Chapter 12

Concentration of Random Variables – Chernoff's Inequality

By Sariel Har-Peled, April 26, 2022[1]

12.1. Concentration of mass and Chernoff's inequality

12.1.1. Example: Binomial distribution

Consider the binomial distribution $\mathrm{Bin}(n, 1/2)$ for various values of n as depicted in Figure 12.1 – here we think about the value of the variable as the number of heads in flipping a fair coin n times. Clearly, as the value of n increases the probability of getting a number of heads that is significantly smaller or larger than $n/2$ is tiny. Here we are interested in quantifying exactly how far can we divert from this expected value. Specifically, if $X \sim \mathrm{Bin}(n, 1/2)$, then we would be interested in bounding the probability $\mathbb{P}[X > n/2 + \Delta]$, where $\Delta = t\sigma_X = t\sqrt{n}/2$ (i.e., we are t standard deviations away from the expectation). For $t > 2$, this probability is roughly 2^{-t}, which is what we prove here.

More surprisingly, if you look only on the middle of the distribution, it looks the same after clipping away the uninteresting tails, see Figure 12.2; that is, it looks more and more like the normal distribution. This is a universal phenomena known the ***central limit theorem*** – every sum of nicely behaved random variables behaves like the normal distribution. We unfortunately need a more precise quantification of this behavior, thus the following.

12.1.2. A restricted case of Chernoff inequality via games

12.1.2.1. Chernoff games

The game. Consider the game where a player starts with $Y_0 = 1$ dollars. At every round, the player can bet a certain amount x (fractions are fine). With probability half she loses her bet, and with probability half she gains an amount equal to her bet. The player is not allowed to go all in – because if she looses then the game is over. So it is natural to ask what her optimal betting strategy is, such that in the end of the game she has as much money as possible.

Is the game pointless? So, let Y_{i-1} be the money the player has in the end of the $(i-1)$th round, and she bets an amount $\psi_i \le Y_{i-1}$ in the ith round. As such, in the end of the ith round, she has

$$Y_i = \begin{cases} Y_{i-1} - \psi_i & \text{LOSE: probability half} \\ Y_{i-1} + \psi_i & \text{WIN: probability half} \end{cases}$$

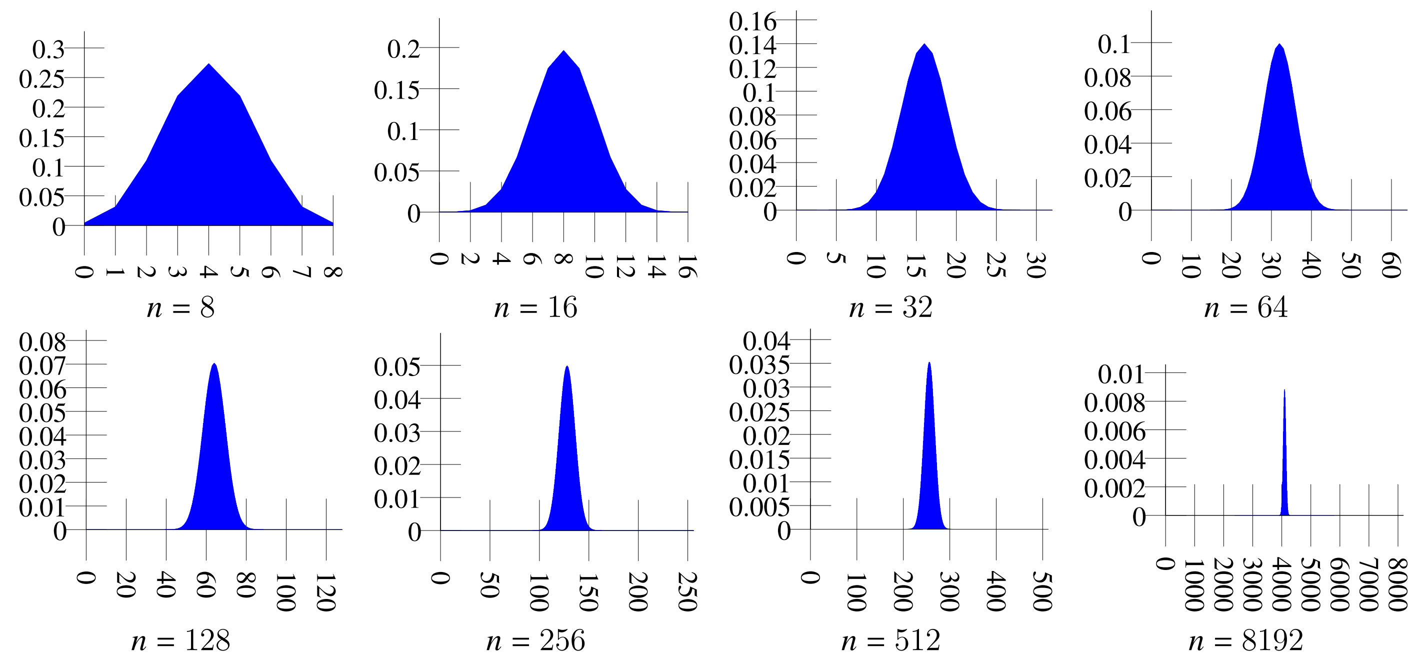

Figure 12.1: The binomial distribution for different values of n. It pretty quickly concentrates around its expectation.

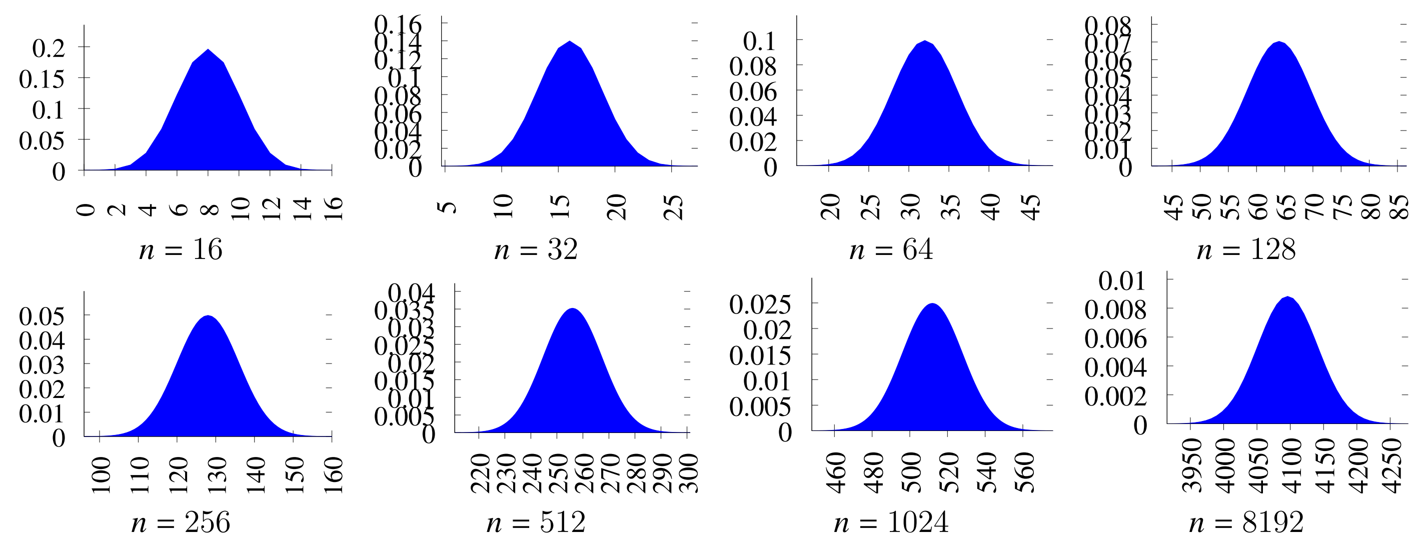

Figure 12.2: The “middle” of the binomial distribution for different values of n. It very quickly converges to the normal distribution (under appropriate rescaling and translation.

