpose gives $ZK = KZK = KZ$. Thus K and Z commute. It also follows that $ZK^2 = ZK$. By multiplying the equality on the left by Z^{-1} we get $K^2 = K$, and K is a projection. Also, the bias is 0 and $\sum_{k\in U} w_k K_{kk} y_k = t_y$.

2 ⇒ 3 Reorder the population by strata. Then the commutant of Z is the set of block diagonal matrices with respect to these stratas, and K is block diagonal. As K is also a projection, each block is actually a projection, and $DSD(K)$ is of fixed size within each stratum.

3 ⇒ 1 As the sampling design restricted to each stratum is of fixed size, and the values $w_k y_k$ are constant, then the linear estimator $\hat{t}_{yw}$ is constant as a sum of constant terms. Finally $\hat{t}_{yw} = E(\hat{t}_{yw}) = \sum_{k\in U} w_k K_{kk} y_k = t_y$. □

Corollary 3.1. *Let y be any variable. Decompose the population U in two subsets: $U_1 = \{k \in U | y_k \neq 0\}$ and $U_2 = \{k \in U | y_k = 0\}$. The total t_y is perfectly estimated by $\hat{t}_{yw}$ based on $DSD(K)$ iff K is a contracting matrix such that $K_{|U_1}$ satisfies the criteria of Theorem 3.1.*

Next example shows that non-Horvitz–Thompson estimators may prove useful.

Example 3.1. Let $U = \{1, \ldots, 6\}$ and $y^T = (0, 0, 2, 4, 8, 8)$. In the context of equal probability sampling with $\pi_k = 1/2$, no determinantal sampling design can produce a perfect Horvitz–Thompson estimator (there are 4 different values of $y_k \pi_k^{-1}$, but there are only 3 blocks). However, $\hat{t}_{yw}$ with

$$\mathbb{S} \sim DSD(K), K = \frac{1}{2}\begin{pmatrix} 1 & z & 0 & 0 & 0 & 0 \\ \bar{z} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}, \quad w^T = (w_1, w_2, 3, 3/2, 2, 2)$$

satisfies $\mathrm{MSE}(\hat{t}_{yw}) = 0$ (for any $z \in \mathbb{C}$, $|z| \leq 1$ and any $w_1, w_2 \in \mathbb{R}$).

Finally, we provide an alternative view on the variance that comes from the general theory of point processes and spatial statistics. The quantity $\sum_{k\neq l} w_k w_l y_k y_l |K_{kl}|^2$ can be interpreted as a weighted measure of global repulsiveness for point processes on a discrete space (Biscio et al., 2016; Lavancier et al., 2015 in the continuous setting). As determinantal point processes are repulsive, we then expect that linear and homogeneous estimators will achieve small variance for certain DSDs compare to other sampling designs. This is validated by our empirical studies in Section 5.3. The general problem of minimization of the MSE will be addressed in Section 5.

3.3. Asymptotic properties of the estimator

The classical settings for the study of asymptotic properties are either the superpopulation models (Deming and Stephan, 1941; Cassel et al., 1977 chapter 4), or the models of nested (finite) populations as described by Isaki and Fuller (1982). We consider this second setting here. In particular, $(U_N, N \in \mathbb{N})$ is a nested sequence of finite populations ($U_N \subseteq U_{N+1}$). The variable of interest y^N may depend on N, $(y^N, N \in \mathbb{N})$ is a sequence of vectors of size N. Also $(w^N, N \in \mathbb{N})$ is a sequence of positive vectors of size N. In all this section, $(\mathcal{P}_N, N \in \mathbb{N})$ is a sequence of determinantal sampling designs on the populations U_N with kernel $(K^N, N \in \mathbb{N})$, whose diagonal terms are positive, and $(\hat{t}^N_{yw}, N \in \mathbb{N})$ is the sequence of associated linear estimators of t_{y^N} with weights w^N. To simplify notations, we consider as before $U_N = \{1, \dots, N\}$, and omit the superscript $(.)^N$, writing y, w, K and $\hat{t}_{yw}$ instead of y^N, w^N, K^N and $\hat{t}^N_{yw} (= \hat{t}^N_{y^N w^N})$.

We focus successively on consistency, central limit theorems and concentration/deviation inequalities.

In this setting, most results about consistency concern the mean square convergence of the Horvitz–Thompson estimator of the mean $m_y = t_y/N$, see Isaki and Fuller (1982), Robinson (1982), Dol et al. (1996) in the case of fixed size sampling designs and Cardot et al. (2010), Chauvet (2014) in the general case. A classical condition within these references is that the sequence $\frac{1}{N}\sum_{k\in U}(\pi_k)^{-2}y_k^2$ is bounded. Using Schur's Theorem Schur (1911) on semidefinite matrices we improve the previous condition for determinantal sampling designs. Theorem 3.2 also applies to other linear homogeneous estimators than the Horvitz–Thompson one. Example 3.1 shows the interest of considering such estimators. More generally, we describe an estimator whose weights result from an optimization problem in Section 5, Theorem 5.1. We pose $\hat{m}_{yw} = \hat{t}_{yw}/N$.

Theorem 3.2 (*Mean-square Convergence*)**.** *Let $\mathbb{S} \sim DSD(K)$. If*

1. $\sum_{k\in U_N} K_{kk}\left(1 - \frac{1}{K_{kk}w_k}\right)^2 = O(1)$,
2. $\frac{1}{N^2}\sum_{k\in U_N} K_{kk}(w_k y_k)^2 \underset{N\to\infty}{\longrightarrow} 0$,

then $(\hat{m}_{yw} - m_y)$ tends to 0 in mean square.
In particular a sufficient condition for the convergence of $(\hat{m}_y^{HT} - m_y)$ towards 0 in mean square is

$$\frac{1}{N^2}\sum_{k\in U_N}\frac{y_k^2}{K_{kk}} \underset{N\to\infty}{\longrightarrow} 0.$$

Proof. By Proposition 3.1

$$\mathrm{MSE}(\hat{t}_{yw}) = (w*y)^T((I_N - K)*\overline{K})(w*y) + [e^T(I_N*K)(w*y) - e^Ty]^2.$$

As the matrices $I, K, I-K$ and $\overline{K}$ are positive semidefinite, it holds that $(I-K)*\overline{K}, I*\overline{K}$ and $K*\overline{K}$ are positive semidefinite by Schur Theorem. Since $(I-K)*\overline{K} = I*\overline{K} - K*\overline{K}$ then it also holds that $(I-K)*\overline{K} \leq I*\overline{K}$ for the partial order on positive semidefinite matrices. It follows that

$$\begin{aligned}(w*y)^T((I-K)*\overline{K})(w*y) &\leq (w*y)^T(I*\overline{K})(w*y)\\ &\leq \sum_{k\in U}(w_k y_k)^2 K_{kk}.\end{aligned}$$

Moreover the bias satisfies

$$\begin{aligned}[e^T(I_N*K)(w*y) - e^Ty]^2 &= \left(\sum_{k\in U}(K_{kk} - \frac{1}{w_k})(w_k y_k)\right)^2\\ &= \left(\sum_{k\in U}(\sqrt{K_{kk}} - \frac{1}{\sqrt{K_{kk}}w_k})(\sqrt{K_{kk}}w_k y_k)\right)^2\\ &\leq \left(\sum_{k\in U}(\sqrt{K_{kk}} - \frac{1}{\sqrt{K_{kk}}w_k})^2\right)\left(\sum_{k\in U}K_{kk}(w_k y_k)^2\right)\end{aligned}$$

by Cauchy–Schwartz-inequality. From these inequalities we get

$$E\left((\frac{\hat{t}_{yw} - t_y}{N})^2\right) \leq \left(1 + \sum_{k\in U_N}K_{kk}(1 - \frac{1}{K_{kk}w_k})^2\right)\frac{1}{N^2}\left(\sum_{k\in U}K_{kk}(w_k y_k)^2\right)$$

which goes to 0 by assumptions. This completes the proof. □

Regarding equal probability determinantal sampling designs with expected size μ ($\pi_k = \mu/N$ for all k) and a bounded variable y, a sufficient condition for convergence of the Horvitz–Thompson estimator of the mean is simply $\mu \to \infty$. More generally

Corollary 3.2. *Let $\mathbb{S} \sim DSD(K)$ and set $\mu = trace(K)$. If*

1. *there exists $c > 0$, such that for all $N \in \mathbb{N}$ and all $k \in U_N$, $c\frac{\mu}{N} \leq K_{kk}$,*
2. *the sequence $(\frac{1}{N}\sum_{k\in U_N} y_k^2, N \in \mathbb{N})$ is bounded,*
3. *the expected size of the samples $\mu \to \infty$.*

Then $(\hat{m}_y^{HT} - m_y) \to 0$ in mean square.

Our first condition is weaker than the one in Cardot et al. (2010): "there exists $\lambda > 0$, $\lambda \leq \min_{k\in U_N}\pi_k$" (take $c = \lambda$ and use that $\mu/N \leq 1$). It is actually strictly weaker because in this corollary, the first order inclusion probabilities ($\pi_k = K_{kk}$) can tend to zero. For instance, we can take $K_{kk} = \log(N)/N$. The second assumption appears for instance in Robinson (1982).

Apart consistency, some authors have considered the existence of central limit theorems for sampling designs. However, this proves generally a difficult task even for means or totals, and existing results either focus on a particular class of sampling designs (equal probability sampling designs: Erdös and Rényi, 1959; Hájek, 1960, rejective Poisson sampling: Hájek, 1964), or assume entropy conditions (Berger, 1998). Assuming only that the determinantal sampling design is "random enough", we obtain a central limit theorem by applying the results of Soshnikov (2000, 2002). These articles contain several theorems on the asymptotic normality of functionals of determinantal point processes. Theorem 1 on linear statistics of bounded measurable functions in Soshnikov (2002) can be applied straightforwardly to the study of determinantal sampling designs and their associated linear homogeneous estimators.

Theorem 3.3 (*Central Limit Theorem*)**.** *Let $\mathbb{S} \sim DSD(K)$. Define for all $N \in \mathbb{N}$ the homogeneous linear estimators*

$$\hat{t}_{yw} = \sum_{k\in\mathbb{S}} w_k y_k \text{ and } \hat{t}_{|y|w} = \sum_{k\in\mathbb{S}} w_k |y_k|.$$

If the variance $var(\hat{t}_{yw}) \to +\infty$ as $N \to \infty$ and if

$$\sup_{k\in U_N} |w_k y_k| = o\big(var(\hat{t}_{yw})\big)^{\epsilon} \text{ and } E(\hat{t}_{|y|w}) = O\big(var(\hat{t}_{yw})\big)^{\delta}$$

for any $\epsilon > 0$ and some $\delta > 0$, then

$$\frac{\hat{t}_{yw} - E(\hat{t}_{yw})}{\sqrt{var(\hat{t}_{yw})}} \overset{law}{\to} \mathcal{N}(0,1).$$

The assumption $var(\hat{t}_{yw}) \to +\infty$ is natural to get a CLT, but a lower bound on the variance is given by the smallest eigenvalue of $(I-K)*K$, that is 0 for instance for fixed size sampling designs. The two other assumptions are more technical. We present a specific case where they are met.

Corollary 3.3. *Let $\mathbb{S} \sim DSD(K)$. If for some $a, b > 0$, $\sup_{k\in U_N}|w_k y_k| = O\left(\log(N)^a\right)$ and $N^b = O\left(var(\hat{t}_{yw})\right)$ then*

$$\frac{\hat{t}_{yw} - E(\hat{t}_{yw})}{\sqrt{var(\hat{t}_{yw})}} \overset{law}{\to} \mathcal{N}(0,1).$$

The condition $\sup_{k\in U_N}|w_k y_k| = O\left(\log(N)^a\right)$ is for instance met in the case of the Horvitz–Thompson estimation of a bounded variable, with $\min_{k\in U_N} K_{kk} \geq c > 0$.

As usual, we can replace the true variance $var(\hat{t}_{yw}^N)$ by any weakly consistent estimator of this variance, using Slutsky theorem. Classical estimators of the variance (Horvitz and Thompson, 1952; Yates and Grundy, 1953; Sen, 1953) need the knowledge and positivity of the second order probabilities. These quantities are perfectly known for determinantal sampling designs.

As previously recorded, from a very different perspective, the work of Berger (1998) proves asymptotic normality for fixed size sampling designs under asymptotically maximal entropy conditions. Recently, the asymptotic normality has also been studied for more general classes of processes (that include the determinantal ones): processes with negative or positive associations (Patterson et al., 2001; Yuan et al., 2003), and processes that satisfy the strong Rayleigh property (Brändén and Jonasson, 2012). We adapt here Theorem 2.4 of Patterson et al. (2001) in the case of the Horvitz–Thompson estimator of the total based on determinantal sampling designs. The variance of the Horvitz–Thompson estimator decomposes as

$$var(\hat{t}_{yHT}) = \overbrace{\sum_{k\in U} y_k^2(K_{kk}^{-1}-1)}^{\text{Poisson contribution}} - \overbrace{2\sum_{k\in U}\sum_{l<k}\frac{y_k y_l}{\pi_k\pi_l}|K_{kl}|^2}^{\text{off-diagonal contribution}}.$$

Set $s^2 = \sum_{k\in U_N} y_k^2(K_{kk}^{-1}-1)$, $r = \sum_{k\in U_N}\sum_{l<k}\frac{y_k y_l}{\pi_k\pi_l}|K_{kl}|^2$ and $C = \sup_{k\in U_N}|\pi_k^{-1}y_k|$.

Theorem 3.4. *Let $\mathbb{S}\sim DSD(K)$. If $s^2\to\infty$, $r=o(s^2)$ and $C=o(s)$, then*

$$\frac{\hat{t}_{yHT}-t_y}{s} \xrightarrow{law} \mathcal{N}(0,1).$$

For instance, the DSD defined in Theorem 4.5 satisfies these properties for any bounded variable $0 < a \le y \le b$ (Lemma A.1). More trivially, so does the Poisson sampling $DSD(D_\Pi)$ (with $0<\alpha\le\pi_k\le\beta<1$).

For processes satisfying the strong Rayleigh property, Pemantle and Peres (2014) recently proved concentration and deviation inequalities that extend those of Lyons (2003) for the number of points of determinantal processes in a subdomain. Their application to sampling theory allows derivation of the following finite distance results.

