L3 Moments, Deviation and Tail Inequalities

谢润烁
2023/11/17

Some contents of this handout are adapted from Prof. Yitong Yin's handout on Concentration of Measure.

1 Moments and Deviations

Definition (Moment). The kth moment of a random variable X is $\mathbb{E}[X^k]$.

Definition (Central Moment). The kth central moment of a random variable X is $\mathbb{E}[(X-\mathbb{E}X)^k]$.

> 矩反应了一个随机变量的特征。除了期望之外，还有一些矩有自己的名称：
>
> > From Wikipedia:
> >
> > If the function is a probability distribution, then the first moment is the **expected value**, the second *central* moment is the **variance**, the third *standardized* moment is the **skewness(偏度)**, and the fourth *standardized* moment is the **kurtosis(峰度)**.

2 Markov's Inequality

Theorem (Markov's Inequality). Let X be a random variable that assumes only nonnegative values. Then, for all $a>0$,

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

3 Chebyshev's Inequality

Theorem (Chebyshev's Inequality). For any $a>0$,

$$\Pr(|X-\mathbb{E}[X]| \geq a) \leq \frac{\mathrm{Var}[X]}{a^2}$$

> Chebyshev's Inequality可以用于证明弱大数定律（也称为辛钦大数定律）。至于为什么称为“*弱*大数定律”，这个涉及几种Convergence Mode，由于和我们的应用关系不大，因此我们就没有深入研究了。

4 The Chernoff Bound

4.1 Moment Generating Function

Definition (Moment Generating Function). The moment generating function of a random variable X is defined as $\mathbf{E}\left[\mathrm{e}^{\lambda X}\right]$ where λ is the parameter of the function.

> 我们使用泰勒展开可以得到：

> $$\begin{aligned}\mathbf{E}\left[\mathrm{e}^{\lambda X}\right] &= \mathbf{E}\left[\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}X^k\right] \\ &= \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\mathbf{E}\left[X^k\right]\end{aligned}$$
>
> 不难看出$\mathbf{E}\left[\mathrm{e}^{\lambda X}\right]$是$\lambda$的函数。

4.2 The Chernoff Bound

Definition (Poisson trials) Poisson trials are independent random variables $X_1, X_2, \ldots$ where each X_i only takes on value 0 or 1.

> Poisson trials只要求独立，不要求同分布。因此，多次独立的Bernoulli trail是Poisson trials的特例。

Theorem (Chernoff bound: the upper tail)

Let $X = \sum_{i=1}^{n} X_i$, where $X_1, X_2, \ldots, X_n$ are independent Poisson trials. Let $\mu = \mathbf{E}[X]$.

Then for any $\delta > 0$,

$$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$

> 证明思路:
>
> 1. $\Pr[X \geq (1+\delta)\mu] = \Pr[e^{\lambda X} \geq e^{\lambda(1+\delta)\mu}]$
>
>   因为Chernoff bound都是通过将Markov inequality应用到Moment generating function得到的，因此我们想要凑出一个$e^{\lambda X}$。
>
> 2. 应用Markov inequality
>
> $$\Pr\left[e^{\lambda X} \geq e^{\lambda(1+\delta)\mu}\right] \leq \frac{\mathbf{E}\left[e^{\lambda X}\right]}{e^{\lambda(1+\delta)\mu}},$$
>
> 3. 估计一下$\mathbf{E}\left[e^{\lambda X}\right]$的大小，我们希望能和已知的$\mu = \mathbf{E}[X]$联系起来
>
> $$\begin{aligned}\mathbf{E}\left[e^{\lambda X}\right] &= \mathbf{E}\left[e^{\lambda \sum_{i=1}^{n} X_i}\right] \\ &= \mathbf{E}\left[\prod_{i=1}^{n} e^{\lambda X_i}\right] \\ &= \prod_{i=1}^{n}\mathbf{E}\left[e^{\lambda X_i}\right]. \qquad \text{(for independent random variables)}\end{aligned}$$
>
>   将$\Pr(X_i = 1)$记为p_i，我们有:

> $$\mu = \mathbf{E}[X] = \mathbf{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} p_i$$
>
> 我们来估算单个随机变量对应的$\mathbf{E}\left[e^{\lambda X_i}\right]$:
>
> $$\begin{aligned}\mathbf{E}\left[e^{\lambda X_i}\right] &= p_i \cdot e^{\lambda \cdot 1} + (1-p_i) \cdot e^{\lambda \cdot 0} \\ &= 1 + p_i(e^{\lambda} - 1) \\ &\leq e^{p_i(e^{\lambda}-1)},\end{aligned}$$
>
> 综合以上的式子，我们可以得到:
>
> $$\begin{aligned}\prod_{i=1}^{n} \mathbf{E}\left[e^{\lambda X_i}\right] &\leq \prod_{i=1}^{n} e^{p_i(e^{\lambda}-1)} \\ &= \exp\left(\sum_{i=1}^{n} p_i(e^{\lambda}-1)\right) \\ &= e^{(e^{\lambda}-1)\mu}.\end{aligned}$$
>
> 4. 将第3步得到的应用到第2步的式子，我们得到了对于任意$\lambda > 0$，有以下的结果
>
> $$\begin{aligned}\Pr[X \geq (1+\delta)\mu] &\leq \frac{\mathbf{E}\left[e^{\lambda X}\right]}{e^{\lambda(1+\delta)\mu}} \\ &\leq \frac{e^{(e^{\lambda}-1)\mu}}{e^{\lambda(1+\delta)\mu}} \\ &= \left(\frac{e^{(e^{\lambda}-1)}}{e^{\lambda(1+\delta)}}\right)^{\mu}\end{aligned}$$
>
> 5. Chernoff bound需要选取一个合适λ来得到一个我们可以使用的式子。比对此处与要证明定理的形式，我们可以令$\lambda = \ln(1+\delta) > 0$，得到
>
> $$\Pr[X \geq (1+\delta)\mu] \leq \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$
>
> 为什么这个地方要这么取λ呢？因为这样取可以使得$\frac{e^{(e^{\lambda}-1)}}{e^{\lambda(1+\delta)}} = e^{e^{\lambda}-1-\lambda(1+\delta)}$最小，毕竟我们想得到一个尽可能紧的界。由于$e^{\lambda} - 1 - \lambda(1+\delta)$是一个关于$\lambda$的函数，我们只需要对$\lambda$求一下导即可验证。