$X_i \in \{-1,+1\}$ $\mathbb{P}[X_i = -1] = \mathbb{P}[X_i = 1] = 1/2$	
$\mathbb{P}\big[Y \geq \Delta\big] \leq \exp\big(-\Delta^2/2n\big)$	Theorem 12.1.7
$\mathbb{P}\big[Y \leq -\Delta\big] \leq \exp\big(-\Delta^2/2n\big)$	Theorem 12.1.7

$X_i \in \{0,1\}$ $\mathbb{P}[X_i = 0] = \mathbb{P}[X_i = 1] = 1/2$
$\mathbb{P}\big[\lvert Y - n/2\rvert \geq \Delta\big] \leq 2\exp\big(-2\Delta^2/n\big)$ Corollary 12.1.9

$X_i \in \{0,1\}$	$\mathbb{P}[X_i = 1] = p_i \qquad \mathbb{P}[X_i = 0] = 1 - p_i$	
$\delta \geq 0$	$P = \mathbb{P}\big[Y > (1+\delta)\mu\big] < \left(e^{\delta}\big/(1+\delta)^{1+\delta}\right)^{\mu}$	Theorem 12.2.1
$\delta \in (0,1)$	$P < \exp\big(-\mu\delta^2/3\big)$	Lemma 12.2.5
$\delta \in (0,4)$	$P < \exp\big(-\mu\delta^2/4\big)$	Lemma 12.2.6
$\delta \in (0,6)$	$P < \exp\big(-\mu\delta^2/5\big)$	Lemma 12.2.7
$\delta \geq 2e - 1$	$P < 2^{-\mu(1+\delta)}$	Lemma 12.2.8
$\delta \geq e^2$	$P < \exp\big(-(\mu\delta/2)\ln\delta\big)$	Lemma 12.2.9
$\delta \geq 0, \varphi \in (0,1]$	$\mathbb{P}\Big[Y > (1+\delta)\mu + \frac{3\ln\varphi^{-1}}{\delta^2}\Big] < \varphi.$	Lemma 12.2.10
$\delta \geq 0$	$\mathbb{P}[Y < (1-\delta)\mu] < \left(e^{-\delta}\big/(1-\delta)^{1-\delta}\right)^{\mu}$	Theorem 12.2.3
	$\mathbb{P}[Y < (1-\delta)\mu] < \exp\big(-\mu\delta^2/2\big)$	Lemma 12.2.4
$\Delta \geq 0$	$\mathbb{P}\big[\,Y - \mu \geq \Delta\,\big] \leq \exp\big(-2\Delta^2/n\big)$ $\mathbb{P}\big[\,Y - \mu \leq -\Delta\,\big] \leq \exp\big(-2\Delta^2/n\big).$	Corollary 12.3.5
$\tau \geq 1$	$\mathbb{P}[Y < \mu/\tau] < \exp\Big(-\big[1 - \frac{1+\ln\tau}{\tau}\big]\mu\Big)$	Theorem 12.2.3

$X_i \in [0,1]$	Arbitrary independent distributions	
$\delta \in [0,1]$	$\mathbb{P}\big[Y \geq (1+\delta)\mu\big] \leq \exp\big(-\delta^2\mu/4\big)$ $\mathbb{P}\big[Y \leq (1-\delta)\mu\big] \leq \exp\big(-\delta^2\mu/2\big).$	Theorem 12.3.6
$\Delta \geq 0$	$\mathbb{P}\big[\,Y - \mu \geq \Delta\,\big] \leq \exp\big(-2\Delta^2/n\big)$ $\mathbb{P}\big[\,Y - \mu \leq -\Delta\,\big] \leq \exp\big(-2\Delta^2/n\big).$	Corollary 12.3.5

$X_i \in [a_i, b_i]$	Arbitrary independent distributions	
$\Delta \geq 0$	$\mathbb{P}\Big[\lvert Y - \mu\rvert \,\geq\, \Delta\Big] \leq 2\exp\left(-\frac{2\,\Delta^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$	Theorem 12.4.3

Table 12.1: Summary of Chernoff type inequalities covered. Here we have n independent random variables $X_1, \ldots, X_n$, $Y = \sum_i X_i$ and $\mu = \mathbb{E}[Y]$.

dollars. This game, in expectation, does not change the amount of money the player has. Indeed, we have

$$\mathbb{E}\Big[Y_i \;\Big|\; Y_{i-1}\Big] = \frac{1}{2}(Y_{i-1} - \psi_i) + \frac{1}{2}(Y_{i-1} + \psi_i) = Y_{i-1}.$$

And as such, we have that $\mathbb{E}\big[Y_i\big] = \mathbb{E}\Big[\mathbb{E}\big[Y_i \big| Y_{i-1}\big]\Big] = \mathbb{E}\big[Y_{i-1}\big] = \cdots = \mathbb{E}\big[Y_0\big] = 1$. In particular, $\mathbb{E}[Y_n] = 1$ – namely, on average, independent of the player strategy she is not going to make any money in this game (and she is allowed to change her bets after every round). Unless, she is lucky[2]...

What about a lucky player? The player believes she will get lucky and wants to develop a strategy to take advantage of it. Formally, she believes that she can win, say, at least $(1+\delta)/2$ fraction of her bets (instead of the predicted $1/2$) – for example, if the bets are in the stock market, she can improve her chances by doing more research on the companies she is investing in[3]. Unfortunately, the player does not know which rounds she is going to be lucky in – so she still needs to be careful.

In a search of a good strategy. Of course, there are many safe strategies the player can use, from not playing at all, to risking only a tiny fraction of her money at each round. In other words, our quest here is to find the best strategy that extracts the maximum benefit for the player out of her inherent luck.

Here, we restrict ourselves to a simple strategy – at every round, the player would bet β fraction of her money, where β is a parameter to be determined. Specifically, in the end of the ith round, the player would have

$$Y_i = \begin{cases} (1-\beta)Y_{i-1} & \text{LOSE} \\ (1+\beta)Y_{i-1} & \text{WIN.} \end{cases}$$

By our assumption, the player is going to win in at least $M = (1+\delta)n/2$ rounds. Our purpose here is to figure out what the value of β should be so that player gets as rich as possible[4]. Now, if the player is successful in $\geq M$ rounds, out of the n rounds of the game, then the amount of money the player has, in the end of the game, is

$$\begin{aligned} Y_n &\geq (1-\beta)^{n-M}(1+\beta)^M = (1-\beta)^{n/2-(\delta/2)n}(1+\beta)^{n/2+(\delta/2)n} = \Big((1-\beta)(1+\beta)\Big)^{n/2-(\delta/2)n}(1+\beta)^{\delta n} \\ &= \Big(1-\beta^2\Big)^{n/2-(\delta/2)n}(1+\beta)^{\delta n} \geq \exp\big(-2\beta^2\big)^{n/2-(\delta/2)n}\exp(\beta/2)^{\delta n} = \exp\big((-\beta^2+\beta^2\delta+\beta\delta/2)n\big). \end{aligned}$$

To maximize this quantity, we choose $\beta = \delta/4$ (there is a better choice, see Lemma 12.1.6, but we use this value for the simplicity of exposition). Thus, we have that $Y_n \geq \exp\left(\left(-\frac{\delta^2}{16} + \frac{\delta^3}{16} + \frac{\delta^2}{8}\right)n\right) \geq \exp\left(\frac{\delta^2}{16}n\right)$, proving the following.

Lemma 12.1.1. *Consider a Chernoff game with n rounds, starting with one dollar, where the player wins in $\geq (1+\delta)n/2$ of the rounds. If the player bets $\delta/4$ fraction of her current money, at all rounds, then in the end of the game the player would have at least* $\exp\big(n\delta^2/16\big)$ *dollars.*

[2]"I would rather have a general who was lucky than one who was good." – Napoleon Bonaparte.

[3]"I am a great believer in luck, and I find the harder I work, the more I have of it." – Thomas Jefferson.

[4]This optimal choice is known as Kelly criterion, see Remark 12.1.3.

Remark 12.1.2. Note, that Lemma 12.1.1 holds if the player wins any $\geq (1+\delta)n/2$ rounds. In particular, the statement does not require randomness by itself – for our application, however, it is more natural and interesting to think about the player wins as being randomly distributed.

