General Info

1. Solve 3 out of 4 questions.
2. Each correct answer is worth 33.3 points.
3. If you have solved more than three questions, please indicate which three you would like to be checked.
4. The exam's duration is 3 hours. If you need more time please ask the attending professor.
5. Good luck!

Useful facts

1. For any vector $x \in \mathbb{R}^d$ we define the p-norm of x as follows:
$$||x||_p = [\sum_{i=1}^{d} (x(i))^p]^{1/p}$$
2. **Markov's inequality:** For any *non-negative* random variable X:
$$\Pr[X > t] \leq E[X]/t.$$
3. **Chebyshev's inequality:** For any random variable X:
$$\Pr[|X - E[X]| > t] \leq \mathrm{Var}[X]/t^2.$$
4. **Chernoff's inequality:** Let $x_1, \ldots, x_n$ be independent $\{0, 1\}$ valued random variables. Each x_i takes the value 1 with probability p_i and 0 else. Let $X = \sum_{i=1}^n x_i$ and let $\mu = E[X] = \sum_{i=1}^n p_i$. Then:
$$\Pr[X > (1+\varepsilon)\mu] \leq e^{-\mu\varepsilon^2/4}$$
$$\Pr[X < (1-\varepsilon)\mu] \leq e^{-\mu\varepsilon^2/2}$$
Or in a another convenient form:
$$\Pr[|X - \mu| > \varepsilon\mu] \leq 2e^{-\mu\varepsilon^2/4}$$
5. **Hoeffding's inequality:** Let $x_1, \ldots, x_n$ be independent random variables taking values in $\{+1, -1\}$ each with probability 1/2, then:
$$\Pr[|\sum_{i=1}^{n} x_i a_i| > t] \leq 2e^{-\frac{t^2}{\sum_{i=1}^n a_i^2}}.$$
6. For any $x \geq 2$ we have:
$$e^{-1} \geq (1 - \frac{1}{x})^x \geq \frac{2}{3}e^{-1}$$
7. For convenience:
$$\frac{3}{5} \leq 1 - e^{-1} \approx 0.632 \leq \frac{2}{3} \quad \text{and} \quad \frac{3}{4} \leq 1 - \frac{2}{3}e^{-1} \approx 0.754 \leq \frac{4}{5}$$

1 Probabilistic inequalities

setup

In this question you will be asked to derive the three most used probabilistic inequalities for a specific random variable. Let $x_1, \ldots, x_n$ be independent $\{-1, 1\}$ valued random variables. Each x_i takes the value 1 with probability 1/2 and -1 else. Let $X = \sum_{i=1}^{n} x_i$.

questions

1. Let the random variable Y be defined as $Y = |X|$. Prove that Markov's inequality holds for Y. Hint: note that Y takes integer values. Also, there is no need to compute $\Pr[Y = i]$.

2. Prove Chebyshev's inequality for the above random variable X. You can use the fact that Markov's inequality holds for any positive variable regardless of your success (or lack of if) in the previous question. Hint: $\mathrm{Var}[X] = E[(X - E[X])^2]$.

3. Argue that
$$\Pr[X > a] = \Pr[\Pi_{i=1}^{n} e^{\lambda x_i} > e^{\lambda a}] \leq \frac{E[\Pi_{i=1}^{n} e^{\lambda x_i}]}{e^{\lambda a}}$$
for any $\lambda \in [0, 1]$. Explain each transition.

4. Argue that:
$$\frac{E[\Pi_{i=1}^{n} e^{\lambda x_i}]}{e^{\lambda a}} = \frac{\Pi_{i=1}^{n} E[e^{\lambda x_i}]}{e^{\lambda a}} = \frac{(E[e^{\lambda x_1}])^n}{e^{\lambda a}}$$
What properties of the random variables x_i did you use in each transition?

5. Conclude that $\Pr[X > a] \leq e^{-\frac{a^2}{2n}}$ by showing that:
$$\exists\ \lambda \in [0, 1]\ \ s.t.\ \ \frac{(E[e^{\lambda x_1}])^n}{e^{\lambda a}} \leq e^{-\frac{a^2}{2n}}$$
Hint: For the hyperbolic cosine function we have $\cosh(x) = \frac{1}{2}(e^x + e^{-x}) \leq e^{x^2/2}$ for $x \in [0, 1]$.

2 Approximating the size of a tree

setup

In this question we will try to approximate the number of leaves in a tree. A binary tree is a graph consisting of internal nodes and n leaves. Each internal node, u, has two children. A left child $l(u)$ and a right child $r(u)$. The only node which does not have a parent is the root of the tree u_{root}. For each node we also denote by $d(u)$ its depth in the tree which is the distance from the root. For example $d(u_{rood}) = 0$ and $d(r(u_{root})) = 1$.

