1 Concentration Inequalities

**Theorem 1.1.** (Markov’ inequality, [Dur96]) Let \( X \) be a nonnegative random variable, then we have
\[
P[X \geq tE[X]] \leq \frac{1}{t}
\]
for any \( t > 0 \).

**Proof.** By definition, we have
\[
E[X] = \int_0^\infty xp(x)dx = \int_0^{tE[X]} xp(x)dx + \int_{tE[X]}^\infty xp(x)dx
\]
\[
\geq \int_0^{tE[X]} 0p(x)dx + \int_{tE[X]}^\infty tE[X]p(x)dx
\]
\[
= tE[X] \int_{tE[X]}^\infty p(x)dx := tE[X]P[X \geq tE[X]].
\]
Dividing both side by \( tE[X] \), then we arrive at \( P[X \geq tE[X]] \leq \frac{1}{t} \). \( \square \)

**Theorem 1.2.** (Chebyshev’s inequality, [Dur96]) Let \( X \) be a random variable, then we have
\[
P[|X - E[X]| \geq t\sqrt{\text{Var}[X]}] \leq \frac{1}{t^2}
\]
for any \( t > 0 \).

**Proof.** Define a new random variable \( Y = (X - E[X])^2 \), thus \( E[Y] = \text{Var}[X] \). By Markov’ inequality, we have
\[
P[Y \geq t^2E[Y]] \leq \frac{1}{t^2}.
\]
Taking square root of \( Y \), then we arrive at \( P[|X - E[X]| \geq t\sqrt{\text{Var}[X]}] \leq \frac{1}{t^2} \). \( \square \)

**Corollary 1.3.** (One side Chebyshev’s inequality) Let \( X \) be a random variable, then we have
\[
P[X \geq E[X] + t\sqrt{\text{Var}[X]}] \leq \frac{1}{1+t^2}
\]
for any \( t > 0 \).

2 Convergence to Gaussian

A random variable with Gaussian distribution or normal distribution \( N(0,1) \) has probability density function as
\[
p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.
\]
Inductively, the probability density function of a multivariate Gaussian distribution \( N(0, I_n) \) (\( I_n \) is an \( n \times n \) identity matrix) is defined as

\[
p(x) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{\|x\|^2}{2}}.
\]

There are three properties regarding Gaussian random variables:

Property 1. \( p(x) \) is rotationally invariant;

Property 2. \( p(x) \) can be viewed as a product of \( n \) single-variate Gaussian distribution probability density functions;

Property 3. Consider two independent Gaussian random variable \( X_1 \sim N(0, 1), X_2 \sim N(0, 1) \), then \( ax_1 + bx_2 \sim N(0, a^2 + b^2) \).

**Theorem 2.1.** (Central Limit Theorem, [Dur96]) Let \( X_1, \ldots, X_n \) be i.i.d. random variables whose summation is \( Y_n \) (i.e., \( Y_n = \sum_{i=1}^n X_i \)), then if \( \text{var}[X_1] \in (0, \infty) \), \( \frac{Y_n}{n} \) converges in distribution to \( N(\mathbb{E}[X_1], \frac{1}{n} \text{Var}[X_1]) \).

**Theorem 2.2.** (Berry-Esseen Theorem, [Ber41]) Let \( X_1, \ldots, X_n \) be independent random variables whose summation is \( Y_n \) (i.e., \( Y_n = \sum_{i=1}^n X_i \)) and \( Z_n \sim N(\mathbb{E}[Y_n], \text{Var}[Y_n]) \), then for any \( t \),

\[
|P(Y_n \leq t) - P(Z_n \leq t)| \leq C \frac{\sum_{i=1}^n \mathbb{E}[|X_i|^3]}{[\text{Var}[Y_n]]^{3/2}},
\]

where \( C \in (0, 1) \) is a constant.

Please note that in [1], we can replace \( \leq \) by \( > \) since \( P(Y_n \leq t) = 1 - P(Y_n > t) \) and \( P(Z_n \leq t) = 1 - P(Z_n > t) \).