Remark 12.1.3. Interestingly, the idea of choosing the best fraction to bet is an old and natural question arising in investments strategies, and the right fraction to use is known as ***Kelly criterion***, going back to Kelly's work from 1956 [Kel56].

12.1.2.2. Chernoff's inequality

The above implies that if a player is lucky, then she is going to become filthy rich[5]. Intuitively, this should be a pretty rare event – because if the player is rich, then (on average) many other people have to be poor. We are thus ready for the kill.

Theorem 12.1.4 (Chernoff's inequality). *Let $X_1, \ldots, X_n$ be n independent random variables, where $X_i = 0$ or $X_i = 1$ with equal probability. Then, for any $\delta \in (0, 1/2)$, we have that*

$$\mathbb{P}\Big[\sum_i X_i \geq (1+\delta)\frac{n}{2}\Big] \leq \exp\Big(-\frac{\delta^2}{16}n\Big).$$

Proof: Imagine that we are playing the Chernoff game above, with $\beta = \delta/4$, starting with 1 dollar, and let Y_i be the amount of money in the end of the ith round. Here $X_i = 1$ indicates that the player won the ith round. We have, by Lemma 12.1.1 and Markov's inequality, that

$$\mathbb{P}\Big[\sum_i X_i \geq (1+\delta)\frac{n}{2}\Big] \leq \mathbb{P}\Big[Y_n \geq \exp\Big(\frac{n\delta^2}{16}\Big)\Big] \leq \frac{\mathbb{E}[Y_n]}{\exp(n\delta^2/16)} = \frac{1}{\exp(n\delta^2/16)} = \exp\Big(-\frac{\delta^2}{16}n\Big). \qquad \blacksquare$$

This is crazy – so intuition maybe? If the player is $(1+\delta)/2$-lucky then she can make a lot of money; specifically, at least $f(\delta) = \exp(n\delta^2/16)$ dollars by the end of the game. Namely, beating the odds has significant monetary value, and this value grows quickly with δ. Since we are in a "zero-sum" game settings, this event should be very rare indeed. Under this interpretation, of course, the player needs to know in advance the value of δ – so imagine that she guesses it somehow in advance, or she plays the game in parallel with all the possible values of δ, and she settles on the instance that maximizes her profit.

Can one do better? No, not really. Chernoff inequality is tight (this is a challenging homework exercise) up to the constant in the exponent. The best bound I know for this version of the inequality has 1/2 instead of 1/16 in the exponent. Note, however, that no real effort was taken to optimize the constants – this is not the purpose of this write-up.

12.1.2.3. Some low level boring calculations

Above, we used the following well known facts.

Lemma 12.1.5. *(A) Markov's inequality. For any positive random variable X and $t > 0$, we have $\mathbb{P}[X \geq t] \leq \mathbb{E}[X]/t$. (B) For any two random variables X and Y, we have that $\mathbb{E}\big[X\big] = \mathbb{E}\big[\mathbb{E}\big[X \,\big|\, Y\big]\big]$. (C) For $x \in (0,1)$, $1 + x \geq e^{x/2}$. (D) For $x \in (0, 1/2)$, $1 - x \geq e^{-2x}$.*

[5] Not that there is anything wrong with that – many of my friends are filthy,

Lemma 12.1.6. *The quantity* $\exp\big((-\beta^2 + \beta^2\delta + \beta\delta/2)n\big)$ *is maximal for* $\beta = \frac{\delta}{4(1-\delta)}$.

Proof: We have to maximize $f(\beta) = -\beta^2 + \beta^2\delta + \beta\delta/2$ by choosing the correct value of β (as a function of δ, naturally). $f'(\beta) = -2\beta + 2\beta\delta + \delta/2 = 0 \iff 2(\delta - 1)\beta = -\delta/2 \iff \beta = \frac{\delta}{4(1-\delta)}$. ■

12.1.3. A proof for −1/+1 case

Theorem 12.1.7. *Let* $X_1, \ldots, X_n$ *be* n *independent random variables, such that* $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$, *for* $i = 1, \ldots, n$. *Let* $Y = \sum_{i=1}^n X_i$. *Then, for any* $\Delta > 0$, *we have*

$$\mathbb{P}\Big[Y \geq \Delta\Big] \leq \exp\big(-\Delta^2/2n\big).$$

Proof: Clearly, for an arbitrary t, to specified shortly, we have

$$\mathbb{P}[Y \geq \Delta] = \mathbb{P}[\exp(tY) \geq \exp(t\Delta)] \leq \frac{\mathbb{E}[\exp(tY)]}{\exp(t\Delta)},$$

the first part follows by the fact that $\exp(\cdot)$ preserve ordering, and the second part follows by the Markov inequality.

Observe that

$$\begin{aligned}
\mathbb{E}[\exp(tX_i)] &= \frac{1}{2}e^t + \frac{1}{2}e^{-t} = \frac{e^t + e^{-t}}{2} \\
&= \frac{1}{2}\left(1 + \frac{t}{1!} + \frac{t^2}{2!} + \frac{t^3}{3!} + \cdots\right) \\
&+ \frac{1}{2}\left(1 - \frac{t}{1!} + \frac{t^2}{2!} - \frac{t^3}{3!} + \cdots\right) \\
&= \left(1 + \quad + \frac{t^2}{2!} + \quad + \cdots + \frac{t^{2k}}{(2k)!} + \cdots\right),
\end{aligned}$$

by the Taylor expansion of $\exp(\cdot)$. Note, that $(2k)! \geq (k!)2^k$, and thus

$$\mathbb{E}[\exp(tX_i)] = \sum_{i=0}^{\infty} \frac{t^{2i}}{(2i)!} \leq \sum_{i=0}^{\infty} \frac{t^{2i}}{2^i(i!)} = \sum_{i=0}^{\infty} \frac{1}{i!}\left(\frac{t^2}{2}\right)^i = \exp\big(t^2/2\big),$$

again, by the Taylor expansion of $\exp(\cdot)$. Next, by the independence of the X_is, we have

$$\mathbb{E}[\exp(tY)] = \mathbb{E}\left[\exp\left(\sum_i tX_i\right)\right] = \mathbb{E}\left[\prod_i \exp(tX_i)\right] = \prod_{i=1}^{n} \mathbb{E}[\exp(tX_i)] \leq \prod_{i=1}^{n} e^{t^2/2} = e^{nt^2/2}.$$

We have $\mathbb{P}[Y \geq \Delta] \leq \dfrac{\exp\big(nt^2/2\big)}{\exp(t\Delta)} = \exp\big(nt^2/2 - t\Delta\big)$.

Next, by minimizing the above quantity for t, we set $t = \Delta/n$. We conclude,

$$\mathbb{P}[Y \geq \Delta] \leq \exp\left(\frac{n}{2}\left(\frac{\Delta}{n}\right)^2 - \frac{\Delta}{n}\Delta\right) = \exp\left(-\frac{\Delta^2}{2n}\right).$$ ■

By the symmetry of Y, we get the following:

Corollary 12.1.8. *Let $X_1, \ldots, X_n$ be n independent random variables, such that $\mathbb{P}[X_i = 1] = \mathbb{P}[X_i = -1] = \frac{1}{2}$, for $i = 1, \ldots, n$. Let $Y = \sum_{i=1}^n X_i$. Then, for any $\Delta > 0$, we have $\mathbb{P}[|Y| \geq \Delta] \leq 2\exp\left(-\Delta^2/2n\right)$.*

Corollary 12.1.9. *Let $X_1, \ldots, X_n$ be n independent* coin flips*, such that $\mathbb{P}[X_i = 0] = \mathbb{P}[X_i = 1] = \frac{1}{2}$, for $i = 1, \ldots, n$. Let $Y = \sum_{i=1}^n X_i$. Then, for any $\Delta > 0$, we have $\mathbb{P}[|Y - n/2| \geq \Delta] \leq 2\exp\left(-2\Delta^2/n\right)$.*

Remark 12.1.10. Before going any further, it is might be instrumental to understand what this inequalities imply. Consider then case where X_i is either zero or one with probability half. In this case $\mu = \mathbb{E}[Y] = n/2$. Set $\delta = t\sqrt{n}$ ($\sqrt{\mu}$ is approximately the standard deviation of X if $p_i = 1/2$). We have by

$$\mathbb{P}\left[\left|Y - \frac{n}{2}\right| \geq \Delta\right] \leq 2\exp\left(-2\Delta^2/n\right) = 2\exp\left(-2(t\sqrt{n})^2/n\right) = 2\exp\left(-2t^2\right).$$

Thus, Chernoff inequality implies exponential decay (i.e., $\leq 2^{-t}$) with t standard deviations, instead of just polynomial (i.e., $\leq 1/t^2$) by the Chebychev's inequality.

