A Note on concentration inequality

Xuanfei Ren*

January 25, 2023

Contents

1 Introduction

This note is a collation of relevant knowledge about measure concentration. Mainly used for our bandit and reinforcement learning study. The most basic part is based on *High-Dimensional Statistics A Non-Asymptotic Viewpoint (2019, Cambridge University Press),* and the rest of the inequalities come from the Internet.

2 Concentration inequality

I will directly list the common knowledge, easy to consult. The specific learning process and proof details are written in later sections.

Markov's inequality(X:non-negative and a finite mean):

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}, \forall t > 0 \tag{2.1}$$

Chebyshev's inequality($Y = (X - \mu)^2$):

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{var(X)}{t^2}, \forall t > 0 \tag{2.2}$$

Extensions of Markov's inequality (X has a central moment of order k):

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\mathbb{E}|X - \mu|^k}{t^k}, \forall t > 0 \tag{2.3}$$

(Chernoff's bounds)Suppose there is some constant $b > 0$ such that the moment generating function $\phi(\lambda) = \mathbb{E}[e^{\lambda(X-\mu)}]$ exist for all $\lambda \leq |b|$:

$$\log \mathbb{P}(X - \mu \geq t) \leq \inf_{\lambda \in [0,b]} (\log \mathbb{E}[e^{\lambda(X-\mu)}] - \lambda t) \tag{2.4}$$

Proposition 2.1. *Let $Z_1, \ldots, Z_n$ be independent Bernoulli variables where for every i,$\mathbb{P}[Z_i = 1] = p_i$, let $p = \Sigma_{k=1}^n p_i$, $Z = \Sigma_{k=1}^n Z_i$. Then, for any $\delta > 0$,*

$$\begin{aligned} \mathbb{P}[Z > (1+\delta)p] \leq e^{-h(\delta)p} &\leq e^{-p\frac{\delta^2}{2+2\delta/3}}, \\ \mathbb{P}[Z < (1-\delta)p] \leq e^{-h(-\delta)p} &\leq e^{-p\frac{\delta^2}{2+2\delta/3}} \end{aligned} \tag{2.5}$$

where $h(a) = (1+a)\log(1+a) - a$.

*University of Science and Technology of China; email: xuanfeiren@mail.ustc.edu.cn or xuanfeir@gmail.com

(Gaussian tail bounds) Suppose X is any $N(\mu, \sigma^2)$ random variable, then:

$$\mathbb{P}(X \geq \mu + t) \leq e^{-\frac{t^2}{2\sigma^2}}, \forall t > 0 \tag{2.6}$$

Definition 2.2 (Sub-Gaussian)**.** *A random variable X with mean μ is sub-Gaussian if there is a positive number σ such that*

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\frac{\sigma^2\lambda^2}{2}}, \forall \lambda \in \mathbb{R} \tag{2.7}$$

Remark 2.3. If $X \in [a, b]$, then it is sub-Gaussian with parameter $\sigma = \frac{b-a}{2}$.

The sub-Gaussian variable satisfies the concentration inequality (2.6) and

$$\mathbb{P}(|X - \mu| \geq t) \leq 2e^{-\frac{t^2}{2\sigma^2}}, \forall t \in \mathbb{R} \tag{2.8}$$

Theorem 2.4 (Hoeffding bound)**.** *Suppose that the variables $X_i, i = 1, \ldots, n$ are independent and X_i has mean μ_i and sub-Gaussian parameter σ_i, then for all $t > 0$, we have*

$$\mathbb{P}\left[\sum_{i=1}^{n}(X_i - \mu_i) \geq t\right] \leq \exp\{-\frac{t^2}{2\Sigma_{i=1}^n \sigma_i^2}\} \tag{2.9}$$

If $X_i \in [a, b]$, then we obtain the bound

$$\begin{aligned} \mathbb{P}\left[\sum_{i=1}^{n}(X_i - \mu_i) \geq t\right] &\leq \exp\{-\frac{2t^2}{n(b-a)^2}\}, \\ \mathbb{P}\left[\sum_{i=1}^{n}|X_i - \mu_i| \geq t\right] &\leq 2\exp\{-\frac{2t^2}{n(b-a)^2}\} \end{aligned} \tag{2.10}$$

Definition 2.5 (sub-exponential)**.** *A random variable X with mean $\mu = \mathbb{E}[X]$ is sub-exponential if there are non-negative parameters (v,α) such that*

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\frac{v^2\lambda^2}{2}} \textit{ for all } |\lambda| < \frac{1}{\alpha} \tag{2.11}$$

Note: Any sub-Gaussian variable is also sub-exponential. However, the converse statement is not true.

