Non-asymptotic bounds for inclusion probabilities in rejective sampling

Simon Ruetz SIMON.RUETZ@UIBK.AC.AT
Karin Schnass KARIN.SCHNASS@UIBK.AC.AT
University of Innsbruck
Technikerstraße 13
6020 Innsbruck, Austria

Abstract

We provide non-asymptotic bounds for first and higher order inclusion probabilities of the rejective sampling model with various size parameters. Further we derive bounds in the semi-definite ordering for matrices that collect (conditional) first and second order inclusion probabilities as their diagonal resp. off-diagonal entries.

Keywords: Rejective sampling, conditional Poisson sampling, inclusion probabilities, matrix of inclusion probabilities, semi-definite order bounds

1. Introduction

In finite population sampling theory the aim is to draw statistical conclusions for population characteristics based on a sample of it. In the landmark paper of Hájek [5], he studied the behaviour of the Horvitz-Thompson estimator under rejective sampling. This proved to be the starting point of many inquiries into rejective sampling [6, 4, 9, 1, 8, 15, 14] and corresponding concentration inequalities [2, 12].

Following Hájek's introduction, we consider a finite population U of size N, essentially meaning $U = [N] := \{1, \dots, N\}$, from which we want to draw a set or sample I of size S. Let $p_1, \dots, p_N$ be drawing probabilities such that $p_i \in (0,1)$ and $\sum_i p_i = S$.

We say that our samples I are drawn from the Poisson sampling model with weights p_i, if the probability of drawing a sample I is given by

$$\mathbb{P}_B(I) = \prod_{i \in I} p_i \prod_{j \notin I} (1 - p_j). \tag{1}$$

The advantage of Poisson sampling is that each index i appears in the sample I independently of all the others since it can be seen as a series of N Bernoulli random variables δ_i, where each δ_i has expectation p_i. The disadvantage is that the sampled sets have varying sizes. Reducing to sets of only one size leads to the corresponding rejective sampling model with weights p_i and parameter S, where the probability of a sample I is given by

$$\mathbb{P}_S(I) := \mathbb{P}_B(I \mid |I| = S) = \begin{cases} c^{-1} \prod_{i \in I} p_i \prod_{j \notin I} (1 - p_j) & \text{if} \quad |I| = S \\ 0 & \text{else} \end{cases}, \tag{2}$$

with $c = \mathbb{P}_B(|I| = S) = \sum_{|I|=S} \prod_{i \in I} p_i \prod_{j \notin I} (1 - p_j)$. Due to its construction, rejective sampling is also known as conditional Poisson sampling. We restrict ourselves to the case

$p_i \in (0,1)$, since for $p_i = 0$ we can just drop the corresponding index i from the population $[N]$ since it gets never picked anyway. Similarly, for $p_i = 1$, we can trivially include the index i in every set I we sample and then just sample $S-1$ indices from the rest of the population $[N] \setminus \{i\}$ — corresponding to rejective sampling of size $S-1$.
Of particular interest are the first order inclusion probabilities which are defined as

$$\pi_i(S) := \sum_{I: i \in I} \mathbb{P}_S(I) = \mathbb{P}_S(i \in I). \tag{3}$$

If S is clear from context or not relevant, we will sometimes write π_i instead. Denoting by $\mathbf{1}_I \in \mathbb{R}^N$ the vector whose i-th entry is 1 if $i \in I$ and zero else, we have for the vector of inclusion probabilities $\pi(S) := \mathbb{E}_S[\mathbf{1}_I] \in \mathbb{R}^N$. For the Poisson sampling model we have $p_i = \mathbb{P}_B(i \in I)$. In general the inclusion probabilities of the rejective and Poisson sampling models are not equal, i.e. $p_i \neq \pi_i$, unless $p_i = c$ for all $i \in [N]$ and for some $c \in [0,1]$. For our short literature review we further define,

$$d := \sum_{i=1}^N p_i(1-p_i) \quad \text{and} \quad \alpha_i := \frac{p_i}{1-p_i}. \tag{4}$$

Hájek [5] first studied the relation between p_i and π_i and showed that

$$\max_{1 \leq i \leq N} |\pi_i/p_i - 1| \to 0 \quad \text{as} \quad d \to \infty. \tag{5}$$

Hence asymptotically, the inclusion probabilities of the rejective and Poisson sampling models are indistinguishable. Nevertheless there are a lot of settings where one is interested in the non-asymptotic relation between the different inclusion probabilities. So it was conjectured in [6] and later shown in [9] that

$$\min_i \alpha_i \leq \min_i \pi_i \qquad \text{and} \qquad \max_i \pi_i \leq \max_i \alpha_i. \tag{6}$$

This already gives control over the extreme points of the two sequences, but still not a general way of relating the two sequences. Using the notion of majorization, it was further shown in [8, 15] that

$$\pi \prec \alpha \quad \text{and} \quad (N^{-1}, \ldots, N^{-1}) = \frac{\pi(N)}{N} \prec \cdots \prec \frac{\pi(S)}{S} \prec \cdots \prec \pi(1) = \alpha, \tag{7}$$

where $\prec$ is defined as follows: for two real vectors $a = (a_1, \ldots, a_N)$ and $b = (b_1, \ldots, b_N)$ we write $a \prec b$, if

$$\sum_{i=1}^N a_i = \sum_{i=1}^N b_i \quad \text{and} \quad \sum_{i=k}^N a_{(i)} \leq \sum_{i=k}^N b_{(i)} \quad k = 2, \ldots N, \tag{8}$$

where $a_{(1)} \leq \cdots \leq a_{(N)}$ and $b_{(1)} \leq \cdots \leq b_{(N)}$ are a and b arranged in increasing order.
Though this generalised the conjectures stated in [5, 9], these results again only give control over extreme values and partial sums of inclusion probabilities. In particular, to the best of our knowledge, non-asymptotic control over the ratio π_i/p_i for non extremal i is missing

in the literature. In this paper we provide non-asymptotic upper and lower bounds for all i via

$$1 - \|p\|_\infty \leq \frac{\pi_i}{p_i} \leq 2. \tag{9}$$

The upper bound is a simple application of [12, Lemma 7], which states that for any event $A \subseteq \mathcal{P}([N])$ such that for all $I, J \subseteq [N]$ it holds

$$[I \in A, I \subseteq J] \implies [J \in A] \tag{10}$$

we have $\mathbb{P}_S(A) \leq 2\mathbb{P}_B(A)$. This result in itself is a generalisation of an earlier result by Hájek, [5]. Applying this to the set $A := \{I \subseteq [N] \mid i \in I\}$ yields the upper bound $\pi_i \leq 2p_i$. In some settings one is interested in the relationship between the inclusion probabilities of rejective sampling of size S and size $S-1$. Equation (7) already gives us control over the extreme points of the inclusion sequences $\pi(S)$ and $\pi(S-1)$ via the relation

$$\min_i \pi_i(S-1) \leq \min_i \pi_i(S) \cdot \frac{S-1}{S} \qquad \text{and} \qquad \max_i \pi_i(S) \cdot \frac{S-1}{S} \leq \max_i \pi_i(S-1)$$