**Example 2.3.** Consider \( X_1, \ldots, X_n \) be i.i.d. random variables where for each \( i \),

\[
X_i = \begin{cases} 
1, & \text{with probability } 1/2 \\
-1, & \text{with probability } 1/2 
\end{cases}
\]

then we have \( \mathbb{E}[Y_n] = 0, \text{Var}[Y_n] = n \). Thus, [1] yields

\[
|P(Y_n \geq t) - P(Z_n \geq t)| \leq C \frac{\sum_{i=1}^n \mathbb{E}[|X_i|^3]}{[\text{Var}[Y_n]]^{3/2}} := C n^{3/2} = C \frac{n}{n^{1/2}}
\]

which will go to zero if \( n \to \infty \).

2.1 Bounding normal distribution

Let \( Z \sim N(0, 1) \). Let us bound the error function of normal distribution defined as for each \( t > 0 \)

\[
\phi(t) := P(Z > t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-\frac{x^2}{2}} dx.
\]

The next lemma shows \( \phi(t) \) can be tightly approximated.
Lemma 2.4. $\phi_{LB}(t) \leq \phi(t) \leq \phi_{UB}(t)$ for any $t > 0$, where $\phi_{UB}(t) := \frac{1}{\sqrt{2\pi} t} \left(1 - \frac{1}{t+1}\right)e^{-\frac{t^2}{2}}$, $\phi_{LB}(t) := \frac{1}{\sqrt{2\pi} t} e^{-\frac{t^2}{2}}$.

Proof.

(1) First of all, to upper bound $\phi(t)$, note that for any $x \geq t$, we have $x/t \geq 1$, thus

$$\phi(t) \leq \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} \frac{x}{t} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi} t} e^{-\frac{t^2}{2}} = \phi_{UB}(t).$$

(2) Now we claim that $\phi_{LB}(t) = \frac{1}{\sqrt{2\pi} t} \left(1 - \frac{1}{t+1}\right)e^{-\frac{t^2}{2}}$ is a lower bound of $\phi(t)$. To prove it, let us define a new function

$$g(t) = \phi(t) - \phi_{LB}(t) = \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} e^{-\frac{x^2}{2}} dx - \frac{1}{\sqrt{2\pi} t} \left(1 - \frac{1}{t^2+1}\right) e^{-\frac{t^2}{2}},$$

whose first derivative is

$$g'(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \left[-1 + \frac{t^2-1}{(t^2+1)^2} + \frac{t^2}{t^2+1}\right] = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \frac{-2t^2}{(t^2+1)^2} < 0.$$

$g(t)$ is monotone decreasing, thus $g(t) \geq \lim_{x \to \infty} g(x) = 0$, i.e. $\phi(t) \geq \phi_{LB}(t)$.

Thus, $\phi(t) \approx \frac{1}{\sqrt{2\pi} t} e^{-\frac{t^2}{2}}$. □

In Example 2.3, we have

$$P(Y_n \geq t\sqrt{n}) \approx P(Z_n \geq t\sqrt{n}) \leq \frac{1}{t\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

which would go to zero if $t = \omega(n)$ such that $\lim_{n \to \infty} \omega(n) = \infty$.

3 Routing with minimum congestion

Given a non-directed graph $G(V, E)$, let us define Congestion($(i, j)$) as number of paths using edge $(i, j) \in E$. Then in this problem, we are trying to find $k$ paths $(s_1, t_1), \ldots, (s_k, t_k)$ with $s_i, t_i \in V$ for all $i$, such that the largest congestion across all the edges are minimized. It turns out that such problem is NP-hard even with $k = 2$.

We now consider relaxing the path into flow (i.e., we are sending one unit of flow from $s_i$ to $t_i$ for each $i$), which is polynomially solvable. The following example illustrates how the relaxation works.

Example 3.1. Consider a graph with four nodes $(s_1, s_2, t_1, t_2)$ and four edges $(s_1, t_2), (t_1, t_2), (s_2, t_1), (s_1, s_2)$ (see Figure 1). In this example, if we solve the original problem the optimal value is 2, because the paths of $(s_1, t_1)$ and $(s_2, t_2)$ share at least one edge. However, if we solve the relaxed one, the optimal value is 1 since for $(s_1, t_1)$ flow, we can send half flow through $s_1 - t_2 - t_1$ and another half flow through $s_1 - s_2 - t_1$, and the same strategy applies to $(s_2, t_2)$ flow. ♦
Now, for each \( l \in \{1, \ldots, k\} \), let us define a set of flow variables \( \{x^l_{ij}\}_{(i,j) \in E} \) which forms a flow with value 1 between \( s_l \) and \( t_l \). Clearly, Congestion((i, j)) can be expressed as a function of \( \{x^l_{ij}\}_{(i,j) \in E} \), i.e., Congestion((i, j)) = \( \sum_{l=1}^{k} x^l_{ij} \). Denote \( \alpha \) to be the maximum congestion across the edges. Thus, our relaxed problem is formulated as follows:

