CONCENTRATION INEQUALITIES

1. Introduction

Concentration inequalities estimate deviation of random variable from its mean value or variance. In this short notes we prove Azuma-Hoeffding inequality.

2. Azuma and Hoeffding inequality

Theorem 2.1 (Azuma inequality)**.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent random variables on a probability space $(\Omega, \mathcal{F}, P)$. Assume that for every $n \in \mathbb{N}$ there exists real number c_n such that*

$$|X_n| \le c_n$$

almost surely and that $\mathbb{E}[X_n] = 0$ for every $n \in \mathbb{N}$. Then

$$P\left(X_0 + X_1 + ... + X_n \ge \lambda\right) \le \exp\left(\frac{-\lambda^2}{2 \cdot \left(c_0^2 + c_1^2 + ... + c_n^2\right)}\right)$$

for every $\lambda \ge 0$.

Lemma 2.1.1. *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, P)$. Suppose that $\mathcal{G} \subseteq \mathcal{F}$ is a σ-subalgebra and c is a positive real number. Assume that $\mathbb{E}[X\,|\,\mathcal{G}] = 0$ and $|X| \le c$ almost surely. Then for every convex function $\phi : \mathbb{R} \to \mathbb{R}$ we have*

$$\mathbb{E}[\phi(X)\,|\,\mathcal{G}] \le \frac{\phi(-c) + \phi(c)}{2}$$

Proof of the lemma. Since ϕ is convex, we derive that

$$\phi(x) \le \frac{c - x}{2c} \cdot \phi(-c) + \frac{c + x}{2c} \cdot \phi(c)$$

for every $x \in [-c, c]$. Hence the inequality

$$\phi(X) \le \frac{c - X}{2c} \cdot \phi(-c) + \frac{c + X}{2c} \cdot \phi(c)$$

holds almost surely. Applying conditional expectation and using the fact that it is a monotone operator, we deduce that

$$\mathbb{E}[\phi(X)\,|\,\mathcal{G}] \le \frac{c - \mathbb{E}[X\,|\,\mathcal{G}]}{2c} \cdot \phi(-c) + \frac{c + \mathbb{E}[X\,|\,\mathcal{G}]}{2c} \cdot \phi(c) = \frac{\phi(-c) + \phi(c)}{2}$$

□

Lemma 2.1.2. *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, P)$. Suppose that $\mathcal{G} \subseteq \mathcal{F}$ is a σ-subalgebra and c is a positive real number. Assume that $\mathbb{E}[X\,|\,\mathcal{G}] = 0$ and $|X| \le c$ almost surely. Then for every $\theta > 0$ we have*

$$\mathbb{E}[e^{\theta X}\,|\,\mathcal{G}] \le \exp\left(\frac{\theta^2 \cdot c^2}{2}\right)$$

Proof of the lemma. Note that

$$\mathbb{E}[e^{\theta X}\,|\,\mathcal{G}] \le \frac{e^{-\theta\cdot c} + e^{\theta \cdot c}}{2} = \cosh(\theta \cdot c)$$

by Lemma 2.1.1. Next observe that

$$\cosh(x) = \frac{e^x + e^{-x}}{2} = \frac{1}{2}\cdot\left(\sum_{n=0}^{+\infty}\frac{x^n}{n!} + \sum_{n=0}^{+\infty}\frac{(-1)^n x^n}{n!}\right) = \sum_{n=0}^{+\infty}\frac{x^{2n}}{(2n)!} \le \sum_{n=0}^{+\infty}\frac{x^{2n}}{2^n\cdot n!} = \exp\left(\frac{x^2}{2}\right)$$

for every $x \in \mathbb{R}$. Hence

$$\mathbb{E}[e^{\theta X}\,|\,\mathcal{G}] \le \exp\left(\frac{\theta^2\cdot c^2}{2}\right)$$

□

Proof of the theorem. Suppose that $\lambda \ge 0$ and $\theta > 0$. We have

$$P\left(X_0 + X_1 + \ldots + X_n \ge \lambda\right) = P\left(e^{\theta\cdot(X_0+X_1+\ldots+X_n)} \ge e^{\theta\cdot\lambda}\right)$$

Applying Markov inequality, we derive that

$$P\left(e^{\theta\cdot(X_0+X_1+\ldots+X_n)} \ge e^{\theta\cdot\lambda}\right) \le e^{-\theta\cdot\lambda}\cdot\mathbb{E}\left[e^{\theta\cdot(X_0+X_1+\ldots+X_n)}\right] = e^{-\theta\cdot\lambda}\cdot\mathbb{E}\left[e^{\theta\cdot X_0+\theta\cdot X_1+\ldots+\theta\cdot X_n}\right]$$

Now let $\mathcal{F}_{n-1}$ be a σ-algebra generated by random variables $X_0,\ldots,X_{n-1}$. According to the standard properties of conditional expectation we have

$$e^{-\theta\cdot\lambda}\cdot\mathbb{E}\left[e^{\theta\cdot X_0+\theta\cdot X_1+\ldots+\theta\cdot X_n}\right] = e^{-\theta\cdot\lambda}\cdot\mathbb{E}\left[\mathbb{E}\left[e^{\theta\cdot X_0+\theta\cdot X_1+\ldots+\theta\cdot X_{n-1}}\cdot e^{\theta\cdot X_n}\,|\,\mathcal{F}_{n-1}\right]\right] =$$

$$= e^{-\theta\cdot\lambda}\cdot\mathbb{E}\left[\mathbb{E}\left[e^{\theta\cdot X_0+\theta\cdot X_1+\ldots+\theta\cdot X_{n-1}}|\,\mathcal{F}_{n-1}\right]\cdot\mathbb{E}\left[e^{\theta\cdot X_n}\,|\,\mathcal{F}_{n-1}\right]\right]$$

