EECS 598-005: Theoretical Foundations of Machine Learning **Fall 2015**

Lecture 4: Hoeffding's Inequality and Martingales

Lecturer: Jacob Abernethy *Scribes: Ruihao Zhu, Editors: Yuan Zhuang*

4.1 Hoeffding's Inequality

In this section we present Hoeffding's Inequality and its proof. To do so, we first go through the Hoeffding's Lemma.

Lemma 4.1 (Hoeffding's Lemma). *For a random variable $a \leq X \leq b$ such that $\mathbf{E}[X] = 0$, we have*

$$\mathbf{E}[\exp(\lambda X)] \leq \exp\left(\frac{\lambda^2(b-a)^2}{8}\right).$$

Hoeffding's Lemma is related to the concept of subgaussian.

Definition 4.2 (subgaussian). *A random variable X is **subgaussian** with parameter σ^2 if*

$$\mathbf{E}[\exp(\lambda X)] \leq \exp\left(\frac{\sigma^2\lambda^2}{2}\right).$$

Note 4.3. *If a random variable X follows a normal distribution with mean 0 and variance σ^2, then*

$$\mathbf{E}[\exp(\lambda X)] = \exp\left(\frac{\sigma^2\lambda^2}{2}\right).$$

We are now ready to get into the Hoeffding's Inequality and its proof (Chernoff Technique).

Theorem 4.4 (Hoeffding's Inequality). *Let $X_1, X_2, \ldots, X_n$ be independent random variables such that, $a_i \leq X_i \leq b_i$ and $\mathbf{E}[X_i] = 0$ for all $i = 1, 2, \ldots, n$. Then, for all $t > 0$*

$$\mathbf{Pr}\left[\sum_{i=1}^n X_i \geq t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (a_i - b_i)^2}\right).$$

Proof: First note that for all $\lambda > 0$, we have

$$\mathbf{Pr}\left[\sum_{i=1}^n X_i \geq t\right] = \mathbf{Pr}\left[\exp\left(\lambda \sum_{i=1}^n X_i\right) \geq \exp(\lambda t)\right].$$

By Markov's Inequality and the independence of all the X_is,

$$\begin{aligned}
\mathbf{Pr}\left[\exp\left(\lambda \sum_{i=1}^n X_i\right) \geq \exp(\lambda t)\right] &\leq \frac{\mathbf{E}\left[\exp\left(\lambda \sum_{i=1}^n X_i\right)\right]}{\exp(\lambda t)} \\
&\leq \exp(-\lambda t) \cdot \mathbf{E}\left[\prod_{i=1}^n \exp(\lambda X_i)\right] \\
&= \exp(-\lambda t) \cdot \prod_{i=1}^n \mathbf{E}\left[\exp(\lambda X_i)\right].
\end{aligned}$$

Applying Hoeffding's Lemma, we have

$$\exp(-\lambda t) \cdot \prod_{i=1}^{n} \mathbf{E}\left[\exp(\lambda X_i)\right] \leq \exp(-\lambda t) \cdot \prod_{i=1}^{n} \left(\exp\left(\lambda^2 (a_i - b_i)^2/8\right)\right)$$
$$= \exp\left(\frac{\sum_{i=1}^{n}(a_i - b_i)^2}{8}\lambda^2 - t\lambda\right).$$

The last term achieves the minimum when $\lambda = 4t/\left(\sum_{i=1}^{n}(a_i^2 - b_i^2)\right)$ so we can conclude that

$$\mathbf{Pr}\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(a_i - b_i)^2}\right). \blacksquare$$

4.2 Martingales

In this section, we introduce the concept of Martingales. Before this, let's first see a motivating example from gambling.