Again, we generalise the upper bound from the extremal to all entries by showing that $\pi_i(S-1) \leq \pi_i(S)$ for all i.
Other interesting quantities are higher inclusion probabilities. We formally define the ℓ-th order inclusion probabilities for all sets with ℓ elements, that is $L = \{i_1, \dots, i_\ell\} \subseteq [N]$, as

$$\pi_L(S) := \pi_{i_1,\dots,i_\ell}(S) := \mathbb{P}_S(L \subseteq I). \tag{11}$$

Second order inclusion probabilities were already studied by Hájek, [5], who derived the asymptotic bound,

$$\pi_{i,j} = \pi_i \pi_j \left[1 - d^{-1}(1-\pi_i)(1-\pi_j) + O(d^{-1})\right] \quad \text{as} \quad d \to \infty, \tag{12}$$

which holds uniformly for all pairs i, j with $i \neq j$. This result was extended to higher order inclusion probabilities by Boistard et. al.,[3], who also sharpened the asymptotic bound to

$$\pi_{i_1,i_2,\dots,i_\ell} = \pi_{i_1}\pi_{i_2}\cdots\pi_{i_\ell}\Big[1 - d^{-1}\sum_{i,j\in L: i<j}(1-\pi_i)(1-\pi_j) + O(d^{-2})\Big] \quad \text{as} \quad d \to \infty, \tag{13}$$

which again holds uniformly in $i_1, i_2, \dots, i_\ell$. We provide non-asymptotic upper bounds for higher order inclusion probabilities in the spirit of (12) and (13), meaning

$$\pi_{L\cup M} \leq \pi_L \pi_M \quad \text{for} \quad L \cap M = \emptyset. \tag{14}$$

Our next results are motivated by applications in sparse approximation and dictionary learning, where rejective sampling is used to model non-uniform distributions of the locations of sparse supports [12, 10, 11]. A simple example for such a sparse signal model would be to fix a $d \times N$ matrix $\Phi = (\phi_1, \dots, \phi_N)$, called dictionary, and model the sparse signals y as

$$y := \Phi_I x_I := \sum_{i \in I} \phi_i x_i,$$

where x is a random vector independent of I, whose entries are i.i.d. centered random variables with unit variance. Questions about the signals quickly turn into questions about the sampling scheme and, in particular, about the matrix $\mathbb{E}[\mathbf{1}_I \mathbf{1}_I^*]$, which collects all first and second order inclusion probabilities as its diagonal resp. off-diagonal entries. For instance if we want to bound the signal spectrum, we have

$$\|\mathbb{E}[yy^*]\| = \|\Phi\, \mathbb{E}[\mathbf{1}_I \mathbf{1}_I^*]\, \Phi^*\|, \tag{15}$$

and if we want to learn the dictionary Φ from the signals y, we often encounter the weighted cross-Gram matrix between Φ and our current guess Ψ, that is $(\Psi^*\Phi) \odot \mathbb{E}[\mathbf{1}_I \mathbf{1}_I^*]$, where $\odot$ denotes the Hadamard (entrywise) product between vectors or matrices.
We derive an interesting relation between first and second order inclusion probabilities with parameters S and $S-1$, which for any $N \times N$ matrix A allows to bound the operator norm of $A \odot \mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]$ in terms involving only A and the vector of first order inclusion probabilities π. To control the full spectrum of the matrices in (15), we finally provide bounds for $\mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]$ in the positive semi-definite ordering of symmetric (Hermitian) matrices, where for two symmetric matrices A, B we have $A \preceq B$ if $B - A$ is positive semi-definite. The bounding matrices again only depend on π.

2. Main

We first provide the non-asymptotic upper and lower bound on the ratio between first order inclusion probabilities of rejective sampling π_i and corresponding weights p_i.

Lemma 1 *Let $\pi_i = \mathbb{P}_S(i \in I)$ be the inclusion probabilities associated to a rejective sampling model with parameter S and weights $p_i \in (0,1)$ with $\sum_i p_i = S$, then we have*

$$1 - \|p\|_\infty \leq \frac{\pi_i}{p_i} \leq 2.$$

Proof The upper bound follows from [12, Lemma 7] so we only need to show the lower bound $c := 1 - \|p\|_\infty \leq \pi_i / p_i$. By definition, we have

$$\pi_i = \mathbb{P}_B(i \in I \mid |I| = S) = \frac{\mathbb{P}_B(\{i \in I\} \cap \{|I| = S\})}{\mathbb{P}_B(|I| = S)} = \frac{\sum_{I:|I|=S, i\in I} \mathbb{P}_B(I)}{\sum_{I:|I|=S} \mathbb{P}_B(I)} \underbrace{\sum_J \mathbb{P}_B(J)}_{=1},$$

and $p_i = \sum_{J: i \in J} \mathbb{P}_B(J)$. So the desired inequality $c \cdot p_i \leq \pi_i$ is equivalent to

$$c \sum_{I:|I|=S} \mathbb{P}_B(I) \sum_{J:i\in J} \mathbb{P}_B(J) \leq \sum_{I:|I|=S, i \in I} \mathbb{P}_B(I) \sum_J \mathbb{P}_B(J).$$

Splitting the sum over I, J into sums over those containing i and those not containing i we see that the inequality above is implied by

$$c \sum_{I:|I|=S, i\notin I} \mathbb{P}_B(I) \sum_{J:i\in J} \mathbb{P}_B(J) \leq \sum_{I:|I|=S, i\in I} \mathbb{P}_B(I) \sum_{J:i\notin J} \mathbb{P}_B(J). \tag{16}$$

Note that for any set I not containing the index i we have

$$\frac{p_i}{1-p_i}\cdot\mathbb{P}_B(I)=\frac{p_i}{1-p_i}\prod_{k\in I}p_k\prod_{k\notin I}(1-p_k)=\prod_{k\in I\cup\{i\}}p_k\prod_{k\notin I\cup\{i\}}(1-p_k)=\mathbb{P}_B(I\cup\{i\}).$$

Multiplying both sides in 16 with $p_i/(1-p_i)$ we get

$$c\sum_{I:|I|=S+1,i\in I}\mathbb{P}_B(I)\sum_{J:i\in J}\mathbb{P}_B(J)\leq\sum_{I:|I|=S,i\in I}\mathbb{P}_B(I)\sum_{J:i\in J}\mathbb{P}_B(J),$$

so it suffices to show that

$$c\sum_{I:|I|=S+1,i\in I}\mathbb{P}_B(I)\leq\sum_{I:|I|=S,i\in I}\mathbb{P}_B(I).$$