Theorem 3.5 (*Deviation and Concentration Inequalities*)**.** *Let $\mathbb{S}\sim DSD(K)$, $\mu = trace(K)$ and $C=\sup_{k\in U}|w_k y_k|$. For all $a>0$,*

$$pr(\hat{t}_{yw}-E(\hat{t}_{yw})>a) \le 3\exp\left(-\frac{a^2}{16\left(aC+2\mu C^2\right)}\right),$$

$$pr(|\hat{t}_{yw}-E(\hat{t}_{yw})|>a) \le 5\exp\left(-\frac{a^2}{16^2\left(aC+2\mu C^2\right)}\right).$$

Moreover, if $DSD(K)$ is of fixed size $\mu=n$, then

$$pr(\hat{t}_{yw}-E(\hat{t}_{yw})>a) \le \exp\left(-\frac{a^2}{8nC^2}\right),$$

$$pr(|\hat{t}_{yw}-E(\hat{t}_{yw})|>a) \le 2\exp\left(-\frac{a^2}{8nC^2}\right).$$

Proof. Function $s\mapsto\sum_{k\in U} w_k y_k 1_{\{k\in s\}}$ is C-Lipschitz for the Hamming distance. Theorems 3.1 and 3.2 of Pemantle and Peres (2014) apply and yield the stated results. □

From this concentration inequality, we derive a new criterion for the convergence in probability of $\hat{t}_{yHT}$:

Corollary 3.4. *Let $\mathbb{S}\sim DSD(K)$. If $\frac{\sqrt{trace(K)}}{N}\sup_{k\in U}|\frac{y_k}{K_{kk}}| \xrightarrow[N\to\infty]{} 0$ then $\left(\hat{m}_{yHT}-m_y\right)\xrightarrow[N\to\infty]{pr} 0$.*

Proof. Let $C=\sup_{k\in U_N}\frac{|y_k|}{K_{kk}}$, $\mu = trace(K)$. It holds that

$$pr(|\hat{t}_{yHT}-t_y|>Na) \le 5\exp\left(-\frac{N^2a^2}{16^2\left(NaC+2\mu C^2\right)}\right).$$

By assumption $C=o(N)$ and $\mu C^2 = o(N^2)$, and the right hand term above tends to 0. □

In the particular case of a bounded variable $|y|\le b$, we have that

$$E(\hat{m}_{yHT}-m_y)^2 \le \frac{C_N}{N}b \le b\frac{\sqrt{trace(K)}}{N}C_N$$

and the assumption of Corollary 3.4 is stronger than the one of Theorem 3.2. This may be disappointing at first sight because convergence in quadratic mean entails convergence in probability by Tchebyshev's inequality. But the improvement of Corollary 3.4 lies in the rate of convergence. Consider for instance an equal probability determinantal sampling design of size n. Applying Tchebyshev's inequality in Theorem 3.2 gives the quadratic rate $P(|\hat{m}_{yHT}-m_y|\ge a)\le\frac{nb^2}{a^2}$, whereas in Corollary 3.4 we have the exponential rate $P(|\hat{m}_{yHT}-m_y|\ge a)\le 5\exp\left(-\frac{a^2n}{16^2(a+2b)}\right)$.

4. Constructing fixed size determinantal sampling designs with prescribed first order inclusion probabilities

It is common in practice to work with fixed size sampling designs with prescribed first order inclusion probabilities. According to Corollary 2.1, constructing such a determinantal sampling design is equivalent to constructing a projection

Table 1
Values of P^{Π}_{kl} : $k > l$.

Values of k	Values of l	
	$l = k_r$	$k_r < l < k_{r+1}$
$k_{r'} < k < k_{r'+1}$	$-\sqrt{\Pi_k}\sqrt{\frac{(1-\Pi_l)(\Pi_l-\alpha_l)}{1-(\Pi_l-\alpha_l)}}\gamma_r^{r'}$	$\sqrt{\Pi_k\Pi_l}\gamma_r^{r'}$
$k = k_{r'+1}$	$-\sqrt{\frac{(1-\Pi_k)\alpha_k}{1-\alpha_k}}\sqrt{\frac{(1-\Pi_l)(\Pi_l-\alpha_l)}{1-(\Pi_l-\alpha_l)}}\gamma_r^{r'}$	$\sqrt{\frac{(1-\Pi_k)\alpha_k}{1-\alpha_k}}\sqrt{\Pi_l}\gamma_r^{r'}$

matrix with a prescribed diagonal. The latter problem is a particular case of the more general issue of constructing Hermitian matrices with prescribed diagonal and spectrum that has re-attracted attention over the last years (Schur, 1911; Horn, 1954; Kadison, 2002; Dhillon et al., 2005; Fickus et al., 2013). Nevertheless, up to now, the effective constructions found in the literature are algorithmic and do not provide a closed form for such the matrices.

Relying on the existing literature, we provide in Section 4.1 a closed-form formula for a matrix P^{Π}, such that $DSD(P^{\Pi})$ is a fixed size sampling designs with first order inclusion probabilities $\pi_k = \Pi_k$, for any prescribed vector of inclusion probabilities Π such that $\sum_{k\in U}\Pi_k$ is an integer (which is obviously a necessary condition and happens to be sufficient, Theorem 4.1). We then discuss the properties of the associated sampling design $DSD(P^{\Pi})$. We finally focus in Sections 4.2 and 4.3 on the particular case of equal probability sampling designs. In Section 4.2 we consider the existence of determinantal sampling designs having the same first and second order probabilities as SRS. The existence or non-existence of such designs will in particular show that there may not exist DSDs with prescribed second-order inclusion probabilities, and that DSDs with real kernels form a proper subset of DSDs with complex kernels. Section 4.3, provides an explicit construction of a parametric family of fixed size and equal probability determinantal sampling designs relying on the Nth primary unit roots, which may prove useful in case of a periodic pattern.

4.1. A general construction

Let Π be a vector of size N such that $0 < \Pi_k < 1$ and $\sum_{k\in U}\Pi_k = n \in \mathbb{N}^*$. Set $k_0 = 0$ and for all integer r such that $1 \le r \le n$, let

- $1 < k_r \le N$ be the integer such that $\sum_{k=1}^{k_r-1}\Pi_k < r$ and $\sum_{k=1}^{k_r}\Pi_k \ge r$,
- $\alpha_{k_r} = r - \sum_{k=1}^{k_r-1}\Pi_k$ and $\alpha_k = \Pi_k$ if $k \neq k_r$,
- $\gamma_r^{r'} = \sqrt{\prod_{j=r+1}^{r'}\frac{(\Pi_{k_j}-\alpha_{k_j})\alpha_{k_j}}{(1-\alpha_{k_j})(1-(\Pi_{k_j}-\alpha_{k_j}))}}$ for $r < r'$, $\gamma_r^{r'} = 1$ otherwise.

Example 4.1. Let $N = 10$, $n = 3$ and $\Pi = (0.2, 0.2, 0.3, 0.3, 0.4, 0.4, 0.3, 0.3, 0.3, 0.3)^T$, we then get:

- $k_1 = 4, k_2 = 7, k_3 = 10$,
- $\alpha_4 = 0.3 = \Pi_4, \alpha_7 = 0.2, \alpha_{10} = 0.3 = \Pi_{10}$.

We define a real symmetric kernel P^{Π} as follows:

- for all $1 \le k \le N, P^{\Pi}_{kk} = \Pi_k$,
- for all $k > l : P^{\Pi}_{kl}$ is computed according to formulas in Table 1.

Theorem 4.1 (*Fixed Size DSD with Prescribed Unequal Probabilities (Construction)*). *The matrix P^{Π} is a real projection matrix, and $DSD(P^{\Pi})$ is a fixed size sampling design with first order inclusion probabilities $\pi_k = \Pi_k$, $1 \le k \le N$.*

The exact knowledge of the coefficients P^{Π}_{kl} enables a precise characterization of the sampling designs so constructed.

Corollary 4.1. *Let P^{Π} be the matrix previously constructed, and $DSD(P^{\Pi})$ the associated sampling design.*

1. *If $(k, l) \in\,]k_r, k_{r+1}[^2$ then $\pi_{kl} = 0$.*
2. *If $j \in]k_r, k_{r+1}[$, $k = k_{r+1}$, $l \in]k_{r+1}, k_{r+2}[$ then $\pi_{jkl} = 0$.*
3. *Set $B_r = [1, k_r]$. Then 1 is an eigenvalue of multiplicity r and 0 an eigenvalue of multiplicity $k_r - r - 1$ of $K_{|B_r}$: the random sample $\mathbb{S}$ has r or $r + 1$ elements in B_r ($r \le \sharp(\mathbb{S} \cap B_r) \le r + 1$).*
4. *If $k - l$ is large then $P^{\Pi}_{kl} \approx 0$, and the events $\{k \in \mathbb{S}\}$ and $\{l \in \mathbb{S}\}$ are asymptotically independent. In practice $\pi_{kl} \approx \Pi_k\Pi_l$ also holds for small values of $k - l$.*
5. *Let $r_1, \ldots, r_H$ be the set of values of $1 \le r \le n$ such that $\sum_{k=1}^{k_r}\Pi_k = r$, and set $r_0 = 0$. Then $DSD(P^{\Pi})$ is stratified with H strata $]k_{r_{h-1}}, k_{r_h}]$.*

Since the proofs of Theorem 4.1 and Corollary 4.1 are important but quite long and technical, we provide them in Appendix B and Appendix C respectively.

In the particular case of equal probability DSDs of size n ($\Pi_k = nN^{-1}$ for all k in U) and when n divides N, according to point 5, the matrix P^{Π} is a block diagonal matrix with n blocks, whose entries are nN^{-1}. The resulting DSD is thus the *1-per-stratum sampling design*. This design is known to be more efficient than systematic sampling of the population in natural order (Fuller, 1970).

In the general case, the construction of P^{Π} actually leads to a partition of the population into intervals

$$U = \bigcup_{1\le r\le n}]k_{r-1}, k_r]$$

such that, if $\mathbb{S} \sim DSD(P^{\Pi})$, then for $r = 1, \dots, n$:

- $\mathbb{S}$ has at most one point into each open interval $]k_{r-1}, k_r[$,
- $\mathbb{S}$ has at least one and at most three points into each closed interval $[k_{r-1}, k_r]$.
- $\mathbb{S}$ has at most two points into each open interval $]k_{r-1}, k_{r+1}[$,

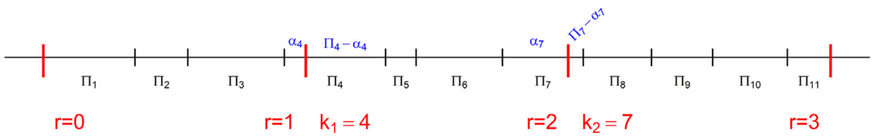
To help understand the way a sample is drawn, Fig. 1 describes the quantities used in Theorem 4.1 and shows examples of unfeasible samples for $n = 3$ and $N = 11$, giving a graphical representation of the previous properties.

We also provide an example of two matrices built by the previous method, that points out that the resulting matrices (and thus the associated determinantal sampling designs) highly depend on the way the population is ordered.

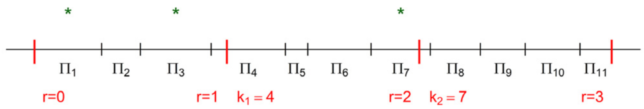
Example 4.2. Let $N = 7$, $n = 4$ and $\Pi = (\frac{1}{2}, \frac{3}{4}, \frac{3}{4}, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5})^T$ and $\Pi' = (\frac{1}{2}, \frac{1}{5}, \frac{3}{4}, \frac{4}{5}, \frac{2}{5}, \frac{3}{5}, \frac{3}{4})^T$. Observe that Π' is a permutation of Π, and that $\Pi_1 + \Pi_2 + \Pi_3 = 2$. Then

$$P^{\Pi} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} & 0 & 0 & 0 & 0 \\ \frac{1}{2\sqrt{2}} & \frac{3}{4} & -\frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{2\sqrt{2}} & -\frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{5} & \frac{\sqrt{2}}{5} & \frac{2}{5\sqrt{3}} & \frac{\sqrt{2}}{5\sqrt{3}} \\ 0 & 0 & 0 & \frac{\sqrt{2}}{5} & \frac{2}{5} & \frac{2\sqrt{2}}{5\sqrt{3}} & \frac{2}{5\sqrt{3}} \\ 0 & 0 & 0 & \frac{2}{5\sqrt{3}} & \frac{2\sqrt{2}}{5\sqrt{3}} & \frac{3}{5} & -\frac{\sqrt{2}}{5} \\ 0 & 0 & 0 & \frac{\sqrt{2}}{5\sqrt{3}} & \frac{2}{5\sqrt{3}} & -\frac{\sqrt{2}}{5} & \frac{4}{5} \end{pmatrix},$$

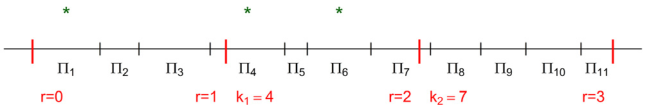
$$P^{\Pi'} = \begin{pmatrix} \frac{1}{2} & \frac{1}{\sqrt{10}} & \frac{\sqrt{3}}{2\sqrt{14}} & \frac{\sqrt{3}}{\sqrt{70}} & \frac{1}{\sqrt{35}} & \frac{1}{\sqrt{65}} & \frac{1}{2\sqrt{26}} \\ \frac{1}{\sqrt{10}} & \frac{1}{5} & \frac{\sqrt{3}}{2\sqrt{35}} & \frac{\sqrt{3}}{5\sqrt{7}} & \frac{\sqrt{2}}{5\sqrt{7}} & \frac{\sqrt{2}}{5\sqrt{13}} & \frac{1}{2\sqrt{65}} \\ \frac{\sqrt{3}}{2\sqrt{14}} & \frac{\sqrt{3}}{2\sqrt{35}} & \frac{3}{4} & -\frac{1}{2\sqrt{5}} & -\frac{1}{\sqrt{30}} & -\frac{\sqrt{7}}{\sqrt{390}} & -\frac{\sqrt{7}}{4\sqrt{39}} \\ \frac{\sqrt{3}}{\sqrt{70}} & \frac{\sqrt{3}}{5\sqrt{7}} & -\frac{1}{2\sqrt{5}} & \frac{4}{5} & -\frac{\sqrt{2}}{5\sqrt{3}} & -\frac{\sqrt{14}}{5\sqrt{39}} & -\frac{\sqrt{7}}{2\sqrt{195}} \\ \frac{1}{\sqrt{35}} & \frac{\sqrt{2}}{5\sqrt{7}} & -\frac{1}{\sqrt{30}} & -\frac{\sqrt{2}}{5\sqrt{3}} & \frac{2}{5} & \frac{2\sqrt{7}}{5\sqrt{13}} & \frac{\sqrt{7}}{\sqrt{130}} \\ \frac{1}{\sqrt{65}} & \frac{\sqrt{2}}{5\sqrt{13}} & -\frac{\sqrt{7}}{\sqrt{390}} & -\frac{\sqrt{14}}{5\sqrt{39}} & \frac{2\sqrt{7}}{5\sqrt{13}} & \frac{3}{5} & -\frac{1}{\sqrt{10}} \\ \frac{1}{2\sqrt{26}} & \frac{1}{2\sqrt{65}} & -\frac{\sqrt{7}}{4\sqrt{39}} & -\frac{\sqrt{7}}{2\sqrt{195}} & \frac{\sqrt{7}}{\sqrt{130}} & -\frac{1}{\sqrt{10}} & \frac{3}{4} \end{pmatrix}.$$



(a) Description of the quantities used in Theorem 4.1.



(b) $\pi_{137} = 0$ and $\pi_{13} = 0$ ($(1,3) \in \,]k_0, k_1[^2$).



(c) $\pi_{146} = 0$ ($1 \in]k_0, k_1[, 4 = k_1, 6 \in]k_1, k_2[$).

Fig. 1. Examples of unfeasible samples $\mathbb{S} \sim DSD(P^{\Pi})$, $n = 3$, $N = 11$.

