ECE 299: STATISTICAL LEARNING THEORY

MAXIM RAGINSKY

Homework 1 Solutions

Assigned February 2, 2011; due February 16, 2011

---

(1) Let $X_1, \ldots, X_n$ be n independent Bernoulli(θ) random variables. Prove the following *multiplicative Chernoff bound*:

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq (1+\gamma)\mu\right) \leq e^{-\gamma^2 n\mu/3}$$

for any $0 \leq \gamma \leq 1$ and any $\theta \leq \mu \leq 1$. You may need the fact that

$$\sum_{k \geq (1+\gamma)\mu n} \binom{n}{k} \mu^k (1-\mu)^{n-k} \leq e^{-\gamma^2 n\mu/3}.$$

Solution. Since the number of times the coin comes up HEADS in n trials is a Binomial(n, θ) random variable, we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq (1+\gamma)\mu\right) = \sum_{k \geq (1+\gamma)\mu n} \binom{n}{k} \theta^k (1-\theta)^{n-k}.$$

Now, the function $\theta \mapsto \theta^k(1-\theta)^{n-k}$ is nondecreasing for $\theta \leq k/n$. The summation above is only over those k, for which $\mu \leq \frac{k}{(1+\gamma)n} \leq \frac{k}{n}$. Since $\theta \leq \mu$ by hypothesis, we can bound

$$\sum_{k \geq (1+\gamma)\nu n} \binom{n}{k} \theta^k (1-\theta)^{n-k} \leq \sum_{k \geq (1+\gamma)\nu n} \binom{n}{k} \mu^k (1-\mu)^{n-k} \leq e^{-\gamma^2 n\mu/3}.$$

(2) Let $X_1, \ldots, X_n$ be $n \geq 2$ real-valued random variables (not necessarily independent). Suppose that there exists some constant $\sigma > 0$, such that for all $s > 0$ and all $i = 1, \ldots, n$

$$\mathbb{E}\left[e^{sX_i}\right] \leq e^{s^2\sigma^2/2}$$

Prove the following:

(a)

$$\mathbb{E}\left[\max_{1 \leq i \leq n} X_i\right] \leq \sigma\sqrt{2\log n}.$$

Hint. Try to bound $e^{s\mathbb{E}[\max_{1 \leq i \leq n} X_i]}$, exploit convexity of the function $\phi(x) = e^{sx}$.

(b)

$$\mathbb{P}\left(\max_{1 \leq i \leq n} X_i \geq \sqrt{2\sigma^2 n}\right) \leq \sqrt{\frac{\log n}{n}}.$$

Solution. (a) For any $s > 0$,

$$\begin{aligned}
e^{s\mathbb{E}\left[\max_{1\le i\le n} X_i\right]} &\le \mathbb{E}\left[e^{s\max_{1\le i\le n} X_i}\right] && \text{(by convexity of the function } \phi(t)=e^{st}\text{)}\\
&= \mathbb{E}\left[\max_{1\le i\le n} e^{sX_i}\right] && \text{(since the function } \phi(t)=e^{st},\ s>0\text{, is increasing)}\\
&\le \mathbb{E}\left[\sum_{i=1}^{n} e^{sX_i}\right]\\
&= \sum_{i=1}^{n} \mathbb{E}\left[e^{sX_i}\right]\\
&\le n e^{s^2\sigma^2/2} && \text{(by assumption on the } X_i\text{'s)}
\end{aligned}$$

Taking logs, we get

$$\mathbb{E}\left[\max_{1\le i\le n} X_i\right] \le \frac{\log n}{s} + \frac{s\sigma^2}{2}.$$

To get the tightest bound, we minimize the right-hand side with respect to s. The optimal choice is $s = \sqrt{2\log n/\sigma^2}$, leading to

$$\mathbb{E}\left[\max_{1\le i\le n} X_i\right] \le \sigma\sqrt{2\log n}.$$

(b) Use Markov's inequality:

$$\mathbb{P}\left(\max_{1\le i\le n} X_i \ge \sqrt{2\sigma^2 n}\right) \le \frac{\mathbb{E}\left[\max_{1\le i\le n} X_i\right]}{\sqrt{2\sigma^2 n}} \le \frac{\sigma\sqrt{2\log n}}{\sigma\sqrt{2n}} = \sqrt{\frac{\log n}{n}}.$$

