# Homework 6 Selected Solutions
# APPM 5650 Fall 2021
# Randomized Algorithms

**Due date**: Monday, Oct 4 2021 — **Instructor**: Prof. Becker
**Theme**: Reading and sampling — **Revision date**: 9/30/2021

**Problem 1:** [READING] Read sections 1 ("Overview") and 2 ("Related Work and Historical Context") from "Finding structure with randomness: Stochastic algorithms for constructing approximate matrix decompositions" by Halko, Martinsson and Tropp (SIAM Review 2011) available at `http://users.cms.caltech.edu/~jtropp/papers/HMT11-Finding-Structure-SIREV.pdf`

**Deliverable**: Write at least 1 paragraph about your response to the reading, e.g., what did you find interesting? Or, according to the authors, when did randomized methods start to have an impact on numerical linear algebra? What do the authors claim is responsible for some of the resistance in adopting randomized techniques?

**Problem 2:** [MATH] Sampling schemes. We would like to sample roughly $k$ items out of $n$. We'll consider only sampling without replacement schemes for this problem, so our samples $\Omega$ can be written as a set (we don't need an ordered list). We'll consider two models:

- SRS (simple random sample, without replacement) which is uniformly sampling all possible subsets of size $k$. We write $\mathbb{P}_{\mathrm{SRS}(k)}[\cdot]$ to denote sampling under this model. The sampled set $\Omega$ is always of size $k$.
- Bernoulli sampling with rate $\rho \approx k/n$. Each entry of $[n]$ is included in $\Omega$ with probability $\rho$, all done independently of the other entries, so $\mathbb{E}[|\Omega|] = \rho n \approx k$. We write $\mathbb{P}_{\mathrm{Ber}(\rho)}[\cdot]$ to denote sampling under this model.

We consider some randomized algorithm, and we wish for it to be successful (meaning, say, that the error is below some given threshold). We let `success` be the event that the algorithm is successful. Our main assumption is that more samples in $\Omega$ is better: specifically, that

$$\mathbb{P}_{\mathrm{SRS}(k)}[\texttt{success}] \le \mathbb{P}_{\mathrm{SRS}(k')}[\texttt{success}] \quad \text{if } \; k \le k' \tag{Assumption}$$

We'll show that if we can control the probability of success under the Bernoulli scheme, then we can also control it under the SRS scheme (hence we could analyze the Bernoulli scheme but do the SRS scheme in practice and still have similar guarantees). Specifically, let $\Omega$ be a sampled set drawn under the Bernoulli scheme with parameter $\rho$, and do the following:

a) Explain why $\mathbb{P}_{\mathrm{Ber}(\rho)}\left[\texttt{success} \,\middle|\, |\Omega| = k\right] = \mathbb{P}_{\mathrm{SRS}(k)}[\texttt{success}]$

**Solution:**

A simple random sample means that all subsets of size $k$ are equally likely. Clearly Bernoulli sampling doesn't favor any one element over another, so since it has size $k$ (due to the conditional probability), it's a simple random sample.

b) Under (Assumption), show that

$$\mathbb{P}_{\mathrm{SRS}(k)}[\texttt{success}] \ge \mathbb{P}_{\mathrm{Ber}(\rho)}[\texttt{success}] - \mathbb{P}_{\mathrm{Ber}(\rho)}[\,|\Omega| > k\,].$$

*Hint*: use the Law of Total Probability, conditioning on the size of $\Omega$.