In the general case, a possible drawback of this general construction is that some of the joint probabilities equal 0 leading to difficulties in estimating the variance. Also, in the case of equal first order inclusion probability nN^{-1}, the algorithm always provides the same matrix (whatever the reordering), which properties (of the associated sampling design) may be difficult to interpret unless n divides N, as explained before. In the next sections we provide other constructions that circumvent these drawbacks, but only in the case of equal first order inclusion probability.

For a given contracting matrix with a prescribed diagonal, we finally describe a method that leads to others matrices with the same diagonal and spectrum.

Theorem 4.2. *Let K be a contracting matrix, $(k, l) \in U^2$ such that $K_{kk} \neq K_{ll}$ for $k \neq l$ and $K_{kl} \neq 0$. Let $W_{kl}(\theta)$ be the unitary operator whose matrix relative to the canonical basis has $\cos\theta$ at the (k, k) and (l, l) entries, $-\sin\theta$ and $\sin\theta$ at the (k, l) and (k, l) entries, respectively, 1 at all other diagonal entries, and 0 at all other off-diagonal entries, where:*

$$t = \frac{2Re(K_{kl})}{K_{kk} - K_{ll}}, \; \cos\theta = \frac{1}{\sqrt{1+t^2}} \text{ and } \sin\theta = t\cos\theta.$$

Then matrix $K' = W_{kl}(\theta) K W_{kl}^T(\theta)$ has the same diagonal and spectrum as K.

Proof. In the two dimensional case, Dhillon et al. (2005) explicitly construct a (real) plane rotation Q_2 so that the diagonal vector of $A' = Q_2 A Q_2^T$ equals a prescribed vector $(a'_1, a'_2)^T$, while having the same spectrum as A. Assuming (without loss of generality) that $a_1 \leq a'_1 \leq a_2 \leq a'_2$, that is to say:

$$Q_2 \begin{pmatrix} a_1 & a_{21}^* \\ a_{21} & a_2^* \end{pmatrix} Q_2^T = \begin{pmatrix} a'_1 & * \\ * & a'_2 \end{pmatrix},$$

where

$$Q_2 = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

$$\begin{aligned} t &= \frac{Rea_{21} \pm \sqrt{(Rea_{21})^2 - (a_1 - a'_1)(a_2 - a'_1)}}{a_2 - a'_1}, \\ \cos\theta &= \frac{1}{\sqrt{1+t^2}}, \\ \sin\theta &= t\sin\theta. \end{aligned} \tag{14}$$

By letting $a_1 = a'_1 = K_{kk}$ and $a_2 = a'_2 = K_{ll} \neq K_{kk}$ in these formulas (with $Re(a_{kl}) \neq 0$), we end up with two rotation matrices: the trivial solution $Q_2 = I_2$ and a second non-trivial given in the theorem. □

Remark 4.1. The rotation $W_{kl}(\theta)$ does not change the modulus of K'_{kl}. The joint inclusion probabilities for (k, l) are then the same for $DSD(K)$ and $DSD(K')$. Nevertheless the other joint probabilities $\pi_{k.}$ or $\pi_{.l}$ might change.

4.2. (N, n)-simple determinantal sampling designs

SRS is not determinantal in general. This negative result does not however settle the question of the existence of a determinantal sampling design with the same first and second order inclusion probabilities as the SRS of size n, that is such that $\pi_k = \frac{n}{N}$ and $\pi_{kl} = \frac{n(n-1)}{N(N-1)}(k \neq l)$. In this section, we prove that such DSDs may or may not exist, depending on the values of n and N and the use of complex kernels.

Definition 4.1 (*(N, n)-Simple Designs*). Let $n \leq N$. A determinantal sampling design is (N, n)-simple if its inclusion probabilities satisfy

$$\pi_k = \frac{n}{N} \text{ and } \pi_{kl} = \frac{n(n-1)}{N(N-1)}.$$

According to Corollary 2.1 such designs (if they exist) are of fixed size, whence their kernel is a specific rank n projection. It appears that such kernels are highly connected with *Equiangular Tight Frames (ETFs)*, see Tropp (2005) and Sustik et al. (2007):

Theorem 4.3 (*(N, n)-Simple Designs and ETFs*). *$DSD(K)$ is a (N, n)-simple sampling design iff $K = \frac{n}{N}\overline{F}^T F$, where $F = (f_1, \dots, f_N)$ is an ETF of $\mathbb{C}^n$.*

The proof is given in Appendix D.

As a consequence of Theorem 4.3, a necessary and sufficient condition for the existence of ETFs would solve the problem of the existence of (N, n)-simple determinantal sampling designs. However, such a condition is not known for the moment. Nevertheless, there exist necessary conditions (recalled in Theorem E.1), and numerical studies compensate for the absence of general existence conditions (Sustik et al., 2007; Casazza et al., 2008).

Table 2 summarizes their results for $n < 9$ and $N < 100$. In the table, the symbol $\mathbb{C}$ indicates that no (N, n)-simple determinantal sampling design with real kernel exists, but that one with complex kernel does exist.

Consequently, it holds that:

1. For a given family of (non-determinantal) sampling designs, there may or may not exist a DSD with the same first and second order inclusion probabilities (for instance, there exists a $(7, 3)$-simple DSD but no $(10, 3)$-simple DSD);
2. There exists a $DSD(C)$, C complex kernel such that no $DSD(R)$, R real kernel has the same first and second order inclusion probabilities (for instance the $(57, 8)$-simple DSD can be realized only using complex kernels). This plaids in favor of using complex kernels.

Example 4.3 (*(6,3)-Simple Determinantal Sampling Design*). Let

$$K = \frac{1}{2}\begin{pmatrix} 1 & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & 1 & -\frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 1 & \frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 1 & \frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 1 & -\frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{5}} & 1 \end{pmatrix}.$$

K is a projection, and $DSD(K)$ is $(6, 3)$-simple. It is not a simple sampling as the samples $\{1, 2, 3\}$ and $\{4, 5, 6\}$ do not have the same probabilities ($\frac{1}{8}(1 - \frac{3}{5} - \frac{2}{5\sqrt{5}})$ and $\frac{1}{8}(1 - \frac{3}{5} + \frac{2}{5\sqrt{5}})$ respectively).

4.3. "Periodic" determinantal sampling designs

In this section, we construct an explicit family of fixed size, equal probability sampling designs, that exhibit some periodic behavior. The kernels involved are special *Toeplitz matrices* constructed upon primitive Nth roots of the unity.

Table 2
Existence of (N, n)-simple determinantal sampling designs, depending on the kernel type (real or complex) for $n < 9$.

n	3	3	4	4	5	5	6	6	6	7	7	7	8	8	8
N	6	7	7	13	10	11	11	16	31	14	15	28	15	29	57
	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{R}$	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{C}$

Theorem 4.4 (*A Parametric Family of Fixed Size, Equal Probability DSDs*)**.** *Let n, r, N be three integers such that $n \leq N$ and $r < N$ with r, N two relatively prime integers. Let $DSD(K^{r,N,n})$ be the determinantal sampling design with kernel $K^{r,N,n}$:*

$$\begin{cases} K_{kl}^{r,N,n} = \dfrac{1}{N}\dfrac{\sin(\frac{nr(k-l)\pi}{N})}{\sin(\frac{r(k-l)\pi}{N})} e^{\frac{ir(n-1)(k-l)\pi}{N}}, \\ K_{kk}^{r,N,n} = \dfrac{n}{N}. \end{cases}$$

$DSD(K^{r,N,n})$ is of fixed size n, and its first and second order inclusion probabilities satisfy

$$\begin{cases} \pi_k^{r,N,n} = \dfrac{n}{N}, \\ \pi_{kl}^{r,N,n} = \dfrac{n^2}{N^2} - \dfrac{1}{N^2}\dfrac{\sin^2\left(\frac{nr(k-l)\pi}{N}\right)}{\sin^2\left(\frac{r(k-l)\pi}{N}\right)} \ (k \neq l). \end{cases}$$

Proof. Let $z = e^{\frac{2i\pi r}{N}}$ be any primitive Nth root of the unity with r, N two relatively prime integers. Set $c = n/N$ and define for all $p = 0, \ldots, n-1$ the vectors $v_p = \frac{\sqrt{c}}{\sqrt{n}}\left((z^p)^1, \ldots, (z^p)^N\right)^T$. They define, by construction, an orthonormal family and $K^{r,N,n} = \sum_{p=0}^{n-1} v_p \overline{v_p}^T = V\overline{V}^T$ is a projection of rank n, where $V = (v_0, \ldots, v_{n-1})$. Its diagonal elements satisfy

$$K_{kk}^{r,N,n} = \sum_{p=0}^{n-1} v_p(k)\overline{v_p}^T(k) = n^{-1}c\sum_{p=0}^{n-1} 1 = c \quad (\text{where } v_p(k) = \frac{\sqrt{c}}{\sqrt{n}}(z^p)^k)$$

for all $k = 1, \ldots, N$. Its off-diagonal elements satisfy

$$K_{kl}^{r,N,n} = \frac{1}{N}\sum_{p=0}^{n-1} z^{(k-l)p} = \frac{1}{N}\frac{1-z^{(k-l)n}}{1-z^{(k-l)}} = \frac{1}{N}\frac{\sin\left(\frac{nr(k-l)\pi}{N}\right)}{\sin\left(\frac{r(k-l)\pi}{N}\right)} e^{\frac{ir(n-1)(k-l)\pi}{N}} \ (k \neq l).$$

The second order inclusion probabilities follow from Eq. (2). □

These designs may alternatively be described as the unitary transform of the (non random) DSD that samples exactly the first n elements of U by the $N - by - N$ unitary Discrete Fourier Transform Matrix (DFT Matrix, see for instance Rao and Yip, 2000; Dickinson and Steiglitz, 1982).

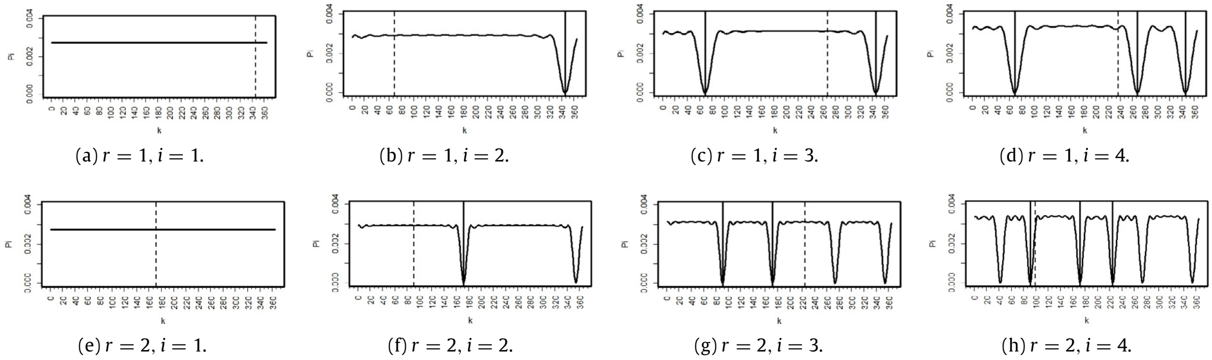
As contracting matrices form a convex set, we can form the mean of the previous matrices. Properties of the resulting DSD are given in Theorem 4.5. The result is actually true whether or not N is prime, but if N is not prime, for some values of r, $K^{r,N,n}$ might not be contracting.

Theorem 4.5. *Let n, N be two integers such that $0 < n < N$, and let $K^{N,n} = \frac{1}{N-1}\sum_{r=1}^{r=N-1} K^{r,N,n}$. Then $DSD(K^{N,n})$ is a determinantal sampling design of random size n with at least one point, such that all subsets of U of same cardinality have the same probability of occurrence. Its kernel actually satisfies*

$$\begin{cases} K_{kl}^{N,n} = \dfrac{N-n}{N(N-1)} \\ K_{kk}^{N,n} = \dfrac{n}{N} \end{cases},$$

and its first and second order inclusion probabilities satisfy

$$\begin{cases} \pi_k^{N,n} = \dfrac{n}{N}, \\ \pi_{kl}^{N,n} = \dfrac{n^2}{N^2} - \dfrac{(N-n)^2}{N^2(N-1)^2}, (k \neq l). \end{cases}$$



(a) $r=1, i=1$. (b) $r=1, i=2$. (c) $r=1, i=3$. (d) $r=1, i=4$.

(e) $r=2, i=1$. (f) $r=2, i=2$. (g) $r=2, i=3$. (h) $r=2, i=4$.

Fig. 2. Inclusion probabilities π_k^i, selected observation (vertical dotted line), previously selected observations (vertical black line) at step i for the first four steps of Algorithm 2.1 with matrices $K^{r,365,15}$: $r=1,2$ defined at Theorem 4.4.

Proof. We first compute a simpler form for the kernel.

$$\begin{aligned}
\frac{1}{N-1}\sum_{r=1}^{N-1} K_{kl}^{r,N,n} &= \frac{1}{N-1}\sum_{r=1}^{N-1}\frac{1}{N}\sum_{p=0}^{n-1}(e^{\frac{2i\pi r}{N}})^{(k-l)p} \\
&= \frac{1}{N}\frac{1}{N-1}\sum_{p=0}^{n-1}\sum_{r=1}^{N-1}(e^{\frac{2i\pi(k-l)p}{N}})^{r} \\
&= \frac{1}{N(N-1)}\left(N-1+\sum_{p=1}^{n-1}(\sum_{r=0}^{N-1} e^{\frac{2i\pi(k-l)p}{N}r}-(n-1))\right) \\
&= \frac{N-n}{N(N-1)}
\end{aligned}$$

since for $p\neq 0$,

$$\sum_{r=0}^{N-1} e^{\frac{2i\pi(k-l)p}{N}r} = \frac{1-(e^{\frac{2i\pi(k-l)p}{N}})^N}{1-e^{\frac{2i\pi(k-l)p}{N}}} = 0.$$

In particular all principal minors of the same size are equal. The characteristic polynomial of $K^{N,n}$ can be computed as a Hurwitz determinant: $p(K^{N,n}) = (1-\lambda)(\frac{n-1}{N-1}-\lambda)^{N-1}$. $K^{N,n}$ is thus a contracting matrix with 1 as maximal eigenvalue, and by Corollary 2.1, $pr(\mathbb{S}=\emptyset)=0$. As $K^{N,n}$ is not a projection, $DSD(K^{N,n})$ is not of fixed size. □

Still by Corollary 2.1, we can also compute the variance of the sample size: $var(\sharp(S)) = (N-1)^{-1}(N-n)(n-1)$. As previously recorded, this DSD provides an example of design that satisfies the conditions of Theorem 3.4 (Lemma A.1).

The kernels $K^{r,N,n}$ exhibit "approximate" periodicity. We thus expect their associated DSDs to have peculiar properties. Indeed, the periodicity of the $K^{r,N,n}$ entails some "exclusion properties" for $DSD(K^{r,N,n})$, as explained by Lemma 4.1 and shown in Fig. 2. But the DSD also exhibits a second "Poissonian" property: apart from these excluded points, the other ones are approximately independent of a given element (Fig. 3).