(Sub-exponential tail bound) Suppose that X is sub-exponential with parameters (v, α), then

$$\mathbb{P}[X - \mu \geq t] \leq \begin{cases} e^{-\frac{t^2}{2v^2}} & \text{if } 0 \leq t \leq \frac{v^2}{\alpha} \\ e^{-\frac{t}{2\alpha}} & \text{if } t > \frac{v^2}{\alpha}. \end{cases} \tag{2.12}$$

Definition 2.6 (Bernstein's condition)**.** *Given a random variable X with mean μ and variance σ^2, we say that Bernstein's condition with parameter b holds if*

$$|\mathbb{E}[(X - \mu)^k]| \leq \frac{1}{2}k!\sigma^2 b^{k-2} \textit{for } k = 2, 3, 4, \ldots. \tag{2.13}$$

Proposition 2.7. *When X satisfies the Bernstein condition, then it is sub-exponential with parameters $(\sqrt{2}\sigma, 2b)$.*

(Bernstein-type bound) For any random variable satisfying the Bernstein's condition, we have

$$\mathbb{E}[e^{\lambda(X-\mu)}] \leq e^{\frac{\lambda^2\sigma^2/2}{1-b|\lambda|}} \text{ for all } |\lambda| < \frac{1}{b},$$

and, moreover, the concentration inequality

$$\mathbb{P}[|X - \mu| \geq t] \leq 2e^{-\frac{t^2}{2(\sigma^2+bt)}} \text{ for all } t > 0. \tag{2.14}$$

There are another version of Bennet's and Bernstein's Inequalities.

Lemma 2.8 (Bennet's inequality). *Let $Z_1, \ldots, Z_m$ be independent random variables with zero mean, and assume that $Z_i 1$ with probability 1. Let*

$$\sigma^2 \geq \frac{1}{m} \sum_{i=1}^{m} \mathbb{E}[Z_i^2].$$

Then for all $\epsilon > 0$,

$$\mathbb{P}[\sum_{i=1}^{m} Z_i > \epsilon] \leq e^{-m\sigma^2 h(\frac{\epsilon}{m\sigma^2})}, \tag{2.15}$$

where h is the same definition as above.

Theorem 2.9 (Bernstein's inequality). *Same as above, assume $|Z_i| < M$ a.s., then*

$$\mathbb{P}[\sum_{i=1}^{m} Z_i > t] \leq \exp\{-\frac{t^2/2}{\Sigma Z_j^2 + Mt/3}\} \tag{2.16}$$

Corollary 2.10 (The sum of sub-exponential variables). *Consider an independent sequence $\{X_k\}_{k=1}^n$ of random variables, such that X_k has mean μ_k, and is sub-exponential with parameters (v_k, α_k), then*

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}(X_k - \mu_k) \geq t\right] \leq \begin{cases} e^{-\frac{nt^2}{2(v_*^2/n)}} & \text{for } 0 \leq t \leq \frac{v_*^2}{n\alpha_*}, \\ e^{-\frac{nt}{2\alpha_*}} & \text{for } t > \frac{v_*^2}{n\alpha_*}, \end{cases} \tag{2.17}$$

Where $\alpha_* := \max_{k=1,\ldots,n} \alpha_k$ and $v_* := \sqrt{\sum_{k=1}^{n} v_k^2}$.

Theorem 2.11 (One-sided Bernstein inequality). *If $X \leq b$ a.s., then*

$$\begin{aligned} &\mathbb{E}[e^{\lambda(X-\mathbb{E}X)}] \leq e^{\frac{\lambda^2 \mathbb{E}X^2}{1-\lambda b/3}} \forall \lambda \in [0, b/3) \\ &\mathbb{P}[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}X) \geq t] \leq \exp\{\frac{-nt^2}{2(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}X_i^2) + \frac{bt}{3}}\}. \end{aligned} \tag{2.18}$$

Theorem 2.12 (Slud's Inequality). *Let X be a (m, p) binomial variable and assume that $p = (1-\epsilon)/2$. Then,*

$$\mathbb{P}[X \geq m/2] \geq \frac{1}{2}(1 - \sqrt{1 - \exp\{-m\epsilon^2/(1-\epsilon^2)\}}). \tag{2.19}$$

