Review on conditional Arzuma-Hoeffding inequality : proof and generalization

Chenlin GU

February 24, 2018

Contents

1 Introduction

In this report, we review the proof of Arzuma-Hoeffding inequality, which is a martingale version of classical Hoeffding inequality.

Theorem 1.1 (Arzuma-Hoeffding)**.** *Given probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}), \mathbb{P})$, and a series of adapted random variables $(X_n)_{n\geq 0}$ satisfying almost surely $a_n \leq X_n \leq b_n$, then $\forall \epsilon > 0$, we have*

$$\mathbb{P}\left[\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mathbb{E}[X_k|\mathcal{F}_{k-1}] > \epsilon\right] \leq \exp\left(-\frac{2\epsilon^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\right)$$

In the case where $(X_n)_{n\geq 0}$ is independent, it becomes a classical Hoeffding inequality. The proof can be divided generally in three steps

1. Apply Markov inequality to the Laplace transform.

2. Estimate the log-Laplace transform by the Hoeffding lemma.

3. Optimize the parameter.

First and third step is more or less easy, however the second step requires a little effort which writes

Lemma 1.1 (Hoeffding)**.** *For any random variable X satisfying $a \leq X \leq b$, then $\forall \lambda > 0$*

$$\log \mathbb{E}[e^{\lambda X}] \leq \lambda \mathbb{E}[X] + \frac{\lambda^2}{8}(b-a)^2$$

In the proof of Arzuma-Hoeffding, it requires a conditional version of this lemma

Lemma 1.2 (Conditional Hoeffding)**.** *For any random variable X satisfying $a \leq X \leq b$, then $\forall \lambda > 0$ and σ-algebra $\mathcal{G}$*

$$\log \mathbb{E}[e^{\lambda X}|\mathcal{G}] \leq \lambda \mathbb{E}[X|\mathcal{G}] + \frac{\lambda^2}{8}(b-a)^2$$

During the course, we give three proofs but we would like know if we can follow the same step as classical Hoeffding lemma.

In the following part, we will prove it by the same strategy applying change of probability under different σ-algebra. Moreover, we will prove a stronger version of Arzuma-Hoeffding says

Theorem 1.2 (Bounded difference Arzuma-Hoeffding)**.** *Given probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}), \mathbb{P})$, and a series of adapted random variables $(X_n)_{n\geq 0}$ satisfying almost surely $G_n + a_n \leq X_n \leq G_n + b_n$ where G_n is predicable($\mathcal{F}_{n-1}$ adapted), we suppose only L^1 condition for $(G_n)_{n\geq 0}, (X_n)_{n\geq 0}$ then $\forall \epsilon > 0$, we have*

$$\mathbb{P}\left[\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mathbb{E}[X_k|\mathcal{F}_{k-1}] > \epsilon\right] \leq \exp\left(-\frac{2\epsilon^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\right)$$

2 Proof of conditional Hoeffding lemma

We recap the proof of classical Hoeffding lemma quickly.

Proof. **Classical Hoeffding lemma** We denote log-Laplace

$$\psi(\lambda) = \log \mathbb{E}[e^{\lambda X}]$$

and a change of probability

$$\frac{d\mathbb{Q}^\lambda}{d\mathbb{P}} = \frac{e^{\lambda X}}{\mathbb{E}[e^{\lambda X}]}$$

, of course this change of probability requires the integrability. Then the first and second derivative of $\psi(\lambda)$ has a very impressing interpretation

$$\begin{aligned} \psi'(\lambda) &= \frac{\mathbb{E}[Xe^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} = \mathbb{E}_{\mathbb{Q}^\lambda}[X] \\ \psi''(\lambda) &= \frac{\mathbb{E}[X^2 e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} - \left(\frac{\mathbb{E}[Xe^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]}\right)^2 = \mathrm{Var}_{\mathbb{Q}^\lambda}[X] \end{aligned}$$