Lemma 4.1 (*Periodicity and Second Order Inclusion Probabilities*). *Consider $DSD(K^{r,N,n})$ and let $k<l\in U$. Pose b the rest of the Euclidean division of $r(l-k)$ by N, $r(l-k)=aN+b$. Then*

$$\pi_{kl}^{r,N,n} = \frac{n^2}{N^2}-\frac{1}{N^2}\frac{\sin^2\left(\frac{nb\pi}{N}\right)}{\sin^2\left(\frac{b\pi}{N}\right)} = \frac{1}{N^2}\left(n^2-U_{n-1}^2\left(\cos\frac{b\pi}{N}\right)\right)$$

where U_{n-1} is the Chebyshev polynomial of the second kind. In particular if $b\ll N$ or $N-b\ll N$ then $\pi_{kl}\cong 0$, otherwise $\pi_{kl}\cong\frac{n^2}{N^2}$.

Proof. The formula follows from the π periodicity of $x\mapsto \sin^2 x$ and one (of the many) definition of the Chebyshev polynomial. We then deduce the other results from the properties of this polynomial (in particular that $U_{n-1}(1)=n$). □

As an application let $U=[1,365]$ be the days of a year and assume you want a sample such that only few selected days in two consecutive weeks have the same weekdays. Then you can choose $r=52$ (weeks/year). Of course a systematic sample

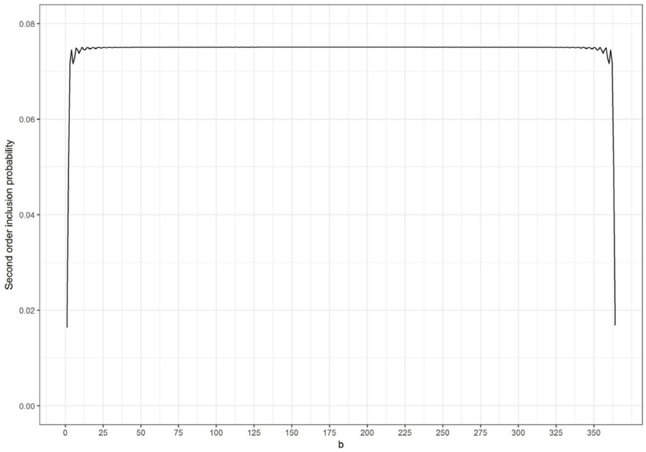


Fig. 3. $DSD(K^{52,365,100})$, second order inclusion probabilities π_{kl} according to $b \in [1:364]$.

Table 3
Trying to avoid consecutive weekdays (gray).

	$DSD(K^{52,365,100})$										SRS									
	Weeks										Weeks									
Weekdays	1	2	3	4	5	6	7	8	9	10	1	2	3	4	5	6	7	8	9	10
Monday		⋆				⋆		⋆				⋆					⋆	⋆		
Tuesday		⋆			⋆			⋆									⋆			
Wednesday		⋆					⋆						⋆	⋆	⋆			⋆		
Thursday					⋆			⋆	⋆		⋆						⋆	⋆		
Friday			⋆	⋆						⋆				⋆				⋆		
Saturday	⋆						⋆			⋆		⋆		⋆	⋆	⋆	⋆		⋆	
Sunday			⋆			⋆									⋆		⋆			

Table 4
RSD of the HT-estimators of the French electricity consummation in 2015 - $DSD(K^{52,365,100})$, SRS and Poisson sampling.

$DSD(K^{52,365,100})$	SRS(365,100)	$DSD(D_{100/365})$ = Poisson sampling
0,738	1,658	8,680

would select perfectly such a sample but with very low entropy, whereas the sample drawn from $DSD(K^{52,365,n})$ remains largely random (high entropy). This is shown by Fig. 2 for $n = 15$ ($r = 1, 2$), and even more evident for $n = 100$ ($r = 52$) (Fig. 3). Indeed in this second case except for two days k, l such that the rest b of the Euclidean division of $r(l-k)$ by N is less than three, $\pi_{k,l} \cong 0,075 = n^2N^{-2}$ which is the second order inclusion probability of Poisson sampling (Fig. 3).

Table 3 plots the first ten weeks of a year for two samples: one drawn according to a $DSD(K^{52,365,100})$ by Algorithm 2.1 and the second one according to a SRS.

We finally compare the efficiency of this DSD to SRS and Poisson sampling on periodic, real data. To this end, we consider the daily values of French electricity consummation in 2015, and compute the variance of the associated Horvitz–Thompson estimators of the annual consummation by Eq. (6). Table 4 shows the corresponding relative standard deviations (RSDs), where $RSD = \frac{\sqrt{v(\hat{t}_y)}}{t_y}$. According to this criteria, taking into account the periodic structure of the data with the $DSD(K^{52,365,100})$ leads to a more accurate estimator than those based on SRS and Poisson sampling (while keeping a high entropy).

5. Optimal strategy

The parametric form of determinantal sampling designs, along with a closed formula for the mean square error (Eq. (11)) allows at least theoretically to search for a pair $(DSD(K), w)$, minimizing the sum of the MSEs for a set of auxiliary variables. This pair (sampling design, vector of weights) is sometimes called a *strategy* in the literature Hájek and Dupac (1981).

It happens that while promising on the theoretical level, solving the optimization problem becomes rapidly unfeasible practically, notably due to the curse of dimensionality. Thus, after examining the theoretical issues in the next section, we then give some practical heuristics.

Table 5
Minimizations problems and their parameter spaces.

Problems	Parameter spaces	Names
DSD of size μ, free weights	$\Theta = \Theta^{\mu}_{c,d} \times \mathbb{R}^N$	Problem P_1
DSD of size μ, Horvitz–Thompson	$\Theta = \{(K, w) \mid K \in \Theta^{\mu}_{c,d}, w_k = K_{kk}^{-1}\}$	Problem P_2
DSD of size μ, fixed weights w	$\Theta = \Theta^{\mu}_{c,d} \times \{w\}$	Problem P_3
DSD with $\pi = \Pi$, fixed weights w	$\Theta = \Theta^{\Pi} \times \{w\}$	Problem P_4
DSD with $\pi = \Pi$, free weights	$\Theta = \Theta^{\Pi} \times \mathbb{R}^N$	Problem P_5
Fixed *DSD*(K), free weights	$\Theta = \{K\} \times \mathbb{R}^N$	Problem P_6

5.1. A generic optimization problem

It is common in the literature to search for sampling designs providing only *representative* or *balanced* samples for a set of Q auxiliary variables, where a sample $\mathbb{S}$ is representative for x if $\sum_{k\in\mathbb{S}}\pi_k^{-1}x_k = t_x$. The underlying idea is that such estimators should perform well on variables of interest correlated with the auxiliary variables. Deville and Tillé (2004) provide a general method, called the cube method, for selecting approximately balanced samples with fixed first order inclusion probabilities and any number of auxiliary variables.

Regarding DSDs, our approach to provide approximately balanced samples is to interpret representativity as follows: let us call a strategy $(DSD(K), w)$ representative for x if $\mathrm{MSE}(\hat{t}_{xw}) = 0$. Note that in our definition, we do not restrict to the Horvitz–Thompson estimator but consider the weights of the linear estimator as a parameter.

Obviously, attaining 0 for the MSE is impossible in most cases, but we can still minimize this MSE. This approach is for instance considered in Fuller (2009) for the Horvitz–Thompson estimator.

Let $x^1, \ldots, x^Q$ be Q auxiliary variables and Θ be a parameter space. From Eq. (6) we define the following generic optimization problem, where $C(K, w)$ is an objective function equal to $\sum_{q=1}^{Q}\mathrm{MSE}(\hat{t}_{x^q w})$ (up to a constant).

Problem 5.1 (*The Generic Minimization Problem*)**.** Find

$$\underset{(K,w)\in\Theta}{\arg\min} C(K, w) = \sum_{q=1}^{Q}(z^q)^T((I_N - K) * \overline{K})z^q + [e^T(K * I_N)z^q - e^T x^q]^2$$

where $z^q = w * x^q$.

We then adjust the parameter space to address the following cases which are of interest for survey sampling: free problem (with fixed average size of the random samples), optimal DSD for the Horvitz–Thompson estimator, optimal DSD for fixed weights, optimal DSD with prescribed first order inclusion probabilities and either free or fixed weights, optimal weights for a fixed DSD. We study more precisely these optimization problems (notably the objective function) in Appendix F (see Table 5)

where

$$\Theta^{\mu}_{c,d} = \{K \mid 0 \le K \le I_N,\ c \le K_{kk} \le d,\ Trace(K) = \mu\}$$
$$\Theta^{\Pi} = \{K^{\Pi} \mid 0 \le K^{\Pi} \le I_N,\ diag(K^{\Pi}) = \Pi\}.$$

We first solve the free problem with sole constraint $Trace(K) = n \ge Q$ (where Q is the number of auxiliary variables). To this end, we consider the non-random sampling design that samples the n first elements of the population. The set of equations $\sum_{1\le k\le n} w_k x_k^q = t_x^q$ is consistent by the Rouché–Cappeli Theorem (Capelli, 1892), and if w is any solution then the associated strategy satisfies $\mathrm{MSE}(\hat{t}_{x^q w}) = 0$ for all $1 \le q \le Q$. This "free problem" thus clearly exhibits the drawback of overfitting. This is the reason people usually either work with the Horvitz–Thompson estimator, or fixed first order inclusion probabilities, or both. Another possibility is to use the constrained parameter space $\Theta^{\mu}_{c,d}$ with $0 < c$ or $d < 1$.

Apart from this trivial case, among these problems, only the last one (Problem P_6, optimal weights) admits an explicit solution. More generally, it admits an explicit solution for any sampling design whose joint probabilities are perfectly known, including the determinantal sampling designs. For such a given sampling design, the solution leads to some calibration estimator, where, unlike Deville and Särndal (1992), the weights do not depend on the sample.

Theorem 5.1 (*Optimal Weights*)**.** *Let $\mathcal{P}$ be a sampling design whose first and second order probabilities are π_k, π_{kl} ($\pi_{kk} = \pi_k$) and $x^1, \ldots, x^Q$ be Q vectors of auxiliary variables. The linear homogeneous estimators that minimize the sum of the Q MSEs correspond to weights w^{opt} in the affine subspace:*

$$w^{opt} \in \left(\left(\sum_{q=1}^{Q} x^q {x^q}^T\right) * \Omega\right)^{\dagger}\left(\left(\sum_{q=1}^{Q} t_{x^q} x^q\right) * \pi\right) + \ker\left(\left(\sum_{q=1}^{Q} x^q {x^q}^T\right) * \Omega\right)$$

where $\Omega = (\pi_{kl})$ is the joint probability matrix of $\mathcal{P}$, π the vector of first order inclusion probabilities, and $M^{\dagger}$ the Moore–Penrose inverse of a matrix M.

Proof.

$$\begin{aligned}\sum_{q=1}^{Q}\mathrm{MSE}(\hat{t}_{x^q w}) &= \sum_{q=1}^{Q}\sum_{k,l}\sum w_k w_l x_k^q x_l^q \Delta_{kl} + \sum_{q=1}^{Q}\left[\sum_{k\in U}(w_k\pi_k - 1)x_k^q\right]^2 \\ &= \sum_{q=1}^{Q}\sum_{k,l} w_k w_l x_k^q x_l^q \Delta_{kl} + \sum_{q=1}^{Q}\sum_{k,l}(w_k\pi_k - 1)x_k^q(w_l\pi_l - 1)x_l^q \\ &= \sum_{q=1}^{Q}\sum_{k,l} w_k w_l x_k^q x_l^q \pi_{kl} - 2\sum_{q=1}^{Q}\sum_{k} w_k t_{x^q}\pi_k x_k^q + \sum_{q=1}^{Q} t_{x^q}^2 \\ &= w^T A w - 2w^T B + C\end{aligned}$$

where $A = (\sum_{q=1}^{Q} x^q x^{q^T}) * \Omega$ and $B = (\sum_{q=1}^{Q} t_{x^q} x^q) * \pi$. Minimizing $\sum_{q=1}^{Q}\mathrm{MSE}(\hat{t}_{x^q w})$ is thus a classical problem of unconstrained quadratic programming. Since $w \mapsto w^T A w - 2w^T B + C$ is nonnegative then $B \in \ker(A)^{\perp}$, and since A is Hermitian (and spaces are finite dimensional), $B \in Im(\overline{A}^T)^{\perp\perp} = Im(A)$, and thus $B = AV$ for some vector V. Then for any w, $w^T A w - 2w^T B + C = (w - V)^T A(w - V) + C - V^T B$ which is minimal for $w^{opt} \in V + \ker(A)$. Finally, as $B = AV$ and $AA^{\dagger}A = A$ then $V' = A^{\dagger}B$ satisfies $B = AV'$. $\square$

Corollary 5.1. *Let $x^1, \ldots, x^Q$ be Q auxiliary variables.*

- *If $Q = 1$ and $x_k \neq 0$ for all $k \in U$ then $w^{opt} = \left((nx_1)^{-1}t_x, \ldots, (nx_N)^{-1}t_x\right)^T$ is an optimal vector of weights for any sampling design $\mathcal{P}$ of fixed size n.*
- *If w^{opt} is an optimal vector of weights for $\mathcal{P}$ then*

$$\sum_{q=1}^{Q}\mathrm{MSE}(\hat{t}_{x^q w}) = \sum_{q=1}^{Q} t_{x^q}^2 - \left(\left(\sum_{q=1}^{Q} t_{x^q} x^q\right) * \pi\right)^T \left(\left(\sum_{q=1}^{Q} x^q x^{q^T}\right) * \Omega\right)^{\dagger} \left(\left(\sum_{q=1}^{Q} t_{x^q} x^q\right) * \pi\right). \tag{15}$$

- *If $\mathcal{P} = DSD(K)$, then $\Omega = (I_N - K) * K + diag(K)diag(K)^T$.*

The optimal weights given by Theorem 5.1 may be far from the values $\pi_k^{-1} = K_{kk}^{-1}$ (indeed, in the first point of Corollary 5.1, they are independent of the first order inclusion probabilities), therefore leading to a possibly highly biased estimator of t_y. A classical way to overcome this issue is either to add a penalization term or a constraint in the minimization problem. For instance, adding the quadratic constraints $(\pi_k w_k - 1)^2 \leq cn^{-1}$ for some positive constant c (and $n = \sum_{k\in U}\pi_k \to \infty$) leads to optimal weights ensuring that the first assumption $\sum_{k\in U} K_{kk}\left(1 - (K_{kk}w_k)^{-1}\right)^2 = O(1)$ of Theorem 3.2 is satisfied.

While appealing, all the other problems are hardly tractable, both theoretically and in practice. On the bright side, we can rewrite all these problems as *semidefinite optimization problems* (see Appendix F), where the parameter space is a *projected spectrahedron*, that is the projection of the intersection of the cone of positive semidefinite matrices and an affine space. And semidefinite optimization (in the convex and linear setting) has become a major field of optimization theory recently (see for instance Blekherman et al., 2013; Vandenberghe and Boyd, 1996). But a significant downside is that the objective functions of the other problems are not convex. And while efficient algorithms exist in the case of a strictly convex objective function, problems are extremely difficult otherwise. For instance, even linear semidefinite optimization is usually NP-hard (Theorem G.1).

