Solution to "Chapter 2: Basic tail and concentration bounds"

Jiaxin Hu

July 9, 2020

1 Summary

Theorem 1.1 (Markov's inequality)**.** *Suppose $X \geq 0$ is a random variable with finite mean, we have*

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

Theorem 1.2 (Chebyshev's inequality)**.** *Suppose $X \geq 0$ is a random variable with finite mean μ and finite variance, we have*

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

Theorem 1.3 (Markov's inequality for polynomial moments)**.** *Suppose the random variable X has a central moment of order k. Applying Markov's inequality to the random variable $|X - \mu|^k$ yields*

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\mathbb{E}[|X - \mu|^k]}{t^k}, \quad \forall\, t > 0.$$

; i.e., … (note the semcolon and period after i.e.)

Theorem 1.4 (Chernoff bound)**.** *Suppose the random variable X has a moment generating function in the neighborhood of 0, i.e. $\varphi_X(\lambda) = \mathbb{E}[e^{\lambda X}] < +\infty, \forall \lambda \in (-b, b), b > 0$. Applying Markov's inequality to the random variable $Y = e^{\lambda(X-\mu)}$ yields* with some b>0

change to "for all"

$$\mathbb{P}((X - \mu) \geq t) \leq \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda t}}.$$

Optimizing the choice of λ for the tightest bound yields the Chernoff bound

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

Theorem 1.5 (Hoeffding bound for bounded variable)**.** *Consider a random variable X with mean $\mu = \mathbb{E}(X)$~~, and~~ such that $X \in [a, b]$ almost surely, where a, b are two constants. Then, for any $\lambda \in \mathbb{R}$, it holds* Assume X is bounded such that

do not use "it", "those",…

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

Particularly, the variable $X \sim subG(\frac{(b-a)^2}{4})$.

Proof. See Exercise 2.4. □

2 Exercises

2.1 Exercise 2.1

where is the equation numbered by (1.1)?

(Tightness of inequalities.) The Markov and Chebyshev's inequalities can not be improved in general.

that attains the equality in (1.1)

(a) Provide a random variable $X \geq 0$ ~~for which Markov's inequality (1.1) is met with equality.~~

(b) Provide a random variable Y for which Chebyshev's inequality (1.2) is met with equality.

... inequality for X is met with equality —> X attains the equality in the ...inequality.

Solution:

(a) Recall the proof of Markov's inequality. For any $t > 0$,

$$\mathbb{E}[X] = \int_0^t x f_X(x)dx + \int_t^{+\infty} x f_X(x)dx \geq \int_t^{+\infty} x f_X(x) \geq t \int_t^{+\infty} f_X(x) = t\mathbb{P}(X \geq t). \quad (1)$$

We prove a random variable that attains the equality in each line of (1)

~~If Markov's inequality meets the equality, the inequality (1) meets the equality.~~

Consider a variable X with distribution $P(X = 0) = 1$. For any $t > 0$, the variable X satisfies

$$\int_0^t x f_X(x)dx = 0 \text{ and } \int_t^{+\infty} x f_X(x)dx = \int_t^{+\infty} t f_X(x)dx.$$

Therefore, for variable X, the Markov's inequality is met with equality.

(b) Chebyshev's inequality follows by applying Markov's inequality to the non-negative random variable $Y = \mathbb{E}(X - \mathbb{E}[X])^2$. Let the distribution of Y be $\mathbb{P}(Y = 0) = 1$. Then the Markov's inequality for Y and the Chebyshev's inequality for X meet the equalities. By transformation, the distribution of random variable X is $\mathbb{P}(X = \mathbb{E}[X]) = 1$. Therefore, for any random variable X with distribution $\mathbb{P}(X = c) = 1, c \in \mathbb{R}$, the Chebyshev's inequality is met with equality.

