New York University Tandon School of Engineering
Computer Science and Engineering

Midterm Practice.

Practice Problems

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 $\mathbb{E}[1/X] = 1/E[X]$? If so, prove this. If not, give an example where the equality does not hold. Not true. Try a $\{-1, 1\}$ random variable. $1/E[X]$ is infinite but $E[1/X]$ 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 $u_1, \ldots, u_{1000}$, 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 $E[X]$? 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.

7. Suppose there is some unknown vector $\boldsymbol{\mu}$. We receive noise perturbed random samples of the form $\mathbf{Y}_1 = \boldsymbol{\mu} + \mathbf{X}_1, \ldots, \mathbf{Y}_k = \boldsymbol{\mu} + \mathbf{X}_k$ 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 $\mathbf{Y}_1, \ldots, \mathbf{Y}_k$ we hope to estimate $\boldsymbol{\mu}$ by $\tilde{\boldsymbol{\mu}} = \frac{1}{k}\sum_{i=1}^{k} \mathbf{Y}_i$.

   (a) Prove that with $k = O(\log d/\epsilon^2)$ samples, $\max_{i=1,\ldots,d} |\boldsymbol{\mu}_i - \tilde{\boldsymbol{\mu}}_i| \leq \epsilon$ with probability 9/10. If we can show that for all $i \in 1, \ldots, d$, $|\boldsymbol{\mu}_i - \tilde{\boldsymbol{\mu}}_i| \leq \epsilon$ with probability $1 - \frac{1}{10d}$ 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 $\tilde{\boldsymbol{\mu}}_i = \boldsymbol{\mu}_i + \frac{1}{k}\sum_{j=1}^{k} [\mathbf{X}_j]_i$, where $[\mathbf{X}_j]_i$ is the i^{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 $\Pr[|\boldsymbol{\mu}_i - \tilde{\boldsymbol{\mu}}_i| \geq \alpha/\sqrt{k}] \leq e^{-O(\alpha^2)}$. Setting $\alpha = O(\sqrt{\log(1/(1/10d)}) = O(\sqrt{\log d})$ and $k = O(\log d/\epsilon^2)$ gives the bound we need. So overall we need $k = O(\log d/\epsilon^2)$ samples.

(b) Prove that with $k = O(d\log d/\epsilon^2)$ samples, $\|\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}\|_2 \leq \epsilon$ 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 $|\boldsymbol{\mu}_i - \tilde{\boldsymbol{\mu}}_{\boldsymbol{i}}| \leq \epsilon/\sqrt{d}$ for all i. This will happen with probability $9/10$ as long as $k = O(\log d/(\epsilon/d)^2) = k = O(d\log d/\epsilon^2)$.

8. Let $\mathbf{\Pi}$ be a random Johnson-Lindenstrauss matrix (e.g. scaled random Gaussians) with $O(\log(1/\delta)/\epsilon^2)$ rows. Prove that with probability $(1-\delta)$,

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

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

9. For two length d binary vectors $\mathbf{q}, \mathbf{y} \in \{0,1\}^d$, consider the hamming similarity:

$$s(\mathbf{q}, \mathbf{y}) = 1 - \frac{\|\mathbf{q} - \mathbf{y}\|_0}{d}.$$

Recall that $\|\mathbf{q} - \mathbf{y}\|_0 = \sum_{i=1}^{d} \mathbb{1}[\mathbf{q}_i \neq \mathbf{y}_i]$. Construct a function h as follows: define the random function $c : \{0,1\}^d \to \{0,1\}$ as $c(\mathbf{x}) = \mathbf{x}[j]$, where j is a uniform random integer in $\{1,\ldots,d\}$. Then, let g be a uniform random hash function from $\{0,1\} \to \{1,\ldots,m\}$. Finally, let:

$$h(\mathbf{x}) = g(c(\mathbf{x})).$$

Prove that h is a locality sensitive hash function for hamming similarity.

To prove h is locality sensitive, we want to argue that $\Pr[h(\mathbf{q}) = h(\mathbf{y})]$ increases as $s(\mathbf{q}, \mathbf{y})$ increases. As in class, we have:

$$Pr[h(\mathbf{q}) = h(\mathbf{y})] = 1 \cdot \Pr[c(\mathbf{q}) = c(\mathbf{y})] + \frac{1}{m} \cdot (1 - \Pr[c(\mathbf{q}) = c(\mathbf{y})])$$

So we need to determine $\Pr[c(\mathbf{q}) = c(\mathbf{y})]$. Since j is chosen randomly, this probability is exactly equal to the percentage of indices i for which $\mathbf{q}_i = \mathbf{y}_i$. The number of equal indices is $d - \|\mathbf{q} - \mathbf{y}\|_0$, so this percentage is $\frac{d - \|\mathbf{q}-\mathbf{y}\|_0}{d} = s(\mathbf{q}, \mathbf{y})$. We conclude that:

$$Pr[h(\mathbf{q}) = h(\mathbf{y})] = s(\mathbf{q}, \mathbf{y}) + \frac{1}{m} \cdot (1 - s(\mathbf{q}, \mathbf{y})),$$

which is an increasing function of $s(\mathbf{q}, \mathbf{y})$. So $s(\mathbf{q}, \mathbf{y})$ is locality sensitive.