In the case of minimization of the EMQ for nonnegative variables x^q, when the weights and the first order inclusion probabilities are fixed (Problem P_4, case usually considered in survey sampling), a striking feature occurs. The objective function is actually concave (see Appendix F), and therefore the solutions are extreme points of the spectrahedron. One of the difficulties in this case is that the extreme points of spectrahedra do not generally admit a simple characterization. Indeed, the problem of deciding whether a given matrix is an extreme point of a given spectrahedron is NP-hard for many spectrahedra. This is for instance the case for the *elliptope of correlation matrices*. Problem P_4 for this particular spectrahedra is studied in Appendix G. A second major issue in the concave case is that many extreme points may be arguments of local/non-global minima, and classical algorithms will be trapped in these local minima, especially as N becomes large.

In practice, existing semidefinite optimization algorithms fail to produce globally optimal solutions when N is large. Indeed, we have seen in Section 4 that producing a projection element in Θ_K^{Π} (projections are extreme points, but not all extreme points are projections) is in itself a difficult task.

In the following, based on algorithmic minimization results for small N ($N \leq 40$) and the theoretical results of the article, we present empirical algorithms to solve Problems P_2 to P_5. The performances of the empirical algorithms are presented in Section 5.3.

5.2. Empirical algorithms

We performed nonlinear semidefinite optimization for Problems P_2, P_3 and P_4 using specific semidefinite optimization algorithms (Polyak, 1992; Tütüncü et al., 2001), for various numbers μ (average sample size), vectors of inclusion probabilities and vectors of weights, integers $N \leq 40$ (size of the population) and auxiliary variables. Our empirical conclusions are triple. When $\mu = n$ is an integer, the minimizer is always a projection. Therefore, in our search for empirical algorithms, we mainly work with projection kernels. When $\Pi_k = \frac{n}{N}$ and n divides N, and for one auxiliary variable only ($q = 1$), the optimal determinantal sampling design for Problem P_4 is the 1-per-stratum sampling design. However, for more than one auxiliary variable ($Q > 1$) the solution is generally not stratified (for $\Pi_k = \frac{n}{N}$ and n divides N).

These results along with the reading of Eq. (8) and Corollary 4.1 suggest that Algorithm 5.1 should produce a low value of $C(K)$ for Problem P_4, at least for 1 nonnegative auxiliary variable. We consider fixed inclusion probabilities Π_k such that $\sum_{k\in U} \Pi_k = n$ and a fixed vector of nonnegative weights w.

Algorithm 5.1 (*Ranking and Projecting Algorithm (RPA, Problem P_4)*)**.** Perform the following steps:

1. Perform a Q multi-dimensional ranking algorithm on the variables $wx^1, \ldots, wk^Q$. This produces a permutation σ on the population. Relabel this new population $\{1, \ldots, N\}$ and update the vector Π accordingly.
2. Construct the projection matrix P^{Π} with diagonal as in Theorem 4.1.

The resulting strategy is $(DSD(P^{\Pi}), w)$ (on the ordered population).

In case of one nonnegative auxiliary variable ($Q = 1$), Algorithm 5.1 actually produces a zero-variance estimator for an auxiliary variable x such that after reordering the population by $w * x$, for all integers $1 \leq r \leq n$, $\sum_{k=1}^{k_r} \Pi_k = r$ for some k_r (Theorem 3.1 and Corollary 4.1). In this case, the population is actually divided into n strata with P^{Π} a projection matrix of rank 1 on each strata. And if this is not the case, then we can nonetheless divide the population into (at most $2n$) strata such that P^{Π} is a rank 1 contraction matrix on each stratum. Therefore, restricted to each stratum, the solution will achieve the minimal variance (Theorem G.2).

The situation is drastically different in case of many auxiliary variables ($Q > 1$) for $\mathbb{R}^Q$ is not globally ordered in this case. The choice of the ranking method is then crucial (see Section 5.3), and the matrix obtained by Algorithm 5.1 may be far from optimal, notably when Q is large. We therefore propose a greedy algorithm, based on the 2-by-2 rotations $W_{kl}(\theta)$ of Theorem 4.2, to improve a given $DSD(K)$. It thus needs a non-trivial initialization matrix $K^0 \in \Theta^{\Pi}$ (for instance, $K^0 = P^{\Pi}$ kernel obtained by Algorithm 5.1).

Algorithm 5.2 (*Constrained Unitary Transform Algorithm (CUTA, Problem P_4)*)**.** Using the notations of Theorem 4.2, for $r = 1$ to R (fixed in advance) do:

1. For each (k, l) in U^2 such that $\Pi_k \neq \Pi_l$ compute the angle θ_{kl}^r of the rotations of Theorem 4.2;
2. Define $(k^r, l^r) = \operatorname{argmin}_{(k,l)\in U^2} C(W_{kl}(\theta_{kl}^r)K^{r-1}W_{kl}^T(\theta_{kl}^r), w)$;
3. Set $K^r = W_{k^r l^r}(\theta_{k^r l^r}^r)K^{r-1}W_{k^r l^r}^T(\theta_{k^r l^r}^r)$ and $r = r + 1$.

If K^0 is a projection matrix, then all matrices K^r are projections.

We now combine Algorithm 5.1 and Theorem 5.1 in the following two steps iterative method, that searches for the best strategy with given first order inclusion probabilities (Problem P_5). We need an initialization strategy $(K^0, w^0) \in \Theta^{\Pi} \times \mathbb{R}^N$ on the population $U = U^0$ (for instance, $K^0 = D_{\Pi}$, $K^0 = P^{\Pi^{\sigma}}$ or $K^0 = K^R$ of Algorithm 5.2, and $w^0 = \Pi^{-1}$ or $w^0 = w^{opt}$).

Algorithm 5.3 (*Iterated RPA for Weights (IRPAW, Problem P_5)*)**.** For $r = 1$ to R (fixed in advance) do:

1. Perform Algorithm 5.1 with weights w^{r-1} and set $K^r = P^{\Pi}$, U^r as the kernel and (reordered) population obtained by the algorithm;
2. Find (one of) the associated optimal weights w^r by Theorem 5.1 and set $r = r + 1$.

Use the strategy $(DSD(K^{r_1}), w^{r_1})$ on population U^{r_1}, where $r_1 = \operatorname{argmin}_{0\leq r\leq R} C(K^r, w^r)$.

Finally, we propose two algorithms to solve Problems P_2 and P_3 (free DSD, Horvitz–Thompson or fixed weights). Algorithm 5.4 is similar to Algorithm 5.3 in that in combines the RPA (Ranking and Projecting Algorithm 5.1) with the use of optimal weights. But now the optimal weights serve to define the first order inclusion probabilities for the next step of the algorithm. In the proposed algorithm, we pose $\Pi \propto 1/w^{opt}$ and stop the algorithm if these are not inclusion probabilities. However, we could make other choices to take into account values outside [0, 1]. We start with an initializing vector $\alpha^0 = 1/\Pi^0$, where Π^0 is a vector of inclusion probabilities summing to n.

Algorithm 5.4 (*Iterated RPA for Kernels (IRPAK, Problems P_2 and P_3)*)**.** For $r = 1$ to R (fixed in advance) do:

1. Set $\Pi_k^r = n\left(\alpha_k^{(r-1)} \sum_{k\in U} \frac{1}{\alpha_k^{(r-1)}}\right)^{-1}$;
2. Perform Algorithm 5.1 with $w^r = 1/\Pi^r$ (Problem P_2) or $w^r = w$ (Problem P_3), and set $K^r = P^{\Pi^r}$, U^r as the kernel and (reordered) population obtained by the algorithm;
3. Find the associated optimal weights α^r by Theorem 5.1;

Table 6
Implementing Algorithms 5.1 to 5.5.

			Benchmark							
Algorithm	π_k	w_k	DSD			Cube	PPS SYS		
			Ranking 1	Ranking 2	Ranking 3		Ranking 1	Ranking 2	Ranking 3
5.1	Π_k	Π_k^{-1}	2957	2221	1355	1170	6052	3580	1540
5.2	Π_k	Π_k^{-1}	1192	1288	1183	–	-	–	–
5.3	Π_k	w_k^{opt}	457	617	261	380	255	202	19
5.4	Π_k^{opt}	$\Pi_k^{opt^{-1}}$	1064	2221	41.6	291	–	-	–
5.5(a)	Π_k^{opt}	$\Pi_k^{opt^{-1}}$	647	1999	41.6	–	-	–	–
5.5(b)	Π_k^{opt}	$\Pi_k^{opt^{-1}}$	417	1934	39.6	–	-	–	–

4. If α^r is positive and $0 \leq n\left(\alpha_k^r \sum_{k\in U} \frac{1}{\alpha_k^r}\right)^{-1} \leq 1$ then set $r = r+1$, otherwise set $(K^s, w^s) = (K^r, w^r)$ for $r+1 \leq s \leq R$ and $r = R+1$.

Use the strategy $(DSD(K^{r_1}), w^{r_1})$ on population U^{r_1}, where $r_1 = \text{argmin}_{0\leq r\leq R}\, C(K^r, w^r)$.

Finally, in a spirit similar to Algorithm 5.2, we can use unitary matrices to search for an optimal DSD with free first order inclusion probabilities but fixed average number of points (Problems P_2 and P_3). To this end, let $W(\rho)$ be a small-dimensional parametric family of unitary matrices (for instance 2-by-2 rotations for certain indexes k, l, or Householder matrices) with $W(0) = I_N$. Let K^0 be a given contracting matrix and let $w_k^0 = 1/K_{kk}^0$ (Problem 2) or $w^0 = w$ (Problem 3).

Algorithm 5.5 (*Free Unitary Transform Algorithm (FUTA, Problems P_2 and P_3)*). For $r = 1$ to R (fixed in advance) do:

1. Compute $\rho^r = \text{argmin}_\rho C(W(\rho)K^{r-1}\overline{W(\rho)}^T, w^{r-1})$;
2. Set $K^r = W(\rho)K^{r-1}\overline{W(\rho)}^T$, $w_k^r = 1/K_{kk}^r$ (Problem 2) or $w^r = w$ (Problem 3);
3. Set $r = r+1$.

Use the strategy $(DSD(K^R), w^R)$.

We conclude this section by comparing the different algorithms. The RPA produces one DSD without relying on actual optimization procedures. Its good performances (see Section 5.3) are explained by the good properties of the matrix defined by Theorem 4.1 for ranked variables, but highly depend on the ranking procedure (Table 6). The Iterated RPAs have in common that they construct a panel of strategies (from RPA) that may be very different at each step, and then choose the best strategy in this panel. There is no reason that the objective function decreases between two steps, but these algorithms explore very different regions of the feasible set. By contrast, the algorithms CUTA and FUTA based on unitary transforms minimize the given criterion at each step of the algorithms. However, CUTA only changes the strategy locally, thus exploring a small region around the initializing strategy. And to work, FUTA needs a small-dimensional parametric family of unitary matrices, and therefore also explores a low-dimensional subset of the parameter space. The resulting improvement in the objective function may be small.

Therefore we advocate for combinations of the previous algorithms: an "exploring" algorithm followed by "local improvements". This is illustrated in Section 5.3.

5.3. Application

The samples for the French household surveys are drawn according to a two-stage cluster sampling design. We consider a simplified version of the first stage. The population consists of geographical Primary Units, and the first order inclusion probabilities of the design are proportional to their numbers of inhabitants: $\Pi_k = nx_k^1/t_{x^1}$. The sampling design is stratified by region and aims at being *representative* for a set of two auxiliary variables: the total amount of unemployment benefit (x^2), and the total amount of taxable incomes (x^3). These variables are normalized so that $t_{x^2} = t_{x^3} = 1000$. We calibrate our studies on the example of Region *Basse-Normandie*, where $N = 148$ and $n = 14$. We aim at finding the optimal strategies (K, w) with w either free or equal to $diag(K)^{-1}$ (Horvitz–Thompson estimator) that minimize the criterion:

$$C(K, w) = \text{MSE}(\hat{t}_{x^2 w}) + \text{MSE}(\hat{t}_{x^3 w}).$$

To do so we implement the previous algorithms successively. Table 6 provides the values of the objective criterion after each algorithm.

All the estimators considered in Table 6 are Horvitz–Thompson estimators except for the third line where the weights are the optimal weights. Such non-Horvitz–Thompson estimators may perform extremely well on the auxiliary variables that served for the optimization procedure, but poorly on another variable of interest due to overlearning and a possibly high bias.

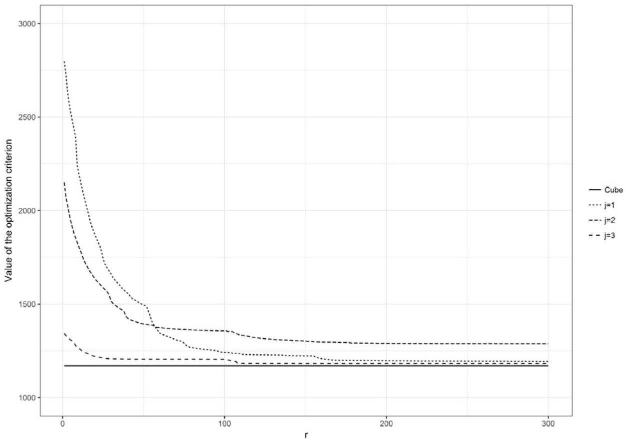


Fig. 4. Algorithm 5.2: Evolution of the optimization criterion, DSDs perform as good as the cube. The curves $j = 1, 2, 3$ correspond respectively to the 3 ranking methods: by x_k^2/Π_k ($j = 1$), by $(x_k^2 + x_k^3)/\Pi_k$ ($j = 2$) and by the Hamilton path ($j = 3$).

We now describe the precise implementation of the algorithms.