Definition 2.13 (Martingale, Martingale difference sequence). *Given a sequence $\{Y_k\}_{k=1}^\infty$ of random variables adapted to a filtration $\{\mathcal{F}_k\}_{k=1}^\infty$, the pair $\{(Y_k, \mathcal{F}_k)\}_{k=1}^\infty$ is a martingale, if for all $k \geq 1$,*

$$\mathbb{E}[|Y_k|] < \infty \text{ and } \mathbb{E}[Y_{k+1}|\mathcal{F}_k] = Y_k.$$

A martingale difference sequence is an adapted sequence $\{(D_k, \mathcal{F}_k)\}_{k=1}^\infty$ such that for all $k \geq 1$,

$$\mathbb{E}[|D_k|] < \infty \text{ and } \mathbb{E}[D_{k+1}|\mathcal{F}_k] = 0.$$

Theorem 2.14 (A general Bernstein-type bound). *Let $\{(D_k, \mathcal{F}_k)\}_{k=1}^\infty$ be a martingale difference sequence, and suppose that $\mathbb{E}[e^{\lambda D_k}|\mathcal{F}_{k-1}] \leq e^{\lambda^2 v_k^2/2}$ almost surely for any $|\lambda| < 1/\alpha_k$. Then the following hold:*

1. *The sum $\Sigma_{k=1}^n D_k$ is sub-exponential with parameters $(\sqrt{\Sigma_{k=1}^n v_k^2}, \alpha_*)$ where $\alpha_* := \max_{k=1,\ldots,n} \alpha_k$.*

2. *The sum satisfies the concentration inequality:*

$$\mathbb{P}\left[|\sum_{k=1}^{n} D_k| \geq t\right] \leq \begin{cases} 2e^{-\frac{t^2}{2\Sigma_{k=1}^n v_k^2}} & \text{if } 0 \leq t \leq \frac{\Sigma_{k=1}^n v_k^2}{2} \\ 2e^{-\frac{t}{2\alpha_*}} & \text{if } t > \frac{\Sigma_{k=1}^n v_k^2}{\alpha_*}. \end{cases} \tag{2.20}$$

Corollary 2.15. *Let X_i be a sequence of i.i.d. random variables such that $|X_i - E[X_i]| \leq b$. Then, it holds that*

$$\mathbb{P}[\sum_{i=1}^{n}(X_i - \mathbb{E}X) \geq t] \leq \exp\{\frac{-t^2}{2n\sigma^2 + \frac{2}{3}bt}\}. \tag{2.21}$$

Corollary 2.16 (Azuma-Hoeffding)**.** *Let $\{(D_k, \mathcal{F}_k)\}_{k=1}^{\infty}$ be a martingale difference sequence for which there are constants $\{(a_k, b_k)\}_{k=1}^{n}$ such that $D_k \in [a_k, b_k]$ almost surely for all $k = 1, \ldots, n$. Then, for all $t \geq 0$,*

$$\mathbb{P}\left[|\sum_{k=1}^{n} D_k| \geq t\right] \leq 2e^{-\frac{2t^2}{\Sigma_{k=1}^{n}(b_k - a_k)^2}}. \tag{2.22}$$

Theorem 2.17 (One-side Azuma-Hoeffding)**.** *Let $X_i \in \mathcal{F}_i$ and $\mathcal{F}_{k-1} \subseteq \mathcal{F}_k$. If it holds that*

$$\mathbb{E}[X_i - \mathbb{E}X_i|\mathcal{F}_{i-1}] = 0, X_i \leq \mathbb{E}X_i + R_i,$$

then it holds that

$$\mathbb{P}[\sum_{i=1}^{n}(X_i - \mathbb{E}X) \geq t] \leq 2\exp\{-\frac{2t^2}{\Sigma_{i=1}^{n} R_i^2}\}.$$

Theorem 2.18 (One-side Azuma-Bernstein)**.** *Let $X_i \in \mathcal{F}_i$ and $\mathcal{F}_{k-1} \subseteq \mathcal{F}_k$. If it holds that*

$$\mathbb{E}[X_i - \mathbb{E}X_i|\mathcal{F}_{i-1}] = 0, X_i \leq \mathbb{E}X_i + R_i, \mathbb{V}[X_i|\mathcal{F}_{i-1}] \leq \sigma_i^2,$$

then it holds that

$$\mathbb{P}[\sum_{i=1}^{n}(X_i - \mathbb{E}X) \geq t] \leq 2\exp\{-\frac{t^2}{2\Sigma_{i=1}^{n}\sigma_i^2 + 2/3Rt}\}.$$

Corollary 2.19 (Bounded differences inequality)**.** *Suppose that f satisfies the bounded difference property with parameters $(L_1, \ldots, L_n)$ and that the random vector $X = (X_1, \ldots, X_n)$ has independent components. Then*

$$\mathbb{P}[|f(X) - \mathbb{E}[f(X)]| \geq t] \leq 2e^{-\frac{2t^2}{\Sigma_{k=1}^{n} L_k^2}} \textit{ for all } t \geq 0, \tag{2.23}$$

where the bounded difference property means if you change only the k th component, the value of the function changes at most L_k.