Indeed we have

$$\begin{aligned}c\sum_{I:|I|=S+1,i\in I}\mathbb{P}_B(I)&=c\sum_{I:|I|=S+1,i\in I}\mathbb{P}_B(I)\sum_{k:k\in I,k\neq i}\frac{1}{S}\\&=c\sum_{I:|I|=S+1,i\in I}\frac{1}{S}\sum_{k:k\in I,k\neq i}\mathbb{P}_B(I\setminus\{k\})\frac{p_k}{1-p_k}\\&\leq\frac{c}{S(1-\|p\|_\infty)}\sum_{\substack{(I,k):|I|=S+1,i\in I\\k\in I,k\neq i}}\mathbb{P}_B(I\setminus\{k\})\cdot p_k\\&=\frac{1}{S}\sum_{J:|J|=S,i\in J}\mathbb{P}_B(J)\sum_{k\notin J}p_k\leq\sum_{J:|J|=S,i\in J}\mathbb{P}_B(J),\end{aligned}$$

where we used that $\sum_{k\notin J}p_k\leq\sum_k p_k=S$. ■

The last lemma tells us that as long as the weights are not too extreme, meaning $\|p\|_\infty\ll 1$, first order inclusion probabilities and weights are comparable. For instance if $\|p\|_\infty\leq 1/2$, we can §switch between the two quantities simply by multiplying with a factor 2. Before we can provide an example where this is convenient we will derive simple bounds relating high order inclusion probabilities of potentially different parameters S to each other.

Lemma 2 *Let $\pi_L(S)=\mathbb{P}_S(L\subseteq I)$ be the inclusion probabilities associated to a rejective sampling model with parameter S and weights $p_i\in(0,1)$, then we have*

$$\pi_L(S-1)\leq\pi_L(S),\tag{a}$$

$$\textit{and}\qquad\pi_{L\cup M}(S)\leq\pi_L(S)\cdot\pi_M(S)\quad\textit{if}\quad L\cap M=\emptyset.\tag{b}$$

For two indices $i\neq j$ we further have

$$\pi_{i,j}(S)=\pi_i(S)\cdot\frac{\pi_j(S-1)-\pi_{i,j}(S-1)}{1-\pi_i(S-1)}.\tag{c}$$

Proof (a) We define $\mathcal{L} = \{I \subseteq [K] : L \subseteq I\}$. Using this together with the definition of $\mathbb{P}_S$ we can rewrite $\pi_L(S-1) \leq \pi_L(S)$ as

$$\frac{\sum_{J:|J|=S-1} \mathbb{1}_{\mathcal{L}}(J) \cdot \mathbb{P}_B(J)}{\sum_{J:|J|=S-1} \mathbb{P}_B(J)} \leq \frac{\sum_{I:|I|=S} \mathbb{1}_{\mathcal{L}}(I) \cdot \mathbb{P}_B(I)}{\sum_{I:|I|=S} \mathbb{P}_B(I)},$$

which is equivalent to

$$\sum_{(I,J):|J|=S-1,|I|=S} \mathbb{1}_{\mathcal{L}}(J) \cdot \mathbb{P}_B(J)\mathbb{P}_B(I) \leq \sum_{(I,J):|J|=S-1,|I|=S} \mathbb{1}_{\mathcal{L}}(I) \cdot \mathbb{P}_B(J)\mathbb{P}_B(I).$$

Now the crucial step, which we will use several times also in the subsequent proofs, is to see that we can partition these sums in a special way. For a pair (I, J), by definition of the Poisson sampling model, we can write $\mathbb{P}_B(I)\mathbb{P}_B(J)$ in the following way

$$\mathbb{P}_B(I)\mathbb{P}_B(J) = \prod_{i\in I} p_i \prod_{j\notin I}(1-p_j) \prod_{i\in J} p_i \prod_{j\notin J}(1-p_j) = \prod_{i\in I\cap J} p_i^2 \prod_{i\in I\triangle J} p_i(1-p_i) \prod_{j\notin I\cup J}(1-p_j)^2,$$

where $I\triangle J$ denotes the symmetric difference of I, J. This implies that if for two pairs (I, J), (I', J') we have

$$I\cap J = I'\cap J' \quad \text{and} \quad I\triangle J = I'\triangle J' \quad \text{then} \quad \mathbb{P}_B(I)\mathbb{P}_B(J) = \mathbb{P}_B(I')\mathbb{P}_B(J').$$

This allows us to define natural partitions on the set of pairs (I, J) such that the probability $\mathbb{P}_B(I)\mathbb{P}_B(J)$ is constant on each partition. Concretely, for any integer $T \in \{1, \dots, S\}$, together with a set $A \subseteq \mathbb{K}$ with $|A| = S-T$ and a set $B \subseteq \mathbb{K} \setminus A$ with $|B| = 2T-1$, we look at the collection of pairs (I, J) with intersection A and symmetric difference B, that is

$$\mathcal{Q}_{A,B} := \{(I, J) : I, J \subseteq \mathbb{K}, |I| = S, |J| = S-1, I\cap J = A, I\triangle J = B\}.$$

Since each pair (I, J) with $|I| = S, |J| = S-1$ can be *uniquely* assigned to a collection $\mathcal{Q}_{A,B}$ and $\mathbb{P}(I)\mathbb{P}(J)$ is constant for all $(I, J) \in \mathcal{Q}_{A,B}$, it is sufficient to show that

$$\sum_{(I,J)\in\mathcal{Q}_{A,B}} \mathbb{1}_{\mathcal{L}}(J) \leq \sum_{(I,J)\in\mathcal{Q}_{A,B}} \mathbb{1}_{\mathcal{L}}(I)$$

or equivalently that

$$|\{(I, J) \in \mathcal{Q}_{A,B} : L \subseteq J\}| \leq |\{(I, J) \in \mathcal{Q}_{A,B} : L \subseteq I\}. \tag{17}$$

If L is not contained in $A\cup B$ there is no valid pair $(I, J) \in \mathcal{Q}_{A,B}$ and the inequality trivially holds. If $L \subseteq A\cup B$ we abbreviate $L_A = L\cap A$ and $L_B = L\cap B$. Since $L_A \subseteq A$, all pairs in $(I, J) \in \mathcal{Q}_{A,B}$ automatically satisfy $L_A \subseteq I$ and $L_A \subseteq J$ so we can rewrite (17) as

$$|\{(I, J) \in \mathcal{Q}_{A,B} : L_B \subseteq I\}| \leq |\{(I, J) \in \mathcal{Q}_{A,B} : L_B \subseteq I\}|. \tag{18}$$

