

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-HOEFFDING INEQUALITY

Theorem 2.1 (Azuma-Hoeffding inequality). *Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of independent random variables on a probability space (Ω, \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 and that $\mathbb{E}[X_n] = 0$ for every $n \in \mathbb{N}$. Then

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

for every $\lambda \geq 0$.

Lemma 2.1.1. *Let X be a random variable on a probability space (Ω, \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 $|X| \leq c$ almost surely. Then for every convex function $\phi : \mathbb{R} \rightarrow \mathbb{R}$ we have*

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

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

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

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

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

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

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

□

Lemma 2.1.2. *Let X be a random variable on a probability space (Ω, \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 $|X| \leq c$ almost surely. Then for every $\theta > 0$ we have*

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

Proof of the lemma. Note that

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

by Lemma 2.1.1. Next observe that

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

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

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

□

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

$$P(X_0 + X_1 + \dots + X_n \geq \lambda) = P(e^{\theta \cdot (X_0 + X_1 + \dots + X_n)} \geq e^{\theta \cdot \lambda})$$

Applying Markov inequality, we derive that

$$P(e^{\theta \cdot (X_0 + X_1 + \dots + X_n)} \geq e^{\theta \cdot \lambda}) \leq e^{-\theta \cdot \lambda} \cdot \mathbb{E}[e^{\theta \cdot (X_0 + X_1 + \dots + X_n)}] = e^{-\theta \cdot \lambda} \cdot \mathbb{E}[e^{\theta \cdot X_0 + \theta \cdot X_1 + \dots + \theta \cdot X_n}]$$

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

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

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

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

and thus we deduce that

$$\begin{aligned} e^{-\theta \cdot \lambda} \cdot \mathbb{E}[e^{\theta \cdot X_0 + \theta \cdot X_1 + \dots + \theta \cdot X_n}] &= e^{-\theta \cdot \lambda} \cdot \mathbb{E}\left[\mathbb{E}[e^{\theta \cdot X_0 + \theta \cdot X_1 + \dots + \theta \cdot X_{n-1}} | \mathcal{F}_{n-1}] \cdot \mathbb{E}[e^{\theta \cdot X_n} | \mathcal{F}_{n-1}]\right] \leq \\ &\leq e^{-\theta \cdot \lambda} \cdot \mathbb{E}\left[\mathbb{E}[e^{\theta \cdot X_0 + \theta \cdot X_1 + \dots + \theta \cdot X_{n-1}} | \mathcal{F}_{n-1}] \cdot \exp\left(\frac{\theta^2 \cdot c_n^2}{2}\right)\right] = e^{-\theta \cdot \lambda} \cdot \mathbb{E}[e^{\theta \cdot X_0 + \theta \cdot X_1 + \dots + \theta \cdot X_{n-1}}] \cdot \exp\left(\frac{\theta^2 \cdot c_n^2}{2}\right) \end{aligned}$$

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

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

Therefore, we deduce that inequality

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

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

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

Hence finally

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

□

Corollary 2.2. Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of independent random variables in probability space (Ω, \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

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

for every $\lambda \geq 0$.

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

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

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

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

Merging these two inequalities we obtain the assertion. \square

Corollary 2.3 (Hoeffding inequality). Let $\{X_n\}_{n \geq 1}$ be a sequence of independent and identically distributed random variables in probability space (Ω, \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

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

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

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

\square