Theorem (Chernoff bound: the lower tail)

Let $X = \sum_{i=1}^{n} X_i$, where $X_1, X_2, \ldots, X_n$ are independent Poisson trials. Let $\mu = \mathbf{E}[X]$.
Then for any $0 < \delta < 1$,

$$\Pr[X \leq (1-\delta)\mu] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}.$$

> lower tail的证明也类似，只要令$\lambda < 0$即可得到正确的不等号方向:

> $$\Pr[X \le (1-\delta)\mu] = \Pr\left[e^{\lambda X} \ge e^{\lambda(1-\delta)\mu}\right]$$

> 前面的Chernoff bound是最一般的形式，但是使用起来不是很方便，于是人们推导出一些没那么一般，但是更加简洁的形式用于分析。

Theorem (Chernoff bound: some useful forms)

Let $X = \sum_{i=1}^{n} X_i$, where $X_1, X_2, \ldots, X_n$ are independent Poisson trials. Let $\mu = \mathbf{E}[X]$. Then

1. for any $0 < \delta < 1$,

$$\begin{aligned}\Pr[X \ge (1+\delta)\mu] &< \exp\left(-\frac{\mu\delta^2}{3}\right);\\ \Pr[X \le (1-\delta)\mu] &< \exp\left(-\frac{\mu\delta^2}{2}\right);\end{aligned} \tag{1}$$

2. for $t \ge 2e\mu$,

$$\Pr[X \ge t] \le 2^{-t}. \tag{2}$$

> 要证明(1)中两条式子，只需要证明对于任意$0 < \delta < 1$：
>
> $$\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \le e^{-\delta^2/3}$$
> $$\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \le e^{-\delta^2/2}$$
>
> 要证明(2)，令$t = (1+\delta)\mu$，则$\delta = t/\mu - 1 \ge 2e - 1$，于是：
>
> $$\begin{aligned}\Pr[X \ge (1+\delta)\mu] &\le \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}\\ &\le \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu}\\ &\le \left(\frac{e}{2e}\right)^{t}\\ &\le 2^{-t}\end{aligned}$$

5 Further Reading

1. *Probability and Computing*

   (a) Ch3 Moments and Deviations
   See 3.1~3.3 for Markov's inequality and Chebyshev's inequality

   (b) Ch4 Chernoff and Hoeffding Bounds
   See 4.2 and 4.3 for detailed proofs of Chernoff bounds and its various forms;
   See 4.5 for the Hoeffding Bound
   See 4.4, 4.6 for some applications of Chernoff and Hoeffding Bounds
   See Ex 4.21 for the steps to prove the runtime of randomized quicksort is $O(n \log n)$ w.h.p

   (c) Ch5 Balls, Bins, and Random Graphs
   See 5.1, 5.2 for the balls into bins model

2. *Randomized Algorithms*

   (a) Ch3 Moments and Deviations

   (b) Ch4 Tail Inequalities

3. *Concentration of Measure for the Analysis of Randomized Algorithms*

   (a) Ch1 Chernoff–Hoeffding Bounds

   (b) Ch2 Applications of the Chernoff–Hoeffding Bounds
   See 2.3 for detailed analysis of the search cost of skip list
   See 2.4 for detailed analysis of the runtime of randomized quicksort is $O(n \log n)$ w.h.p

   > 这本书关于快排的证明可能有些小问题，大家需要谨慎地看

4. Slides and handouts from Prof. Yitong Yin

   (a) Moment and Deviation from *Probability and Mathematical Statistics*

   (b) Concentration of Measure from *Probability and Mathematical Statistics*

   (c) slides and handouts of *Concentration of Measure* from *Advanced Algorithms*

5. Lecture notes from other professors

   (a) Quick Sort with High Probability By *Sariel Har-Peled*

   (b) Lectures 18, 19, 20 & 21 (Randomized Algorithms & High Probability Bounds) from CSE 548: Analysis of Algorithms by Rezaul A. Chowdhury

   (c) Exercises of Chapter 4 by *Chuang-Chieh Lin*
   You can find the answer to the Ex 4.21 of *Probability and Computing*