\[
\begin{align*}
\min \quad & \alpha \\
\text{s.t.} \quad & \sum_{l=1}^{k} x^l_{ij} \leq \alpha, \forall (i, j) \in E \\
& \{x^l_{ij}\}_{(i,j) \in E} \text{ forms a flow with value 1 between } s_l \text{ and } t_l \\
& 0 \leq x^l_{ij} \leq 1, \forall (i, j) \in E, l \in l \in \{1, \ldots, k\}
\end{align*}
\]

which can be solved as a linear program (LP). Let \( \{x^*_{ij}\}_{(i,j) \in E} \), \( \alpha^* \) be the optimal solution.

Now we are ready to prove main result in this section.

**Theorem 3.2.** With probability at least \( \frac{1}{2} \), there exists a routing in \( G(V, E) \) whose maximum congestion is no larger than \( \max((2e - 1)\alpha^*, 2 \ln n) \leq (2e - 1)\alpha^* + 2 \ln n \).

**Proof.**

(1) First of all, for each \( l \in \{1, \ldots, k\} \), we can decompose the optimal flow between \( s_l \) and \( t_l \) by path. This can be done iteratively: at iteration \( \tau \), suppose we find a path \( P_\tau(s_l, t_l) \) from \( s_l \) to \( t_l \), let \( f_\tau \) be the minimum edge flow on this path, then reduce the flow of each edge on this path by \( f_\tau \); go to next iteration until we cannot find any path with positive flow. This procedure takes at most \( |E| \) iterations as at each iteration at least one edge is removed. Hence, after flow decomposition, we arrive at the following equality

\[
\text{flow}(s_l, t_l) = \sum_{\tau} f^\tau_{ij} P_\tau(s_l, t_l), \forall l \in \{1, \ldots, k\}
\]

and \( \sum_{\tau} f^\tau_{ij} = 1 \) with \( f^\tau_{ij} \in [0, 1] \) for all \( \tau \). We see that for each \( l \in \{1, \ldots, k\} \), \( \{f^\tau_{ij}\}_\tau \) can be viewed as a probability distribution of paths \( \{P_\tau(s_l, t_l)\}_\tau \). This inspires us to define independent random variables \( \{Y^l_i\}_i \) for each edge \((i, j)\) as

\[
Y^l_i = \begin{cases} 
1, & \text{if } (i, j) \text{ is used for } (s_l, t_l) \text{ with probability } p_l \\
0, & \text{otherwise}
\end{cases}
\]

Then, Congestion((i, j)) := \( Y = \sum_{l=1}^{k} Y^l_i \) and let

\[
\mu := E[Y] = \sum_{l=1}^{k} E[Y^l_i] = \sum_{l=1}^{k} p_l := \sum_{l=1}^{k} x^*_{ij} \leq \alpha^*.
\]
(2) Now we would like to bound $Y$. Given $\delta, t > 0$, then we have

$$P[Y \geq (1 + \delta)\mu] = P[e^{tY} \geq e^{(1 + \delta)\mu}]$$

(e$^x$ is monotone increasing)

$$\leq \frac{E[e^{tY}]}{e^{(1 + \delta)\mu}}$$

(Markov’s inequality)

$$= \frac{E[e^{t\sum_{i=1}^{k} Y_i}]}{e^{(1 + \delta)\mu}}$$

($Y = \sum_{i=1}^{k} Y_i$)

$$= \prod_{i=1}^{k} E[e^{tY_i}]/e^{(1 + \delta)\mu}$$

(indipendence of $\{Y_i\}$)

$$\leq \prod_{i=1}^{k} \left(1 - p_i + p_i e^t\right) e^{t\mu} = \prod_{i=1}^{k} \left[1 + p_i(e^t - 1)\right]/e^{(1 + \delta)\mu}$$

(definition of $\{Y_i\}$)

$$\leq \prod_{i=1}^{k} p_i(e^t - 1)$$

($\mu = \sum_{i=1}^{k} p_i$).