- Algorithm 5.1: we set $\pi_k = \Pi_k$ and $w_k = \Pi_k^{-1}$. Since there is no total order in $\mathbb{R}^2$, we test three different ranking methods: by x_k^2/Π_k ($j = 1$), by $(x_k^2 + x_k^3)/\Pi_k$ ($j = 2$) and by an Hamilton path going through the points whose coordinates are $(x_k^2/\Pi_k, x_k^3/\Pi_k)$ ($j = 3$). The benchmark consists of two other popular sampling designs: the fixed size cube method balanced on (x^1, x^2), and the Systematic PPS Sampling (ordered by the same three ranking methods). The cube method performs best than the DSDs and than PPS. The results for the former and the latter highly depend on the ranking method, the Hamilton path being the best.
- Algorithm 5.2: we use the matrices obtained at the previous step to initialize the algorithm and set $R = 300$ and $w_k = \Pi_k^{-1}$. The choice of R is arbitrary but large enough to exhibit convergence of the algorithm (see Fig. 4). This algorithm is not relevant for the benchmark methods. Whether it be for ranking method $j = 1$ or $j = 3$, the DSDs now perform as good as the cube (Fig. 4).
- Algorithm 5.3: we set $R = 100$ and observe that for all the scenarii the minimal value of the objective function stabilizes rapidly. The initializing matrix K^0 is the kernel obtained at the previous step and $w^0 = w^{opt}$ is the optimal vector of weights for this kernel (for each ranking method). For the benchmark methods, we only compute the optimal weights by Theorem 5.1, Ω being estimated by Monte Carlo methods. The very good results for the PPS method 3 is due to the very low entropy of this design leading to a low number of different feasible samples. The conditions for finding weights leading to balanced samples are close to those of the Rouché–Cappeli Theorem. Within high entropy sampling designs, DSD with ranking method 3 now performs better than the cube. The results for the ranking method 2 are less good due to the fact that, with this method, ranking using $w^0 = w^{opt}$ has led to the same ordered population than with $w = \Pi^{-1}$, so that the algorithm stabilized at the first step.
- Algorithm 5.4: we set $R = 100$ and initialize the algorithm by $\alpha = 1/\Pi$. For the benchmark methods we do as Algorithm 5.4 but with the cube method or PPS replacing Algorithm 5.1. For all the scenarii, after fluctuations (and high decrease for $j = 3$), the function $r \mapsto c(K^r, w^r)$ becomes non-decreasing. The result for the DSD ranked by the Hamilton path is now close to 0 and *much* better than the cube equivalent. Unfortunately the method is less efficient for the two other DSDs: the objective function only decreases at step 1 for $j = 1$ and never for $j = 2$. The choice of the ranking method is then crucial for the DSDs. Finally the method does not work well for the PPS since at the very first step of the algorithm some of the weights are negative.
- Algorithm 5.5: we set $R = 1$ and K^0 is the kernel obtained at the previous step and $w_k^0 = 1/K_{kk}^0$. We test two parametric families of unitary matrices: Householder matrices and two-by-two rotations (5.5(a) and 5.5(b)). This algorithm is not relevant for the benchmark methods. For the Householder method, ρ is a unitary vector of $\mathbb{R}^{148}$ or $\mathbb{C}^{148}$ and $W(\rho) = I_{148} - 2\rho\overline{\rho}^T$. For the two-by-two rotations method, we set $(k_{(i)}, l_{(i)})$ the ith most contributing term of the objective criterion $C(K^0, w^0)$, written as a sum on $(k, l) \in U^2$ of positive terms (see Fig. 5). We set $\rho = (\theta_1, \ldots, \theta_i, \ldots, \theta_{150})$ and $W(\rho) = \prod_{i=150}^{i=1} W_{k_{(i)},l_{(i)}}(\theta_i)$. The two-by-two rotations method performs better. However its computation time is very much larger than the Householder method. For the Householder method the results are similar whether ρ is real or complex.

Finally, we illustrate in Fig. 6(a) and 6(b) the Gaussian behavior of the Horvitz–Thompson estimators based on DSDs, using 10000 samples drawn by Algorithm 2.1 using the DSD that results from the implementation of Algorithm 5.4.

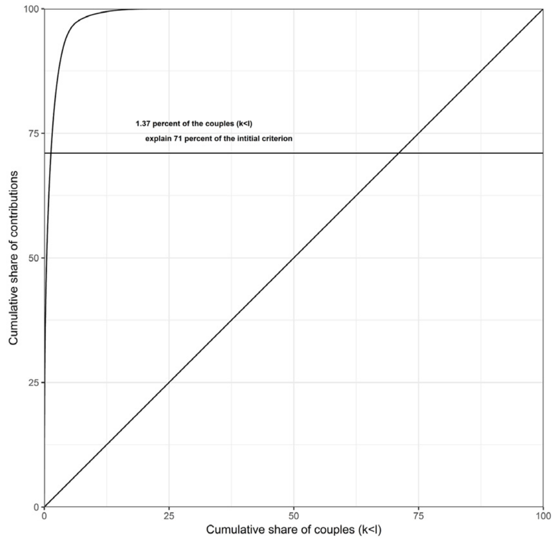
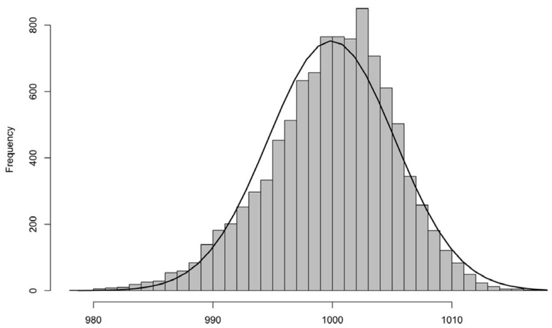
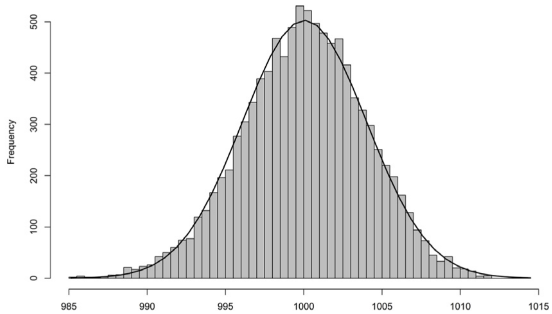


Fig. 5. Initializing Algorithm 5.5(a): Lorenz curve of the contribution to the initial criterion. From Eq. (9), we compute for each couple (k, l) its contribution to the criterion as $100 * \sum_{q=1}^{q=Q} (w_k^0 x_k^q - w_l^0 x_l^q)^2 |K_{kl}^0|^2 / C(K^0, w^0)$. We then build the associated Lorenz curve. According to this curve, $1.37\% (= 150/(148 * 147/2))$ of the couples (k, l) of U^2, $k < l$, explain 71% of the initial criterion.



(a) Law of $\hat{t}_{x^2 HT}$.



(b) Law of $\hat{t}_{x^3 HT}$.

Fig. 6. Ranking method 3, Algorithm 5.4: law of $\hat{t}_{x^2 HT}$ and $\hat{t}_{x^3 HT}$.

6. Conclusion and perspectives

This article provides theoretical and empirical evidence that determinantal sampling designs may be useful in sampling theory. Indeed, they offer the possibility to use powerful results from different domains of mathematics (probability theory, frame theory, matrix theory, semidefinite optimization). More practically, while the DSDs are indexed by the very large family of Hermitian contracting matrix, we show that the construction of a single DSD (Theorem 4.1) can be of real practical interest (Section 5). Nevertheless various directions of research remain to be further explored. The major issue concerns the possibility to use practically DSDs on large population size N. Indeed, while their theoretical asymptotic relevance is exhibited in Section 3, our parametric family of DSDs is of size N^2 and sparse descriptions have to be found. In the particular case of Theorem 4.1, we could work directly on the rank one decompositions of the matrices and compute the rotations on these decompositions. This has also the benefit of avoiding Step 1 of the sampling algorithm (that precisely consists in finding the rank one decomposition). More generally, the computation time of the sampling algorithm (with or without Step 1) should be compared to classical algorithms. Specific to the optimization procedures is the relevance of multidimensional ranking algorithms for many auxiliary variables (based for instance on clustering or multidimensional Hamiltonian paths). Another possibility would be to use other constructions than Theorem 4.1 in the empirical algorithms or to use more efficient unitary matrices. Also, completely different algorithms could be designed.

Acknowledgments

This research has been conducted as part of the project Labex MME-DII (ANR11-LBX-0023-01). We acknowledge Xu Kai from the French National School of Statistics and Economic Administration (ENSAE) for his helpful implementation of semidefinite optimization with MATLAB. We also thank Martin Brady from the Australian Bureau of Statistics (ABS) and Antoine Chambaz from Modal'X (Université Paris Descartes) for their useful comments. Finally, we would like to acknowledge that this article has greatly benefited from insightful comments and very relevant suggestions made by one of the reviewers. We are very grateful.

Appendix A. $DSD(K^{N,n})$ satisfies the assumptions of Theorem 3.4

Let $DSD(K)$ be a sampling design such that $0 < K_{kk} = d < 1$ for all $k \in U$ and $|K_{kl}|^2 = c$ for all $k \neq l$. Let also y be a bounded variable, $0 < a \leq |y| \leq b$. Set $s^2 = \sum_{k\in U_N} y_k^2(K_{kk}^{-1} - 1)$, $r = \sum_{k\in U_N}\sum_{l<k} \frac{y_k y_l}{\pi_k \pi_l}|K_{kl}|^2$ and $C = \sup_{k\in U_N}|\pi_k^{-1} y_k|$.

Then $s^2 \geq \sum k \in U_N a^2(d^{-1}-1) = Na^2(d^{-1}-1)$. Second, $|r| \leq \sum_{k\in U_N}\sum_{l<k}\frac{b^2}{d^2}c = \frac{b^2 c}{d^2}\frac{N(N-1)}{2}$, so that $s^{-2}|r| \leq \frac{b^2}{a^2}\frac{c}{d(1-d)}\frac{N-1}{2}$. And third, $C = \sup_{k\in U_N}|\pi_k^{-1}y_k| \leq bd^{-1}$ so that $s^{-2}C^2 \leq \frac{b^2}{a^2}\frac{d^2 N}{d(1-d)n^2}$.

We use these computations to deduce:

Lemma A.1. *Let $DSD(K^{N,n})$ be the determinantal sampling design defined in Theorem 4.5 and y be a bounded variable such that $0 < a \leq |y| \leq b$. If $0 < \alpha \leq nN^{-1} \leq \beta < 1$ then $s^2 \to \infty$, $r = o(s^2)$ and $C = o(s)$.*

Proof. As $K_{kk}^{N,n} = \frac{n}{N}$ for all $k \in U$ and $|K_{kl}^{N,n}|^2 = \frac{(N-n)^2}{N^2(N-1)^2}$ for all $k \neq l$ then

$$\begin{aligned}
s^2 &\geq \frac{N(N-n)}{n}a^2\\
|r| &\leq \frac{b^2(N-n)^2}{n^2(N-1)^2}\frac{N(N-1)}{2}\\
&\leq \frac{b^2(N-n)^2 n^2(N-1)}{2N^2}\\
s^{-2}|r| &\leq \frac{b^2}{a^2}\frac{N-n}{n(N-1)}\\
C &\leq \frac{bN}{n}\\
s^{-2}C^2 &\leq \frac{b^2}{a^2}\frac{N}{n(N-n)}
\end{aligned}$$

under the assumption $0 < \alpha \leq nN^{-1} \leq \beta < 1$ we deduce that

$$\begin{aligned}
s^2 &\geq \frac{1-\beta}{\beta}Na^2 \to \infty\\
s^{-2}|r| &\leq \frac{b^2}{a^2}(\alpha N)^{-1} \to 0\\
s^{-2}C^2 &\leq \frac{b^2}{a^2}(\alpha(1-\beta)N)^{-1} \to 0. \quad \square
\end{aligned}$$

Similar computations also show that Poisson sampling with $0 < \alpha \leq \pi_k \leq \beta < 1$ also satisfy the assumptions of Theorem 3.4, whereas (N, n)-simple sampling designs may not satisfy the assumptions of Theorem 3.4.

Appendix B. Proof of Theorem 4.1

Let P_0 be a $(N \times N)$ rank-n projection matrix whose entries $P_0(k, l)$ are 0 apart from (k_r+1, k_r+1) entries, $r = 0, \ldots, n-1$ whose values are 1. We aim at transforming in $N - 1$ steps the diagonal of P_0, without changing its spectrum, to finally end up with a projection matrix $P_{N-1} = P^{\Pi}$ whose diagonal is Π. We pose $T_k = \sum_{i=k}^{k_{r+1}}\alpha_i$ for $k_r < k \leq k_{r+1}$. Table 7 shows the expected diagonal entries at each step.

We use the unitary operator $R_q(\theta)$ whose matrix relative to the canonical basis has $\sin\theta_q$ at the (q, q) and $(q+1, q+1)$ entries, $-\cos\theta_q$ and $\cos\theta_q$ at the $(q, q+1)$ and $(q+1, q)$ entries, respectively, 1 at all other diagonal entries, and 0 at all other off-diagonal entries. We then build the sequence $P_q = R_q(\theta_q)P_{q-1}R_q^T(\theta_q)$, $q = 1, \ldots, N-1$. Let $P_q(k, l)$ be P_q k,l entry.

Table 7
Diagonal entries of P_q.

Step	Diagonal entries												
q	1	2	3	...	k_1-1	k_1	k_1+1	...	k_r-1	k_r	k_r+1	...	N
0	1	0	0	...	0	0	1	...	0	0	1	...	0
1	Π_1	T_2	0	...	0	0	1	...	0	0	1	...	0
2	Π_1	Π_2	T_3	...	0	0	1	...	0	0	1	...	0
$\vdots$	Π_1	Π_2	Π_3	...	0	0	1	...	0	0	1	...	0
k_1-1	Π_1	Π_2	Π_3	...	Π_{k_1-1}	a_{k_1}	1	...	0	0	1	...	0
k_1	Π_1	Π_2	Π_3	...	Π_{k_1-1}	Π_{k_1}	T_{k_1+1}	...	0	0	1	...	0
$\vdots$	Π_1	Π_2	Π_3	...	Π_{k_1-1}	Π_{k_1}	Π_{k_1+1}	...	$\vdots$	$\vdots$	$\vdots$	...	0
$N-1$	Π_1	Π_2	Π_3	...	Π_{k_1-1}	Π_{k_1}	Π_{k_1+1}	...	Π_{k_r-1}	Π_{k_r}	Π_{k_r+1}	...	$a_N=\Pi_N$

Table 8
Values of θ_q.

	$\exists r \vert q = k_r$	$k_r < q < k_{r+1}$
$\sin\theta_q$	$\sqrt{\frac{1-\Pi_q}{1-\alpha_q}}$	$\sqrt{\frac{\Pi_q}{T_q}}$
$\cos\theta_q$	$\sqrt{\frac{\Pi_q-\alpha_q}{1-\alpha_q}}$	$\sqrt{\frac{T_{q+1}}{T_q}}$

Relying on the general relations between the entries of P_q and those of P_{q-1} provided by Kadison (2002), we deduce that θ_q is a solution of an equation depending on whether there exists r such that $q = k_r$:

$$\sin^2\theta_q = \frac{1-\Pi_q}{1-\alpha_q}, (\exists r|q=k_r)$$
$$\sin^2\theta_q = \frac{\Pi_q}{T_q}, (\nexists r|q=k_r).$$

The definitions of α_q and T_q ensure the existence of at least one value for θ_q and as a consequence the existence of a real idempotent self-adjoint with a prescribed diagonal Π. Kadison's Theorem 7 consists precisely of the latter result. We go one step further and provide a closed-form for this matrix. To do so, we choose among the solutions for θ_q the one given in Table 8.

We then iteratively use of the relations between then entries P_q and those of P_{q-1} to deduce an explicit formulas for $P_{N-1}(k,l) = P^{\Pi}_{kl}$:

$$P^{\Pi}_{kl} = \sin\theta_k(\alpha_l - 1)\sin\theta_l \prod_{q=l}^{k-1}\cos\theta_q, (l = k_r \le k)$$
$$P^{\Pi}_{kl} = \sin\theta_k T_l \sin\theta_l \prod_{q=l}^{k-1}\cos\theta_q, (l \ne k_r \le k).$$

For instance, Tables 9 and 10 show how the entries $P_q(k,k_1)$ and $P_q(k,k_1+1)$ are fixed across the process leading to the final corresponding entries of P^{Π}.