12.2. The Chernoff Bound — General Case

Here we present the Chernoff bound in a more general settings.

Theorem 12.2.1. *Let $X_1, \ldots, X_n$ be n independent variables, where $\mathbb{P}\left[X_i = 1\right] = p_i$ and $\mathbb{P}\left[X_i = 0\right] = q_i = 1 - p_i$, for all i. Let $X = \sum_{i=1}^b X_i$. $\mu = \mathbb{E}\left[X\right] = \sum_i p_i$. For any $\delta > 0$, we have*

$$\mathbb{P}\left[X > (1+\delta)\mu\right] < \left(e^{\delta}\Big/(1+\delta)^{1+\delta}\right)^{\mu}.$$

Proof: We have $\mathbb{P}[X > (1+\delta)\mu] = \mathbb{P}\left[e^{tX} > e^{t(1+\delta)\mu}\right]$. By the Markov inequality, we have:

$$\mathbb{P}\left[X > (1+\delta)\mu\right] < \frac{\mathbb{E}\left[e^{tX}\right]}{e^{t(1+\delta)\mu}}$$

On the other hand,

$$\mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[e^{t(X_1+X_2\ldots+X_n)}\right] = \mathbb{E}\left[e^{tX_1}\right]\cdots\mathbb{E}\left[e^{tX_n}\right].$$

Namely,

$$\mathbb{P}[X > (1+\delta)\mu] < \frac{\prod_{i=1}^n \mathbb{E}\left[e^{tX_i}\right]}{e^{t(1+\delta)\mu}} = \frac{\prod_{i=1}^n\left((1-p_i)e^0 + p_i e^t\right)}{e^{t(1+\delta)\mu}} = \frac{\prod_{i=1}^n\left(1 + p_i(e^t - 1)\right)}{e^{t(1+\delta)\mu}}.$$

Let $y = p_i(e^t - 1)$. We know that $1 + y < e^y$ (since $y > 0$). Thus,

$$\begin{aligned}
\mathbb{P}[X > (1+\delta)\mu] &< \frac{\prod_{i=1}^n \exp(p_i(e^t-1))}{e^{t(1+\delta)\mu}} = \frac{\exp\left(\sum_{i=1}^n p_i(e^t-1)\right)}{e^{t(1+\delta)\mu}} \\
&= \frac{\exp\left((e^t-1)\sum_{i=1}^n p_i\right)}{e^{t(1+\delta)\mu}} = \frac{\exp\left((e^t-1)\mu\right)}{e^{t(1+\delta)\mu}} = \left(\frac{\exp\left(e^t-1\right)}{e^{t(1+\delta)}}\right)^{\mu} \\
&= \left(\frac{\exp(\delta)}{(1+\delta)^{(1+\delta)}}\right)^{\mu},
\end{aligned}$$

if we set $t = \log(1+\delta)$. ■

12.2.1. The lower tail

We need the following low level lemma.

Lemma 12.2.2. *For $x \in [0,1)$, we have $(1-x)^{1-x} \geq \exp(-x + x^2/2)$.*

Proof: For $x \in [0,1)$, we have, by the Taylor expansion, that $\ln(1-x) = -\sum_{i=1}^{\infty}(x^i/i)$. As such, we have

$$(1-x)\ln(1-x) = -(1-x)\sum_{i=1}^{\infty}\frac{x^i}{i} = -\sum_{i=1}^{\infty}\frac{x^i}{i} + \sum_{i=1}^{\infty}\frac{x^{i+1}}{i} = -x + \sum_{i=2}^{\infty}\left(\frac{x^i}{i-1} - \frac{x^i}{i}\right) = -x + \sum_{i=2}^{\infty}\frac{x^i}{i(i-1)}.$$

This implies that $(1-x)\ln(1-x) \geq -x + x^2/2$, which implies the claim by exponentiation. ■

Theorem 12.2.3. *Let $X_1, \ldots, X_n$ be n independent random variables, where $\mathbb{P}\left[X_i = 1\right] = p_i$, $\mathbb{P}\left[X_i = 0\right] = q_i = 1 - p_i$, for all i. For $X = \sum_{i=1}^{n} X_i$, its expectation is $\mu = \mathbb{E}\left[X\right] = \sum_i p_i$. We have that*

$$\mathbb{P}\left[X < (1-\delta)\mu\right] < \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu}.$$

For any positive $\tau > 1$, we have that $\mathbb{P}\left[X < \mu/\tau\right] \leq \exp\left(-\left(1 - \frac{1+\ln\tau}{\tau}\right)\mu\right)$.

Proof: We follow the same proof template seen already. For $t = -\ln(1-\delta) > 0$, we have $\mathbb{E}[\exp(-tX_i)] = (1-p_i)e^0 + p_i e^{-t} = 1 - p_i + p_i(1-\delta) = 1 - p_i\delta \leq \exp(-p_i\delta)$. As such, we have

$$\mathbb{P}\left[X < (1-\delta)\mu\right] = \mathbb{P}\left[-X > -(1-\delta)\mu\right] = \mathbb{P}\left[\exp(-tX) > \exp(-t(1-\delta)\mu)\right] \leq \frac{\prod_{i=1}^{n}\mathbb{E}[\exp(-tX_i)]}{\exp(-t(1-\delta)\mu)}$$
$$\leq \frac{\exp\left(-\sum_{i=1}^{n} p_i\delta\right)}{\exp(-t(1-\delta)\mu)} = \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu}.$$

For the last inequality, set $\delta = 1 - 1/\tau$, and observe that

$$\mathbb{P}\left[X < (1-\delta)\mu\right] \leq \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu} = \left[\frac{\exp(-1+1/\tau)}{(1/\tau)^{1/\tau}}\right]^{\mu} = \exp\left(-\left(1 - \frac{1+\ln\tau}{\tau}\right)\mu\right).$$ ■

Lemma 12.2.4. *Let $X_1, \ldots, X_n \in \{0,1\}$ be n independent random variables, with $p_i = \mathbb{P}\left[X_i = 1\right]$, for all i. For $X = \sum_{i=1}^{n} X_i$, and $\mu = \mathbb{E}\left[X\right] = \sum_i p_i$, we have that $\mathbb{P}\left[X < (1-\delta)\mu\right] < \mathrm{Exp}\left(-\mu\delta^2/2\right)$.*

Proof: This alternative simplified form of Theorem 12.2.3, follows readily from Lemma 12.2.2, since

$$\mathbb{P}\left[X < (1-\delta)\mu\right] \leq \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu} \leq \left[\frac{e^{-\delta}}{\mathrm{Exp}(-\delta+\delta^2/2)}\right]^{\mu} \leq \mathrm{Exp}(-\mu\delta^2/2).$$ ■

12.2.2. A more convenient form of Chernoff's inequality

Lemma 12.2.5. *Let $X_1, \ldots, X_n$ be n independent Bernoulli trials, where $\mathbb{P}[X_i = 1] = p_i$, and $\mathbb{P}[X_i = 0] = 1 - p_i$, for $i = 1, \ldots, n$. Let $X = \sum_{i=1}^{b} X_i$, and $\mu = \mathbb{E}[X] = \sum_i p_i$. For $\delta, \in (0, 1)$, we have*

$$\mathbb{P}\left[X > (1+\delta)\mu\right] < \exp\left(-\mu\delta^2/3\right).$$

Proof: By Theorem 12.2.1, it is sufficient to prove, for $\delta \in [0, 1]$, that

$$\left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \leq \exp\left(-\frac{\mu\delta^2}{c}\right) \iff \mu\left(\delta - (1+\delta)\ln(1+\delta)\right) \leq -\mu\delta^2/c$$
$$\iff f(\delta) = \delta^2/c + \delta - (1+\delta)\ln(1+\delta) \leq 0.$$

