

New York University Tandon School of Engineering  
Computer Science and Engineering

Midterm Practice.

## Practice Problems

### Random variables and concentration.

1. Show that for any random variable  $X$ ,  $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$ .  $\mathbb{E}[X^2] - \mathbb{E}[X]^2 = \text{Var}[X]$ , and variance must be non-negative.
2. Show that for independent  $X$  and  $Y$  with  $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ ,  $\text{Var}[X \cdot Y] = \text{Var}[X] \cdot \text{Var}[Y]$ . Since  $X$  and  $Y$  are independent  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = 0$ . So  $\text{Var}[XY] = \mathbb{E}[X^2Y^2] = \mathbb{E}[X^2]\mathbb{E}[Y^2]$ . The last step follows from that fact that  $X^2$  and  $Y^2$  are independent because  $X$  and  $Y$  are. Finally, since  $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ , we have  $\mathbb{E}[X^2] = \text{Var}[X]$  and  $\mathbb{E}[Y^2] = \text{Var}[Y]$ . Plugging in gives the proof.
3. Given a random variable  $X$ , can we conclude that  $\mathbb{E}[1/X] = 1/\mathbb{E}[X]$ ? If so, prove this. If not, give an example where the equality does not hold. Not true. Try a  $\{-1, 1\}$  random variable.  $1/\mathbb{E}[X]$  is infinite but  $\mathbb{E}[1/X]$  is not.
4. Indicate whether each of the following statements is **always** true, **sometimes** true, or **never** true. Provide a short justification for your choice.
  - (a)  $\Pr[X = s \text{ and } Y = t] > \Pr[X = s]$ . ALWAYS SOMETIMES **NEVER**.  $\Pr[X = s] = \Pr[X = s \text{ and } Y = t] + \Pr[X = s \text{ and } Y \neq t] \leq \Pr[X = s \text{ and } Y = t]$ .
  - (b)  $\Pr[X = s \text{ or } Y = t] \leq \Pr[X = s] + \Pr[Y = t]$ . **ALWAYS** SOMETIMES NEVER  
This is the union bound.
  - (c)  $\Pr[X = s \text{ and } Y = t] = \Pr[X = s] \cdot \Pr[Y = t]$ . ALWAYS **SOMETIMES** NEVER This is true for independent random variables, but not if we don't have independence.
5. Assume there are 1000 registered users on your site  $u_1, \dots, u_{1000}$ , and in a given day, each user visits the site with some probability  $p_i$ . The event that any user visits the site is independent of what the other users do. Assume that  $\sum_{i=1}^{1000} p_i = 500$ .
  - (a) Let  $X$  be the number of users that visit the site on the given day. What is  $\mathbb{E}[X]$ ? By linearity of expectation, 500.
  - (b) Apply a Chernoff bound to show that  $\Pr[X \geq 600] \leq .01$ . Setting  $\epsilon = .2$  and applying the bound from class we have  $\Pr[X \geq 600] \leq e^{-.2^2 500 / 2.2} = .0001$ .
  - (c) Apply Markov's inequality and Chebyshev's inequality to bound the same probability. How do they compare? By Markov's  $\Pr[X \geq 600] \leq \frac{1}{1.2} = .833$ . For Chebyshev's we need a variance calculation. One thing we can say is that  $\text{Var}[X] = \sum_{i=1}^{1000} \text{Var}[p_i] \leq \sum_{i=1}^{1000} p_i = 500$ . By Chebyshev,  $\Pr[|X - 500| \geq k\sqrt{500}] \leq \frac{1}{k^2}$ . Setting  $k = 100/\sqrt{500} = 4.47$  gives  $\frac{1}{k^2} = .05$ .
6. Give an example of a random variable and a deviation  $t$  where Markov's inequality gives a tighter bound than Chebyshev's inequality. Take a uniform  $\{0, 1\}$  random variable and bound  $\Pr[X \geq 1]$ . Markov's will give the tight  $1/2$  bound, but Chebyshev's gives a vacuously weak bound.

