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

Jiaxin Hu

July 12, 2020

1 Summary

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

$$\mathbb{P}(X \geq t) \leq \frac{E[X]}{t}, \quad \forall\ t > 0. \tag{1}$$

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

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

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*

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

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

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

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

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

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

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

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

Proof. See Exercise 2.4. $\square$

2 Exercises

2.1 Exercise 2.1

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

(a) Provide a random variable $X \geq 0$ that attains the equality in Markov's inequality (1).

(b) Provide a random variable Y that attains the equality in Chebyshev's inequality (2).

Solution:

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

$$\mathbb{E}[X] = \int_0^t x f_X(x)dx + \int_t^{+\infty} x f_X(x)dx \geq \int_t^{+\infty} x f_X(x) \geq t \int_t^{+\infty} f_X(x) = t\mathbb{P}(X \geq t). \quad (3)$$

Below, given a constant $t > 0$, we construct a random variable that attains the equality in each line of equation (3).

Consider a variable X with distribution $P(X = t) = 1$. 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 given t, the variable X attains the equality of Markov's inequality.

(b) Chebyshev's inequality follows by applying Markov's inequality to the non-negative random variable $Y = (X - \mathbb{E}[X])^2$. Let the distribution of Y be $\mathbb{P}(Y = t^2) = 1$. Then the Markov's inequality for Y and the Chebyshev's inequality for X meet the equalities. Consider the variable Z with distribution $\mathbb{P}(Z = c + t) = \mathbb{P}(Z = c - t) = 1/2$ for any $c \in \mathbb{R}$. For Z, the transformation probability satisfies $\mathbb{P}((Z - \mathbb{E}[Z])^2 = t^2) = 1$. Therefore, the variable Z attains the equality of Chebyshev's inequality.

a few

Remark 1 (Tightness of Markov's inequality)**.** Only few variables ~~can~~ attain the equalities in Markov's and Chebyshev's inequalities. In research, we should ~~be careful if there is a~~ concentration
pay attention to
bound ~~that is~~ tighter than Markov's inequality.
bounds

(you should apply the writing principle whenever you write, in email/paper/report…)

2.2 Exercise 2.2

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

$$\phi'(z) + z\phi(z) = 0, \quad (4)$$

and

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

Proof. First, we prove the equation (4).
The pdf of standard normal distribution is

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

The equation (4) follows by taking the derivative of $\phi(z)$.

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

Next, we prove the equation (5).
We write the upper tail probability of the standard normal variable as~~:~~ no colon

$$\mathbb{P}(Z \geq z) = \int_z^{+\infty} \phi(t)dt = \int_z^{+\infty} -\frac{1}{t}\phi'(t)dt = \frac{1}{z}\phi(z) - \int_z^{+\infty} \frac{1}{t^2}\phi(t)dt, \tag{6}$$

where the second equality follows by the equation (5). Applying the equation (5) to the last term in equation (6) yields

$$\int_z^{+\infty} \frac{1}{t^2}\phi(t)dt = \int_z^{+\infty} \frac{1}{t^3}\phi'(t)dt = -\frac{1}{z^3}\phi(z) + \int_z^{+\infty} \frac{3}{t^4}\phi(t)dt \geq -\frac{1}{z^3}\phi(z) \tag{7}$$

Plugging the equation (7) into the equation (6) , we have $\mathbb{P}(Z \geq z) \geq \phi(z)(\frac{1}{z} - \frac{1}{z^3})$. On the other hand, applying the equation (5) again to the equation (7) yields

$$\int_z^{+\infty} \frac{3}{t^4}\phi(t)dt = \int_z^{+\infty} -\frac{3}{t^5}\phi'(t)dt = \frac{3}{z^5}\phi(z) - \int_z^{+\infty} \frac{15}{t^6}\phi(t)dt \leq \frac{3}{z^5}\phi(z). \tag{8}$$

Combing ~~with the~~ equations (6), (7), and (8), we have $\mathbb{P}(Z \geq z) \leq \phi(z)(\frac{1}{z} - \frac{1}{z^3} + \frac{3}{z^5})$. $\square$

Remark 2. Direct calculation of tail probability for a univariate normal variable is hard. Equation (5) provides a numerical approximation to the tail probability. Particularly, the tail probability decays at the rate of $z^{-1}e^{-z^2/2}$ as $z \to +\infty$. The decay rate is faster than polynomial rate $\mathcal{O}(z^{-\alpha})$, for any $\alpha \geq 1$.