We have

$$f'(\delta) = 2\delta/c - \ln(1+\delta). \qquad \text{and} \qquad f''(\delta) = 2/c - \frac{1}{1+\delta}.$$

For $c = 3$, we have $f''(\delta) \leq 0$ for $\delta \in [0, 1/2]$, and $f''(\delta) \geq 0$ for $\delta \in [1/2, 1]$. Namely, $f'(\delta)$ achieves its maximum either at 0 or 1. As $f'(0) = 0$ and $f'(1) = 2/3 - \ln 2 \approx -0.02 < 0$, we conclude that $f'(\delta) \leq 0$. Namely, f is a monotonically decreasing function in $[0, 1]$, which implies that $f(\delta) \leq 0$, for all δ in this range, thus implying the claim. ■

Lemma 12.2.6. *Let $X_1, \ldots, X_n$ be n independent Bernoulli trials, where $\mathbb{P}[X_i = 1] = p_i$, and $\mathbb{P}[X_i = 0] = 1 - p_i$, for $i = 1, \ldots, n$. Let $X = \sum_{i=1}^{b} X_i$, and $\mu = \mathbb{E}[X] = \sum_i p_i$. For $\delta \in (0, 4)$, we have*

$$\mathbb{P}\left[X > (1+\delta)\mu\right] < \exp\left(-\mu\delta^2/4\right),$$

Proof: Lemma 12.2.5 implies a stronger bound, so we need to prove the claim only for $\delta \in (1, 4]$. Continuing as in the proof of Lemma 12.2.5, for case $c = 4$, we have to prove that

$$f(\delta) = \delta^2/4 + \delta - (1+\delta)\ln(1+\delta) \leq 0,$$

where $f''(\delta) = 1/2 - \frac{1}{1+\delta}$.

For $\delta > 1$, we have $f''(\delta) > 0$. Namely $f(\cdot)$ is convex for $\delta \geq 1$, and it achieves its maximum on the interval $[1, 4]$ on the endpoints. In particular, $f(1) \approx -0.13$, and $f(4) \approx -0.047$, which implies the claim. ■

Lemma 12.2.7. *Let $X_1, \ldots, X_n$ be n independent random variables, where $\mathbb{P}[X_i = 1] = p_i$, and $\mathbb{P}[X_i = 0] = 1 - p_i$, for $i = 1, \ldots, n$. Let $X = \sum_{i=1}^{b} X_i$, and $\mu = \mathbb{E}[X] = \sum_i p_i$. For $\delta \in (0, 6)$, we have*

$$\mathbb{P}\left[X > (1+\delta)\mu\right] < \exp\left(-\mu\delta^2/5\right),$$

Proof: Lemma 12.2.6 implies a stronger bound, so we need to prove the claim only for $\delta \in (4, 5]$. Continuing as in the proof of Lemma 12.2.5, for case $c = 5$, we have to prove that

$$f(\delta) = \delta^2/5 + \delta - (1+\delta)\ln(1+\delta) \leq 0,$$

where $f''(\delta) = 2/5 - \frac{1}{1+\delta}$. For $\delta \geq 4$, we have $f''(\delta) > 0$. Namely $f(\cdot)$ is convex for $\delta \geq 4$, and it achieves its maximum on the interval $[4, 6]$ on the endpoints. In particular, $f(4) \approx -0.84$, and $f(6) \approx -0.42$, which implies the claim. ■

Lemma 12.2.8. *Let $X_1, \ldots, X_n$ be n independent Bernoulli trials, where $\mathbb{P}[X_i = 1] = p_i$, and $\mathbb{P}[X_i = 0] = 1 - p_i$, for $i = 1, \ldots, n$. Let $X = \sum_{i=1}^{b} X_i$, and $\mu = \mathbb{E}\big[X\big] = \sum_i p_i$. For $\delta > 2e - 1$, we have $\mathbb{P}\big[X > (1+\delta)\mu\big] < 2^{-\mu(1+\delta)}$.*

Proof: By Theorem 12.2.1, we have

$$\left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} \leq \left(\frac{e}{1+2e-1}\right)^{(1+\delta)\mu} \leq 2^{-(1+\delta)\mu},$$

since $\delta > 2e - 1$. ∎

Lemma 12.2.9. *Let $X_1, \ldots, X_n$ be n independent Bernoulli trials, where $\mathbb{P}[X_i = 1] = p_i$, and $\mathbb{P}[X_i = 0] = 1 - p_i$, for $i = 1, \ldots, n$. Let $X = \sum_{i=1}^{b} X_i$, and $\mu = \mathbb{E}\big[X\big] = \sum_i p_i$. For $\delta > e^2$, we have $\mathbb{P}\big[X > (1+\delta)\mu\big] < \exp\Big(-\frac{\mu\delta \ln \delta}{2}\Big)$.*

Proof: Observe that

$$\mathbb{P}\Big[X > (1+\delta)\mu\Big] < \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} = \exp\Big(\mu\delta - \mu(1+\delta)\ln(1+\delta)\Big). \tag{12.1}$$

As such, we have

$$\mathbb{P}\Big[X > (1+\delta)\mu\Big] < \exp\Big(-\mu(1+\delta)\big(\ln(1+\delta) - 1\big)\Big) \leq \exp\left(-\mu\delta \ln \frac{1+\delta}{e}\right) \leq \exp\left(-\frac{\mu\delta\ln\delta}{2}\right),$$

since for $x \geq e^2$ we have that $\dfrac{1+x}{e} \geq \sqrt{x} \iff \ln\dfrac{1+x}{e} \geq \dfrac{\ln x}{2}$. ∎

12.2.2.1. Bound when the expectation is small

Lemma 12.2.10. *Let $X_1, \ldots, X_n$ be n independent Bernoulli trials, where $\mathbb{P}[X_i = 1] = p_i$, and $\mathbb{P}[X_i = 0] = 1 - p_i$, for $i = 1, \ldots, n$. Let $Y = \sum_{i=1}^{b} X_i$, and $\mu = \mathbb{E}[Y] = \sum_i p_i$. For $\delta \in (0,1]$, and $\varphi \in (0,1]$, we have*

$$\mathbb{P}\left[Y > (1+\delta)\mu + \frac{3\ln\varphi^{-1}}{\delta^2}\right] < \varphi.$$

Proof: Let $\xi = \delta + \frac{3\ln\varphi^{-1}}{\mu\delta^2}$. If $\xi \geq 2e - 1 \approx 4.43$, by Lemma 12.2.8, we have

$$\alpha = \mathbb{P}\left[Y > (1+\delta)\mu + \frac{3\ln\varphi^{-1}}{\delta^2}\right] = \mathbb{P}\big[Y > (1+\xi)\mu\big] \leq 2^{-\mu(1+\xi)} < \varphi,$$

since $-\mu(1+\xi) > -\mu\xi > \mu\frac{3\ln\varphi^{-1}}{\mu\delta^2} > \log_2\varphi^{-1}$, since $\delta \in (0,1]$.