Therefore,

$$\begin{aligned}
\psi(\lambda) &= \psi(0) + \int_0^\lambda \psi'(t)dt \\
&= \psi(0) + \lambda\psi'(0) + \int_0^\lambda \int_0^t \psi''(s)dsdt \\
&= \psi(0) + \lambda\psi'(0) + \int_0^\lambda \int_0^t \mathrm{Var}_{\mathbb{Q}^s}[X]dsdt \\
&\leq \lambda\mathbb{E}[X] + \frac{\lambda^2}{8}(b-a)^2
\end{aligned}$$

In the last step, we use the fact that the variance of X is bounded by $\frac{(b-a)^2}{4}$ under any probability space $\mathbb{Q}$ since

$$\mathrm{Var}_{\mathbb{Q}}[X] = \mathbb{E}_{\mathbb{Q}}\left[(X - \mathbb{E}_{\mathbb{Q}}[X])^2\right] \leq \mathbb{E}_{\mathbb{Q}}\left[(X - \frac{(a+b)}{2})^2\right] \leq \frac{(b-a)^2}{4}$$

$\square$

In fact, we can follow the same procedure as above. However, we have to recall some basic property of conditional expectation and change of probability under different different σ-algebra.

Proposition 2.1. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and another σ-algebra $\mathcal{G} \subset \mathcal{F}$. We also has another probability defined on $\mathcal{F}$ such that*

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = L$$

To reduce the notation, we write $\mathbb{E}_{\mathbb{Q}}[\cdot]$ if we apply expectation under probability $\mathbb{Q}$ while keep $\mathbb{E}[\cdot]$ for the expectation under $\mathbb{P}$. We recall the following properties about the conditional expectation.

1. ***Conditional expectation*** *$\mathbb{E}[X|\mathcal{G}]$ is a $\mathcal{G}$-measurable random variable such that*
$$\forall A \in \mathcal{G}, \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A]$$
. The conditional expectation also follows the monotony i.e if $X \leq Y, \mathbb{P} - p.s$ then $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$.

2. ***Change of probability*** *Under another σ-algebra $\mathcal{G}$, we have*
$$\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{G}} = \mathbb{E}_{\mathbb{P}}[L|\mathcal{G}]$$

3. ***Bayes formula*** *We can do conditional expectation under probability $\mathbb{Q}$*
$$\mathbb{E}_{\mathbb{Q}} = \frac{1}{\mathbb{E}[L|\mathcal{G}]}\mathbb{E}[LX|\mathcal{G}]$$
(With convention $\frac{0}{0} = 0$)

Proof. 1. is just the definition of conditional expectation.

2. $\forall A \in \mathcal{G}$,

$$Q[A] = \mathbb{E}_Q[\mathbf{1}_A] = \mathbb{E}[L\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[L|\mathcal{G}]\mathbf{1}_A]$$

this implies that $\frac{d\mathbb{Q}}{d\mathbb{P}}|_{\mathcal{G}} = \mathbb{E}_{\mathbb{P}}[L|\mathcal{G}]$.

3. We check the definition of conditional expectation under the probability $\mathbb{Q}$ applying the second property.

$$\begin{aligned}
\mathbb{E}_Q\left[\frac{1}{\mathbb{E}[L|\mathcal{G}]}\mathbb{E}[LX|\mathcal{G}]\mathbf{1}_A\right] &= \mathbb{E}[\mathbb{E}[LX|\mathcal{G}]\mathbf{1}_A] \\
&= \mathbb{E}[LX\mathbf{1}_A] \\
&= \mathbb{E}_Q[X\mathbf{1}_A]
\end{aligned}$$

□

With these propositions, we can prove the conditional Hoeffding lemma in the same way of classical Hoeffding lemma.