Since we need to have $A\cup L_B \subseteq J$, in case $|A\cup L_B| = |A| + |L_B| > S$ the left hand side in (17) is zero and the inequality holds. Finally, if $k = |L_B| \leq S - |A| = T$, we can still choose

$T-k-1$ out of the $2T-k-1$ resp. $T-k$ out of the $2T-k-1$ remaining elements in B to fill I resp. J and create a valid pair. Since

$$\binom{2T-k-1}{T-k-1} \leq \binom{2T-k-1}{T-k}$$

which completes the proof of (a). **(a)✓**

(b) We define $\mathcal{L} = \{I \subseteq [K] : L \subseteq I\}$ and $\mathcal{M} = \{I \subseteq [K] : M \subseteq I\}$. Using this together with the definition of $\mathbb{P}_S$ we can rewrite $\pi_{L\cup M}(S) \leq \pi_L(S) \cdot \pi_M(S)$ as

$$\frac{\sum_{I:|I|=S} \mathbb{1}_{\mathcal{M}}(I) \cdot \mathbb{1}_{\mathcal{L}}(I) \cdot \mathbb{P}_B(I)}{\sum_{I:|I|=S} \mathbb{P}_B(I)} \leq \frac{\sum_{J:|J|=S} \mathbb{1}_{\mathcal{M}}(J) \cdot \mathbb{P}_B(J)}{\sum_{J:|J|=S} \mathbb{P}_B(J)} \cdot \frac{\sum_{I:|I|=S} \mathbb{1}_{\mathcal{L}}(I) \cdot \mathbb{P}_B(I)}{\sum_{I:|I|=S} \mathbb{P}_B(I)},$$

which is equivalent to

$$\sum_{(I,J):|I|=|J|=S} \mathbb{1}_{\mathcal{M}}(I) \cdot \mathbb{1}_{\mathcal{L}}(I) \cdot \mathbb{P}_B(I)\mathbb{P}_B(J) \leq \sum_{(I,J):|I|=|J|=S} \mathbb{1}_{\mathcal{M}}(J) \cdot \mathbb{1}_{\mathcal{L}}(I) \cdot \mathbb{P}_B(I)\mathbb{P}_B(J).$$

We now use a similar decomposition as before. For $T \in \{0, \dots, S\}$, $A \subseteq \mathbb{K}$ with $|A| = S-T$ and $B \subseteq \mathbb{K} \setminus A$ with $|B| = 2T$, we again let A be the intersection and B the symmetric difference of the sets I and J respectively and for any combination A, B define

$$\mathcal{Q}_{A,B} := \{(I, J) : I, J \subseteq \mathbb{K}, |I| = |J| = S, I \cap J = A, I \triangle J = B\}.$$

Since $\mathbb{P}(I)\mathbb{P}(J)$ is constant for all $(I, J) \in \mathcal{Q}_{A,B}$ and every pair (I, J) is contained in exactly one of those sets, it is sufficient to show that

$$\sum_{(I,J)\in\mathcal{Q}_{A,B}} \mathbb{1}_{\mathcal{M}}(I) \cdot \mathbb{1}_{\mathcal{L}}(I) \leq \sum_{(I,J)\in\mathcal{Q}_{A,B}} \mathbb{1}_{\mathcal{M}}(J) \cdot \mathbb{1}_{\mathcal{L}}(I)$$

or equivalently that

$$|\{(I, J) \in \mathcal{Q}_{A,B} : M \subseteq I, L \subseteq I\}| \leq |\{(I, J) \in \mathcal{Q}_{A,B} : M \subseteq J, L \subseteq I\}|.$$

If $L \cup M$ is not contained in $A \cup B$ there is no valid pair $(I, J) \in \mathcal{Q}_{A,B}$ and the inequality trivially holds. If $(L \cup M) \subseteq A \cup B$ we abbreviate $L_A = L \cap A$, $L_B = L \cap B$, $M_A = M \cap A$ and $M_B = M \cap B$. Since $(L_A \cup M_A) \subseteq A$, all pairs in $(I, J) \in \mathcal{Q}_{A,B}$ automatically satisfy $(L_A \cup M_A) \subseteq I$, $M_A \subseteq J$ and $L_A \subseteq I$ so we can rewrite the inequality we want to show as

$$|\{(I, J) \in \mathcal{Q}_{A,B} : M_B \subseteq I, L_B \subseteq I\}| \leq |\{(I, J) \in \mathcal{Q}_{A,B} : M_B \subseteq J, L_B \subseteq I\}|. \tag{19}$$

Since we need to have $(A \cup L_B \cup M_B) \subseteq I$, in case $|A \cup L_B \cup M_B| = |A| + |L_B| + |M_B| > S$ the left hand side in (19) is zero and the inequality trivially holds. Finally, if $k = |L_B| + |M_B| \leq S - |A| = T$, we can still choose $T-k$ out of the $2T-k$ resp. $T - |L_B|$ out of the $2T-k$ remaining elements in B to fill I resp. J and create a valid pair. Since $|L_B| \leq k$, we have

$$\binom{2T-k}{T-k} \leq \binom{2T-k}{T-|L_B|},$$

meaning the inequality in (19) is again satisfied, which completes the proof of (b). **(b)**✓

(c) We want to show that $[1-\pi_i(S-1)]\cdot\pi_{i,j}(S) = \pi_i(S)\cdot[\pi_j(S-1)-\pi_{i,j}(S-1)]$. Recalling that for any set J not containing the index i we have

$$\frac{p_i}{1-p_i}\cdot\mathbb{P}_B(J) = \mathbb{P}_B(J\cup\{i\}),$$

we get

$$\begin{aligned}
\pi_{i,j}(S) &= \frac{\sum_{I:|I|=S}\mathbb{1}_I(i)\mathbb{1}_I(j)\cdot\mathbb{P}_B(I)}{\sum_{I:|I|=S}\mathbb{P}_B(I)}\cdot\frac{\sum_{I:|I|=S,i\in I}\mathbb{P}_B(I)}{\sum_{I:|I|=S,i\in I}\mathbb{P}_B(I)}\\
&= \frac{\sum_{I:|I|=S,i\in I}\mathbb{1}_I(j)\cdot\mathbb{P}_B(I)}{\sum_{|I|=S,i\in I}\mathbb{P}_B(I)}\cdot\pi_i(S)\\
&= \frac{p_i}{1-p_i}\cdot\frac{1-p_i}{p_i}\cdot\frac{\sum_{J:|J|=S-1,i\notin J}\mathbb{1}_J(j)\cdot\mathbb{P}_B(J)}{\sum_{J:|J|=S-1,i\notin J}\mathbb{P}_B(J)}\cdot\pi_i(S)\cdot\frac{\sum_{J:|J|=S-1}\mathbb{P}_B(J)}{\sum_{J:|J|=S-1}\mathbb{P}_B(J)}\\
&= \pi_i(S)\cdot\frac{\sum_{J:|J|=S-1,i\notin J}\mathbb{1}_J(j)\cdot\mathbb{P}_B(J)}{\sum_{J:|J|=S-1}\mathbb{P}_B(J)}\cdot\frac{\sum_{J:|J|=S-1}\mathbb{P}_B(J)}{\sum_{J:|J|=S-1,i\notin J}\mathbb{P}_B(J)}.
\end{aligned}$$

