# Chapter 2 Concentration of sums of independent random variables

Desh Raj

May 31, 2018

**Exercise 2.1.4.**

**Solution**    Given, $g \sim \mathcal{N}(0,1)$. On differentiating, we get $g'(t) = -tg(t)$.

$$
\begin{aligned}
\mathbb{E}\,X^2 \mathbb{1}_{X>t} &= \int_t^\infty x^2 g(x) dx \\
&= \int_t^\infty x(xg(x))dx \\
&= x\int_t^\infty xg(x)dx + \int_t^\infty g(x)dx && \text{(Integration by parts)}\\
&= -xg(x)\Big]_t^\infty + \mathbb{P}(X > t) \\
&= tg(t) + \mathbb{P}(X > t) && \text{(Taking limit of the first term)}
\end{aligned}
$$

This is the desired result.

**Exercise 2.2.3.**  Bounding the hyperbolic cosine

**Solution**    We have,

$$
\begin{aligned}
\cosh(x) &= \frac{e^x + e^{-x}}{2} \\
&= \frac{1}{2}\left( \sum_{i=0}^\infty \frac{x^i}{i!} + \sum_{i=0}^\infty \frac{(-x)^i}{i!} \right) && \text{(Using Taylor series expansion)}\\
&= \sum_{i=0}^\infty \frac{x^{2i}}{(2i)!} \geq \sum_{i=0}^\infty \frac{x^{2i}}{2^i i!}
\end{aligned}
$$

(Consider the denominator of the $i^{th}$ term. Each product term in $(2i)!$ is strictly greater than that in $2^i i!$.)

$$
= \exp\left(\frac{x^2}{2}\right). \qquad\qquad \text{(Again using Taylor series expansion)}
$$

**Exercise 2.2.7.**

**Solution**    Given, $X_1, \ldots, X_N$ are independent random variables such that $X_i \in [m_i, M_i]$. Let $S_N = X_1 + \ldots + X_N$. Then, for $t > 0$, we have

$$\mathbb{P}\{S_N - \mathbb{E}\, S_N \geq t\} = \mathbb{P}\{e^{s(S_N - \mathbb{E}\, S_N)} \geq e^{st}\} \qquad \text{(For some } s > 0)$$

$$\leq \frac{\mathbb{E}\, e^{s(S_N - \mathbb{E}\, S_N)}}{e^{st}} \qquad \text{(Using Markov's inequality)}$$

$$= e^{-st} \prod_{i=1}^{N} \mathbb{E}\left[e^{s(X_i - \mathbb{E}\, X_i)}\right]$$

$$\leq e^{-st} \prod_{i=1}^{N} \exp\left(\frac{s^2(M_i - m_i)^2}{8}\right) \qquad \text{(Using Hoeffding's lemma, since } \mathbb{E}(X_i - \mathbb{E}\, X_i) = 0)$$

$$= \exp\left(-st + \frac{s^2}{8}\sum_{i=1}^{N}(M_i - m_i)^2\right).$$

Let $g(s) = \exp\left(-st + \frac{s^2}{8}\sum_{i=1}^{N}(M_i - m_i)^2\right)$. $g(s)$ achieves its minimum at $s = \frac{4t}{\sum_{i=1}^{N}(M_i - m_i)^2}$. Substituting this value in the inequality, we get

$$\mathbb{P}\{S_N - \mathbb{E}\, S_N \geq t\} \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{N}(M_i - m_i)^2}\right).$$

**Exercise 2.2.8.**

**Solution**  Let $X_i$ denote the Bernoulli random variable indicating that the answer is wrong in the $i^{th}$ turn. Then $\mathbb{E}\, X_i = \frac{1}{2} - \delta$, and $X_i \in [0, 1]$. We have to bound the probability that the final answer is wrong with probability $\epsilon$, i.e., $\mathcal{P}(\text{majority of decisions are wrong}) \leq \epsilon$. From Hoeffding's inequality, we have

$$\mathcal{P}\left(\sum_{i=1}^{N}(X_i - \mathbb{E}\, X_i) \geq t\right) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^{N}(M_i - m_i)^2}\right)$$

$$\Rightarrow \mathcal{P}\left(\sum_{i=1}^{N}\left(X_i - \frac{1}{2} + \delta\right) \geq t\right) \leq \exp\left(\frac{-2t^2}{N}\right) \qquad \text{(Since } M_i = 1 \text{ and } m_i = 0)$$

To get a wrong answer finally, we require that at least half of all answers must be wrong, i.e.

