CS 7545: Machine Learning Theory **Fall 2019**

Lecture 4: Chernoff-type Deviation Bounds

Lecturer: Jacob Abernethy *Scribes: Alex Oshin and Kunal Chawla*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

4.1 Previously...

Last lecture we learned Markov's Inequality, which will be used to help prove certain deviation bounds discussed in this lecture. As a review:

Theorem 4.1 *(Markov's Inequality) For any random variable $X \geq 0$,*

$$Pr(X > t) \leq \frac{\mathbb{E}[X]}{t} \tag{4.1}$$

In fact, Markov's Inequality is a deviation bound in and of itself. But we can do better...

4.2 Gaussian and Sub-Gaussian Random Variables

Definition 4.2 (Gaussian) *We say a random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is a **Gaussian**, or normally distributed, random variable. Its probability density function is defined as*

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{4.2}$$

Gaussian random variables are useful because their tails are exponentially bounded (they decay at a rate roughly e^{-x^2}).

Fact 4.3 *Let $X \sim \mathcal{N}(0, \sigma^2)$. Then for all $t > 0$,*

$$\mathbb{E}[\exp(tX)] = e^{\frac{t^2\sigma^2}{2}} \tag{4.3}$$

Note: This is not related to the probability density function, but is instead a bound on the moment-generating function of X.

Definition 4.4 (sub-Gaussian) *A random variable X, with $\mathbb{E}[X] = 0$, is **sub-Gaussian** with variance proxy σ^2 if for all $t > 0$,*

$$\mathbb{E}[\exp(tX)] \leq e^{\frac{t^2\sigma^2}{2}} \tag{4.4}$$

4.3 Hoeffding's Inequality

Lemma 4.5 (Hoeffding's Lemma) *Let X be a bounded random variable with $a \leq X \leq b$ and $\mathbb{E}[X] = 0$. Then X is sub-Gaussian with variance proxy $\frac{(b-a)^2}{4}$, and*

$$\mathbb{E}[\exp(tX)] \leq \exp\left(\frac{t^2(b-a)^2}{8}\right) \tag{4.5}$$

Proof: The proof of this lemma is available in *Foundations of Machine Learning* (2nd ed.), pg. 437. ■

In essence, Hoeffding's Lemma states that a bounded random variable is bounded by a Gaussian random variable with zero mean and variance $\frac{(b-a)^2}{4}$.

Claim 4.6 *Let X be a sub-Gaussian random variable with $\mathbb{E}[X] = 0$ and variance proxy σ^2. Then,*

$$Pr(|X| > t) \leq 2\exp(-\frac{t^2}{2\sigma^2}) \tag{4.6}$$

We will not prove this claim, but instead prove a more general case for a sum of random variables, better known as Hoeffding's Inequality.

Theorem 4.7 (Hoeffding's Inequality) *Let $X_1, \ldots, X_n$ be independent random variables with $a_i \leq X_i \leq b_i$ and $\mathbb{E}[X_i] = \mu_i$. Let $\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$ and $\mu := \frac{1}{n}\sum_{i=1}^n \mu_i$. Then,*

$$Pr(|\bar{X}_n - \mu| > t) \leq 2\exp\left(-\frac{2n^2t^2}{\sum_{i=1}^n (a_i - b_i)^2}\right) \tag{4.7}$$

The factor of 2 multiplied by the exponential in the inequality of (4.7) is motivated by the fact that the absolute value of the probability on the left hand side can be split up into disjoint events:

$$\begin{aligned}
Pr(|\bar{X}_n - \mu| > t) &= Pr(\bar{X}_n - \mu > t) \text{ or } Pr(\bar{X}_n - \mu < -t) \\
&= Pr(\bar{X}_n - \mu > t) + Pr(\bar{X}_n - \mu < -t) \\
&\leq \exp\left(-\frac{2n^2t^2}{\sum_{i=1}^n (a_i - b_i)^2}\right) + \exp\left(-\frac{2n^2t^2}{\sum_{i=1}^n (a_i - b_i)^2}\right)
\end{aligned}$$

Since both sides are symmetric and bounded by the same exponential, we need only to prove one side, and then multiply the final bound by 2.