Further rewriting the fractions in the expression above yields

$$\frac{\sum_{J:|J|=S-1,i\notin J}\mathbb{1}_J(j)\cdot\mathbb{P}_B(J)}{\sum_{J:|J|=S-1}\mathbb{P}_B(J)} = \underbrace{\frac{\sum_{J:|J|=S-1}\mathbb{1}_J(j)\cdot\mathbb{P}_B(I)}{\sum_{J:|J|=S-1}\mathbb{P}_B(J)}}_{\pi_j(S-1)} - \underbrace{\frac{\sum_{J:|J|=S-1}\mathbb{1}_J(i)\mathbb{1}_J(j)\cdot\mathbb{P}_B(J)}{\sum_{J:|J|=S-1}\mathbb{P}_B(J)}}_{\pi_{i,j}(S-1)}$$

as well as

$$\frac{\sum_{J:|J|=S-1}\mathbb{P}_B(J)}{\sum_{J:|J|=S-1,i\notin J}\mathbb{P}_B(J)} = \frac{\sum_{J:|J|=S-1}\mathbb{P}_B(J)}{\sum_{J:|J|=S-1}\mathbb{P}_B(J)-\sum_{J:|J|=S-1,i\in J}\mathbb{P}_B(J)} = \frac{1}{1-\pi_i(S-1)},$$

which completes the proof of (c). **(c)**✓

■

The last statement of the lemma might seem rather arbitrary, however it leads to the following convenient way of bounding quantities such as $\|(\Psi^*\Phi)\odot\mathbb{E}[\mathbf{1}_I\mathbf{1}_I^*]\|$.

Theorem 3 *Let $\mathbb{E}_S$ be the expectation according to the rejective sampling probability with parameter S and weights $p_i\in(0,1)$. Further let $\pi\in\mathbb{R}^N$ be the vector of first order inclusion probabilities and D_π be the $N\times N$ matrix with π on the diagonal and zero else. Then for any $N\times N$ matrix A we have*

$$\|A\odot\mathbb{E}_S[\mathbf{1}_I\mathbf{1}_I^*]\| \le \frac{1+\|\pi\|_\infty}{(1-\|\pi\|_\infty)^2}\cdot\|D_\pi[A-\operatorname{diag}(A)]D_\pi\| + \|\operatorname{diag}(A)D_\pi\|.$$

In order to prove the theorem, we need the following corollary of Schur's product theorem. For convenience we include its short proof.

Corollary 4 *Let A and B be two square matrices of the same dimension. If A is positive-semidefinite (p.s.d.), then*

$$\|A\odot B\| \le \|\operatorname{diag}(A)\|\cdot\|B\|.$$

Proof The matrix

$$\begin{pmatrix} \|B\| \cdot (\mathbb{I} \odot A) & A \odot B \\ (A \odot B)^* & \|B\| \cdot (\mathbb{I} \odot A) \end{pmatrix} = \begin{pmatrix} A & A \\ A & A \end{pmatrix} \odot \begin{pmatrix} \|B\| \cdot \mathbb{I} & B \\ B^* & \|B\| \cdot \mathbb{I} \end{pmatrix}$$

is p.s.d., since the right hand side of the equation is a Hadamard product of two p.s.d. matrices which is by Schur's product theorem also p.s.d. By Theorem 7.7.9 in [7] there thus exists a contraction C, meaning $\|C\| \leq 1$, such that

$$A \odot B = \|B\| \, (\mathbb{I} \odot A)^{1/2} \, C \, (\mathbb{I} \odot A)^{1/2},$$

and hence $\|A \odot B\| \leq \|\mathbb{I} \odot A\| \cdot \|B\| = \|\operatorname{diag}(A)\| \cdot \|B\|$. ■

Proof [of Theorem 3] We first note that since $\mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]$ has $\pi(S)$ on the diagonal, a simple application of the triangle inequality yields

$$\begin{aligned} \|A \odot \mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]\| &\leq \|(A - \operatorname{diag}(A)) \odot \mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]\| + \|\operatorname{diag}(A) \odot \mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]\| \\ &= \|(A - \operatorname{diag}(A)) \odot \mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]\| + \|\operatorname{diag}(A) D_{\pi(S)}\|, \end{aligned}$$

which already proves the theorem for $S = 1$, where all off-diagonal entries of $\mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]$ are zero, meaning the first norm term is zero. In case $S \geq 2$ it remains to show that for $H = A - \operatorname{diag}(A)$ we have $\|H \odot \mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]\| \leq c \cdot \|D_{\pi(S)} H D_{\pi(S)}\|$ with constant c as above. Using the abbreviation $\Delta_{\pi(S)} = I - D_{\pi(S)}$ we know from Lemma 2(c) that

$$\begin{aligned} (H \odot \mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*])_{ij} = H_{ij} \cdot \pi_{i,j}(S) &= \frac{\pi_i(S)}{1 - \pi_i(S-1)} \cdot H_{ij} \cdot [\pi_j(S-1) - \pi_{i,j}(S-1)] \\ &= \left(\Delta^{-1}_{\pi(S-1)} D_{\pi(S)} H D_{\pi(S-1)} - \Delta^{-1}_{\pi(S-1)} D_{\pi(S)} H \odot \mathbb{E}_{S-1}[\mathbf{1}_I \mathbf{1}_I^*] \right)_{ij}. \end{aligned}$$

Next Lemma 2(a) tells us that $D_{\pi(S-1)} \preceq D_{\pi(S)}$, which leads to

$$\begin{aligned} \|H \odot \mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]\| &\leq \|\Delta^{-1}_{\pi(S-1)}\| \cdot \|D_{\pi(S)} H D_{\pi(S-1)} - D_{\pi(S)} H \odot \mathbb{E}_{S-1}[\mathbf{1}_I \mathbf{1}_I^*]\| \\ &\leq (1 - \|\pi(S)\|_\infty)^{-1} \cdot \left(\|D_{\pi(S)} H D_{\pi(S)}\| + \|D_{\pi(S)} H \odot \mathbb{E}_{S-1}[\mathbf{1}_I \mathbf{1}_I^*]\| \right). \end{aligned} \tag{20}$$