Theorem 2.20 (Lipchitz bound)**.** *Let $X = (X_1, \ldots, X_n)$ be a vector of i.i.d standard Gaussian variables, and let $f : \mathbb{R}^n \to \mathbb{R}$ be L-Lipchitz with respect to the Euclidean norm. Then the variale $f(X) - \mathbb{E}[f(X)]$ is sub-Gaussian with parameter at most L, and hence*

$$\mathbb{P}[|f(X) - \mathbb{E}[f(X)]| \geq t] \leq 2e^{-\frac{t^2}{2L^2}} \textit{ for all } t \geq 0 \tag{2.24}$$

Using the corollary above we can derive the χ^2-concentration:

$$\mathbb{P}[Y \geq n(1+t)] \leq e^{-\frac{nt^2}{18}} \text{ for all } t \in [0, 3], \tag{2.25}$$

where $Y := \Sigma_{k=1}^{n} Z_k^2$ follows a χ^2-distribution with n degrees of freedom.

Proposition 2.21. *Let $Z \sim \chi_k^2$, then for all $\epsilon > 0$ we have*

$$\mathbb{P}[Z \leq (1-\epsilon)k] \leq e^{-\frac{\epsilon k^2}{6}}$$

, and for all $\epsilon \in (0, 3)$ we have

$$\mathbb{P}[Z \geq (1+\epsilon)k] \leq e^{-\frac{\epsilon k^2}{6}}.$$

Finally, for all $\epsilon \in (0, 3)$

$$\mathbb{P}[(1-\epsilon)k \leq Z \leq (1+\epsilon)k] \geq 1 - 2e^{-\frac{\epsilon k^2}{6}}.$$

3 Useful details

Here are some details about the concentration.

Theorem 3.1 (Equivalent characterizations of sub-Gaussian variables)**.** *Given any zero-mean random variable X, the following properties are equivalent:*

1. *There is a constant $\sigma > 0$ such that*

$$\mathbb{E}[e^{\lambda X}] \leq e^{\frac{\lambda^2 \sigma^2}{2}} \text{ for all } \lambda \in \mathbb{R}.$$

2. *There is a constant $c > 0$ and Gaussian random variable $Z \sim N(0, \tau^2)$ such that*

$$\mathbb{P}[|X| > s] \leq c\mathbb{P}[|Z| > s] \text{ for all } s \geq 0.$$

3. *There is a constant $\theta \geq 0$ such that*

$$\mathbb{E}[X^{2k}] \leq \frac{(2k)!}{2^k k!}\theta^{2k} \text{ for all } k = 1, 2, \ldots$$

4. *There is a constant $\sigma > 0$ such that*

$$\mathbb{E}\left[e^{\frac{\lambda X^2}{2\sigma^2}} \leq \frac{1}{\sqrt{1-\lambda}}\right] \text{ for all } \lambda \in [0, 1).$$

Theorem 3.2 (Equivalent characterizations of sub-exponential variables)**.** *For a zeromean random variable X, the following statements are equivalent:*

1. *There are non-negative numbers (v, α) such that*

$$\mathbb{E}[e^{\lambda X}] \leq e^{\frac{v^2 \lambda^2}{2}} \text{for all } |\lambda| < \frac{1}{\alpha}$$

2. *There is a positive number $c_0 > 0$ such that $\mathbb{E}[e^{\lambda X}] < \infty$ for all $|\lambda| \leq c_0$.*
3. *There are constants $c_1, c_2 > 0$ such that*

$$\mathbb{P}[|X| \geq t] \leq c_1 e^{-c_2 t} \text{ for all } t > 0$$

4. *The quantity $\gamma := \sup_{k \geq 2} \left[\frac{\mathbb{E}[X^k]}{k!}\right]^{1/k}$ is finite.*

I plan to study uniform convergence about concentration next. In particular, the uniform convergence is required for the class of *UCB* algorithms.