Proof:

$$\begin{aligned}
Pr\left(\bar{X}_n - \mu > t\right) &= Pr\left(exp\left(s\left(\bar{X}_n - \mu\right)\right) > exp\left(st\right)\right) \text{ for any } s > 0 \\
&\leq \frac{E\left[exp\left(s\left(\bar{X}_n - \mu\right)\right)\right]}{exp\left(st\right)} \text{ by Markov's Inequality (4.1)} \\
&= exp\left(-st\right) E\left[exp\left(\frac{s}{n}\sum_{i=1}^{n}\left(X_i - \mu_i\right)\right)\right] \\
&= exp\left(-st\right) E\left[\prod_{i=1}^{n}\left(exp\left(\frac{s}{n}\left(X_i - \mu_i\right)\right)\right)\right] \\
&= exp\left(-st\right)\prod_{i=1}^{n} E\left[exp\left(\frac{s}{n}\left(X_i - \mu_i\right)\right)\right] \text{ By independence of random variables } X_i, i \in \{1, n\}
\end{aligned}$$

Since all $X_i, i \in \{1, n\}$ are bounded and hence sub-Guassian,

$$\begin{aligned}
&\leq exp\left(-st\right)\prod_{i=1}^{n} exp\left(\frac{s^2\left(b_i - a_i\right)^2}{8n^2}\right) \text{ by Hoeffding's Lemma (4.5), } a_i \leq X_i \leq b_i \\
&= exp\left(\frac{s^2\sum_{i=1}^{n}\left(b_i - a_i\right)^2}{8n^2} - st\right)
\end{aligned}$$

Since this is true for all $s > 0$, let's substitute the value for which this function attains maximum. Let $\alpha = \sum_{i=1}^{n} (b_i - a_i)^2 / 8n^2 > 0$ and $\beta = t > 0$. Consider the maximum value of $f(s) = \alpha s^2 - \beta s$. Let $f'(s) = 2\alpha s - \beta = 0 \implies s = \beta/2\alpha > 0$. Also $f''(s) = 2\alpha > 0$.

$f(s) = 2\alpha \left(\frac{\beta}{2\alpha}\right)^2 - \beta\left(\frac{\beta}{2\alpha}\right) = -\frac{\beta^2}{4\alpha}$ when $s = \frac{\beta}{2\alpha}$. So $f(s) \leq -\frac{\beta^2}{4\alpha}\ \forall s > 0$.

$$\text{So } exp\left(\frac{s^2\sum_{i=1}^{n}(b_i - a_i)^2}{8n^2} - st\right) \leq exp\left(\frac{-t^2}{4\frac{\sum_{i=1}^{n}(b_i-a_i)^2}{8n^2}}\right) \quad \forall s > 0$$

$$\implies Pr\left(\bar{X}_n - \mu > t\right) \leq exp\left(\frac{s^2\sum_{i=1}^{n}(b_i - a_i)^2}{8n^2} - st\right) \leq exp\left(\frac{-t^2}{4\frac{\sum_{i=1}^{n}(b_i-a_i)^2}{8n^2}}\right) \quad \forall s > 0$$

$$\implies Pr\left(\bar{X}_n - \mu > t\right) \leq exp\left(\frac{-2n^2t^2}{4\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

■

Claim 4.8 *Let $X_1, \ldots, X_n$ be independent random variables such that $0 \leq X_i \leq 1,\ i \in \{1, n\}$. Then, for a $0 < \delta < 1$, we have, with probability at least $1 - \delta$*

$$|\frac{\sum_{i=1}^{n} X_i}{n} - E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right]| \leq \sqrt{\frac{log\left(\frac{2}{\delta}\right)}{2n}} \tag{4.8}$$

Proof: This is actually just another way to write Hoeffding's Inequality. To establish the equivalency,

$$\begin{aligned}
&\text{let } \delta = 2\exp\left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) \\
&\implies log(\delta/2) = -\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2} \\
&\implies log(2/\delta) = \frac{2n^2t^2}{\sum_{i=1}^{n}(1 - 0)^2} \text{ since } 0 \leq X_i \leq 1 \forall i \in \{1, n\} \\
&\implies \left(\frac{log(2/\delta)}{2}\right) = \frac{n^2t^2}{n} \\
&\implies t = \sqrt{\frac{log\left(\frac{2}{\delta}\right)}{2n}}
\end{aligned}$$

Substituting this value of t in (4.7)

$$Pr(|\bar{X}_n - \mu| > t) \leq \exp\left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(a_i - b_i)^2}\right)$$

$$\implies Pr\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right]\right| > \sqrt{\frac{log\left(\frac{2}{\delta}\right)}{2n}}\right) \leq \delta$$

$$\implies Pr\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right]\right| \leq \sqrt{\frac{log\left(\frac{2}{\delta}\right)}{2n}}\right) > 1 - \delta$$

■