

Homework 1 Sample Solutions

Problem 1: Optimal Homework Strategies

Write $s = k/3$ and $\Delta = p - q$. The number of assignments receiving score p is then $H \sim \text{Bin}(n, s)$. Let T_r denote the total after dropping the r lowest grades.

(a) Since $T_0 = nq + \Delta H$ and $\mathbb{E}H = ns$,

$$\mathbb{E}T_0 = n(q + \Delta s) = n \left(\frac{k}{3}p + \left(1 - \frac{k}{3}\right)q \right).$$

(b) Dropping one grade removes a q unless every grade is p . Thus

$$T_1 = (n-1)q + \Delta(H - \mathbf{1}_{\{H=n\}}), \quad \mathbb{E}T_1 = (n-1)q + \Delta(ns - s^n).$$

For $n \geq 2$, dropping two grades removes one p when $H = n-1$, two when $H = n$, and none otherwise. Therefore

$$\mathbb{E}T_2 = (n-2)q + \Delta(ns - n(1-s)s^{n-1} - 2s^n).$$

(c) Set $I = \mathbf{1}_{\{H=n\}}$. We have $\mathbb{E}I = s^n$ and $HI = nI$, so $\text{Cov}(H, I) = n(1-s)s^n$. Consequently,

$$\begin{aligned} \text{Var}(T_1) &= \Delta^2(\text{Var}(H) + \text{Var}(I) - 2\text{Cov}(H, I)) \\ &= \Delta^2[ns(1-s) + s^n(1-s^n) - 2n(1-s)s^n]. \end{aligned}$$

Random scores. Even if the distributions of the scores are independent, the mean and covariance of the score distributions will not suffice to determine the answer. The key quantity that matters the distribution of the *minimum* of n copies of a random variable. This in general can not be determined by the mean and variance of the variable, but depends heavily on its tail statistics and concentration properties.

(You would not be expected to write this part) In general, this has a closed form pdf, but not a nice one: if p is the pdf of a single copy of a random variable, then the minimum of n copies of the random variable has pdf

$$p_n(x) = np(x)(1 - F(x))^{n-1},$$

where $F(x) = \int_{-\infty}^x p(x)dx$ is the cumulative density function.

Problem 2: Approximate Matrix Multiplication

Let $\xi_i = \mathbf{1}_{\{i \in J\}}$, $M_i = a_i b_i^\top$, and $2 < \omega < 3$ be as in the problem.

(a) By linearity, $\mathbb{E}X = c_p \sum_i \mathbb{E}[\xi_i] M_i = c_p p AB$, so $\boxed{c_p = 1/p}$ for $p > 0$.

(b) Given $m = |J| \geq 1$, compute $X = p^{-1} A_{[:,J]} B_{[J,:]}$. Partition both length- n dimensions into blocks of size m (padding as needed). This requires $\lceil n/m \rceil^2$ products of $m \times m$ matrices, costing $O(n^2 m^{\omega-2})$. The expected runtime is then (asymptotically) $\mathbb{E}_m[n^2 m^{\omega-2}]$. In particular, since $\mathbb{E}m = np$, concavity and Jensen's inequality give

$$\boxed{\mathbb{E}[\text{runtime}] = O(n^2 (np)^{\omega-2})}.$$

(c) Observe that $\mathbb{E}\|X - AB\|_F^2 = \sum_{j,k} \mathbb{E}[(X - AB)_{jk}^2]$, so we can consider every j, k separately. Write $(X - AB)_{jk} = \sum_i (\xi_i/p - 1)(M_i)_{jk}$. Note then that $\mathbb{E}[(X - AB)_{jk}^2]$ is just the variance of sum of independent random variables, and thus,

$$\mathbb{E}[(X - AB)_{jk}^2] = \sum_{i=1}^n \text{Var}[\xi_i (M_i)_{jk}] = \frac{1-p}{p} \sum_{i=1}^n (M_i)_{jk}^2,$$

and thus

$$\mathbb{E}\|X - AB\|_F^2 = \frac{1-p}{p} \sum_i \sum_{j,k} (M_i)_{jk}^2 = \frac{1-p}{p} \sum_i \|M_i\|_F^2.$$

Using $\|M_i\|_F^2 = \|a_i\|_2^2 \|b_i\|_2^2$ gives

$$\mathbb{E}\|X - AB\|_F^2 = \frac{1-p}{p} \sum_i \|a_i\|_2^2 \|b_i\|_2^2 \leq \boxed{\frac{C(1-p)}{np}},$$

since $\|a_i\|_2^2 \leq C/n$ and $\sum_i \|b_i\|_2^2 = 1$.