We define a random walk on a tree as the process of starting at the root and then randomly moving to one of the children until we hit a leaf. More precisely:

1. $u \leftarrow u_{root}$
2. while u is an internal node
3.   w.p. 1/2
4. $u \leftarrow l(u)$
5.   otherwise
6. $u \leftarrow r(u)$
7. return u

questions

1. Let the leaf u be at depth $d(u)$. Calculate the probability, $p(u)$, that the random walk outputs u?

2. Let x be the output leaf of a random walk and let $f(x) = 2^{d(x)}$ be a function defined on the leaves. Compute the value of:

$$E_{x \sim w}[f(x)]$$

   where $x \sim w$ denotes that x is chosen according to the distribution on the leaves generated by the random walk.

3. We say that a tree is c-balanced if $d(u) \leq \log_2 n + c$ for all leaves in the tree. Show that for a c-balanced tree

$$\mathrm{Var}_{x \sim w}[f(x)] \leq 2^c n^2$$

4. Let $Y = \frac{1}{s}\sum_{i=1}^{s} f(x_i)$ where x_i are output nodes of s independent random walks on the tree. Compute $E[Y]$ **and** show that $\mathrm{Var}[Y] \leq 2^c n^2/s$.

5. Use Chebyshev's inequality to find a value for s such that for two constants $\varepsilon \in [0, 1]$ and $\delta \in [0, 1]$:

$$\Pr[|Y - n| > \varepsilon n] < \delta.$$

   s should be a function of c, ε and δ.

3 Multi-armed bandit

setup

In the following question we will devise an algorithm for finding the best slot machine in the casino. This problem is named Multi-armed bandit since a single slot machine is also known as a one-armed bandit.[1] We are presented with m machines $M_1, \ldots, M_m$. The probability of wining at machine M_i is p_i and we assume each game is independent of all others. The best machine to play is, of course, the one maximizing the winning probability. Our strategy is to first play each machine k times regardless of the outcomes. Then, pick the machine which won the largest number of times.[2]

questions

1. Let w_i denote the number of wins at machine M_i after having played k rounds. Compute $E[w_i]$.

2. Without loss of generality, let the best machine be M_1, i.e., $\forall i \;\; p_1 \geq p_i$. We call a machine M_i "good" if $p_i \geq (1-\alpha)p_1$ and "bad" if $p_i < (1-\alpha)p_1$. Argue that if for the best machine $w_1 \geq E[w_1] - \alpha k/2$ and for all bad machines $w_i < E[w_1] - \alpha k/2$ then we will pick a "good" machine.

3. Bound from above the probability that for the best machine

   $$w_1 < E[w_1] - \alpha k/2$$

   for any $\alpha \leq p_1$.

4. Bound from above the probability that for a "bad" machine

   $$w_i \geq E[w_1] - \alpha k/2.$$

   Do not be confused: on the right hand side of the equation we have w_1 and not w_i.

5. Using the results of 3 and 4 give an upper bound on the value of k which will guarantee that we pick a "good" machine with probability at least $1-\delta$. Remember that you might have $m-1$ bad machines.

---

[1] Slot machines are known as one-armed bandits because slot machines were originally operated by a lever on the side of the machine (the one arm) and because of their ability to leave the gamer penniless.

[2] This problem comes up often in data mining scenarios in which we estimate probabilities of events based on observations in the data.

4 Weak random projections

setup

In this question we will construct a simple and weak version of random projections. That is, given two vectors $x, y \in \mathbb{R}^d$ we will find two new vectors $x', y' \in \mathbb{R}^k$ such that from x' and y' we could approximate the value of $||x - y||$. The idea is to define k vectors $r_i \in \mathbb{R}^d$ such that each $r_i(j)$ takes a value in $\{+1, -1\}$ uniformly at random. Setting $x'(i) = r_i^T x$ and $y'(i) = r_i^T y$ the questions will lead you through arguing that $\frac{1}{k}||x' - y'||_2^2 \approx ||x - y||_2^2$.

questions

1. Let $z = x - y$, and $z' = x' - y'$. Show that $z'(\ell) = r_\ell^T z$ for any index $\ell \in [1, \ldots, k]$.

2. Show that $E[\frac{1}{k}||z'||_2^2] = E[(z'(\ell))^2] = ||z||_2^2$.

3. Show that

$$\text{Var}[(z'(\ell))^2] \leq 4||z||_2^4.$$

Hint: for any vector w we have $||w||_4 \leq ||w||_2$.

4. From 3 (even if you did not manage to show it) claim that

$$\text{Var}[\frac{1}{k}||z'||_2^2] \leq 4||z||_2^4/k.$$

5. Use 3 and Chebyshev's inequality do obtain a value for k for which:

$$(1 - \varepsilon)||x - y||_2^2 \leq \frac{1}{k}||x' - y'||_2^2 \leq (1 + \varepsilon)||x - y||_2^2$$

with probability at least $1 - \delta$.