2.3 Exercise 2.3

Lemma 2 (Polynomial bound and Chernoff bound)**.** *Consider a non-negative variable $X \geq 0$. Suppose that the moment generating function of X, $\varphi_X(\lambda)$, exists in the neighborhood of $\lambda = 0$. Given some $\delta > 0$, for any integer $k \in \mathbb{Z}_+$, we have*

$$\inf_{k\in\mathbb{Z}_+} \frac{\mathbb{E}[|X|^k]}{\delta^k} \leq \inf_{\lambda>0} \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda\delta}}. \tag{9}$$

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

$$e^{\lambda X} = \sum_{k=0}^{+\infty} \frac{X^k\lambda^k}{k!}, \quad \forall\lambda \in \mathbb{R}. \tag{10}$$

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

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

for all, not for any. (do not make same mistakes twice)

Therefore, the moment $\mathbb{E}[|X|^k]$ exists for ~~any~~ $k \in \mathbb{Z}_+$. Noted that $\mathbb{E}[|X|^k] \geq 0$, we have

$$\mathbb{E}[|X|^k] \leq \sum_{k=0}^{+\infty} \mathbb{E}[|X|^k], \quad \forall k \in \mathbb{Z}_+. \tag{11}$$

Applying power series (10) to the right hand side of equation (9) yields

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

Combine A with B… “combine” is vt.

Combing with equation (11), we have

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

□

make them consistent

Remark 3. Applying different functions $g(X)$ to the Markov's inequality leads to distinct bounds for the tail probability of variable X. Equation (9) implies that optimized polynomial bound is tighter than the Chernoff bound, provided that the moment generating function of X exsits. Actually, the optimization on polynomial bound can be released. Then, any polynomial bound is better than Chernoff bound.

why do you mean by “release the optimization?”

2.4 Exercise 2.4

Capital

“a” bounded variable

In Exercise 2.4, we prove the theorem 1.5, the Hoeffding bound for bounded variable.

Proof. Suppose X is a bounded random variable ~~that~~ where $X \in [a, b]$ almost surely. Let $\mu = \mathbb{E}[X]$. Define the function

for all

$$g(\lambda) = \log \mathbb{E}[e^{\lambda X}], \quad \cancel{\forall}\lambda \in \mathbb{R}.$$

Applying Taylor Expansion to $g(\lambda)$ at 0, we have

$$g(\lambda) = g(0) + g'(0)\lambda + \frac{g''(\lambda_0)}{2}\lambda^2, \quad \text{where } \lambda_0 = t\lambda, \text{ for some } t \in [0, 1]. \tag{12}$$

In equation (12), the term $g(0) = \log \mathbb{E}[e^0] = 0$. By power series (10), we calculate the first derivative $g'(\lambda)$ as following.

$$\begin{aligned} g'(\lambda) &= \left(\log \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^k]\right)' \\ &= \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^{(k+1)}] \Bigg/ \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^k] \\ &= \frac{\mathbb{E}[Xe^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} \end{aligned} \tag{13}$$

period

Therefore, $g'(0) = \mathbb{E}[X] = \mu$. Taking the derivative to equation (13), we calculate the second-order derivative $g^{''}(\lambda)$ as following.

$$g^{''}(\lambda) = \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^{(k+2)}] \Big/ \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^k] - \left(\sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^{(k+1)}] \Big/ \sum_{k=0}^{n} \frac{\lambda^k}{k!}\mathbb{E}[X^k] \right)^2$$
$$= \frac{\mathbb{E}[X^2 e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} - \left(\frac{\mathbb{E}[X e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} \right)^2$$

period

omit “can”

The second-order derivative $g''(\lambda)$ can be interpreted as the variance of X with a re-weighted distribution $dP' = e^{\lambda X}/\mathbb{E}[e^{\lambda X}]dP_X$, where P_X is the distribution of X. Taking integral ~~to~~ dP', we have

take integral of 1 with respect to dP’

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

which implies that the function P' is indeed a distribution. Under any distribution, there is always an upper bound for the variance of random variable X:

this statement is wrong. X has to be bounded.

omit “there is“

You should make sure every single sentence, if stated by itself, is mathematically correct.

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

what is a, b?