Example Each day a bookie offers a bet: you pay \$$b$ and you have a 50% chance of receiving \$$2b$ and a 50% chance of losing your money. Let Z_i be gambler's net gain on day i and X_i can be interpreted as the indicator variable for the outcome of the bet (*i.e.,* the r.v. X takes the values 1 and -1 with equal probability). We analyze the following two strategies:

- **Independent betting strategy:** always betting \$$c$, and the gambler's net gain on day n is
$$Z_n = \sum_{i=1}^{n} cX_i.$$
- **Martingale strategy:** On day n, bet δZ_{n-1}, where $\delta \in [0, 1]$. The change of wealth on day n can then be expressed recursively as
$$Z_n = Z_{n-1} + \delta Z_{n-1} X_{n-1}$$

Definition 4.5 (Martingales)**.** *A **martingale** sequence of random variables $Z_0, Z_1, \ldots, Z_n$ satisfies*

$$\mathbf{E}[Z_{i+1} | Z_0, \ldots, Z_i] = Z_i$$

for all $i = 0, 1, \ldots, n-1$.

Note 4.6. *We call $X_1, X_2, \ldots, X_n$ a martingale difference sequence if $Z_i = \sum_{j=1}^{i} X_j$ is a martingale sequence of random variables.*

One important inequality related to Martingales is Azuma's Inequality, which is similar to Hoeffding's Inequality.

Theorem 4.7 (Azuma's Inequality)**.** *Let $Z_0, Z_1, \ldots, Z_n$ be a martingale sequence of random variables such that for all i, there exists a constant c_i such that $|Z_i - Z_{i-1}| < c_i$, then*

$$\mathbf{Pr}[Z_n - Z_0 \geq t] \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^{n} c_i^2}\right).$$

Proof: The proof is modelled on that of Hoeffding's Inequality. First, using Markov's inequality and some algebra we have

$$\begin{aligned}
\mathbf{Pr}[Z_n - Z_0 \geq t] &= \mathbf{Pr}[\exp(\lambda(Z_n - Z_0)) \geq \exp(\lambda t)] \\
&\leq \exp(-\lambda t) \cdot \mathbf{E}[\exp(\lambda(Z_n - Z_0))] \\
&= \exp(-\lambda t) \cdot \mathbf{E}\left[\exp\left(\lambda \sum_{i=1}^{n} (Z_i - Z_{i-1})\right)\right] \\
&= \exp(-\lambda t) \cdot \mathbf{E}\left[\prod_{i=1}^{n} \exp(\lambda(Z_i - Z_{i-1}))\right].
\end{aligned}$$

We now we can always include additional conditional expectation so it follows that

$$\mathbf{Pr}[Z_n - Z_0 \geq t] \leq \exp(-\lambda t) \cdot \mathbf{E}\left[\mathbf{E}\left[\prod_{i=1}^{n} \exp(\lambda(Z_i - Z_{i-1})) | Z_0, Z_1, \ldots, Z_{n-1}\right]\right].$$

Since $\prod_{i=1}^{n} \exp(\lambda(Z_i - Z_{i-1}))$ is a constant once we condition on $Z_0, \cdots, Z_{n-1}$, we can take it out of the expectation so

$$\mathbf{Pr}[Z_n - Z_0 \geq t] \leq \exp(-\lambda t) \cdot \mathbf{E}\left[\left(\prod_{i=1}^{n-1} \exp(\lambda(Z_i - Z_{i-1}))\right) \mathbf{E}[\exp(\lambda(Z_n - Z_{n-1})) | Z_0, Z_1, \ldots, Z_{n-1}]\right]$$

Now, since (Z_i) is a Martingale, we know that $\mathbb{E}[Z_n - Z_{n-1} \mid Z_0, \cdots, Z_{n-1}] = 0$. Also, $|Z_n - Z_{n-1}| \leq c_n$ so using Hoeffding's lemma we have

$$\mathbf{Pr}[Z_n - Z_0 \geq t] \leq \exp(-\lambda t) \exp(\lambda^2 c_n^2 / 2) \cdot \mathbf{E}\left[\left(\prod_{i=1}^{n-1} \exp(\lambda(Z_i - Z_{i-1}))\right)\right].$$

It then follows from induction that

$$\mathbf{Pr}[Z_n - Z_0 \geq t] \leq \exp\left(\frac{\sum_{i=1}^{n} c_i^2}{2} \lambda^2 - t\lambda\right)$$

Finally, letting $\lambda = \frac{t}{\sum_{i=1}^{n} c_i^2}$ we get

$$\mathbf{Pr}[Z_n - Z_0 \geq t] \leq \exp\left(-\frac{t^2}{2 \sum_{i=1}^{n} c_i^2}\right)$$

■