Note that by minimizing $e^t - 1 - t(1 + \delta)$ over $t > 0$, the minimizer is $t^* := \ln(1 + \delta)$. Substitute $t = \ln(1 + \delta)$, the above inequality yields

$$P[Y \geq (1 + \delta)\mu] \leq \left[\frac{e^{\ln(1 + \delta) - 1}}{2^{(1 + \delta)/\mu}}\right]\mu$$

($t = \ln(1 + \delta)$)

$$\leq \frac{1}{2^{-(1 + \delta)/\mu}}$$

(if $\delta \geq 2e - 1$)

$$\leq \frac{1}{n^2}$$

(if $(1 + \delta)\mu \geq 2 \ln n$).

As there are at most $n^2$ edges, thus according to union bound, the probability that the maximum congestion is no larger than $\max((2e - 1)\alpha^*, 2 \ln n) \leq (2e - 1)\alpha^* + 2 \ln n$ is at least $\frac{1}{2}$.

This example arises an important concentration bound.

**Theorem 3.3.** (Chernoff bound, [Che52]) Let $X_1, \ldots, X_n \in [0, 1]$ be independent random variables, and $X := \sum_{i=1}^{n} X_i$, then we have $P[X > (1 + \delta)E[X]] \leq e^{-\frac{\delta^2}{2}E[X]}$ for any $\delta > 0$.

Moreover, if $\delta \in (0, 1)$, we have $P[X < (1 - \delta)E[X]] \leq e^{-\frac{\delta^2}{2}E[X]}$

A more general inequality is

**Theorem 3.4.** (Hoeffding inequality, [Hoe63]) Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in [a_i, b_i]$, and $X := \sum_{i=1}^{n} X_i$, then we have $P[X > E[X] + t] \leq e^{-\frac{2t^2}{\sum_{i=1}^{n}(a_i - b_i)^2}}$ and $P[X < E[X] - t] \leq e^{-\frac{2t^2}{\sum_{i=1}^{n}(a_i - b_i)^2}}$ for any $t > 0$. 

5
4 Johnson-Lindenstrauss Lemma

In this section, let us consider $m$ vectors $u_1, \ldots, u_m \in \mathbb{R}^n$ and we would like to demonstrate that for any $\epsilon \in (0, 1)$, there exists an integer $k$ and $v_1, \ldots, v_m \in \mathbb{R}^k$ such that $(1 - \epsilon)||u_i - u_j||_2^2 \leq ||v_i - v_j||_2^2 \leq (1 + \epsilon)||u_i - u_j||_2^2$ for all $i, j$. Johnson-Lindenstrauss Lemma tells us that $k = O(\log(m)/\epsilon^2)$ suffices.

First of all, let us define random projection as follows. Let $R \in \mathbb{R}^{n \times k}$ be an entry-wise independent Gaussian random matrix, where $R_{ij} \sim N(0, 1)$, then the linear projection of $u_i$ is defined as $v_i = \frac{1}{\sqrt{k}}R^7 u_i$ for each $i$. Now we are ready to show Johnson-Lindenstrauss Lemma.

**Theorem 4.1.** (*Johnson-Lindenstrauss Lemma, [JL84]*) Given $\epsilon \in (0, 1)$, let $u \in \mathbb{R}^n$ and $v = \frac{1}{\sqrt{k}}R^7 u$, then

$$P[||u||_2^2 - ||v||_2^2 > \epsilon||u||_2^2] \leq 2e^{-\frac{1}{8}(\epsilon^2 - 2\epsilon^3/3)}.$$  \hspace{1cm} (2)

In particular, we choose $k = O(\log(m)/\epsilon^2)$, then with probability at least $\frac{9}{10}, (1 - \epsilon)||u_i - u_j||_2^2 \leq ||v_i - v_j||_2^2 \leq (1 + \epsilon)||u_i - u_j||_2^2$ for all $i, j$.

**Proof.** First of, if $u = 0$, then (2) holds trivially. Now we assume that $u \neq 0$, thus without loss of generality, we can normalize it as $||u||_2 = 1$.