**Solution:**

$$
\begin{aligned}
\mathbb{P}_{\mathrm{Ber}(\rho)}\left[\texttt{success}\right] &= \sum_{j=0}^{n} \mathbb{P}_{\mathrm{Ber}(\rho)}\left[\texttt{success} \,\middle|\, |\Omega| = j)\right] \cdot \mathbb{P}_{\mathrm{Ber}(\rho)}\left[\,|\Omega| = j\right] \\
&= \sum_{j=0}^{n} \mathbb{P}_{\mathrm{SRS}(j)}\left[\texttt{success}\right] \cdot \mathbb{P}_{\mathrm{Ber}(\rho)}\left[\,|\Omega| = j\right] \text{ via the previous problem} \\
&= \sum_{j=k+1}^{n} \underbrace{\mathbb{P}_{\mathrm{SRS}(j)}\left[\texttt{success}\right]}_{\leq 1} \cdot \mathbb{P}_{\mathrm{Ber}(\rho)}\left[\,|\Omega| = j\right] + \sum_{j=0}^{k} \mathbb{P}_{\mathrm{SRS}(j)}\left[\texttt{success}\right] \cdot \mathbb{P}_{\mathrm{Ber}(\rho)}\left[\,|\Omega| = j\right] \\
&\leq \sum_{j=k+1}^{n} \mathbb{P}_{\mathrm{Ber}(\rho)}\left[\,|\Omega| = j\right] + \sum_{j=0}^{k} \mathbb{P}_{\mathrm{SRS}(j)}\left[\texttt{success}\right] \cdot \mathbb{P}_{\mathrm{Ber}(\rho)}\left[|\Omega| = j\right] \\
&= \mathbb{P}_{\mathrm{Ber}(\rho)}\left[\,|\Omega| > k\right] + \sum_{j=0}^{k} \mathbb{P}_{\mathrm{SRS}(j)}\left[\texttt{success}\right] \cdot \mathbb{P}_{\mathrm{Ber}(\rho)}\left[|\Omega| = j\right] \\
&\leq \mathbb{P}_{\mathrm{Ber}(\rho)}\left[\,|\Omega| > k\right] + \mathbb{P}_{\mathrm{SRS}(k)}\left[\texttt{success}\right] \underbrace{\sum_{j=0}^{k} \cdot \mathbb{P}_{\mathrm{Ber}(\rho)}\left[|\Omega| = j\right]}_{\mathbb{P}_{\mathrm{Ber}(\rho)}[\,|\Omega| \leq k] \leq 1} \\
&\leq \mathbb{P}_{\mathrm{Ber}(\rho)}\left[\,|\Omega| > k\right] + \mathbb{P}_{\mathrm{SRS}(k)}\left[\texttt{success}\right]
\end{aligned}
$$

which, after re-arranging, gives the desired result.

c) For any $\delta > 0$, set $\epsilon = \sqrt{\frac{\ln(\delta^{-1})}{2n}}$ and set $\rho = k/n - \epsilon$, so we expect $\Omega$ to have size $k + \epsilon n$. Show $\mathbb{P}_{\mathrm{SRS}(k)}\left[\texttt{success}\right] \geq \mathbb{P}_{\mathrm{Ber}(\rho)}\left[\texttt{success}\right] - \delta$. *Hint*: use a concentration inequality.

**Solution:**

Let $\rho = k/n + \epsilon$, then

$$
\begin{aligned}
\mathbb{P}_{\mathrm{Ber}(\rho)}\left[\,|\Omega| > k\right] &= \mathbb{P}_{\mathrm{Ber}(\rho)}\left[\frac{1}{n}|\Omega| > \frac{k}{n}\right] \\
&= \mathbb{P}_{\mathrm{Ber}(\rho)}\left[\frac{1}{n}|\Omega| > \rho + \epsilon\right] \\
&\leq \exp(-2n\epsilon^2) \quad \text{via Hoeffding's inequality} \\
&= \delta
\end{aligned}
$$

and then apply the previous problem. Specifically, to apply Hoeffding, we let $X_i$ be 1 if we select entry $i$ (so this is distributed $\sim$ Bernouilli$(\rho)$), and so $(X_i)$ is iid, and furthermore $|\Omega| = \sum_{i=1}^{n} X_i$.

[NO WORK NECESSARY] Note that similar arguments can also be used to show that $\mathbb{P}_{\mathrm{Ber}(\rho)}\left[\texttt{success}\right] \geq \mathbb{P}_{\mathrm{SRS}(k)}\left[\texttt{success}\right](1-\delta)$ for $k = n\rho - \epsilon$ for a similar $\epsilon$ to the previous problem, meaning that if you analyzed SRS sampling but wanted to do Bernouilli sampling in practice, you also get guarantees.

This problem somewhat follows the Appendix in *Robust Principal Component Analysis?* by Candès, Li, Ma and Wright 2009 (but don't look *before* attempting the above problems; also, we fix a typo and hence need to adjust several results).