Applying the inequality above to H^* and using the symmetry of $\mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]$ we also get

$$\|H \odot \mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]\| \leq (1 - \|\pi(S)\|_\infty)^{-1} \cdot \left(\|D_{\pi(S)} H D_{\pi(S)}\| + \|H D_{\pi(S)} \odot \mathbb{E}_{S-1}[\mathbf{1}_I \mathbf{1}_I^*]\| \right). \tag{21}$$

For $S = 2$, the matrix $\mathbb{E}_{S-1}[\mathbf{1}_I \mathbf{1}_I^*]$ is again a diagonal matrix, meaning the second norm term vanishes and we are done. For $S > 2$ we simply apply the inequality in (21) to $\bar{H} \odot \mathbb{E}_{S-1}[\mathbf{1}_I \mathbf{1}_I^*]$ with $\bar{H} = D_{\pi(S)} H$, leading to

$$\begin{aligned} \|D_{\pi(S)} H \odot \mathbb{E}_{S-1}[\mathbf{1}_I \mathbf{1}_I^*]\| &= \|\bar{H} \odot \mathbb{E}_{S-1}[\mathbf{1}_I \mathbf{1}_I^*]\| \\ &\leq (1 - \|\pi(S-1)\|_\infty)^{-1} \cdot \left(\|D_{\pi(S-1)} \bar{H} D_{\pi(S-1)}\| + \|\bar{H} D_{\pi(S-1)} \odot \mathbb{E}_{S-2}[\mathbf{1}_I \mathbf{1}_I^*]\| \right) \\ &\leq (1 - \|\pi(S)\|_\infty)^{-1} \cdot \left(\|\pi(S)\|_\infty \|D_{\pi(S)} H D_{\pi(S)}\| + \|D_{\pi(S)} H D_{\pi(S)} \odot \mathbb{E}_{S-2}[\mathbf{1}_I \mathbf{1}_I^*]\| \right). \end{aligned}$$

Inserting the inequality above into (21) yields

$$\|H \odot \mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]\| \le (1 - \|\pi(S)\|_\infty)^{-2} \cdot \left(\|D_{\pi(S)} H D_{\pi(S)}\| + \|D_{\pi(S)} H D_{\pi(S)} \odot \mathbb{E}_{S-2}[\mathbf{1}_I \mathbf{1}_I^*]\|\right)$$

and since $\mathbb{E}_{S-2}[\mathbf{1}_I \mathbf{1}_I^*]$ has $\pi(S-2)$ on its diagonal the result follows from Corollary 4 and Lemma 2(a). ■

The theorem allows for instance to bound the weighted Gram matrix of a dictionary Φ using only the first order inclusion probabilities π or alternatively the weights p, as

$$\begin{aligned}\|(\Phi^*\Phi) \odot \mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]\| &\le \frac{1 + \|\pi\|_\infty}{(1 - \|\pi\|_\infty)^2} \left(\|\Phi D_\pi\|^2 + 2\|\pi\|_\infty\right) \\ &\le \frac{1 + 2\|p\|_\infty}{(1 - 2\|p\|_\infty)^2} \left(4\|\Phi D_p\|^2 + 4\|p\|_\infty\right).\end{aligned}$$

If we want to control not only the largest singular value of the weighted Gram matrix, but the full spectrum, we need the following result, which provides bounds for $\mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*]$ in the semi-definite order.

Lemma 5 *Let $\mathbb{E}_S$ be the expectation according to the rejective sampling probability with parameter S and weights $p_i \in (0,1)$. Further let $\pi \in \mathbb{R}^N$ be the vector of first order inclusion probabilities and D_π be the $N \times N$ matrix with π on the diagonal and zero else. Then we have*

$$\mathbb{E}_S[\mathbf{1}_I \mathbf{1}_I^*] \preceq \pi\pi^* + 2D_\pi. \tag{a}$$

Further, defining for $L \subseteq [K]$ with $|L| < S$ the set $\mathcal{L} := \{I \subseteq [K] : L \subseteq I\}$, we have

$$\mathbb{E}_S\left[\mathbf{1}_{I\setminus L} \mathbf{1}_{I\setminus L}^* \cdot \mathbb{1}_{\mathcal{L}}(I)\right] \preceq \mathbb{E}_{S-|L|}[\mathbf{1}_I \mathbf{1}_I^*] \cdot \prod_{\ell \in L} \frac{\pi_\ell}{1 - \pi_\ell}. \tag{b}$$

Proof **(a)** We want to show that $\mathbb{E}_S(\mathbf{1}_I \mathbf{1}_I^*) \preceq \pi\pi^* + 2D_\pi$ or equivalently

$$\frac{\sum_{I:|I|=S} \mathbf{1}_I \mathbf{1}_I^* \mathbb{P}_B(I)}{\sum_{I:|I|=S} \mathbb{P}_B(I)} \preceq \frac{\sum_{(I,J):|I|=|J|=S} \mathbf{1}_I \mathbf{1}_J^* \mathbb{P}_B(I)\mathbb{P}_B(J)}{(\sum_{I:|I|=S} \mathbb{P}_B(I))^2} + \frac{2 \cdot \sum_{I:|I|=S} \operatorname{diag}(\mathbf{1}_I)\mathbb{P}_B(I)}{\sum_{I:|I|=S} \mathbb{P}_B(I)}.$$

Multiplying both sides by $(\sum_{I:|I|=S} \mathbb{P}_B(I))^2$ we therefore have to show that

$$\begin{aligned}\sum_{(I,J):|I|=|J|=S} \mathbf{1}_I \mathbf{1}_I^* \mathbb{P}_B(I)\mathbb{P}_B(J) \preceq &\sum_{(I,J):|I|=|J|=S} \mathbf{1}_I \mathbf{1}_J^* \mathbb{P}_B(I)\mathbb{P}_B(J) \\ &+ 2 \sum_{(I,J):|I|=|J|=S} \operatorname{diag}(\mathbf{1}_I)\mathbb{P}_B(I)\mathbb{P}_B(J).\end{aligned} \tag{22}$$

We now use the same partition as in the proof of Lemma 2(b), that is, for $T \in \{0, \dots, S\}$, $A \subseteq \mathbb{K}$ with $|A| = S - T$ and $B \subseteq \mathbb{K} \setminus A$ with $|B| = 2T$, we define

$$\mathcal{Q}_{A,B} := \{(I,J) : I, J \subseteq \mathbb{K}, |I| = |J| = S, I \cap J = A, I \triangle J = B\}.$$