2.2 Exercise 2.2

Lemma 1 (Standard normal distribution)**.** *Let $\phi(z)$ be the density function of a standard normal $Z \sim N(0, 1)$ variable. Then,* varaible

$$\phi'(z) + z\phi(z) = 0, \quad (2)$$

and

$$\phi(z)(\frac{1}{z} - \frac{1}{z^3}) \leq \mathbb{P}(Z \geq z) \leq \phi(z)(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5}), \quad \textit{for all } z > 0. \quad (3)$$

Proof. First, we prove the equation (2). The pdf of standard normal distribution $\phi(z)$ satisfies

. (break line). The equation (2) follows from the following calculation

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}); \quad \phi'(z) = -z\frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2}) = -z\phi(z).$$

(principle: proof should go through even if all maths are deleted. Ask yourself whether should put these two equations in a single line?)

Next, we prove the equation (3). Using equation (2), we have

$$\begin{aligned}\mathbb{P}(Z \geq z) &= \int_z^{+\infty} \phi(t)dt = \int_z^{+\infty} -\frac{1}{t}\phi'(t)dt = \frac{1}{z}\phi(z) - \int_z^{+\infty} \frac{1}{t^2}\phi(t)dt \\ &= \frac{1}{z}\phi(z) + \int_z^{+\infty} \frac{1}{t^3}\phi'(t)dt = \frac{1}{z}\phi(z) - \frac{1}{z^3}\phi(z) + \int_z^{+\infty} \frac{3}{t^4}\phi(t)dt\end{aligned}$$

5 steps in total. Principle:
single-line expression: # steps <= 3
multiple-line expressions: # steps <= 3.

If # steps > 4, break into chunks, and add explanations in words between chunks.

should not use both

Since $\frac{3}{t^4}\phi(t) \geq 0$, therefore $\mathbb{P}(Z \geq z) \geq \phi(z)(\frac{1}{z} - \frac{1}{z^3})$. On the other hand,

$$\int_z^{+\infty} \frac{3}{t^4}\phi(t)dt = \int_z^{+\infty} -\frac{3}{t^5}\phi'(t)dt = \frac{3}{z^5}\phi(z) - \int_z^{+\infty} \frac{15}{t^6}\phi(t)dt \leq \frac{3}{z^5}\phi(z).$$

Therefore, $\mathbb{P}(Z \geq z) \leq \phi(z)(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5})$. □

why? Say combining ** with **, or plugging ** into **

2.3 Exercise 2.3

Consider a nonnegative random variable X>0. Suppose that

Lemma 2 (Polynomial bound and Chernoff bound)**.** *~~Suppose $X \geq 0$, and~~ that the moment generating function of X exists in the neighborhood of 0. Given some $\delta > 0$ ~~and integer $k \in \mathbb{Z}_+$,~~ we have*

phi_X(t) in the neighborhood of t=0 Why is k given?

$$\inf_{k \in \mathbb{Z}_+} \frac{\mathbb{E}[|X|^k]}{\delta^k} \leq \inf_{\lambda > 0} \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda\delta}}.$$

neither sides of the equality depends on k.

Consequently, an optimized bound based on polynomial moments is always at least as good as the Chernoff upper bound.

Proof. By power series, we have

$$e^{\lambda X} = \sum_{k=0}^{+\infty} \frac{X^k \lambda^k}{k!}, \quad \forall \lambda \in \mathbb{R} \tag{4}$$

period.

Since the moment generating function $\varphi_X(\lambda)$ exists in the neighbor hood of 0, there exists a constant $b > 0$ such that

$$\inf_{\lambda > 0} \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda\delta}} = \inf_{\lambda \in (0,b)} \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda\delta}} < +\infty.$$

where is this line used in your proof? Say it explicitly.