Proof. **Conditional Hoeffding lemma** We denote now

$$\psi(\lambda) = \log \mathbb{E}[e^{\lambda X}|\mathcal{G}]$$

then we have

$$\begin{aligned}
\psi'(\lambda) &= \frac{\mathbb{E}[Xe^{\lambda X}|\mathcal{G}]}{\mathbb{E}[e^{\lambda X}|\mathcal{G}]} \\
\psi''(\lambda) &= \frac{\mathbb{E}[X^2e^{\lambda X}|\mathcal{G}]}{\mathbb{E}[e^{\lambda X}|\mathcal{G}]} - \left(\frac{\mathbb{E}[Xe^{\lambda X}|\mathcal{G}]}{\mathbb{E}[e^{\lambda X}|\mathcal{G}]}\right)^2
\end{aligned}$$

We have to interpret these derivative. In fact, by defining

$$\frac{d\mathbb{Q}^\lambda}{d\mathbb{P}} = L_\lambda = \frac{e^{\lambda X}}{\mathbb{E}[e^{\lambda X}|\mathcal{G}]}$$

we see easily it's a density since $\mathbb{E}[L_\lambda|\mathcal{G}] = 1$ so $\mathbb{E}[L_\lambda] = \mathbb{E}[\mathbb{E}[L_\lambda|\mathcal{G}]] = 1$. Then we check the third property

$$\mathbb{E}_{\mathbb{Q}^\lambda}[X|G] = \frac{1}{\mathbb{E}[L_\lambda|\mathcal{G}]}\mathbb{E}[L_\lambda X|\mathcal{G}] = \frac{\mathbb{E}[Xe^{\lambda X}|\mathcal{G}]}{\mathbb{E}[e^{\lambda X}|\mathcal{G}]}$$

Therefore, we obtain the same interpretation as change of probability applying in conditional expectation.

$$\begin{aligned}
\psi'(\lambda) &= \mathbb{E}_{\mathbb{Q}^\lambda}[X|\mathcal{G}] \\
\psi''(\lambda) &= \mathbb{E}_{\mathbb{Q}^\lambda}[X^2|\mathcal{G}] - (\mathbb{E}_{\mathbb{Q}^\lambda}[X|\mathcal{G}])^2 \\
&= \mathbb{E}_{\mathbb{Q}^\lambda}\left[(X - \mathbb{E}_{\mathbb{Q}^\lambda}[X|\mathcal{G}])^2|\mathcal{G}\right] := \mathrm{Var}_{\mathbb{Q}^\lambda}[X]
\end{aligned}$$

□

The second one has a similar bound estimation of conditional variance by applying the bias-mean decomposition, i.e $\forall Y$ a $\mathcal{G}$-measurable and $\forall \mathbb{Q}$ a probability

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}[(X-Y)^2|\mathcal{G}] &= \mathbb{E}_{\mathbb{Q}}\left[(X-\mathbb{E}_{\mathbb{Q}}[X|\mathcal{G}])^2+(\mathbb{E}_{\mathbb{Q}}[X|\mathcal{G}]-Y)^2+2(X-\mathbb{E}_{\mathbb{Q}}[X|\mathcal{G}])\times(\mathbb{E}_{\mathbb{Q}}[X|\mathcal{G}]-Y)|\mathcal{G}\right] \\
&= \mathbb{E}_{\mathbb{Q}}\left[(X-\mathbb{E}_{\mathbb{Q}}[X|\mathcal{G}])^2|\mathcal{G}\right]+(\mathbb{E}_{\mathbb{Q}}[X|\mathcal{G}]-Y)^2 \\
&\geq \mathbb{E}_{\mathbb{Q}}\left[(X-\mathbb{E}_{\mathbb{Q}}[X|\mathcal{G}])^2|\mathcal{G}\right]
\end{aligned}$$

On apply $Y=\frac{a+b}{2}$ for $\mathbb{Q}^\lambda$ and obtain that

$$\begin{aligned}
\psi''(\lambda) &= \mathbb{E}_{\mathbb{Q}^\lambda}[(X-\mathbb{E}_{\mathbb{Q}^\lambda}[X|\mathcal{G}])^2|\mathcal{G}] \\
&\leq \mathbb{E}_{\mathbb{Q}^\lambda}[(X-\frac{a+b}{2})^2|\mathcal{G}] \\
&\leq \mathbb{E}_{\mathbb{Q}^\lambda}\left[\left(\frac{b-a}{2}\right)^2|\mathcal{G}\right]=\frac{(b-a)^2}{4}
\end{aligned}$$

The rest part is just like the classical Hoeffding lemma.