Note that

$$k||v||_2^2 = \sum_{l=1}^k (R_{il}u)^2 = \sum_{l=1}^k Y_i^2 = Y.$$  \hspace{1cm}

We also have

$$E[Y] = kE[||v||_2^2] = E[u^T RR^T u] = u^T E[RR^T]u = k$$

where the last equality is due to $||u||_2^2 = 1$ and

$$E[RR^T]_{ij} = \begin{cases} \sum_{l=1}^k E[R_{il}R_{jl}] = \sum_{l=1}^k E[R_{il}^2] = k, & \text{if } i = j \\ \sum_{l=1}^k E[R_{il}R_{jl}] = \sum_{l=1}^k E[R_{il}E[R_{jl}]] = 0, & \text{otherwise} \end{cases}.$$  \hspace{1cm}

Thus, (2) is equivalent to show

$$P[|Y - E[Y]| > \epsilon E[Y]] \leq 2e^{-\frac{1}{8}(\epsilon^2 - 2\epsilon^3/3)}.$$  \hspace{1cm} (3)

Indeed, if (3) holds, then choose $k$ such that $2e^{(-\epsilon^2 + \epsilon^3/3)k} < \frac{1}{10m^2}$, i.e. $k = O(\log(m)/\epsilon^2)$, and according to union bound, the probability that $(1 - \epsilon)||u_i - u_j||_2^2 \leq ||v_i - v_j||_2^2 \leq (1 + \epsilon)||u_i - u_j||_2^2$ for all $i, j$ is at least $\frac{9}{10}$.

We will prove (3) in the next lemma. \hfill \Box

**Lemma 4.2.** Suppose $Y_1, \ldots, Y_k$ are i.i.d. $N(0, 1)$ random variables and $Y := \sum_{l=1}^k Y_l^2$, then (3) holds.
Proof. Note that

\[ P[|Y - E[Y]| > \epsilon E[Y]] \leq P[Y > (1 + \epsilon)E[Y]] + P[Y < (1 - \epsilon)E[Y]]. \]

We will first bound \( P[Y > (1 + \epsilon)E[Y]] \). Indeed, for any \( t > 0 \),

\[ P[Y > (1 + \epsilon)E[Y]] = P[e^{tY} > e^{t(1+\epsilon)k}] \quad (E[Y] = k) \]

\[ \leq \frac{E[e^{tY}]}{e^{t(1+\epsilon)k}} \quad (\text{Markov's inequality}) \]

\[ = \prod_{l=1}^{k} \frac{E[e^{tY^2}]}{e^{t(1+\epsilon)k}} \quad (Y = \sum_{l=1}^{k} Y^2_l \text{ and by independence of } \{Y_i\}) \]

\[ = \left[ \frac{1}{\sqrt{1 - 2t e^{t(1+\epsilon)}}} \right]^k \quad (Y_i \sim N(0,1) \text{ and } E[e^{tY^2}] = \frac{1}{\sqrt{1 - 2t}} \text{ by Claim 4.3}) \]

Claim 4.3. Let \( X \sim N(0,1) \), then \( E[e^{tX^2}] = \frac{1}{\sqrt{1 - 2t}} \) for any \( t \in (0, \frac{1}{2}) \).

Proof. It can be shown by direct calculation

\[ E[e^{tX^2}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx^2} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{1 - 2t}} \left[ \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1 - 2t}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{1 - 2t}} dx \right] \]

\[ = \frac{1}{\sqrt{1 - 2t}} \]

where the last inequality is because \( \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1 - 2t}} e^{-\frac{x^2}{1 - 2t}} \) is the density function of Gaussian random variable \( N(0, \frac{1}{1 - 2t}) \). \( \square \)

We would like to maximize \((1 - 2t)e^{2t(1+\epsilon)}\) over \( t > 0 \), whose maximizer is \( t^* = \frac{\epsilon}{2(1+\epsilon)} \). Thus, \( P[Y > (1 + \epsilon)E[Y]] \) can be further bounded as

\[ P[Y > (1 + \epsilon)E[Y]] \leq \left[ \frac{1}{(1 - 2t)e^{2t(1+\epsilon)}} \right]^\frac{1}{2} = \left[ \frac{1 + \epsilon}{e^\epsilon} \right]^\frac{1}{2} \quad (\text{set } t = \frac{\epsilon}{2(1+\epsilon)}) \]

\[ \leq e^{\left( -\frac{\epsilon}{2} + \frac{\epsilon^2}{2} \right)} \quad (\text{set } 1 + \epsilon \leq e^{-\frac{\epsilon^2}{2} + \frac{\epsilon^2}{2}}, \forall \epsilon \in (0,1)). \]

By the similar reasoning, we can also bound

\[ P[Y < (1 - \epsilon)E[Y]] \leq e^{(-\epsilon^2 + 2\epsilon^3/3)} \]

Hence, (3) holds. \( \square \)
References