Taking the expectation on both sides of the power series (4) yields

this argument alone cannot lead to your conclusion in blue

$$\mathbb{E}[e^{\lambda X}] = \sum_{k=0}^{+\infty} \frac{\mathbb{E}[|X|^k]\lambda^k}{k!} < +\infty, \quad \forall \lambda \in (0, b).$$

(the moment … exists for any k)

Therefore, the moment $\mathbb{E}[|X|^k] < +\infty$, $\forall k \in \mathbb{Z}_+$ exists. Applying the power serious to $e^{\lambda\delta}$, we obtain the result

$$\inf_{k \in \mathbb{Z}_+} \frac{\mathbb{E}[|X|^k]}{\delta^k} \leq \sum_{k=0}^{+\infty} \frac{\mathbb{E}[|X|^k]}{\delta^k} = \sum_{k=0}^{+\infty} \frac{\frac{\mathbb{E}[|X|^k]\lambda^k}{k!}}{\frac{\lambda^k\delta^k}{k!}} = \inf_{\lambda > 0} \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda\delta}}.$$

There are three steps in this line. □

Only the second step in red relies on your argument “applying the power series to …”.

You should make your argument and conclusion consistent within each step.

2.4 Exercise 2.4

Rewrite. For example, you could say ``the conclusion follows by …., where the second equality uses the power series to … “

In Exercise 2.4, we prove theorem 1.5, Hoeffding bound for bounded variable. Consider a random variable X with mean $\mu = \mathbb{R}[X]$ and such that $a \leq X \leq b$ almost surely. Define the function

capital letter

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

for all lambda in R

We apply Taylor Expansion of $\varphi(\lambda)$ at 0.

“such that” means “in order to”. Does the sentence make sense with “in order to” inserted?

$$\varphi(\lambda) = \varphi(0) + \varphi'(0)\lambda + \frac{\varphi''(\lambda_0)}{2}\lambda^2, \quad \lambda_0 = t\lambda, \text{ for some } t \in [0, 1]. \tag{5}$$

In equation (5), the term $\varphi(0) = \log \mathbb{E}[e^0] = 0$. For the first-order derivative $\varphi'(\lambda)$, we apply the power series and have

$$
\begin{aligned}
\varphi'(\lambda) &= \left(\log \mathbb{E}\left[\sum_{k=0}^{n} \frac{\lambda^k X^k}{k!}\right]\right)' = \left(\log \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^k]\right)' \\
&= \sum_{k=1}^{n} \frac{k\lambda^k}{k!}\mathbb{E}[X^k] \Bigg/ \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^k] = \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^{(k+1)}] \Bigg/ \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^k] \\
&= \frac{\mathbb{E}[Xe^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]}
\end{aligned}
$$

Therefore, $\varphi'(0) = \mathbb{E}[X] = \mu$. For the second-order derivative $\varphi^{''}(\lambda)$,

$$
\begin{aligned}
\varphi^{''}(\lambda) &= \left(\sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^{(k+1)}] \Bigg/ \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^k]\right)' \\
&= \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^{(k+2)}] \Bigg/ \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^k] - \left(\sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^{(k+1)}] \Bigg/ \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^k]\right)^2 \\
&= \frac{\mathbb{E}[X^2 e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} - \left(\frac{\mathbb{E}[Xe^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]}\right)^2 \\
&\triangleq \mathbb{E}_\lambda[X^2] - (\mathbb{E}_\lambda[X])^2
\end{aligned}
$$

Therefore, the second-order derivative $\varphi''(\lambda)$ can be interpreted as the variance of X with a re-weighted distribution $dP' = \frac{e^{\lambda X}}{\mathbb{E}[e^{\lambda X}]}dP_X$, where P_X is the distribution of X. Since

$$
\int_{-\infty}^{+\infty} dP' = \int_{-\infty}^{+\infty} \frac{e^{\lambda X}}{\mathbb{E}[e^{\lambda X}]} dP_X = 1,
$$

the function P' is indeed a distribution. Under any distribution, we always have

$$
var(X) = var(X - \frac{a+b}{2}) \leq \mathbb{E}[(X - \frac{a+b}{2})^2] = \frac{(b-a)^2}{4}.
$$

