Hoeffding's Inequality and Martingales

Chenye Yang

11/26/2022

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 and Markov's Inequality.

Lemma 1 (Hoeffding's Lemma). For a random variable $a \le X \le b$ such that $E[X] = 0$, we have

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

Proof: Note that $\exp(\lambda x)$ is a convex function, we have:

$$\exp(\lambda X) \le \frac{b-X}{b-a}\exp(\lambda a) + \frac{X-a}{b-a}\exp(\lambda b)$$

Take the expectation,

$$\begin{aligned} E[\exp(\lambda X)] &\le \frac{b-E[X]}{b-a}\exp(\lambda a) + \frac{E[X]-a}{b-a}\exp(\lambda b) \\ &= \frac{b}{b-a}\exp(\lambda a) + \frac{-a}{b-a}\exp(\lambda b) \\ &= \exp(L(\lambda(b-a))) \end{aligned}$$

where

$$L(h) = \frac{ha}{b-a} + \ln\left(1 + \frac{a - a\exp(h)}{b-a}\right)$$

Then we have

$$L(0) = L'(0) = 0 \qquad L''(h) = -\frac{ab\exp(h)}{(b-a\exp(h))^2}$$

By inequality of arithmetic and geometric means (AM–GM inequality) $\frac{x+y}{2} \ge \sqrt{xy}$,

$$\begin{aligned} & -ab\exp(h) = (b)\cdot(-a\exp(h)) \le \left(\frac{b-a\exp(h)}{2}\right)^2 \\ \Longrightarrow \quad & L''(h) \le \frac{1}{4}, \text{ for all } h \end{aligned}$$

From Tayler's theorem, for $\theta \in [0,1]$,

$$L(h) = L(0) + hL'(0) + \frac{1}{2}h^2L''(h\theta) = \frac{1}{2}h^2L''(h\theta) \le \frac{1}{8}h^2, \text{ for all } h$$

Therefore, by the mono-increasing of $\exp(x)$,

$$E[\exp(\lambda X)] \le \exp(L(\lambda(b-a))) \le \exp\left(\frac{1}{8}\lambda^2(b-a)^2\right)$$

Proved.

Theorem 1 (Markov's Inequality) X is a non-negative random variable, $a > 0$,

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Proof: By definition of $E[X]$,

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} xf(x)dx = \int_{0}^{\infty} xf(x)dx = \int_{0}^{a} xf(x)dx + \int_{a}^{\infty} xf(x)dx \\ &\geq \int_{a}^{\infty} xf(x)dx \\ &\geq \int_{a}^{\infty} af(x)dx = a\int_{a}^{\infty} f(x)dx = aP(X \geq a) \\ \implies \quad & \frac{E[X]}{a} \geq P(X \geq a) \end{aligned}$$

Proved.

Theorem 2 (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 $E[X_i] = 0$ for all $i = 1, 2, \ldots, n$. Then, for all $t > 0$,

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

Proof: For all $\lambda > 0$, by the monotonically increasing of $\exp(\cdot)$, we have

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

By Markov's inequality and the independence of all X_i:

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

Apply Hoeffding's Lemma, we have

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

The last term achieves the minimum when $\lambda = 4t/(\sum_{i=1}^{n}(a_i - b_i)^2)$, take λ as this, and we get

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

Proved.

2 Martingales

In this section, we introduce the concept of Martingales and show the related Azuma-Hoeffding Inequality.

Definition 1 (Martingales). A basic definition of a discrete-time martingale is a discrete-time stochastic process (i.e., a sequence of random variables) $X_1,\ X_2,\ X_3$ that satisfies for any time n,

$$\begin{aligned} &E[|X_n|] < \infty \\ &E[X_{n+1} \mid X_1, \ldots, X_n] = X_n \end{aligned}$$

That is, the conditional expected value of the next observation, given all the past observations, is equal to the most recent observation.

Definition 2 (Martingale Difference Sequence). A martingale difference sequence (MDS) is related to the concept of the martingale. A stochastic series X is an MDS if its expectation with respect to the past is zero. Formally, consider an adapted sequence $\{X_t, \mathcal{F}_t\}_{-\infty}^{\infty}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. X_t is an MDS if it satisfies the following two conditions:

$$\begin{aligned} &E\left[|X_t|\right] < \infty \\ &E\left[X_t | \mathcal{F}_{t-1}\right] = 0, a.s. \end{aligned}$$

for all t. By construction, this implies that if Y_t is a martingale, then $X_t = Y_t - Y_{t-1}$ will be an MDS—hence the name.

Theorem 3 (Azuma-Hoeffding 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

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

Proof: By Markov Inequality,

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

By using the iterated expectation property that

$$E[g(X,Y)] = E_Y[E_{X|Y}[g(X,Y) \mid Y]]$$

We have

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

Since $\prod_{i=1}^{n-1} \exp\left(\lambda(Z_i - Z_{i-1})\right)$ is a constant once given $Z_0, Z_1, \ldots, Z_{n-1}$, we can take it out of the expectation:

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

Because (Z_i) is a Martingale sequence, we have

$$E[Z_n - Z_{n-1} \mid Z_0, Z_1, \ldots, Z_{n-1}] = E[Z_n \mid Z_0, Z_1, \ldots, Z_{n-1}] - Z_{n-1} = Z_{n-1} - Z_{n-1} = 0$$

Also, $|Z_n - Z_{n-1}| \leq c_n$. Then we can apply Hoeffding's lemma:

$$E\left[\exp\left(\lambda(Z_n - Z_{n-1})\right) \mid Z_0, Z_1, \ldots, Z_{n-1}\right] \leq \exp\left(\frac{\lambda^2(2c_n)^2}{8}\right) = \exp\left(\frac{\lambda^2 c_n^2}{2}\right)$$

Then

$$\begin{aligned} P[Z_n - Z_0 \geq t] &\leq \exp(-\lambda t) E\left[\prod_{i=1}^{n-1} \exp\left(\lambda(Z_i - Z_{i-1})\right) \cdot E\left[\exp\left(\lambda(Z_n - Z_{n-1})\right) \mid Z_0, Z_1, \ldots, Z_{n-1}\right]\right] \\ &\leq \exp(-\lambda t) \exp\left(\frac{\lambda^2 c_n^2}{2}\right) E\left[\prod_{i=1}^{n-1} \exp\left(\lambda(Z_i - Z_{i-1})\right)\right] \end{aligned}$$

By induction,

$$\begin{aligned} P[Z_n - Z_0 \geq t] &\leq \exp(-\lambda t) \exp\left(\frac{\lambda^2 c_n^2}{2}\right) E\left[\prod_{i=1}^{n-1} \exp\left(\lambda(Z_i - Z_{i-1})\right)\right] \\ &\leq \exp(-\lambda t) \exp\left(\frac{\lambda^2 c_n^2}{2}\right) \exp\left(\frac{\lambda^2 c_{n-1}^2}{2}\right) E\left[\prod_{i=1}^{n-2} \exp\left(\lambda(Z_i - Z_{i-1})\right)\right] \\ &\ldots \\ &\leq \exp(-\lambda t) \exp\left(\frac{\lambda^2 \sum_{i=1}^{n}(c_i^2)}{2}\right) \end{aligned}$$

Letting $\lambda = \frac{t}{\sum_{i=1}^{n}(c_i^2)}$, we get

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

Proved.