Lastly, we consider the case $k_r < l < k_{r+1}, k_{r'} < k < k_{r'+1}, r < r'$ and provide the corresponding formulas given in Table 1, the other cases being similar.

$$\begin{aligned} P^{\Pi}_{kl} &= \sin\theta_k T_l \sin\theta_l \prod_{q=l}^{k-1}\cos\theta_q \\ &= \underbrace{\sqrt{\frac{\Pi_k}{T_k}}}_{\sin\theta_k} \underbrace{\sqrt{\Pi_l T_l}}_{T_l\sin\theta l} \underbrace{\prod_{q=l}^{k_{r+1}-1}\sqrt{\frac{T_{q+1}}{T_q}}}_{\prod_{q=l}^{k_{r+1}-1}\cos\theta_q} \underbrace{\sqrt{\frac{\Pi_{k_{r+1}}-\alpha_{k_{r+1}}}{1-\alpha_{k_{r+1}}}}}_{\cos\theta_{k_{r+1}}} \underbrace{\prod_{q=k_{r+1}+1}^{k_{r+2}-1}\sqrt{\frac{T_{q+1}}{T_q}}}_{\prod_{q=k_{r+1}+1}^{k_{r+2}-1}\cos\theta_q} \underbrace{\sqrt{\frac{\Pi_{k_{r+2}}-\alpha_{k_{r+2}}}{1-\alpha_{k_{r+2}}}}}_{\cos\theta_{k_{r+2}}} \\ &\quad \dots \prod_{q=k_{r'-1}+1}^{k_{r'}-1}\sqrt{\frac{T_{q+1}}{T_q}}\sqrt{\frac{\Pi_{k_{r'}}-\alpha_{k_{r'}}}{1-\alpha_{k_{r'}}}}\prod_{q=k_{r'}+1}^{k-1}\sqrt{\frac{T_{q+1}}{T_q}} \end{aligned}$$

Table 9
Building step by step the values $P_q(k, k_1)$, $k \geq k_1$.

Line	Step (q)			
k	$k_1 - 1$	k_1	$k_1 + 1$	$k_1 + 2$
k_1	a_{k_1}	$\underline{\Pi_{k_1}}$	$\underline{\Pi_{k_1}}$	$\underline{\Pi_{k_1}}$
$k_1 + 1$	0	$\cos\theta_{k_1}(\alpha_{k_1} - 1)\sin\theta_{k_1}$	$\underline{\sin\theta_{k_1+1}\cos\theta_{k_1}(\alpha_{k_1} - 1)\sin\theta_{k_1}}$	$\underline{\sin\theta_{k_1+1}\cos\theta_{k_1}(\alpha_{k_1} - 1)\sin\theta_{k_1}}$
$k_1 + 2$	0	0	$\cos\theta_{k_1+1}\cos\theta_{k_1}(\alpha_{k_1} - 1)\sin\theta_{k_1}$	$\underline{\sin\theta_{k_1+2}\cos\theta_{k_1+1}\cos\theta_{k_1}(\alpha_{k_1} - 1)\sin\theta_{k_1}}$
$k_1 + 3$	0	0	0	$\cos\theta_{k_1+2}\cos\theta_{k_1+1}\cos\theta_{k_1}(\alpha_{k_1} - 1)\sin\theta_{k_1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Underlying a formula indicates that it will remain unchanged for the rest of the process.

Table 10
Building step by step the values $P_q(k, k_1 + 1)$, $k \geq k_1$.

Line	Step(q)			
k	k_1	$k_1 + 1$	$k_1 + 2$	$k_1 + 3$
$k_1 + 1$	T_{k_1+1}	$\underline{\Pi_{k_1+1}}$	$\underline{\Pi_{k_1+1}}$	$\underline{\Pi_{k_1+1}}$
$k_1 + 2$	0	$\cos\theta_{k_1+1}\sin\theta_{k_1+1}T_{k_1+1}$	$\underline{\sin\theta_{k_1+2}\cos\theta_{k_1+1}\sin\theta_{k_1+1}T_{k_1+1}}$	$\underline{\sin\theta_{k_1+2}\cos\theta_{k_1+1}\sin\theta_{k_1+1}T_{k_1+1}}$
$k_1 + 3$	0	0	$\cos\theta_{k_1+2}\cos\theta_{k_1+1}\sin\theta_{k_1+1}T_{k_1+1}$	$\underline{\sin\theta_{k_1+3}\cos\theta_{k_1+2}\cos\theta_{k_1+1}\sin\theta_{k_1+1}T_{k_1+1}}$
$k_1 + 4$	0	0	0	$\cos\theta_{k_1+3}\cos\theta_{k_1+2}\cos\theta_{k_1+1}\sin\theta_{k_1+1}T_{k_1+1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Underlying a formula indicates that it will remain unchanged for the rest of the process.

$$
\begin{aligned}
&= \sqrt{\Pi_l T_l}\sqrt{\frac{\Pi_k}{T_k}}\sqrt{\frac{T_{k_{r+1}}}{T_l}}\sqrt{\frac{\Pi_{k_{r+1}} - \alpha_{k_{r+1}}}{1 - \alpha_{k_{r+1}}}}\sqrt{\frac{T_{k_{r+2}}}{T_{k_{r+1}+1}}}\sqrt{\frac{\Pi_{k_{r+2}} - \alpha_{k_{r+2}}}{1 - \alpha_{k_{r+2}}}} \\
&\quad \cdots \sqrt{\frac{T_{k_{r'}}}{T_{k_{r'-1}+1}}}\sqrt{\frac{\Pi_{k_{r'}} - \alpha_{k_{r'}}}{1 - \alpha_{k_{r'}}}}\sqrt{\frac{T_k}{T_{k_{r'}+1}}} \\
&= \sqrt{\Pi_k \Pi_l}\prod_{j=r+1}^{r'}\sqrt{\frac{(\Pi_{k_j} - \alpha_{k_j})\alpha_{k_j}}{(1 - \alpha_{k_j})T_{k_j+1}}}.
\end{aligned}
$$

Appendix C. Proof of Corollary 4.1

The first two points follow from the calculations of the respective 2×2 and 3×3 determinant.

$$
P^{\Pi}_{|jkl} = \begin{pmatrix}
\Pi_j & \sqrt{\frac{(1 - \Pi_k)\alpha_k}{1 - \alpha_k}\Pi_j} & \sqrt{\Pi_l \Pi_j \frac{(\Pi_k - \alpha_k)\alpha_k}{(1 - \alpha_k)(1 - (\Pi_k - \alpha_k))}} \\
\sqrt{\frac{(1 - \Pi_k)\alpha_k}{1 - \alpha_k}\Pi_j} & \Pi_k & -\sqrt{\Pi_l \frac{(1 - \Pi_k)(\Pi_k - \alpha_k)}{(1 - (\Pi_k - \alpha_k))}} \\
\sqrt{\Pi_l \Pi_j \frac{(\Pi_k - \alpha_k)\alpha_k}{(1 - \alpha_k)(1 - (\Pi_k - \alpha_k))}} & -\sqrt{\Pi_l \frac{(1 - \Pi_k)(\Pi_k - \alpha_k)}{(1 - (\Pi_k - \alpha_k))}} & \Pi_l
\end{pmatrix}.
$$

Applying Sarrus's rule, we get:

$$
\begin{aligned}
&\det(P^{\Pi}_{|jkl}) = \\
&\Pi_j\Pi_k\Pi_l - 2\frac{\Pi_j\Pi_l(1 - \Pi_k)(\Pi_k - \alpha_k)\alpha_k}{(1 - \alpha_k)(1 - (\Pi_k - \alpha_k))} - \frac{\Pi_j\Pi_l(1 - \Pi_k)\alpha_k}{1 - \alpha_k} - \frac{\Pi_j\Pi_l(1 - \Pi_k)(\Pi_k - \alpha_k)}{(1 - (\Pi_k - \alpha_k))} - \frac{\Pi_l\Pi_j\Pi_k(\Pi_k - \alpha_k)a_k}{(1 - \alpha_k)(1 - (\Pi_k - \alpha_k))} = \\
&\Pi_j\Pi_k\Pi_l - \frac{\Pi_j\Pi_l(\Pi_k - \alpha_k)\alpha_k(2 - \Pi_k) + \Pi_j\Pi_l(1 - \Pi_k)\alpha_k(1 - (\Pi_k - \alpha_k)) + \Pi_j\Pi_l(1 - \Pi_k)(\Pi_k - \alpha_k)(1 - \alpha_k)}{(1 - \alpha_k)(1 - (\Pi_k - \alpha_k))} = \\
&\Pi_j\Pi_k\Pi_l - \frac{\Pi_j\Pi_l(\Pi_k - \alpha_k)(1 - (\Pi_k - \alpha_k)) + \Pi_j\Pi_l(1 - \Pi_k)\alpha_k(1 - (\Pi_k - \alpha_k))}{(1 - \alpha_k)(1 - (\Pi_k - \alpha_k))} = \\
&\Pi_j\Pi_k\Pi_l - \frac{\Pi_j\Pi_k\Pi_l(1 - (\Pi_k - \alpha_k))(1 - \alpha_k)}{(1 - (\Pi_k - \alpha_k))(1 - \alpha_k)} = 0.
\end{aligned}
$$

We finally consider points 3 to 5. It holds that $P^{\Pi}_{|B_r}$ has the same spectrum as $(P_0)_{|B_r}$ except one 0 of $(P_0)_{|B_r}$ that may increase proving point 3. Point 4 follows from the expression of P^{Π}_{kl} as a product of cosines. Finally, point 5 follows from the implication: if $\sum_1^{k_r} \Pi_k = r$ then $\Pi_{k_r} - \alpha_{k_r} = 0$, and thus $P^{\Pi}_{kl} = 0$ for k, l in different strata.

Appendix D. Proof of Theorem 4.3

Let $DSD(K)$ be a (N, n)-simple DSD. Applying Eqs. (1) and (2), we get that K satisfies

$$\begin{cases} K_{kk} &= \dfrac{n}{N}, \\ |K_{kl}|^2 &= \dfrac{n(N-n)}{N^2(N-1)} (k \neq l). \end{cases}$$

Let F be a $(n \times N)$ matrix such that $V = (\frac{n}{N})^{1/2}\overline{F}^T$ is an orthonormal basis of the range of K ($K = V\overline{V}^T = \frac{n}{N}\overline{F}^T F$). It holds that:

1. For all $l \in 1, \dots, N$, $\sum_{k=1}^n F_{kl}^2 = 1$,
2. There exists a nonnegative α such that $|\sum_{j=1}^n \overline{F_{jk}} F_{jl}| = \alpha = \sqrt{\frac{N-n}{n(N-1)}}$ $(k \neq l)$,
3. $F\overline{F}^T = \frac{N}{n} I_n$.

But these properties are exactly those defining an ETF (Tropp, 2005; Sustik et al., 2007).

Appendix E. Existence of (N, n)-simple DSDs

Theorem E.1. *Let $1 < n < N-1$ be two integers.*

1. *There exists a $(N, n)-$ simple determinantal sampling design only if $N \leq min\{n^2, (N-n)^2\}$* (Tropp, 2005).
2. *There exists a $(N, n)-$ simple determinantal sampling design with a real kernel K only if* $N \leq \min\{\frac{n(n+1)}{2}, \frac{(N-n)(N-n+1)}{2}\}$ (Sustik et al., 2007 Theorem C).
3. *When $N \neq 2n$, a necessary condition of the existence of a (N, n)-simple determinantal sampling design with real kernel K is that the following two quantities be odd integers:*

$$\alpha = \sqrt{\frac{n(N-1)}{N-n}}, \ \beta = \sqrt{\frac{(N-n)(N-1)}{n}}.$$

When $N = 2n$, it is necessary that n be odd and that $N - 1$ be the sum of two squares (Sustik et al., 2007 Theorem A and Casazza et al., 2008 Theorem 4.1).

This a only a small part of a rich literature on the subject, going from strongly regular graphs (Waldron 2009) to Gauss sums and finite field theory Strohmer (2008).

Appendix F. Optimization problems

Consider the free optimization problem P_1, where the only constraint is the average size of the sample, $E(\sharp\mathbb{S}) = \mu$. The parameter space is then $\Theta = \Theta^{\mu}_{c,d} \times \mathbb{R}^N$ where Θ_{K^μ} is the set of contracting matrices of trace μ (with $0 < \mu \leq N$)

$$\Theta^{\mu}_{c,d} = \{K \mid 0 \leq K \leq I_N,\ c \leq K_{kk} \leq d,\ Trace(K) = \mu\}$$

which is a projected spectrahedron.

We can then plug in this problem the optimal weights obtained in Theorem 5.1 and use Corollary 5.1 to deduce that (K, w) solves Problem 5.1 for $\Theta = \Theta_{K^\mu} \times \mathbb{R}^N$ iff w belongs to the affine subspace described in Theorem 5.1 and K solves:

Problem F.1. Find

$$\underset{K \in \Theta^{\mu}_{c,d}}{\arg\min} - u(K)^T (V * \Omega(K))^{\dagger} u(K)$$

where $u(K) = (\sum_{q=1}^Q t_{x^q} x^q) * diag(K)$, $V = \sum_{q=1}^{Q} x^q {x^q}^T$ and $\Omega(K) = (I_N - K) * K + diag(K) diag(K)^T$.

that is therefore a semidefinite minimization problem with non-convex objective function.

A variation of the problem is to consider that all the first order inclusion probabilities are fixed in advance, and still search for an optimal strategy (Problem P_5).

Once again by Theorem 5.1 this rewrites as:

Problem F.2. Find

$$\arg\min_{K\in\Theta^{\Pi}} -u^T(V*\Omega(K))^{\dagger}u$$

where $u=(\sum_{q=1}^{Q}t_{x^q}x^q)*\Pi$, $V=\sum_{q=1}^{Q}x^q{x^q}^T$ and $\Omega(K)=D_{\Pi}-K*K+\Pi\Pi^T$.

where Θ^{Π} is the set of contracting matrices of diagonal Π:

$\Theta^{\Pi}=\{K^{\Pi}\mid 0\leq K^{\Pi}\leq I_N, diag(K^{\Pi})=\Pi\}$.

Simpler but still neither convex nor concave is the objective function of the free problem restricted to the Horvitz–Thompson estimator (Problem P_2):

Problem F.3. Find

$$\arg\min_{K\in\Theta^{\mu}}\sum_{q=1}^{Q}(z^q)^T((I_N-K)*\overline{K})z^q$$

where $z^q=[diag(K)]^{-1}*x^q$.

or the objective function of Problem P_3 (fixed weights):

Problem F.4. Find

$$\arg\min_{K\in\Theta^{\mu}} C(K)=\sum_{q=1}^{Q}(z^q)^T((I_N-K)*\overline{K})z^q+[e^T(K*I_N)z^q-e^Tx^q]^2$$

where $z^q=w*x^q$.

Finally, Problem P_4 (fixed first order inclusion probabilities, fixed weights) can be rewritten as:

Problem F.5. Find

$$\arg\min_{K\in\Theta^{\Pi}} -\sum_{q=1}^{Q}(z^q)^T(K*\overline{K})z^q$$

where $z^q=w*x^q$.

(This problem includes the case of the Horvitz–Thompson estimator, take $w_k=\Pi_k^{-1}$).
The objective function is concave for nonnegative auxiliary variables x^q.