If $\xi \leq 6$, then by Lemma 12.2.7, we have

$$\alpha = \mathbb{P}\big[Y > (1+\xi)\mu\big] \leq \exp\big(-\mu\xi^2/5\big) \leq \varphi,$$

since

$$-\frac{\mu}{5}\xi^2 = -\frac{\mu}{5}\left(\delta + \frac{3\ln\varphi^{-1}}{\mu\delta^2}\right)^2 > -\frac{\mu}{5}\left(2\cdot\delta\cdot\frac{3\ln\varphi^{-1}}{\mu\delta^2}\right) = -\frac{6}{5}\cdot\frac{\ln\varphi}{\delta} > -\ln\varphi.$$

∎

Example 12.2.11. Let $X_1, \ldots, X_n$ be n independent Bernoulli trials, where $\mathbb{P}[X_i = 1] = p_i$, and $\mathbb{P}[X_i = 0] = 1 - p_i$, for $i = 1, \ldots, n$. Let $Y = \sum_{i=1}^{b} X_i$, and $\mu = \mathbb{E}[Y] = \sum_i p_i$. Assume that $\mu \leq 1/2$. Setting $\delta = 1$, We have, for $t > 6$, that

$$\mathbb{P}[Y > 1 + t] \leq \mathbb{P}\left[Y > (1+\delta)\mu + \frac{3\ln\exp(t/3)}{\delta^2}\right] \leq \exp(-t/3),$$

by Lemma 12.2.10.

12.3. A special case of Hoeffding's inequality

In this section, we prove yet another version of Chernoff inequality, where each variable is randomly picked according to its own distribution in the range $[0,1]$. We prove a more general version of this inequality in Section 12.4, but the version presented here does not follow from this generalization.

Theorem 12.3.1. *Let $X_1, \ldots, X_n \in [0,1]$ be n independent random variables, let $X = \sum_{i=1}^n X_i$, and let $\mu = \mathbb{E}[X]$. We have that* $\mathbb{P}\Big[X - \mu \geq \eta\Big] \leq \left(\dfrac{\mu}{\mu+\eta}\right)^{\mu+\eta}\left(\dfrac{n-\mu}{n-\mu-\eta}\right)^{n-\mu-\eta}$.

Proof: Let $s \geq 1$ be some arbitrary parameter. By the standard arguments, we have

$$\gamma = \mathbb{P}\Big[X \geq \mu + \eta\Big] = \mathbb{P}\Big[s^X \geq s^{\mu+\eta}\Big] \leq \frac{\mathbb{E}\big[s^X\big]}{s^{\mu+\eta}} = s^{-\mu-\eta}\prod_{i=1}^n \mathbb{E}\big[s^{X_i}\big].$$

By calculations, see Lemma 12.3.7 below, one can show that $\mathbb{E}\big[s^{X_1}\big] \leq 1 + (s-1)\,\mathbb{E}[X_i]$. As such, by the AM-GM inequality[6], we have that

$$\prod_{i=1}^n \mathbb{E}\big[s^{X_i}\big] \leq \prod_{i=1}^n \Big(1 + (s-1)\,\mathbb{E}[X_i]\Big) \leq \left(\frac{1}{n}\sum_{i=1}^n \Big(1 + (s-1)\,\mathbb{E}[X_i]\Big)\right)^n = \left(1 + (s-1)\frac{\mu}{n}\right)^n.$$

Setting $s = \dfrac{(\mu+\eta)(n-\mu)}{\mu(n-\mu-\eta)} = \dfrac{\mu n - \mu^2 + \eta n - \eta\mu}{\mu n - \mu^2 - \eta\mu}$ we have that

$$1 + (s-1)\frac{\mu}{n} = 1 + \frac{\eta n}{\mu n - \mu^2 - \eta\mu}\cdot\frac{\mu}{n} = 1 + \frac{\eta}{n-\mu-\eta} = \frac{n-\mu}{n-\mu-\eta}.$$

As such, we have that

$$\gamma \leq s^{-\mu-\eta}\prod_{i=1}^n \mathbb{E}\big[s^{X_i}\big] = \left(\frac{\mu(n-\mu-\eta)}{(\mu+\eta)(n-\mu)}\right)^{\mu+\eta}\left(\frac{n-\mu}{n-\mu-\eta}\right)^n = \left(\frac{\mu}{(\mu+\eta)}\right)^{\mu+\eta}\left(\frac{n-\mu}{n-\mu-\eta}\right)^{n-\mu-\eta}. \qquad \blacksquare$$

Remark 12.3.2. Setting $s = (\mu+\eta)/\mu$ in the proof of Theorem 12.3.1, we have

$$\mathbb{P}\Big[X - \mu \geq \eta\Big] \leq \left(\tfrac{\mu}{\mu+\eta}\right)^{\mu+\eta}\left(1 + \left(\tfrac{\mu+\eta}{\mu} - 1\right)\tfrac{\mu}{n}\right)^n = \left(\tfrac{\mu}{\mu+\eta}\right)^{\mu+\eta}\left(1 + \tfrac{\eta}{n}\right)^n.$$

Corollary 12.3.3. *Let $X_1, \ldots, X_n \in [0,1]$ be n independent random variables, let $\overline{X} = \sum_{i=1}^n X_i/n$, $p = \mathbb{E}\Big[\overline{X}\Big] = \mu/n$ and $q = 1 - p$. Then, we have that $\mathbb{P}\Big[\overline{X} - p \geq t\Big] \leq \exp\big(nf(t)\big)$, for*

$$f(t) = (p+t)\ln\frac{p}{p+t} + (q-t)\ln\frac{q}{q-t}. \qquad (12.2)$$

Theorem 12.3.4. *Let $X_1, \ldots, X_n \in [0,1]$ be n independent random variables, let $\overline{X} = (\sum_{i=1}^n X_i)/n$, and let $p = \mathbb{E}[X]$. We have that $\mathbb{P}\Big[\overline{X} - p \geq t\Big] \leq \exp\big(-2nt^2\big)$ and $\mathbb{P}\Big[\overline{X} - p \leq -t\Big] \leq \exp\big(-2nt^2\big)$.*

[6]The inequality between arithmetic and geometric means: $(\sum_{i=1}^n x_i)/n \geq \sqrt[n]{x_1 \cdots x_n}$.

Proof: Let $p = \mu/n$, $q = 1 - p$, and let $f(t)$ be the function from Eq. (12.2), for $t \in (-p, q)$. Now, we have that

$$f'(t) = \ln\frac{p}{p+t} + (p+t)\frac{p+t}{p}\left(-\frac{p}{(p+t)^2}\right) - \ln\frac{q}{q-t} - (q-t)\frac{q-t}{q}\frac{q}{(q-t)^2} = \ln\frac{p}{p+t} - \ln\frac{q}{q-t}$$
$$= \ln\frac{p(q-t)}{q(p+t)}.$$

As for the second derivative, we have

$$f''(t) = \frac{q(p+t)}{p(q-t)} \cdot \frac{p}{q} \cdot \frac{(p+t)(-1) - (q-t)}{(p+t)^2}. = \frac{-p-t-q+t}{(q-t)(p+t)} = -\frac{1}{(q-t)(p+t)} \leq -4.$$

Indeed, $t \in (-p, q)$ and the denominator is minimized for $t = (q-p)/2$, and as such $(q-t)(p+t) \leq \Big(2q - (q-p)\Big)\Big(2p + (q-p)\Big)/4 = (p+q)^2/4 = 1/4$.

Now, $f(0) = 0$ and $f'(0) = 0$, and by Taylor's expansion, we have that $f(t) = f(0) + f'(0)t + \frac{f''(x)}{2}t^2 \leq -2t^2$, where x is between 0 and t.