3 Generalization to bounded difference case

We prove the generalized case. We study a case where $G+a\leq X\leq G+b$ where G is $\mathcal{G}$-measurable. Then we apply Hoeffding lemma to $X-G$ and obtain

$$\log\mathbb{E}[e^{\lambda(X-G)}|\mathcal{G}]\leq\lambda\mathbb{E}[X-G|\mathcal{G}]+\frac{\lambda^2}{8}(b-a)^2$$

However, since we don't know $\mathbb{E}[e^{\lambda X}]<\infty$, we cannot expect a common version of conditional Hoeffding lemma. One correct way to see it may be defining

$$\mathbb{E}[e^{\lambda(X-G)}|\mathcal{G}]=\lim_{N\to\infty}e^{-\lambda G}\mathbb{E}[e^{\lambda X}\mathbf{1}_{|G|\leq N}|\mathcal{G}] \tag{1}$$

Since $e^{\lambda X}\mathbf{1}_{|G|\leq N}$ is always bounded, the definition always makes sense and the limit is a convergence monotone.(But the limit may not be L^1.)

We check this definition. $\forall A\in\mathcal{G}$

$$\begin{aligned}
&\mathbb{E}\left[\lim_{N\to\infty}e^{-\lambda G}\mathbb{E}[e^{\lambda X}\mathbf{1}_{|G|\leq N}|\mathcal{G}]\mathbf{1}_A\right] \\
=\ &\mathbb{E}\left[\lim_{N\to\infty}\mathbb{E}[e^{\lambda(X-G)}\mathbf{1}_{|G|\leq N}\mathbf{1}_A|\mathcal{G}]\right] \\
=\ &\mathbb{E}\left[\mathbb{E}[\lim_{N\to\infty}e^{\lambda(X-G)}\mathbf{1}_{|G|\leq N}\mathbf{1}_A|\mathcal{G}]\right] \text{ (Monotone convergence)} \\
=\ &\mathbb{E}\left[\mathbb{E}[e^{\lambda(X-G)}\mathbf{1}_A|\mathcal{G}]\right] \\
=\ &\mathbb{E}\left[e^{\lambda(X-G)}\mathbf{1}_A\right]
\end{aligned}$$

Therefore, we get an asymptotic conditional Hoeffding lemma

$$\lim_{N\to\infty} \log \mathbb{E}[e^{\lambda X}\mathbf{1}_{|G|\le N}|\mathcal{G}] \le \lambda\mathbb{E}[X|\mathcal{G}] + \frac{\lambda^2}{8}(b-a)^2$$

This is good, but it helps us so little since we could not apply an asymptotic version of Hoeffding lemma to the proof. However, this inspires us that the lemma works once we truncate the random variable. We give two proofs to the same conclusion by going over the problem of integrability applying truncation.

Remark. The second comes first after I notice that the above asymptotic conditional Hoeffding formula helps so little. However, when I review the note of the course, I believe that Prof.Stoltz may want not only well defined the conditional expectation, but also a modified Hoeffding lemma that works in a general case. Thus I revise the formula and follow the same idea but get another proof.

Proof. (**Proof 1 : Truncate the random variable directly**) We modify a little the asymptotic conditional formula and get the following lemma.