(d) Choose $p = \min\{1, C/(n\varepsilon^2\delta)\}$. If $p < 1$, Markov's inequality gives

$$\Pr(\|X - AB\|_F \geq \varepsilon) \leq \frac{C(1-p)}{np\varepsilon^2} \leq \delta.$$

If $p = 1$, the answer is exact. By (b), the expected runtime is

$$\boxed{O\left(n^2 \left(\min\left\{n, \frac{C}{\varepsilon^2\delta}\right\}\right)^{\omega-2}\right)}.$$

(e) Without the assumption, only $\|a_i\|_2^2 \leq 1$ is guaranteed. Taking $C = n$ above gives variance at most $(1-p)/p$ and runtime $O(n^\omega)$ for $0 < \varepsilon, \delta < 1$, so there is no guaranteed asymptotic speedup.

Intuitively, this is because we are sampling large-norm rows and small-norm rows uniformly, but large-norm rows are much more significant and important to keep. This motivates *importance sampling*, where one should sample larger norm rows with higher probability.

(You would not be expected to write this part) Specifically, it turns out the right way is to sample independently using squared column norms:

$$p_i = \min\left\{1, \frac{\|a_i\|_2^2}{\varepsilon^2\delta}\right\}, \quad X = \sum_{i:p_i>0} \frac{\xi_i}{p_i} a_i b_i^\top.$$

If you run through the same calculations as before, it turns out you get the following expected runtime:

$$O\left(n^2\left(\min\left\{n, \frac{1}{\varepsilon^2\delta}\right\}\right)^{\omega-2}\right).$$

Problem 3: The Moment Method

Let $\bar{X} = n^{-1} \sum_i X_i$.

(a) If any tuple has a term which only appears once, by independence, its expectation is 0. Thus only tuples in which every used index occurs at least twice can contribute; in particular, they use at most $k/2$ distinct indices. Some of these terms may also be zero.

(c) Each individual term has absolute value at most 1, since

$$\left| \mathbb{E} \left[\prod_{j=1}^k X_{i_j} \right] \right| \leq \prod_{i:m_i > 0} \mathbb{E} |X_i|^{m_i} \leq 1.$$

(b) Every surviving tuple has a set of $\leq k/2$ positions containing the first instance of every distinct index that appears in the tuple. Choose these positions in $\binom{k}{\leq k/2} \leq 2^k$ ways, assign them values in at most $n^{k/2}$ ways, and assign each of the at most k remaining positions one of their at most $k/2$ values. This overcounts, but gives

$$\#\{\text{nonzero terms}\} \leq 2^k n^{k/2} k^{k/2} \leq 2^k (nk)^{k/2} = \boxed{(4nk)^{k/2}}.$$

Thus $A = 4$ suffices.

(d) Combining the count and the bound on each term,

$$\mathbb{E} |\bar{X}|^k = \mathbb{E} [\bar{X}^k] \leq \left(\frac{2k}{n} \right)^{k/2}.$$

For $t > 0$, Markov's inequality therefore gives

$$\Pr(|\bar{X}| \geq t) \leq \min \left\{ 1, \left(\frac{2k}{nt^2} \right)^{k/2} \right\}.$$

(e) Let $k = Cnt^2$, for some constant C sufficiently large. Plugging this value of k into the above expression yields

$$\Pr(|\bar{X}| \geq t) \leq (2C)^{-k/2} \leq \exp \left(-\log(2C) \cdot \frac{C}{2} nt^2 \right),$$

which recovers Theorem 3.1 with value $C' = -\log(2C) \cdot \frac{C}{2}$.