The first bound now readily follows from plugging this bound into Corollary 12.3.3. The second bound follows by considering the random variants $Y_i = 1 - X_i$, for all i, and plugging this into the first bound. Indeed, for $\overline{Y} = 1 - \overline{X}$, we have that $q = \mathbb{E}\Big[\overline{Y}\Big]$, and then $\overline{X} - p \leq -t \iff t \leq p - \overline{X} \iff t \leq 1 - q - (1 - \overline{Y}) = \overline{Y} - q$. Thus, $\mathbb{P}\Big[\overline{X} - p \leq -t\Big] = \mathbb{P}\Big[\overline{Y} - q \geq t\Big] \leq \exp(-2nt^2)$. ■

Corollary 12.3.5. *Let $X_1, \ldots, X_n \in [0,1]$ be n independent random variables, let $Y = \sum_{i=1}^n X_i$, and let $\mu = \mathbb{E}[X]$. For any $\Delta > 0$, we have $\mathbb{P}\big[\, Y - \mu \geq \Delta \,\big] \leq \exp(-2\Delta^2/n)$ and $\mathbb{P}\big[\, Y - \mu \leq -\Delta \,\big] \leq \exp(-2\Delta^2/n)$.*

Proof: For $\overline{X} = Y/n$, $p = \mu/n$, and $t = \Delta/n$, by Theorem 12.3.4, we have

$$\mathbb{P}\big[Y - \mu \geq \Delta\big] = \mathbb{P}\big[\overline{X} - p \geq t\big] \leq \exp(-2nt^2) = \exp(-2\Delta^2/n). \qquad ■$$

Theorem 12.3.6. *Let $X_1, \ldots, X_n \in [0,1]$ be n independent random variables, let $X = (\sum_{i=1}^n X_i)$, and let $\mu = \mathbb{E}[X]$. We have that $\mathbb{P}\big[X - \mu \geq \varepsilon\mu\big] \leq \exp(-\varepsilon^2\mu/4)$ and $\mathbb{P}\big[X - \mu \leq -\varepsilon\mu\big] \leq \exp(-\varepsilon^2\mu/2)$.*

Proof: Let $p = \mu/n$, and let $g(x) = f(px)$, for $x \in [0,1]$ and $xp < q$. As before, computing the derivative of g, we have

$$g'(x) = pf'(xp) = p\ln\frac{p(q-xp)}{q(p+xp)} = p\ln\frac{q-xp}{q(1+x)} \leq p\ln\frac{1}{1+x} \leq -\frac{px}{2},$$

since $(q-xp)/q$ is maximized for $x = 0$, and $\ln\frac{1}{1+x} \leq -x/2$, for $x \in [0,1]$, as can be easily verified[7]. Now, $g(0) = f(0) = 0$, and by integration, we have that $g(x) = \int_{y=0}^{x} g'(y)\mathrm{d}y \leq \int_{y=0}^{x}(-py/2)\mathrm{d}y = -px^2/4$. Now, plugging into Corollary 12.3.3, we get that the desired probability $\mathbb{P}\big[X - \mu \geq \varepsilon\mu\big]$ is

$$\mathbb{P}\Big[\overline{X} - p \geq \varepsilon p\Big] \leq \exp\big(nf(\varepsilon p)\big) = \exp\Big(ng(\varepsilon)\Big) \leq \exp(-pn\varepsilon^2/4) = \exp(-\mu\varepsilon^2/4).$$

[7]Indeed, this is equivalent to $\frac{1}{1+x} \leq e^{-x/2} \iff e^{x/2} \leq 1 + x$, which readily holds for $x \in [0,1]$.

As for the other inequality, set $h(x) = g(-x) = f(-xp)$. Then

$$h'(x) = -pf'(-xp) = -p\ln\frac{p(q+xp)}{q(p-xp)} = p\ln\frac{q(1-x)}{q+xp} = p\ln\frac{q-xq}{q+xp} = p\ln\left(1-x\frac{p+q}{q+xp}\right)$$
$$= p\ln\left(1-x\frac{1}{q+xp}\right) \le p\ln(1-x) \le -px,$$

since $1-x \le e^{-x}$. By integration, as before, we conclude that $h(x) \le -px^2/2$. Now, plugging into Corollary 12.3.3, we get $\mathbb{P}\left[X-\mu \le -\varepsilon\mu\right] = \mathbb{P}\left[\overline{X}-p \le -\varepsilon p\right] \le \exp\left(nf(-\varepsilon p)\right) \le \exp\left(nh(\varepsilon)\right) \le \exp\left(-np\varepsilon^2/2\right) \le \exp\left(-\mu\varepsilon^2/2\right)$. ■

12.3.1. Some technical lemmas

Lemma 12.3.7. *Let $X \in [0,1]$ be a random variable, and let $s \ge 1$. Then $\mathbb{E}\left[s^X\right] \le 1 + (s-1)\,\mathbb{E}[X]$.*

Proof: For the sake of simplicity of exposition, assume that X is a discrete random variable, and that there is a value $\alpha \in (0,1/2)$, such that $\beta = \mathbb{P}[X=\alpha] > 0$. Consider the modified random variable X', such that $\mathbb{P}[X'=0] = \mathbb{P}[X=0]+\beta/2$, and $\mathbb{P}[X'=2\alpha] = \mathbb{P}[X=\alpha]+\beta/2$. Clearly, $\mathbb{E}[X] = \mathbb{E}[X']$. Next, observe that $\mathbb{E}\left[s^{X'}\right] - \mathbb{E}\left[s^X\right] = (\beta/2)(s^{2\alpha}+s^0) - \beta s^\alpha \ge 0$, by the convexity of s^x. We conclude that $\mathbb{E}\left[s^X\right]$ achieves its maximum if takes only the values 0 and 1. But then, we have that $\mathbb{E}\left[s^X\right] = \mathbb{P}[X=0]s^0 + \mathbb{P}[X=1]s^1 = (1-\mathbb{E}[X]) + \mathbb{E}[X]\,s = 1+(s-1)\,\mathbb{E}[X]$, as claimed. ■

12.4. Hoeffding's inequality

In this section, we prove a generalization of Chernoff's inequality. The proof is considerably more tedious, and it is included here for the sake of completeness.

Lemma 12.4.1. *Let X be a random variable. If $\mathbb{E}[X] = 0$ and $a \le X \le b$, then for any $s > 0$, we have $\mathbb{E}\left[e^{sX}\right] \le \exp\left(s^2(b-a)^2/8\right)$.*

Proof: Let $a \le x \le b$ and observe that x can be written as a convex combination of a and b. In particular, we have

$$x = \lambda a + (1-\lambda)b \qquad \text{for} \qquad \lambda = \frac{b-x}{b-a} \in [0,1]\,.$$

Since $s > 0$, the function $\exp(sx)$ is convex, and as such

$$e^{sx} \le \frac{b-x}{b-a}e^{sa} + \frac{x-a}{b-a}e^{sb},$$

since we have that $f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y)$ if $f(\cdot)$ is a convex function. Thus, for a random variable X, by linearity of expectation, we have

$$\mathbb{E}\left[e^{sX}\right] \le \mathbb{E}\left[\frac{b-X}{b-a}e^{sa} + \frac{X-a}{b-a}e^{sb}\right] = \frac{b-\mathbb{E}[X]}{b-a}e^{sa} + \frac{\mathbb{E}[X]-a}{b-a}e^{sb}$$
$$= \frac{b}{b-a}e^{sa} - \frac{a}{b-a}e^{sb},$$

since $\mathbb{E}[X] = 0$.

Next, set $p = -\dfrac{a}{b-a}$ and observe that $1 - p = 1 + \dfrac{a}{b-a} = \dfrac{b}{b-a}$ and

$$-ps(b-a) = -\left(-\frac{a}{b-a}\right)s(b-a) = sa.$$

As such, we have

$$\begin{aligned}\mathbb{E}\left[e^{sX}\right] &\leq (1-p)e^{sa} + pe^{sb} = \left(1 - p + pe^{s(b-a)}\right)e^{sa}\\ &= \left(1 - p + pe^{s(b-a)}\right)e^{-ps(b-a)}\\ &= \exp\left(-ps(b-a) + \ln\left(1 - p + pe^{s(b-a)}\right)\right) = \exp(-pu + \ln(1 - p + pe^{u})),\end{aligned}$$

for $u = s(b-a)$. Setting

$$\phi(u) = -pu + \ln(1 - p + pe^{u}),$$

we thus have $\mathbb{E}\left[e^{sX}\right] \leq \exp(\phi(u))$. To prove the claim, we will show that $\phi(u) \leq u^2/8 = s^2(b-a)^2/8$.