Lemma 3.1. *We suppose that $X, G \in L^1(\Omega)$ and G is $\mathcal{G}$-measurable and $G + a \le X \le G + b$. Then for all $\lambda > 0$, we have*

$$\mathbb{E}[e^{\lambda(X-G)}|\mathcal{G}] = \lim_{N\to\infty} \mathbb{E}[e^{\lambda(X-G)\mathbf{1}_{|G|\le N}}|\mathcal{G}] \tag{2}$$

This expression gives us that

$$\lim_{N\to} \log \mathbb{E}[e^{\lambda X\mathbf{1}_{|G|\le N}}|\mathcal{G}] \le \lambda\mathbb{E}[X|\mathcal{G}] + \frac{\lambda^2}{8}(b-a)^2 \tag{3}$$

This lemma is just a revised version of the previous one. We give its proof easily.

Proof. $\forall A \in \mathcal{G}$, we have

$$\begin{aligned}
& \mathbb{E}\left[\lim_{N\to\infty} \mathbb{E}[e^{\lambda(X-G)\mathbf{1}_{|G|\le N}}|\mathcal{G}]\mathbf{1}_A\right] \\
= \ & \mathbb{E}\left[\lim_{N\to\infty} \mathbb{E}[e^{\lambda(X-G)\mathbf{1}_{|G|\le N}}\mathbf{1}_A|\mathcal{G}]\right] \\
= \ & \mathbb{E}\left[\mathbb{E}[\lim_{N\to\infty} e^{\lambda(X-G)\mathbf{1}_{|G|\le N}}\mathbf{1}_A|\mathcal{G}]\right] \text{ (Dominated convergence theorem)} \\
= \ & \mathbb{E}\left[\mathbb{E}[e^{\lambda(X-G)\mathbf{1}_{|G|\le N}}\mathbf{1}_A|\mathcal{G}]\right] \\
= \ & \mathbb{E}\left[e^{\lambda(X-G)\mathbf{1}_{|G|\le N}}\mathbf{1}_A\right]
\end{aligned}$$

Since $\mathbb{E}[e^{\lambda X \mathbf{1}_{|G|\leq N}}|\mathcal{G}]$ is always well defined, we have

$$\begin{aligned}
\log \mathbb{E}[e^{\lambda(X-G)}|\mathcal{G}] &= \lim_{N\to\infty} \log \mathbb{E}[e^{\lambda(X-G)\mathbf{1}_{|G|\leq N}}|\mathcal{G}] \\
&= \lim_{N\to\infty} \mathbb{E}[e^{\lambda X \mathbf{1}_{|G|\leq N}}|\mathcal{G}] - \lambda G \mathbf{1}_{|G|\leq N} \\
&= \lim_{N\to\infty} \mathbb{E}[e^{\lambda X \mathbf{1}_{|G|\leq N}}|\mathcal{G}] - \lambda G
\end{aligned}$$

Therefore, we get

$$\lim_{N\to\infty} \log \mathbb{E}[e^{\lambda X \mathbf{1}_{|G|\leq N}}|\mathcal{G}] \leq \lambda \mathbb{E}[X|\mathcal{G}] + \frac{\lambda^2}{8}(b-a)^2$$

□

We apply this asymptotic conditional inequality to the Arzuma-Hoeffding inequality.

$$\begin{aligned}
& \mathbb{P}[\sum_{k\leq n}(X_k - \mathbb{E}[X_k|\mathcal{F}_{k-1}]) > \epsilon] \\
= & \lim_{\forall k\leq n, N_k\to\infty} \mathbb{P}[\sum_{k\leq n}(X_k \mathbf{1}_{|G_k|\leq N_k} - \mathbb{E}[X_k|\mathcal{F}_{k-1}]) > \epsilon] \\
= & \lim_{\forall k\leq n, N_k\to\infty} e^{-\lambda\epsilon} \mathbb{E}[e^{\lambda \sum_{k\leq n}(X_k \mathbf{1}_{|G_k|\leq N_k} - \mathbb{E}[X_k|\mathcal{F}_{k-1}])}] \\
= & \lim_{\forall k\leq n, N_k\to\infty} e^{-\lambda\epsilon} \mathbb{E}\left[e^{\lambda \sum_{k\leq n-1}(X_k \mathbf{1}_{|G_k|\leq N_k} - \mathbb{E}[X_k|\mathcal{F}_{k-1}])} \mathbb{E}[e^{\lambda X_n \mathbf{1}_{|G_n|\leq N_n} - \mathbb{E}[X_n|\mathcal{F}_{n-1}]}|\mathcal{F}_{n-1}]\right] \\
& \text{(Using asymptotic Hoeffding lemma)} \\
= & \lim_{\forall k\leq n, N_k\to\infty} e^{-\lambda\epsilon} e^{\frac{\lambda^2}{8}\sum_{k\leq n}(b_k-a_k)^2+\delta} (\delta \to 0)
\end{aligned}$$