Since the sum of positive semi-definite matrices is positive semi-definite it suffices to show that for all possible choices of A, B we have

$$\sum_{(I,J)\in\mathcal{Q}_{A,B}} \mathbf{1}_I\mathbf{1}_I^* \preceq \sum_{(I,J)\in\mathcal{Q}_{A,B}} \mathbf{1}_I\mathbf{1}_J^* + 2\sum_{(I,J)\in\mathcal{Q}_{A,B}} \operatorname{diag}(\mathbf{1}_I).$$

For A, B fixed we abbreviate $Q = \sum_{(I,J)\in\mathcal{Q}_{A,B}} \mathbf{1}_I\mathbf{1}_I^*$ and $\bar{Q} = \sum_{(I,J)\in\mathcal{Q}_{A,B}} \mathbf{1}_I\mathbf{1}_J^*$. Note that $\sum_{(I,J)\in\mathcal{Q}_{A,B}} \operatorname{diag}(\mathbf{1}_I) = \operatorname{diag}(Q)$, so the inequality above is equivalent to showing that

$$0 \preceq \bar{Q} - Q + 2\operatorname{diag}(Q). \tag{23}$$

For the entries of these matrices we have

$$Q_{ij} = |\{(I,J)\in\mathcal{Q}_{A,B} : i,j\in I\}| \quad \text{resp.} \quad \bar{Q}_{ij} = |\{(I,J)\in\mathcal{Q}_{A,B} : i\in I, j\in J\}|.$$

In case $i, j \in A$ we obviously have $Q_{ij} = \bar{Q}_{ij} = |\mathcal{Q}_{A,B}|$, so

$$Q\odot(\mathbf{1}_A\mathbf{1}_A^*) = \bar{Q}\odot(\mathbf{1}_A\mathbf{1}_A^*).$$

In particular, this means that (23) holds trivially for $T = 0$, where $B = \emptyset$.
In case that $i \in A, j \in B$ we have $Q_{ij} = \bar{Q}_{ij} = \binom{2T-1}{T-1} =: d_T$ and therefore

$$Q\odot(\mathbf{1}_A\mathbf{1}_B^* + \mathbf{1}_B\mathbf{1}_A^*) = \bar{Q}\odot(\mathbf{1}_A\mathbf{1}_B^* + \mathbf{1}_B\mathbf{1}_A^*).$$

It only remains to check what happens for $i, j \in B$. The case $T = 0$ is already settled, thus we assume $T \geq 1$. On the diagonal we get $\bar{Q}_{ii} = 0$ while $Q_{ii} = \binom{2T-1}{T-1} = d_T$. We have $Q_{ij} = \binom{2T-2}{T-2} =: q_T$ and $\bar{Q}_{ij} = \binom{2T-2}{T-1} =: \bar{q}_T$. In summary

$$\bar{Q} - Q + 2\operatorname{diag}(Q) = (\bar{q}_T - q_T)\cdot\mathbf{1}_B\mathbf{1}_B^* - (\bar{q}_T - q_T + d_T)\cdot\operatorname{diag}(\mathbf{1}_B) + 2\operatorname{diag}(Q). \tag{24}$$

Since $\bar{q}_T \geq q_T$ the matrix $(\bar{q}_T - q_T)\cdot\mathbf{1}_B\mathbf{1}_B^*$ is positive semi-definite. Finally, as $\binom{2T-2}{T-2} + \binom{2T-2}{T-1} = \binom{2T-1}{T-1}$ we have $\bar{q}_T + q_T = d_T$. Thus for all $T \geq 1$, we have

$$\begin{aligned}(\bar{q}_T - q_T + d_T)\operatorname{diag}(\mathbf{1}_B) &= 2\cdot\bar{q}_T\operatorname{diag}(\mathbf{1}_B)\\ &\preceq 2\cdot d_T\operatorname{diag}(\mathbf{1}_B) = 2\operatorname{diag}(Q)\operatorname{diag}(\mathbf{1}_B) \preceq 2\operatorname{diag}(Q),\end{aligned}$$

showing that also the remaining terms in (24) are positive semi-definite, which completes the proof of (a). **(a)**✓

(b) We will prove the statement by induction. Let $\hat{L}$ be a set of size $T \leq S-2$ and $\hat{\mathcal{L}} = \{I \subseteq [K] : \hat{L} \subseteq I\}$. We first show that for $k \notin \hat{L}$ and $L = \hat{L}\cup\{k\}$ we have

$$(1-\pi_k(S))\cdot\mathbb{E}_S\big[\mathbf{1}_{I\setminus L}\mathbf{1}_{I\setminus L}^*\cdot\mathbb{1}_{\mathcal{L}}(I)\big] \preceq \pi_k(S)\cdot\mathbb{E}_{S-1}\big[\mathbf{1}_{I\setminus\hat{L}}\mathbf{1}_{I\setminus\hat{L}}^*\cdot\mathbb{1}_{\hat{\mathcal{L}}}(I)\big]. \tag{25}$$

We again use that for any set J not containing the index k we have

$$\frac{p_k}{1-p_k}\cdot\mathbb{P}_B(J) = \mathbb{P}_B(J\cup\{k\}).$$

Thus expanding the expectation we get

$$\begin{aligned}
\mathbb{E}_S\big[\mathbf{1}_{I\setminus L}\mathbf{1}_{I\setminus L}^* \cdot \mathbb{1}_{\mathcal{L}}(I)\big] &= \frac{\sum_{I:|I|=S, L\subseteq I} \mathbb{P}_B(I)(\mathbf{1}_{I\setminus L}\mathbf{1}_{I\setminus L}^*)}{\sum_{I:|I|=S} \mathbb{P}_B(I)} \cdot \frac{\sum_{I:|I|=S,k\in I} \mathbb{P}_B(I)}{\sum_{I:|I|=S,k\in I} \mathbb{P}_B(I)} \\
&= \frac{\sum_{I:|I|=S, L\subseteq I} \mathbb{P}_B(I)(\mathbf{1}_{I\setminus L}\mathbf{1}_{I\setminus L}^*)}{\sum_{I:|I|=S,k\in I} \mathbb{P}_B(I)} \cdot \frac{\sum_{I:|I|=S,k\in I} \mathbb{P}_B(I)}{\sum_{I:|I|=S} \mathbb{P}_B(I)} \\
&= \frac{\sum_{J:|J|=S-1, k\notin J, \hat{L}\subseteq J} \mathbb{P}_B(J)(\mathbf{1}_{J\setminus \hat{L}}\mathbf{1}_{J\setminus \hat{L}}^*)}{\sum_{J:|J|=S-1,k\notin J} \mathbb{P}_B(J)} \cdot \pi_k(S) \\
&\preceq \frac{\sum_{J:|J|=S-1, \hat{L}\subseteq J} \mathbb{P}_B(J)(\mathbf{1}_{J\setminus \hat{L}}\mathbf{1}_{J\setminus \hat{L}}^*)}{\sum_{J:|J|=S-1,k\notin J} \mathbb{P}_B(J)} \cdot \frac{\sum_{I:|I|=S-1} \mathbb{P}_B(I)}{\sum_{I:|I|=S-1} \mathbb{P}_B(I)} \cdot \pi_k(S) \\
&= \frac{\sum_{J:|J|=S-1, \hat{L}\subseteq J} \mathbb{P}_B(J)(\mathbf{1}_{J\setminus \hat{L}}\mathbf{1}_{J\setminus \hat{L}}^*)}{\sum_{I:|I|=S-1} \mathbb{P}_B(I)} \cdot \frac{\sum_{I:|I|=S-1} \mathbb{P}_B(I)}{\sum_{J:|J|=S-1,k\notin J} \mathbb{P}_B(J)} \cdot \pi_k(S) \\
&= \mathbb{E}_{S-1}\big[\mathbf{1}_{I\setminus \hat{L}}\mathbf{1}_{I\setminus \hat{L}}^* \cdot \mathbb{1}_{\hat{\mathcal{L}}}(I)\big] \cdot \frac{\sum_{I:|I|=S-1} \mathbb{P}_B(I)}{\sum_{J:|J|=S-1,k\notin J} \mathbb{P}_B(J)} \cdot \pi_k(S).
\end{aligned}$$