To see that, expand $\phi(u)$ about zero using Taylor's expansion. We have

$$\phi(u) = \phi(0) + u\phi'(0) + \frac{1}{2}u^2\phi''(\theta) \tag{12.3}$$

where $\theta \in [0, u]$, and notice that $\phi(0) = 0$. Furthermore, we have

$$\phi'(u) = -p + \frac{pe^{u}}{1 - p + pe^{u}},$$

and as such $\phi'(0) = -p + \frac{p}{1-p+p} = 0$. Now,

$$\phi''(u) = \frac{(1 - p + pe^{u})pe^{u} - (pe^{u})^2}{(1 - p + pe^{u})^2} = \frac{(1-p)pe^{u}}{(1 - p + pe^{u})^2}.$$

For any $x, y \geq 0$, we have $(x+y)^2 \geq 4xy$ as this is equivalent to $(x-y)^2 \geq 0$. Setting $x = 1 - p$ and $y = pe^{u}$, we have that

$$\phi''(u) = \frac{(1-p)pe^{u}}{(1 - p + pe^{u})^2} \leq \frac{(1-p)pe^{u}}{4(1-p)pe^{u}} = \frac{1}{4}.$$

Plugging this into Eq. (12.3), we get that

$$\phi(u) \leq \frac{1}{8}u^2 = \frac{1}{8}(s(b-a))^2 \quad \text{and} \quad \mathbb{E}\left[e^{sX}\right] \leq \exp(\phi(u)) \leq \exp\left(\frac{1}{8}(s(b-a))^2\right),$$

as claimed. ■

Lemma 12.4.2. *Let X be a random variable. If $\mathbb{E}[X] = 0$ and $a \leq X \leq b$, then for any $s > 0$, we have*

$$\mathbb{P}[X > t] \leq \frac{\exp\left(\frac{s^2(b-a)^2}{8}\right)}{e^{st}}.$$

Proof: Using the same technique we used in proving Chernoff's inequality, we have that

$$\mathbb{P}[X > t] = \mathbb{P}\left[e^{sX} > e^{st}\right] \le \frac{\mathbb{E}\left[e^{sX}\right]}{e^{st}} \le \frac{\exp\left(\frac{s^2(b-a)^2}{8}\right)}{e^{st}}.$$ ■

Theorem 12.4.3 (Hoeffding's inequality). *Let $X_1, \ldots, X_n$ be independent random variables, where $X_i \in [a_i, b_i]$, for $i = 1, \ldots, n$. Then, for the random variable $S = X_1 + \cdots + X_n$ and any $\eta > 0$, we have*

$$\mathbb{P}\left[\left|S - \mathbb{E}[S]\right| \ge \eta\right] \le 2\exp\left(-\frac{2\,\eta^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Proof: Let $Z_i = X_i - \mathbb{E}[X_i]$, for $i = 1, \ldots, n$. Set $Z = \sum_{i=1}^n Z_i$, and observe that

$$\mathbb{P}[Z \ge \eta] = \mathbb{P}\left[e^{sZ} \ge e^{s\eta}\right] \le \frac{\mathbb{E}[\exp(sZ)]}{\exp(s\eta)},$$

by Markov's inequality. Arguing as in the proof of Chernoff's inequality, we have

$$\mathbb{E}[\exp(sZ)] = \mathbb{E}\left[\prod_{i=1}^n \exp(sZ_i)\right] = \prod_{i=1}^n \mathbb{E}[\exp(sZ_i)] \le \prod_{i=1}^n \exp\left(\frac{s^2(b_i - a_i)^2}{8}\right),$$

since the Z_is are independent and by Lemma 12.4.1. This implies that

$$\mathbb{P}[Z \ge \eta] \le \exp(-s\eta) \prod_{i=1}^n e^{s^2(b_i - a_i)^2/8} = \exp\left(\frac{s^2}{8}\sum_{i=1}^n (b_i - a_i)^2 - s\eta\right).$$

The upper bound is minimized for $s = 4\eta/\left(\sum_i (b_i - a_i)^2\right)$, implying

$$\mathbb{P}[Z \ge \eta] \le \exp\left(-\frac{2\eta^2}{\sum (b_i - a_i)^2}\right).$$

The claim now follows by the symmetry of the upper bound (i.e., apply the same proof to $-Z$). ■

12.5. Bibliographical notes

Some of the exposition here follows more or less the exposition in [MR95]. Exercise 12.6.1 (without the hint) is from [Mat99]. McDiarmid [McD89] provides a survey of Chernoff type inequalities, and Theorem 12.3.6 and Section 12.3 is taken from there (our proof has somewhat weaker constants).

A more general treatment of such inequalities and tools is provided by Dubhashi and Panconesi [DP09].

12.6. Exercises

Exercise 12.6.1 (Chernoff inequality is tight.). Let $S = \sum_{i=1}^{n} S_i$ be a sum of n independent random variables each attaining values $+1$ and -1 with equal probability. Let $P(n, \Delta) = \mathbb{P}[S > \Delta]$. Prove that for $\Delta \leq n/C$,

$$P(n, \Delta) \geq \frac{1}{C} \exp\left(-\frac{\Delta^2}{Cn}\right),$$

where C is a suitable constant. That is, the well-known Chernoff bound $P(n, \Delta) \leq \exp(-\Delta^2/2n))$ is close to the truth.

Exercise 12.6.2 (Chernoff inequality is tight by direct calculations.). For this question use only basic argumentation – do not use Stirling's formula, Chernoff inequality or any similar "heavy" machinery.

(A) Prove that $\displaystyle\sum_{i=0}^{n-k} \binom{2n}{i} \leq \frac{n}{4k^2} 2^{2n}$.

Hint: Consider flipping a coin $2n$ times. Write down explicitly the probability of this coin to have at most $n - k$ heads, and use Chebyshev inequality.

(B) Using (A), prove that $\binom{2n}{n} \geq 2^{2n}/4\sqrt{n}$ (which is a pretty good estimate).

(C) Prove that $\displaystyle\binom{2n}{n+i+1} = \left(1 - \frac{2i+1}{n+i+1}\right)\binom{2n}{n+i}$.

(D) Prove that $\displaystyle\binom{2n}{n+i} \leq \exp\left(\frac{-i(i-1)}{2n}\right)\binom{2n}{n}$.

(E) Prove that $\displaystyle\binom{2n}{n+i} \geq \exp\left(-\frac{8i^2}{n}\right)\binom{2n}{n}$.

(F) Using the above, prove that $\displaystyle\binom{2n}{n} \leq c\frac{2^{2n}}{\sqrt{n}}$ for some constant c (I got $c = 0.824...$ but any reasonable constant will do).

(G) Using the above, prove that

$$\sum_{i=t\sqrt{n}+1}^{(t+1)\sqrt{n}} \binom{2n}{n-i} \leq c 2^{2n} \exp\left(-t^2/2\right).$$

In particular, conclude that when flipping fair coin $2n$ times, the probability to get less than $n - t\sqrt{n}$ heads (for t an integer) is smaller than $c' \exp\left(-t^2/2\right)$, for some constant c'.

(H) Let X be the number of heads in $2n$ coin flips. Prove that for any integer $t > 0$ and any $\delta > 0$ sufficiently small, it holds that $\mathbb{P}[X < (1-\delta)n] \geq \exp\left(-c''\delta^2 n\right)$, where c'' is some constant. Namely, the Chernoff inequality is tight in the worst case.

Exercise 12.6.3 (Tail inequality for geometric variables). Let $X_1, \ldots, X_m$ be m independent random variables with geometric distribution with probability p (i.e., $\mathbb{P}[X_i = j] = (1-p)^{j-1}p$). Let $Y = \sum_i X_i$, and let $\mu = \mathbb{E}[Y] = m/p$. Prove that $\mathbb{P}\left[Y \geq (1+\delta)\mu\right] \leq \exp\left(-m\delta^2/8\right)$.

References

[DP09] D. P. Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[Kel56] J. L. Kelly. *A new interpretation of information rate*. *Bell Sys. Tech. J.*, 35(4): 917–926, 1956.

[Mat99] J. Matoušek. *Geometric discrepancy*. Vol. 18. Algorithms and Combinatorics. Springer, 1999.

[McD89] C. McDiarmid. *Surveys in Combinatorics*. Cambridge University Press, 1989. Chap. On the method of bounded differences.

[MR95] R. Motwani and P. Raghavan. *Randomized algorithms*. Cambridge, UK: Cambridge University Press, 1995.