Then we optimize the parameter and get the desired result. □

Proof. (**Proof 2 : Truncation by conditional probability**)

We denote $A_k = \{|G_k| \leq N_k\}$ and $B_n = \bigcap_{k\leq n} A_k$ and we define a change of probability

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{\mathbf{1}_{B_n}}{\mathbb{P}[B_n]}$$

. One may worry the probability of B_n. In fact,

$$\mathbb{P}[B_n] = 1 - \mathbb{P}[B_n^C], \mathbb{P}[B_n^C] = \mathbb{P}\left[\bigcup_{k\leq n} A_k^C\right] \leq \sum_{k\leq n} \mathbb{P}[|G_k| > N_k]$$

So we make N_k big enough, we have $\mathbb{P}[B_n] > 0$. Under the probability $\mathbb{Q}$, $(G_k)_{k\leq n}$ are bounded, so are $(X_k)_{k\leq n}$. So, we apply the Arzuma-Hoeffding inequality under the probability $\mathbb{Q}$ and obtain

$$\mathbb{Q}\left[\sum_{k=1}^n X_k - \sum_{k=1}^n \mathbb{E}_{\mathbb{Q}}[X_k|\mathcal{F}_{k-1}] > \epsilon\right] \leq \exp\left(-\frac{2\epsilon^2}{\sum_{k=1}^n (b_k-a_k)^2}\right)$$

This is

$$\begin{aligned}\mathbb{P}\left[\left\{\sum_{k=1}^n X_k - \sum_{k=1}^n \mathbb{E}_{\mathbb{Q}}[X_k|\mathcal{F}_{k-1}] > \epsilon\right\} \cap B_n\right] &\leq \mathbb{P}[B_n] \exp\left(-\frac{2\epsilon^2}{\sum_{k=1}^n (b_k - a_k)^2}\right) \\ &\leq \exp\left(-\frac{2\epsilon^2}{\sum_{k=1}^n (b_k - a_k)^2}\right)\end{aligned}$$

We would like pass to the limit. The dominated function is easy since function of indicator $\mathbf{1}_{\left\{\sum_{k=1}^n X_k - \sum_{k=1}^n \mathbb{E}_{\mathbb{Q}}[X_k|\mathcal{F}_{k-1}] > \epsilon\right\}} \mathbf{1}_{B_n} < 1$. Observing that when we pass $N_k \to \infty, \forall k \leq n$, we have $\mathbf{1}_{B_n} \to 1$ and

$$\mathbb{E}_{\mathbb{Q}}[X_k|\mathcal{F}_{k-1}] = \frac{\mathbb{E}[\mathbf{1}_{B_n} X_k|\mathcal{F}_{k-1}]}{\mathbb{E}[\mathbf{1}_{B_n}|\mathcal{F}_{k-1}]} \to \mathbb{E}[X_k|\mathcal{F}_{k-1}]$$

This passage to limit is established by the Bayes formula and the conditional dominated convergence.

In conclusion, when we pass $N_k \to \infty, \forall k \leq n$, we obtain the exact Arzuma-Hoeffding inequality. $\square$