Hence, the second-order derivative $g''(\lambda) \leq \frac{(b-a)^2}{4}$. We plug the results of g' and $g^{''}$ into the equation (12). Then,

$$g(\lambda) = g(0) + g'(0)\lambda + \frac{g^{''}(\lambda_0)}{2}\lambda^2 \leq 0 + \lambda\mu + \frac{(b-a)^2}{8}\lambda^2. \tag{14}$$

Taking the exponential on both sides of the inequality (14), we have

$$\mathbb{E}[e^{\lambda X}] = \exp(g(\lambda)) \leq e^{\mu\lambda + \frac{(b-a)^2}{8}\lambda^2}. \tag{15}$$

The equation (15) implies that X is a sub-Gaussian variable with at most $\sigma = \frac{(b-a)}{2}$. $\square$

Remark 4. For any bounded random variable X ~~with domain~~ supported on $[a, b]$, X is a sub-gaussian variable with parameter at most $\sigma^2 = (b-a)^2/4$. Any ~~propositions~~ properties for sub-Gaussian variables can be applied to bounded variables.

2.5 Exercise 2.5

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

$$\mathbb{E}[e^{\lambda X}] \leq e^{\frac{\lambda^2\sigma^2}{2} + \mu\lambda}, \quad \cancel{\forall} \text{ for all } \lambda \in \mathbb{R}. \tag{16}$$

(I have made same correction in your earlier notes)

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

denoted …

Proof. By equation (16), the moment generating function of X, $\varphi_X(\lambda)$, exists in the neighborhood of $\lambda = 0$. Hence, the mean and variance of X exist. Applying power series on both sides of equation (16) yields

(check the rule 6 in BJORN POONEN’s note that you sent me)

$$\lambda\mathbb{E}[X] + \frac{\lambda^2}{2}\mathbb{E}[X^2] + o(\lambda^2) \leq \mu\lambda + \frac{\lambda^2\sigma^2 + \lambda^2\mu^2}{2} + o(\lambda^2). \tag{17}$$

for all \lambda in neighborhood of 0

; —> .

Dividing by $\lambda > 0$ on both sides of equation (17) and letting $\lambda \to 0^+$, we have $\mathbb{E}(X) \leq \mu$; dividing by $\lambda < 0$ on both sides of equation (17) and letting $\lambda \to 0^-$, we have $\mathbb{E}(X) \geq \mu$. Therefore, the mean $\mathbb{E}[X] = \mu$. Then, we divide $2/\lambda^2$ on both sides of equation (17). The term $\mathbb{E}[X]\lambda$ and $\mu\lambda$ are cancelled. 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: Let σ^2_{min} denote the smallest possible σ satisfying the inequality (16). Is it true that $var(X) = \sigma^2_{min}$?

necessarily

Solution: The statement that $var(X) = \sigma^2_{min}$ is not ~~always~~ true. Recall the function $g(\lambda)$ in exercise 2.4. By the results in exercise 2.4, the equation (16) is equal to

for all

(check the latex expression)

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

the subject of the sub-clause should be the same as the main clause

Does it make sense to say "we satisfy the ...."

the

("we" is the subject of the main clause)

where $g^{''}(\lambda)$ is the variance of X with with ~~a~~ re-weighted distribution $dP' = e^{\lambda X}/\mathbb{E}[e^{\lambda X}]dP_X$, where P_X is the distribution of X and $g^{''}(0) = var(X)$. Therefore, $\max_\lambda g^{''}(\lambda) = \sigma^2_{min}$. To satisfy the equality $var(X) = \sigma^2_{min}$, we need to show that $\max_\lambda g^{''}(\lambda) = g^{''}(0)$ holds for X.

However, the statement $\max_\lambda g^{''}(\lambda) = g^{''}(0)$ is not always true. A counter example is below. Consider a random variable $X \sim Ber(1/3)$. The variance of X is $var(X) = 2/9$. Let $\lambda = 1$. The re-weighted distribution dP' is

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

The variance of X with dP' is $2/3\mathbb{E}[e^X] \times e/3\mathbb{E}[e^X] = 0.2442 > 2/9$. Therefore, for this variable X, $g^{''}(0) < g^{''}(1) \leq \max_\lambda g^{''}(\lambda)$.

Remark 5. Parameters of a sub-Gaussian distribution provide the exact value of the mean, $\mathbb{E}[X] = \mu$, and an upper bound of the variance, $var(X) \leq \sigma^2$. For any variable X whose moment generating function exists, the tail distribution of X can be bounded by a sub-Gaussian distribution with a proper choice of σ^2.