Since $\mathbb{E}\left[X_n\,|\,\mathcal{F}_{n-1}\right] = 0$ and $|X_n| \le c_n$ almost surely, by Lemma 2.1.2 we have

$$\mathbb{E}\left[e^{\theta\cdot X_n}\,|\,\mathcal{F}_{n-1}\right] \le \exp\left(\frac{\theta^2\cdot c_n^2}{2}\right)$$

and thus we deduce that

$$e^{-\theta\cdot\lambda}\cdot\mathbb{E}\left[e^{\theta\cdot X_0+\theta\cdot X_1+\ldots+\theta\cdot X_n}\right] = e^{-\theta\cdot\lambda}\cdot\mathbb{E}\left[\mathbb{E}\left[e^{\theta\cdot X_0+\theta\cdot X_1+\ldots+\theta\cdot X_{n-1}}|\,\mathcal{F}_{n-1}\right]\cdot\mathbb{E}\left[e^{\theta\cdot X_n}\,|\,\mathcal{F}_{n-1}\right]\right] \le$$

$$\le e^{-\theta\cdot\lambda}\cdot\mathbb{E}\left[\mathbb{E}\left[e^{\theta\cdot X_0+\theta\cdot X_1+\ldots+\theta\cdot X_{n-1}}|\,\mathcal{F}_{n-1}\right]\cdot\exp\left(\frac{\theta^2\cdot c_n^2}{2}\right)\right] = e^{-\theta\cdot\lambda}\cdot\mathbb{E}\left[e^{\theta\cdot X_0+\theta\cdot X_1+\ldots+\theta\cdot X_{n-1}}\right]\cdot\exp\left(\frac{\theta^2\cdot c_n^2}{2}\right)$$

for every $n \in \mathbb{N}$. Hence by easy induction

$$e^{-\theta\cdot\lambda}\cdot\mathbb{E}\left[e^{\theta\cdot X_0+\theta\cdot X_1+\ldots+\theta\cdot X_n}\right] \le \exp\left(-\theta\cdot\lambda\right)\cdot\exp\left(\frac{\theta^2\cdot(c_0^2+c_1^2+\ldots+c_n^2)}{2}\right)$$

Therefore, we deduce that inequality

$$P\left(X_0 + X_1 + \ldots + X_n \ge \lambda\right) \le \exp\left(\frac{c_0^2+c_1^2+\ldots+c_n^2}{2}\cdot\theta\cdot\left(\theta - \frac{2\cdot\lambda}{c_0^2+c_1^2+\ldots+c_n^2}\right)\right)$$

holds for every $\theta > 0$. The right hand side of the inequality is continuous for every $\theta \in [0,+\infty)$ and attains global minimum for

$$\theta = \frac{\lambda}{c_0^2+c_1^2+\ldots+c_n^2} \in [0,+\infty)$$

Hence finally

$$P\left(X_0 + X_1 + \ldots + X_n \ge \lambda\right) \le \exp\left(\frac{-\lambda^2}{2\cdot\left(c_0^2+c_1^2+\ldots+c_n^2\right)}\right)$$

□

Corollary 2.2. *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent random variables in probability space $(\Omega, \mathcal{F}, P)$. Assume that for every $n \in \mathbb{N}$ there exists real number c_n such that*

$$|X_n| \leq c_n$$

almost surely. Then

$$P\left(|X_0 + X_1 + ... + X_n| \geq \lambda\right) \leq 2 \cdot \exp\left(\frac{-\lambda^2}{2 \cdot \left(c_0^2 + c_1^2 + ... + c_n^2\right)}\right)$$

for every $\lambda \geq 0$.

Proof. Fix $\lambda \geq 0$. According to Theorem 2.1 we have

$$P\left(X_0 + X_1 + ... + X_n \geq \lambda\right) \leq \exp\left(\frac{-\lambda^2}{2 \cdot \left(c_0^2 + c_1^2 + ... + c_n^2\right)}\right)$$

Applying Theorem 2.1 to a sequence $\{-X_n\}_{n\in\mathbb{N}}$ we derive

$$P\left(X_0 + X_1 + ... + X_n \leq -\lambda\right) \leq \exp\left(\frac{-\lambda^2}{2 \cdot \left(c_0^2 + c_1^2 + ... + c_n^2\right)}\right)$$

Merging these two inequalities we obtain the assertion. □

Corollary 2.3 (Hoeffding inequality)**.** *Let $\{X_n\}_{n\geq 1}$ be a sequence of independent and identically distributed random variables in probability space $(\Omega, \mathcal{F}, P)$. Assume that there exists a positive real number c such that*

$$|X_1| \leq c$$

almost surely and let m be the expected value of X_1. Then

$$P\left(\left|\frac{X_1 + ... + X_n}{n} - m\right| \geq \lambda\right) \leq 2 \cdot \exp\left(\frac{-\lambda^2 \cdot n}{2 \cdot (c + |m|)^2}\right)$$

for every $\lambda \geq 0$.

Proof. Write $Z_n = X_n - \mathbb{E}[X_n] = X_n - m$. Then $\{Z_n\}_{n\geq 1}$ are independent and $|Z_n| \leq c + |m|$. Fix $\lambda \geq 0$. Then applying Corollary 2.2 we derive that

$$P\left(\left|\frac{X_1 + ... + X_n}{n} - m\right| \geq \lambda\right) = P\left(|Z_1 + ... + Z_n| \geq n \cdot \lambda\right) \leq 2 \cdot \exp\left(\frac{-\lambda^2 \cdot n^2}{2 \cdot n \cdot (c + |m|)^2}\right) = 2 \cdot \exp\left(\frac{-\lambda^2 \cdot n}{2 \cdot (c + |m|)^2}\right)$$

□