$$\sum_{i=1}^{N} X_i \geq \frac{N}{2}$$

$$\Rightarrow \sum_{i=1}^{N}\left(X_i - \frac{1}{2} + \delta\right) \geq \delta N$$

Plugging $t = \delta N$ in earlier inequality, we get

$$\mathcal{P}\left(\sum_{i=1}^{N}\left(X_i - \frac{1}{2} + \delta\right) \geq \delta N\right) \leq \exp\left(-2N\delta^2\right)$$

For this probbility to be bounded by $\epsilon$, we require that

$$\exp(-2N\delta^2) \le \epsilon$$
$$\Rightarrow -2N\delta^2 \le \log(\epsilon) \qquad \text{(Taking log both sides)}$$
$$\Rightarrow N \ge \frac{\log(\epsilon^{-1})}{2\delta^2}.$$

**Exercise 2.2.9.**

**Solution   Part 1**

We want to bound the probability that the sample mean deviates from the true mean by a quantity greater than $\epsilon$. Using Chebyshev's inequality, we have

$$\mathcal{P}\left( \|\frac{1}{N}\sum_{i=1}^{N} X_i - \mu\| \ge \epsilon \right) \le \frac{\text{var}\left( \frac{1}{N}\sum_{i=1}^{N} X_i \right)}{\epsilon^2} \qquad (1)$$

Since the samples are drawn i.i.d, we have

$$\text{var}(X_1 + \ldots + X_N) = \text{var}(X_1) + \ldots + \text{var}(X_N)$$
$$\Rightarrow \text{var}(X_1 + \ldots + X_N) = N\sigma^2 \qquad \text{(Since var}(X_i)=\sigma^2)$$
$$\Rightarrow \frac{1}{N^2}\text{var}(X_1 + \ldots + X_N) = \frac{\sigma^2}{N} \qquad \text{(Dividing both sides by } \sigma^2)$$
$$\Rightarrow \text{var}\left( \frac{1}{N}\sum_{i=1}^{N} X_i \right) = \frac{\sigma^2}{N}$$

Plugging this value in 1, we get

$$\mathcal{P}\left( \|\frac{1}{N}\sum_{i=1}^{N} X_i - \mu\| \ge \epsilon \right) \le \frac{\sigma^2}{N\epsilon^2}$$

We want that an $\epsilon$-accurate estimate should be computed with probability at least $3/4$, i.e., the error probability should be at most $1/4$. Using this bound, we get $\frac{\sigma^2}{N\epsilon^2} \le \frac{1}{4}$, which implies $N = \mathcal{O}(\frac{\sigma^2}{\epsilon^2})$.

**Part 2**

Using Theorem 2.2.6 with $X_i \in [-\sigma, \sigma]$, we get

$$\mathcal{P}\left( \frac{1}{N}\sum_{i=1}^{N} X_i - \mu \ge \epsilon \right) = \mathcal{P}\left( \sum_{i=1}^{N}(X_i - \mu) \ge N\epsilon \right) \qquad \text{(Multiplying both sides by } N)$$
$$\le \exp\left( -\frac{2N^2\epsilon^2}{\sum_{i=1}^{N}(2\sigma)^2} \right) \qquad \text{(From Theorem 2.2.6)}$$
$$= \exp\left( -\frac{2N^2\epsilon^2}{4N\sigma^2} \right)$$
$$= \exp\left( -\frac{N\epsilon^2}{2\sigma^2} \right)$$

We want that an $\epsilon$-accurate estimate should be computed with probability at least $1 - \delta$, i.e., the error probability should be at most $\delta$. This means that

$$\exp\left(-\frac{N\epsilon^2}{2\sigma^2}\right) \leq \delta$$

$$\Rightarrow -\frac{N\epsilon^2}{2\sigma^2} \leq \log(\delta) \qquad \text{(Taking log both sides)}$$

$$\Rightarrow N = \mathcal{O}(\log(\delta^{-1})\frac{\sigma^2}{\epsilon^2}),$$

which is the desired result.

**Exercise 2.2.10.**

**Solution   Part 1**

Let $f_{X_i}$ denote the probability density function. Since $X_i$ is non-negative, it's MGF is given as

$$M_{X_i}(-t) = \int_0^\infty e^{-tx} f_{X_i}(x)dx$$

Since $e^{-tx} \geq 0$, $e^{-tx}f_{X_i} \leq e^{-tx}\max\|f_{X_i}\| = e^{-tx}$ (since $\max\|f_{X_i}\| = 1$). Using this in above equation, we get

$$M_{X_i}(-t) \leq \int_0^\infty e^{-tx}dx = \frac{1}{t}$$

**Part 2**

We have

$$\mathcal{P}\left(\sum_{i=1}^N X_i \leq \epsilon N\right) = \mathcal{P}\left(\sum_{i=1}^N -X_i \geq -\epsilon N\right)$$