Back to the equation (5),

$$
\varphi(\lambda) = \varphi(0) + \varphi'(0)\lambda + \frac{\varphi^{''}(\lambda_0)}{2}\lambda^2 \leq 0 + \lambda\mu + \frac{(b-a)^2}{8}\lambda^2
$$

Taking exponential on both sides of the inequality, we have

$$
\mathbb{E}[e^{\lambda X}] = \exp(\varphi(\lambda)) \leq e^{\mu\lambda + \frac{(b-a)^2}{8}\lambda^2}. \tag{6}
$$

The equation (6) implies that X is a sub-Gaussian variable with at most $\sigma = \frac{(b-a)}{2}$.

2.5 Exercise 2.5

Lemma 3 (Sub-Gaussian bounds and means/variance)**.** *Consider a random variable X such that*

$$
\mathbb{E}[e^{\lambda X}] \leq e^{\frac{\lambda^2\sigma^2}{2} + \mu\lambda}, \quad \forall \lambda \in \mathbb{R}. \tag{7}
$$

Then, $\mathbb{E}[X] = \mu$ *and* $var(X) \leq \sigma^2$.

Proof. First, by equation (7), the moment generating function of X exists, and thus the mean and variance of X exist. Applying power series on both sides of equation (7),

$$\lambda\mathbb{E}[X] + \frac{\lambda^2}{2}\mathbb{E}[X^2] + o(\lambda^2) \leq \mu\lambda + \frac{\lambda^2\sigma^2 + \lambda^2\mu^2}{2} + o(\lambda^2). \tag{8}$$

Dividing by $\lambda > 0$ on both sides of equation (8) and letting $\lambda \to 0^+$, we have $\mathbb{E}(X) \leq \mu$; dividing dividing by $\lambda < 0$ on both sides of equation (8) and letting $\lambda \to 0^-$, we have $\mathbb{E}(X) \geq \mu$. Therefore, the mean $\mathbb{E}[X] = \mu$. Similarly, we subtract $\mathbb{E}[X]\lambda$ and $\mu\lambda$ and divide $\frac{2}{\lambda^2}$ on both sides of equation (8). We have $\mathbb{E}[X^2] \leq \sigma^2 + \mu^2$, and thus $var(X) \leq \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \sigma^2$. $\square$

Question: Suppose the smallest possible σ satisfying the inequality (7) is chosen. Is it true that $var(X) = \sigma^2$?

Solution: The statement that $var(X) = \sigma^2$ is not always true. Recall the function $\varphi(\lambda)$ in exercise 2.4. By the results in exercise 2.4, the equation (7) is equal to

$$\varphi''(\lambda) \leq \sigma^2, \quad \forall\lambda \in \mathbb{R},$$

where $\varphi''(\lambda)$ is the variance of X with with a re-weighted distribution $dP' = \frac{e^{\lambda X}}{\mathbb{E}[e^{\lambda X}]}dP_X$, where P_X is the distribution of X. If the statement that $var(X) = \sigma^2$ is true, then $\max_\lambda \varphi''(\lambda) = \varphi''(0)$, which is not always true. A counter example is below.

Consider a random variable $X \sim Ber(1/3)$ with $var(X) = 2/9$. Let $\lambda = 1$. The re-weighted distribution dP' is

$$P'(X=0) = \frac{2}{3\mathbb{E}[e^X]}; \quad P'(X=1) = \frac{e}{3\mathbb{E}[e^X]}, \quad \text{where } \mathbb{E}[e^X] = \frac{2}{3} + \frac{e}{3}.$$

Therefore, the variance of X with dP' is $\frac{2}{3\mathbb{E}[e^X]} \times \frac{e}{3\mathbb{E}[e^X]} = 0.2442 > 2/9$. Therefore, the smallest possible σ^2 is strictly larger than $var(X)$ in this case.