Now all that remains to do in order to prove (25) is to bound the fraction above. Writing out the expression in the denominator we get

$$\begin{aligned}
\frac{\sum_{I:|I|=S-1} \mathbb{P}_B(I)}{\sum_{J:|J|=S-1,k\notin J} \mathbb{P}_B(J)} &= \frac{\sum_{I:|I|=S-1} \mathbb{P}_B(I)}{\sum_{I:|I|=S-1} \mathbb{P}_B(I) - \sum_{I:|I|=S-1,k\in I} \mathbb{P}_B(I)} \\
&= \frac{1}{1-\mathbb{P}_{S-1}(k\in I)} \leq \frac{1}{1-\mathbb{P}_{S}(k\in I)} = \frac{1}{1-\pi_k(S)}.
\end{aligned}$$

By induction and using again the bound from Lemma 2(a) that $\pi_k(S-1) \leq \pi_k(S)$ we finally get

$$\mathbb{E}_S\big[\mathbf{1}_{I\setminus L}\mathbf{1}_{I\setminus L}^* \cdot \mathbb{1}_{\mathcal{L}}(I)\big] \prod_{\ell\in L}(1-\pi_\ell(S)) \preceq \mathbb{E}_{S-|L|}[\mathbf{1}_I\mathbf{1}_I^*] \cdot \prod_{\ell\in L}\pi_\ell(S).$$

which completes the proof of (b). **(b)✓**

■

Again we give an application example for the derived result. If we have a collection of sparse signals y, whose supports follow a rejective sampling model, we know from (15) that

$$\begin{aligned}
\|\mathbb{E}[yy^*]\| = \|\Phi\mathbb{E}_S[\mathbf{1}_I\mathbf{1}_I^*]\Phi^*\| &\leq \|\Phi\pi\pi^*\Phi^*\| + 2\|\Phi D_\pi \Phi^*\| \\
&= \|\Phi\pi\|_2^2 + 2\|\Phi D_{\sqrt{\pi}}\|^2 \leq (S+2)\|\Phi D_{\sqrt{\pi}}\|^2,
\end{aligned}$$

where in the last inequality we have used that $\Phi\pi = \Phi D_{\sqrt{\pi}}\sqrt{\pi}$ and that $\|\sqrt{\pi}\|_2^2 = \|\pi\|_1 = S$.

3. Discussion

We have derived non-asymptotic bounds for inclusion probabilities and matrices that collect (conditional) first and second order inclusion probabilities as their diagonal resp. off-diagonal entries. Most results are motivated by problems in sparse modelling and dictionary learning and so we have provided example applications throughout the text. More applications can for instance be found in [11, 13], where we derive convergence results for two popular dictionary learning algorithms (MOD and K-SVD) under the rejective sampling model. However, we think that the proof-techniques developed in this text are of independent interest, as they provide an easy way to analyse the relation between rejective and Poisson sampling in the non-asymptotic regime.

Acknowledgments

This work was supported by the Austrian Science Fund (FWF) under Grant no. Y760.

References

[1] N. Aires. Algorithms to find exact inclusion probabilities for conditional Poisson sampling and Pareto πps sampling designs. *Methodology and Computing in Applied Probability*, 1(4):457–469, 1999.

[2] P. Bertail and S. Clémençon. Sharp exponential inequalities in survey sampling: conditional Poisson sampling schemes. *Bernoulli*, 25, 2016.

[3] H. Boistard, H. P. Lopuhaä, and A. Ruiz-Gazen. Approximation of rejective sampling inclusion probabilities and application to high order correlations. *Electronic Journal of Statistics*, 6:1967–1983, 2012.

[4] J. Dupačová. A note on rejective sampling. In J. Jurečková, editor, *Contributions to Statistics (J. Hájek Memorial Volume)*, pages 72–78. Reidel: Holland and Academia: Prague, 1979.

[5] J. Hájek. Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics*, 35(4):1491–1523, 1964.

[6] J. Hájek. *Sampling from a Finite Population.* Marcel Dekker Inc., New York, 1981.

[7] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge University Press, New York, 2nd edition, 2013.

[8] S. Kochar and R. Korwar. On random sampling without replacement from a finite population. *Annals of the Institute of Statistical Mathematics*, 53(3):631–646, 2001.

[9] H. Milbrodt. Comparing inclusion probabilities and drawing probabilities for rejective sampling and successive sampling. *Statistics and Probability Letters*, 14(3):243–246, 1992.

[10] S. Ruetz. Adapted variable density subsampling for compressed sensing. *arXiv:2206.13796*, 2022.

[11] S. Ruetz. *Compressed Sensing and Dictionary Learning with Non-Uniform Support Distribution.* PhD thesis, University of Innsbruck, 2022.

[12] S. Ruetz and K. Schnass. Submatrices with non-uniformly selected random supports and insights into sparse approximation. *SIAM Journal on Matrix Analysis and Applications*, 42(3):1268–1289, 2021.

[13] S. Ruetz and K. Schnass. Convergence of MOD and aK-SVD for dictionary learning. *in preparation*, 2023.

[14] P. K. Sen. The Hájek asymptotics for finite population sampling and their ramifications. *Kybernetika*, 31(3):251–268, 1995.

[15] Y. Yaming. On the inclusion probabilities in some unequal probability sampling plans without replacement. *Bernoulli*, 18(1):279–289, 2012.