$$= \mathcal{P}\left(\sum_{i=1}^N -\frac{X_i}{\epsilon} \geq -N\right)$$

$$= \mathcal{P}\left(\exp(\sum_{i=1}^N -\frac{X_i}{\epsilon} \geq \exp(-N)\right) \qquad \text{(Since exp is monotonously increasing)}$$

$$\leq e^N \mathbb{E}\exp\left(\sum_{i=1}^N -\frac{X_i}{\epsilon}\right) \qquad \text{(Using Markov inequality)}$$

$$= e^n \prod_{i=1}^N \mathbb{E}\exp\left(-\frac{X_i}{\epsilon}\right)$$

$$\leq e^N \prod_{i=1}^N \epsilon \qquad \text{(Using result in part 1)}$$

$$= (e\epsilon)^N.$$

4

**Exercise 2.3.2.**

**Solution**   Since $t < \mu$, so the function $f(x) = \left(\frac{t}{\mu}\right)^x$ is monotonically decreasing. So we can write

$$
\begin{aligned}
\mathcal{P}(S_N \leq t) &= \mathcal{P}\left(\left(\frac{t}{\mu}\right)^{S_N} \geq \left(\frac{t}{\mu}\right)^t\right) \\
&\leq \frac{\mathbb{E}\left(\frac{t}{\mu}\right)^{S_N}}{\left(\frac{t}{\mu}\right)^t} && \text{(By Markov's inequality)}
\end{aligned}
$$

We have

$$
\begin{aligned}
\mathbb{E}\,\alpha^{S_N} = \mathbb{E}\,\alpha^{\sum_{i=1}^{N} X_i} &= \mathbb{E}\prod_{i=1}^{N}\alpha^{X_i} \\
&= \prod_{i=1}^{N}\left[p_i\alpha^1 + (1-p_i)\alpha^0\right] \\
&\quad \text{(Since } X_i \text{ can only take values 1 and 0 with probability } p_i \text{ and } 1-p_i\text{)} \\
&= \prod_{i=1}^{N}(1 + (\alpha-1)p_i) \\
&\leq \prod_{i=1}^{N}\exp((\alpha-1)p_i) && \text{(Using } 1+x \leq e^x\text{)} \\
&= \exp\left(\sum_{i=1}^{N}((\alpha-1)p_i)\right) \\
&= \exp[(\alpha-1)\mu]
\end{aligned}
$$

Using this inequality in above, we get

$$
\begin{aligned}
\mathcal{P}(S_N \leq t) &\leq \exp(t-\mu)\left(\frac{\mu}{t}\right)^t \\
&= e^{-\mu}\left(\frac{e\mu}{t}\right)^t,
\end{aligned}
$$

which is the desired result.

**Exercise 2.3.3.**

**Solution**   Since $X \sim \mathrm{Pois}(\lambda)$, by Poisson limit theorem, we can say that $X$ is the sum of $N$ Bernoulli random variables $X_1, \ldots, X_N$, for some large $N$, where $\mathbb{E}\,X_i = \lambda$. As such, we can use the Chernoff concentration bound for the sum of Bernoulli random variables to bound $X$, hence giving the desired result.

**Exercise 2.3.5.**

**Solution**    First, using $t = (1 + \delta)\mu$ in Chernoff bounds for upper tails, we get

$$\mathcal{P}\{X - \mu \geq \delta\mu\} \leq e^{-\mu} \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu}. \tag{2}$$

Now, using $t = (1 - \delta)\mu$ in Chernoff bounds for lower tails (proved in previous problem), we get

$$\mathcal{P}\{X - \mu \leq -\delta\mu\} \leq e^{-\mu} \left(\frac{e}{1-\delta}\right)^{(1-\delta)\mu}. \tag{3}$$

Adding 2 and 3, we get

$$\mathcal{P}\{\|X - \mu\| \geq \delta\mu\} \leq e^{-\mu} \left[ \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} + \left(\frac{e}{1-\delta}\right)^{(1-\delta)\mu} \right]$$

We can bound the terms inside the bracket on the RHS as follows.