Appendix G. Minimization over sampling designs of average size (less than) one. The elliptope case

We first consider equal-probability determinantal sampling designs of average size one. In this case, the parameter space for Problem F.5 is $\Theta_e=\{0\leq K\leq I_N, K_{kk}=\frac{1}{N}\}$, the spectrahedron of positive semidefinite matrices of diagonal $\frac{1}{N}$ (this set is homothetic to the set of correlation matrices, also known as the elliptope, which is the set of positive semidefinite matrices of diagonal 1). The literature on the elliptope and linear optimization over it is abundant, see for instance Ycart (1985), Grone et al. (1990), Laurent and Poljak (1995, 1996), Kurowicka and Cooke (2003) and Laurent and Varvitsiotis (2014). It is known that (for real matrices):

Theorem G.1 (*Linear Optimization Over the Elliptope*)**.**

1. *For any integer k such that* $\binom{k+1}{2}\leq N$, *there exists a matrix of rank k that is an extreme point of* Θ_e (Grone et al., 1990 Theorem 2).
2. *The vertices of* Θ_e *(extreme points where the normal cone to* Θ_e *is of rank N) are the projections of* Θ_e *(rank 1 matrices).*
3. *It is NP-hard to decide whether the optimum of linear optimization problem* $\max_{K\in\sigma}\langle A,K\rangle$ *is reached at a vertex.*

Otherwise stated, the minimization of a linear function over the elliptope can be considerably hard, and the solution may not be a projection matrix.

Surprisingly, for this particular set (equal probability determinantal sampling designs of average size 1), the quadratic problem is much more simpler than the linear one. Actually, the minimization Problem 5.1 for all unequal-probability sampling designs of average size less than 1 (not only the determinantal ones) admits a simple solution.

Theorem G.2 (*Optimal Sampling Design, Average Size Less Than* 1)**.** *Let Π be a vector of inclusion probabilities such that $\sum_{k\in U}\Pi_k \leq 1$, and $x^1,\ldots,x^Q$ be nonnegative variables. There exists a unique sampling design that minimize $\sum_{q=1}^{Q}\mathrm{MSE}(\hat{t}_{x^q w})$ within all sampling designs with fixed first order inclusion probabilities $\pi_k = \Pi_k$. It is the determinantal sampling design $\mathcal{P} = DSD(K^{\Pi})$, where K^{Π} is any rank 1 matrix with the prescribed diagonal. This sampling design $\mathcal{P}$ consists in sampling no element with probability $1-\sum_{k\in U}\Pi_k$, and the single element k with probability Π_k.*

Proof. Let $\mathcal{P}$ be any sampling design with fixed first order inclusion probabilities $\pi_k = \Pi_k$. As for $k \neq l$, $\Delta_{kl} \geq -\pi_k\pi_l$ then

$$\begin{aligned}\sum_{q=1}^{Q} var(\hat{t}_{x^q w}) &= \sum_{q=1}^{Q}\left[\sum_{k\in U}(w_k x_k^q)^2(\Pi_k-\Pi_k^2) + \sum_{k\neq l\in U} w_k x_k^q w_l x_l^q \Delta_{kl}\right]\\ &\geq \sum_{q=1}^{Q}\left[\sum_{k\in U}(w_k x_k^q)^2(\Pi_k-\Pi_k^2) - \sum_{k\neq l\in U} w_k x_k^q w_l x_l^q \pi_k\pi_l\right],\end{aligned}$$

with equality iff for $k \neq l$, $\Delta_{kl} = -\pi_k\pi_l$ that is $\pi_{kl} = 0$. The only sampling design that satisfies these equalities is $\mathcal{P}$, which is thus the optimal design.

Consider now $K^{\Pi} = bb^T$ a rank one matrix with the prescribed diagonal. Then $\|b\|^2 = \sum_{k\in U}\Pi_k \leq 1$, and K^{Π} is a contraction of rank 1. It follows that $DSD(K^{\Pi})$ exists, and has no more than 1 element by Corollary 2.1, so that $\pi_{kl} = 0, k \neq l$. Finally $DSD(K^{\Pi})$ achieves this lower bound, and $DSD(K^{\Pi}) = \mathcal{P}$. □

If $\sum_{k\in U}\Pi_k = 1$ (in particular if $\Pi_k = \frac{1}{N}$) we get the following corollaries:

Corollary G.1 (*Minimization Over the Elliptope*)**.** *Assume the variables $x^1,\ldots,x^Q$ are nonnegative. Then the solutions of Problem F.5 over Θ_e are the rank one projections with diagonal $\frac{1}{N}$ (vertices of Θ_e).*

More generally, the solutions of Problem 5.1 over Θ with $\sum_{k\in U}\Pi_k = 1$ are the rank one projections with diagonal Π_k.

Corollary G.2 (*SRS*(1) *is Optimal*)**.** *The sampling design with equal first order inclusion probabilities $\pi_k = \frac{1}{N}$ that minimizes the sum of the* MSEs *for nonnegative variables is the SRS of size* 1*, which is determinantal.*

Nonnegativity is crucial in the previous results. Consider the following example:

Example G.1. Let $U = \{1,2\}$, $x_1 = -1$, $x_2 = 1$ and $\Pi_1 = \Pi_2 = \frac{1}{2}$. Then the variance of the Horvitz–Thompson estimator for any equal probability sampling design of average size one that satisfies the Sen–Yates–Grundy conditions is $var(\hat{t}_{xHT}) = 2 - 8\Delta_{1,2} \geq 2$, which is the variance of the estimator under Poisson sampling $=$ DSD $\begin{pmatrix}1/2 & 0\\ 0 & 1/2\end{pmatrix}$. But the matrix $K = \begin{pmatrix}1/2 & 0\\ 0 & 1/2\end{pmatrix} = 1/2\begin{pmatrix}1/2 & 1/2\\ 1/2 & 1/2\end{pmatrix} + 1/2\begin{pmatrix}1/2 & -1/2\\ -1/2 & 1/2\end{pmatrix}$ is not extremal.

For more complex spectrahedra ($\sum_k \Pi_k > 1$), a characterization of the solutions of Problem F.5 is unknown. In particular, the question whether the solutions are always projections for integer sums remains open.

References

Berger, Y.G., 1998. Rate of convergence for asymptotic variance of the Horvitz–Thompson estimator. J. Statist. Plann. Inference 74 (1), 149–168.
Biscio, C.A.N., Lavancier, F., et al., 2016. Quantifying repulsiveness of determinantal point processes. Bernoulli 22 (4), 2001–2028.
Blekherman, G., Parrilo, P.A., Thomas, R.R., 2013. Semidefinite Optimization and Convex Algebraic Geometry, Vol. 13. SIAM.
Borcea, J., Brändén, P., Liggett, T., 2009. Negative dependence and the geometry of polynomials. J. Amer. Math. Soc. 22 (2), 521–567.
Borodin, A., 2009. Determinantal point processes, arXiv preprint arXiv:0911.1153.
Brändén, P., Jonasson, J., 2012. Negative dependence in sampling. Scand. J. Stat. 39 (4), 830–838.
Capelli, A., 1892. Sopra la compatibilita o incompatibilita di piu equazioni di primo grado fra piu incognite. Riv. Mat. 2, 54–58.
Cardot, H., Chaouch, M., Goga, C., Labruère, C., 2010. Properties of design-based functional principal components analysis. J. Statist. Plann. Inference 140 (1), 75–91.
Casazza, P.G., Redmond, D., Tremain, J.C., 2008. Real equiangular frames. In: CISS. Citeseer, pp. 715–720.
Cassel, C.-M., Särndal, C.E., Wretman, J.H., 1977. Foundations of Inference in Survey Sampling. Wiley.
Chauvet, G., 2014. A note on the consistency of the Narain-Horvitz-Thompson estimator, arXiv preprint arXiv:1412.2887.
De Klerk, E., 2006. Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications, Vol. 65. Springer Science & Business Media.
Deming, W.E., Stephan, F.F., 1941. On the interpretation of censuses as samples. J. Amer. Statist. Assoc. 36 (213), 45–49.
Deville, J.-C., 2012. Comment conserver l'équilibre dans un sondage: la quête du Graal et la suite. In: Septième Colloque Francophone sur Les Sondages. Rennes.
Deville, J.-C., Särndal, C.-E., 1992. Calibration estimators in survey sampling. J. Amer. Statist. Assoc. 87 (418), 376–382.
Deville, J.-C., Tillé, Y., 2004. Efficient balanced sampling: the cube method. Biometrika 91 (4), 893–912.
Dhillon, I.S., Heath Jr., R.W., Sustik, M.A., Tropp, J.A., 2005. Generalized finite algorithms for constructing Hermitian matrices with prescribed diagonal and spectrum. SIAM J. Matrix Anal. Appl. 27 (1), 61–71.
Dickinson, B., Steiglitz, K., 1982. Eigenvectors and functions of the discrete Fourier transform. IEEE Trans. Acoust. Speech Signal Process. 30 (1), 25–31.
Dol, W., Steerneman, T., Wansbeek, T., 1996. Matrix algebra and sampling theory: The case of the Horvitz-Thompson estimator. Linear Algebra Appl. 237, 225–238.

Erdös, P., Rényi, A., 1959. On the central limit theorem for samples from a finite population. Publ. Math. Inst. Hung. Acad. Sci. 4, 49–61.
Fickus, M., Mixon, D.G., Poteet, M.J., Strawn, N., 2013. Constructing all self-adjoint matrices with prescribed spectrum and diagonal. Adv. Comput. Math. 39 (3–4), 585–609.
Fuller, W.A., 1970. Sampling with random stratum boundaries. J. R. Stat. Soc. Ser. B Stat. Methodol. 209–226.
Fuller, W.A., 2009. Some design properties of a rejective sampling procedure. Biometrika asp042.
Grone, R., Pierce, S., Watkins, W., 1990. Extremal correlation matrices. Linear Algebra Appl. 134, 63–70.
Hájek, J., 1960. Limiting distributions in simple random sampling from a finite population. Publ. Math. Inst. Hung. Acad. Sci. 5 (361), 74.
Hájek, J., 1964. Asymptotic theory of rejective sampling with varying probabilities from a finite population. Ann. Math. Statist. 1491–1523.
Hájek, J., Dupac, V., 1981. Sampling from a Finite Population. Marcel Dekker Inc.
Horn, A., 1954. Doubly stochastic matrices and the diagonal of a rotation matrix. Amer. J. Math. 76 (3), 620–630.
Horn, R.A., Johnson, C.R., 1991. Topics in Matrix Analysis. Cambridge Univ. Press.
Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. J. Amer. Statist. Assoc. 47 (260), 663–685.
Hough, J.B., Krishnapur, M., Peres, Y., Virág, B., et al., 2006. Determinantal processes and independence. Probab. Surv. 3, 206–229.
Hough, J.B., Krishnapur, M., Peres, Y., Virág, B., 2009. Zeros of Gaussian Analytic Functions and Determinantal Point Processes, Vol. 51. American Mathematical Society, Providence, RI.
Isaki, C.T., Fuller, W.A., 1982. Survey design under the regression superpopulation model. J. Amer. Statist. Assoc. 77 (377), 89–96.
Kadison, R.V., 2002. The Pythagorean theorem: I. The finite case. Proc. Natl. Acad. Sci. 99 (7), 4178–4184.
Kulesza, A., 2012. Learning with Determinantal Point Processes (Ph.D. thesis), University of Pennsylvania.
Kulesza, A., Taskar, B., 2011. Learning determinantal point processes. In: Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence. AUAI Press, pp. 419–427.
Kurowicka, D., Cooke, R., 2003. A parameterization of positive definite matrices in terms of partial correlation vines. Linear Algebra Appl. 372, 225–251.
Laurent, M., Poljak, S., 1995. On a positive semidefinite relaxation of the cut polytope. Linear Algebra Appl. 223, 439–461.
Laurent, M., Poljak, S., 1996. On the facial structure of the set of correlation matrices. SIAM J. Matrix Anal. Appl. 17 (3), 530–547.
Laurent, M., Varvitsiotis, A., 2014. A new graph parameter related to bounded rank positive semidefinite matrix completions. Math. Program. 145 (1–2), 291–325.
Lavancier, F., Møller, J., Rubak, E., 2015. Determinantal point process models and statistical inference. J. R. Stat. Soc. Ser. B Stat. Methodol. 77 (4), 853–877.
Lyons, R., 2003. Determinantal probability measures. Publ. Math. L'Inst. Hautes Études Sci. 98, 167–212.
Macchi, O., 1975. The coincidence approach to stochastic point processes. Adv. Appl. Probab. 83–122.
Patterson, R.F., Smith, W.D., Taylor, R.L., Bozorgnia, A., 2001. Limit theorems for negatively dependent random variables. Nonlinear Anal. TMA 47 (2), 1283–1295.
Pemantle, R., Peres, Y., 2014. Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures. Combin. Probab. Comput. 23 (01), 140–160.
Polyak, R., 1992. Modified barrier functions (theory and methods). Math. Program. 54 (1–3), 177–222.
Rao, K.R., Yip, P.C., 2000. The Transform and Data Compression Handbook, Vol. 1. CRC press.
Robinson, P., 1982. On the convergence of the Horvitz-Thompson estimator. Austral. J. Statist. 24 (2), 234–238.
Särndal, C.-E., Swensson, B., Wretman, J., 2003. Model Assisted Survey Sampling. Springer Science & Business Media.
Scardicchio, A., Zachary, C.E., Torquato, S., 2009. Statistical properties of determinantal point processes in high-dimensional Euclidean spaces. Phys. Rev. E 79 (4), 041108.
Schur, J., 1911. Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen.. J. Reine Angew. Math. 140, 1–28.
Sen, A.R., 1953. On the estimate of the variance in sampling with varying probabilities. J. Indian Soc. Agric. Statist. 5 (1194), 127.
Soshnikov, A., 2000. Determinantal random point fields. Russian Math. Surveys 55 (5), 923–975.
Soshnikov, A., 2002. Gaussian limit for determinantal random point fields. Ann. Probab. 171–187.
Strohmer, T., 2008. A note on equiangular tight frames. Linear Algebra Appl. 429 (1), 326–330.
Sustik, M.A., Tropp, J.A., Dhillon, I.S., Heath, R.W., 2007. On the existence of equiangular tight frames. Linear Algebra Appl. 426 (2), 619–635.
Tillé, Y., 2011. Sampling Algorithms. Springer.
Tropp, J.A., 2005. Complex equiangular tight frames. In: Optics & Photonics 2005. International Society for Optics and Photonics.
Tütüncü, R., Toh, K., Todd, M., 2001. SDPT3—a Matlab software package for semidefinite-quadratic-linear programming, version 3.0. Web Page. http://www.math.nus.edu.sg/~mattohkc/sdpt3.html.
Vandenberghe, L., Boyd, S., 1996. Semidefinite programming. SIAM Rev. 38 (1), 49–95.
Waldron, S., 2009. On the construction of equiangular frames from graphs. Linear Algebra Appl. 431 (11), 2228–2242.
Yates, F., Grundy, P., 1953. Selection without replacement from within strata with probability proportional to size. J. R. Stat. Soc. Ser. B Stat. Methodol. 253–261.
Ycart, B., 1985. Extreme points in convex sets of symmetric matrices. Proc. Amer. Math. Soc. 95 (4), 607–612.
Yuan, M., Su, C., Hu, T., 2003. A central limit theorem for random fields of negatively associated processes. J. Theoret. Probab. 16 (2), 309–323.