### Hashing, Dimensionality Reduction, High Dimensional Vectors

1. Suppose there is some unknown vector  $\mu$ . We receive noise perturbed random samples of the form  $\mathbf{Y}_1 = \mu + \mathbf{X}_1, \dots, \mathbf{Y}_k = \mu + \mathbf{X}_k$  where each  $\mathbf{X}_i$  is a random vector with each of its entries distributed as an independent random normal  $\mathcal{N}(0, 1)$ . From our samples  $\mathbf{Y}_1, \dots, \mathbf{Y}_k$  we hope to estimate  $\mu$  by  $\tilde{\mu} = \frac{1}{k} \sum_{i=1}^k \mathbf{Y}_i$ .

- (a) How many samples  $k$  do we require so that  $\max |\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}| \leq \epsilon$  with probability 9/10? If we can show that for all  $i \in 1, \dots, n$ ,  $|\mu_i - \tilde{\mu}_i| \leq \epsilon$  with probability  $1 - \frac{1}{10n}$  then by a union bound we will have that  $\max |\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}| \leq \epsilon$  with probability  $1 - \frac{1}{10}$ . So we focus on this simpler problem. Notice that  $\tilde{\mu}_i = \mu_i + \frac{1}{k} \sum_{j=1}^k [\mathbf{X}_j]_i$ , where  $[\mathbf{X}_j]_i$  is the  $i^{\text{th}}$  entry of  $\mathbf{X}_j$ , which is a norm random variable. Since the sum of random normals is norm, we have that  $\frac{1}{k} \sum_{j=1}^k [\mathbf{X}_j] \sim \frac{1}{k} \mathcal{N}(0, k) = \mathcal{N}(0, 1/k)$ . Applying the Gaussian tail bound from lecture, we have thus have that  $\Pr[|\mu_i - \tilde{\mu}_i| \geq \alpha/\sqrt{k}] \leq e^{-O(\alpha^2)}$ . Setting  $\alpha = O(\sqrt{\log(1/(1/10n))}) = O(\sqrt{\log n})$  and  $k = O(\log n/\epsilon^2)$  gives the bound we need. So overall we need  $k = O(\log n/\epsilon^2)$  samples.
- (b) How many samples  $k$  do we require so that  $\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}\|_2 \leq \epsilon$  with probability 9/10? The proof is essentially the same. To ensure that  $\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}\|_2 \leq \epsilon$ , it suffices to have  $|\mu_i - \tilde{\mu}_i| \leq \sqrt{\epsilon/n}$  for all  $i$ . This will happen with probability 9/10 as long as  $k = O(\log n/(\sqrt{\epsilon/n})^2) = k = O(n \log n/\sqrt{\epsilon}^2)$ .
2. Let  $\boldsymbol{\Pi}$  be a random Johnson-Lindenstrauss matrix (e.g. scaled random Gaussians) with  $O(\log(1/\delta)/\epsilon^2)$  rows. Prove that with probability  $(1 - \delta)$ ,

$$\min_{\mathbf{x}} \|\boldsymbol{\Pi} \mathbf{A} \mathbf{x} - \boldsymbol{\Pi} \mathbf{b}\|_2^2 \leq (1 + \epsilon) \min_{\mathbf{x}} \|\mathbf{A} \mathbf{x} - \mathbf{b}\|_2^2$$

Let  $\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{A} \mathbf{x} - \mathbf{b}\|_2^2$ . Then by the distributional Johnson-Lindenstrauss lemma we have  $\|\boldsymbol{\Pi} \mathbf{A} \mathbf{x}^* - \boldsymbol{\Pi} \mathbf{b}\|_2^2 = \|\boldsymbol{\Pi}(\mathbf{A} \mathbf{x}^* - \mathbf{b})\|_2^2 \leq (1 + \epsilon) \|\mathbf{A} \mathbf{x}^* - \mathbf{b}\|_2^2$ . And then we have that  $\min_{\mathbf{x}} \|\boldsymbol{\Pi} \mathbf{A} \mathbf{x} - \boldsymbol{\Pi} \mathbf{b}\|_2^2 \leq \|\boldsymbol{\Pi} \mathbf{A} \mathbf{x}^* - \boldsymbol{\Pi} \mathbf{b}\|_2^2$ , which gives the result.