(3) Let $X^n = (X_1, \ldots, X_n)$ be an n-tuple of independent random variables taking values in some space X. The *Hamming distance* between any n-tuples $x^n, y^n \in \mathsf{X}^n$ is defined as the number of coordinates in which x^n and y^n differ:

$$d(x^n, y^n) \triangleq \sum_{i=1}^{n} \mathbf{1}_{\{x_i \ne y_i\}}.$$

If B is an arbitrary (measurable) subset of X^n, then the Hamming distance between an n-tuple $x^n \in \mathsf{X}^n$ and B is defined as

$$d(x^n, B) \triangleq \min_{y^n \in B} d(x^n, y^n).$$

Use McDiarmid's inequality to prove the following fact: if

$$\varepsilon \ge \sqrt{\frac{1}{2n}\log\frac{1}{\mathbb{P}(B)}},$$

then

$$\mathbb{P}\left(d(X^n, B) \ge n\varepsilon\right) \le \exp\left(-2n\left(\varepsilon - \sqrt{\frac{1}{2n}\log\frac{1}{\mathbb{P}(B)}}\right)^2\right),$$

where $\mathbb{P}(B) = \mathbb{P}(X^n \in B)$ is the probability of the set $B \in \mathsf{X}^n$ under the joint distribution of X^n.

Hint. You may find the following fact handy: $d(x^n, B) \le 0$ if and only if $x^n \in B$.

Solution. Consider any $x^n \in \mathsf{X}^n$, as well as any $x_i' \in \mathsf{X}$. Let $x^n_{(i)}$ denote x^n with the ith component replaced by x_i'. Then

$$\begin{aligned} d(x^n, B) - d(x^n_{(i)}, B) &= \min_{y^n B} d(x^n, y^n) - \min_{y^n \in B} d(x^n_{(i)}, y^n) \\ &\leq \max_{y^n \in B} \Big[d(x^n, y^n) - d(x^n_{(i)}, y^n) \Big] \\ &= \max_{y^n \in B} \big[\mathbf{1}_{\{x_i \neq y_i\}} - \mathbf{1}_{\{x_i' \neq y_i\}} \big] \\ &\leq 1. \end{aligned}$$

This holds for every $i = 1, \ldots, n$. Hence, the function $g(X^n) = d(X^n, B)$ has the bounded differences property with $c_1 = \ldots = c_n = 1$. McDiarmid's inequality then gives, for any $t > 0$:

$$\mathbb{P}\Big(\mathbb{E}d(X^n, B) - d(X^n, B) \geq t \Big) \leq e^{-2t^2/n}$$

If we use $t = \mathbb{E}d(X^n, B)$, then we get

$$\mathbb{P}\Big(d(X^n, B) \leq 0 \Big) \leq e^{-2(\mathbb{E}d(X^n, B))^2/n}.$$

Since $d(X^n, B) \leq 0$ if and only if $X^n \in B$, the left-hand side of the above inequality is equal to $\mathbb{P}(B)$. Therefore, we get

$$\mathbb{E}d(X^n, B) \leq \sqrt{\frac{n}{2} \log \frac{1}{\mathbb{P}(B)}}$$

Now, using the other direction of McDiarmid's inequality, we have

$$\mathbb{P}\Big(d(X^n, B) - \mathbb{E}d(X^n, B) \geq t \Big) \leq e^{-2t^2/n},$$

which, together with our bound on $\mathbb{E}d(X^n, B)$, gives

$$\mathbb{P}\left(d(X^n, B) \geq t + \sqrt{\frac{n}{2} \log \frac{1}{\mathbb{P}(B)}} \right) \leq e^{-2t^2/n}.$$

Now let $t = n\varepsilon - \sqrt{\frac{n}{2} \log \frac{1}{\mathbb{P}(B)}}$. By our hypothesis on ε, $t \geq 0$. Substituting this choice of t into the above probability bound, we obtain the desired result.