$$
\begin{aligned}
\left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu} &= \exp\left[(1+\delta)\mu \log\left(\frac{e}{1+\delta}\right)\right] \\
&= \exp\left[\mu(1+\delta)(1 - \log(1+\delta))\right] \\
&\leq \exp\left[\mu(1+\delta)(1 - \frac{\delta}{1+\delta/2})\right] && \text{(Since } \log(1+x) \geq \tfrac{x}{1+x/2}) \\
&= \exp\left[\mu(1+\delta)\frac{1 - \frac{\delta}{2}}{1 + \frac{\delta}{2}}\right] \\
&= \exp\left[\mu\left(1 - \frac{\delta^2}{2+\delta}\right)\right] \\
&\leq \exp\left[\mu\left(1 - \frac{\delta^2}{3}\right)\right] && \text{(Since } 2 + \delta \leq 3)
\end{aligned}
$$

and

$$
\begin{aligned}
\left(\frac{e}{1-\delta}\right)^{(1-\delta)\mu} &= \exp\left[(1-\delta)\mu \log\left(\frac{e}{1-\delta}\right)\right] \\
&= \exp\left[\mu(1-\delta)(1 - \log(1-\delta))\right] \\
&\leq \exp\left(\mu(1-\delta)(1 - \frac{-\delta}{\sqrt{1-\delta}})\right) && \text{(Since } \log(1-x) \geq \tfrac{-x}{\sqrt{1-x}}) \\
&= \exp\left[\mu(1 - \delta + \delta\sqrt{1-\delta})\right] \\
&\leq \exp\left[\mu\left(1 - \frac{\delta^2}{3}\right)\right] && \text{(Since } \sqrt{1-\delta} \leq 1 - \tfrac{\delta}{2})
\end{aligned}
$$

Using these bounds, we get

$$
\begin{aligned}
\mathcal{P}\{\|X - \mu\| \geq \delta\mu\} &\leq 2\exp\left[-\mu\left(1 - \frac{\delta^2}{3}\right)\right] \\
&\leq 2e^{-c\mu\delta^2}.
\end{aligned}
$$

**Exercise 2.3.6.**

**Solution** Since $X \sim \text{Pois}(\lambda)$, by Poisson limit theorem, we can say that $X$ is the sum of $N$ Bernoulli random variables $X_1, \ldots, X_N$, for some large $N$, where $\mathbb{E} X_i = \lambda$. Using the result in Exercise 2.3.5, we get

$$\mathcal{P}\{\|X - \lambda\| \geq \delta\lambda\} \leq 2\exp(-c\lambda\delta^2). \tag{4}$$

Now substituting $t = \delta\lambda$ in 4, we get the desired result.

**Exercise 2.3.8.**

**Solution** We will first show that the sum of independent Poisson distributions is a Poisson distribution. Let $X$ and $Y$ be Poisson distributions with means $\lambda_1$ and $\lambda_2$, respectively. Their MGFs are given as $M_X(s) = e^{-\lambda_1(1-s)}$ and $M_Y(s) = e^{-\lambda_2(1-s)}$. Then, we have

$$M_{X+Y}(s) = M_X(s)M_Y(s) = e^{-(\lambda_1+\lambda_2)(1-s)}.$$

Hence, $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$.

It is given that $X \sim \text{Pois}(\lambda)$. Therefore, $X$ can be considered as a sum of $\lambda$ Poisson distributions with mean 1, using the above result. Now consider the distribution $Z = \frac{X-\lambda}{\sqrt{\lambda}}$. We can directly apply the Central Limit Theorem to $Z$ to obtain the desired result, since $\mathbb{E} X = \lambda$ and $\text{var}(X) = \lambda$.

**Exercise 2.4.2.**

**Solution** Consider a fixed vertex $i$. Using Markov inequality, we can bound the probability that its degree $d_i$ is larger than $\mathcal{O}(\log n)$, as

$$\mathcal{P}\{d_i \geq \mathcal{O}(\log n)\} \leq \frac{\mathbb{E} d_i}{\mathcal{O}(\log n)} = c$$

We can now use the union bound to bound the probability that such a vertex exists as follows

$$\mathcal{P}\{\exists i : d_i \geq \mathcal{O}(\log n)\} = \sum_{i=1}^{n} \mathcal{P}\{d_i \geq \mathcal{O}(\log n)\} \leq cn$$

We can choose a constant $c$ such that this probability is very low, and therefore, the probability of its counterpart is very high.

**Exercise 2.4.3.**

**Solution** Similar to the previous exercise, we can show that if the expected degree is $d = \mathcal{O}(1)$, then with a high probability, all vertices of $G$ have degrees $\mathcal{O}(1)$. Furthermore, we know from asymptotic notations that $\mathcal{O}(1) \leq \mathcal{O}\left(\frac{\log n}{\log \log n}\right)$. Hence, the result is satisfied.

**Exercise 2.4.4.**

**Solution**

**Exercise 2.4.5.**

**Solution**