Concentration inequalities and applications

董海辰 518030910417

June 30, 2020

Contents

Introduction

In this article, I summarized some of the famous concentration inequalities with some proofs by myself. I also research for the origins and some interesting stories of these inequalities, like mis-naming. Then I give a coin flipping example to show the power of these inequalities together with intuitive figures. Finally I share and rearrange what I've read in a paper in mechanism design, in which uses the inequalities widely.

1 Concentration inequalities

In this section, I will make a brief summary of some concentration inequalities, and focus most on how sum of independent random variables can deviates from some value, especially the sum of their expectations.

1: Problem

Let $X_1, X_2, \cdots, X_n$ be independent random variables. Let $X = \sum_{i=1}^{n} X$. How does X deviate from $E(X)$?

Besides the inequalities themselves, I would also like to explore the interesting history and stories behind them.

1.1 Markov's inequality, Chebyshev's inequality and Chernoff bound

2: Markov's inequality

If X is a non-negative random variable and $a > 0$, then

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

Proof.

$$E[X] \geq E[X \cdot 1_{X \geq a}] \geq P[X \geq a] \cdot a.$$

□

It is named after the Russian mathematician Andrey Markov, although it appeared earlier in the work of Pafnuty Chebyshev (Markov's teacher).

3: Chebyshev's inequality

If X has finite expected value and variance, then for every constant $a > 0$,

$$P[|X - E[X]| \geq a] \leq \frac{Var[X]}{a^2}.$$

Proof.

$$\begin{aligned} Var[X] &= E[(X-\mu)^2] \\ &\geq E[(X-\mu)^2 \text{ when } |X-\mu| > a\ ] \\ &\geq a^2 \cdot P[|X-\mu| > a]. \end{aligned}$$

□

According to [Wikipedia, 2020b], the theorem is named after Russian mathematician Pafnuty Chebyshev, although it was first formulated by his friend and colleague Irénée-Jules Bienaymé. The theorem was first stated without proof by Bienaymé in 1853 and later proved by Chebyshev in 1867. His student Andrey Markov provided another proof in his 1884 Ph.D. thesis.

By slightly modify 2: Markov's inequality, changing $X \geq t$ into $\exp(X\lambda) \geq \exp(t\lambda)$ for some $\lambda > 0$, we can sometimes get a tighter bound.

4: Chernoff bound

Let X be an random variable with $E[\exp(tX)]$ exists, then

$$P[X \geq t] = P[\exp(\lambda X) \geq \exp(\lambda t)] \leq \min_{\lambda>0} \frac{E[\exp(\lambda X)]}{\exp(t\lambda)}.$$

Proof. 2: Markov's inequality here works for every $\lambda > 0$ and we can choose different λ in different situations. □

Chernoff publishes this result in [Chernoff et al., 1952]. He showed in his Theorem 1:

> S_n is the sum of n independent observations $X_1, X_2, \cdots, X_n$ on a chance variable X with memonet generating function $M(t) = E(\exp(tX))$ and cumulative distribution function $F(x) = P(X \leq x)$. Let
>
> $$m(a) = \inf E(\exp(t(X-a))) = \inf \exp(-at)M(t).$$
>
> If $E(x) < \infty$ and $x \geq E(X)$, then
>
> $$P(S_n \geq na) \leq [m(a)]^n.$$

One more thing quite interesting is that in [Chernoff, 2014], Chernoff admitted that it is actually Herman Rubin who produced this result.

> In working on an artificial example, I discovered that I was using the Central Limit Theorem for large deviations where it did not apply. This led me to derive the asymptotic upper and lower bounds that were needed for the tail probabilities. Rubin claimed he could get these bounds with much less work and I challenged him. He produced a rather simple argument, using the Markov inequality, for the upper bound. Since that seemed to be a minor lemma in the ensuing paper I published (Chernoff, 1952), I neglected to give him credit. I now consider it a serious error in judgment, especially because his result is stronger, for the upper bound, than the asymptotic result I had derived. $\cdots\cdots$ My paper permitted him to modify his results and led to a great deal of publicity in the computer science literature for the so-called Chernoff bound which was really Rubin' s result.

For example, if we would like to use 4: Chernoff bound to find $P[X \geq (1+\delta)\mu]$ where $\mu = E[X]$. $X_i = 1$ with probability p_i and otherwise 0. We can do the following construction.

Solution.

$$
\begin{aligned}
P[X \geq (1+\delta)\mu] &= P[\exp(\lambda X) \geq \exp(\lambda(1+\delta)\mu)] \\
&\leq \frac{E[\exp \lambda X]}{\exp(\lambda(1+\delta)\mu)} \\
&= \exp(-\lambda(1+\delta)\mu) \prod_{i\in[n]} E[\exp(\lambda X_i)] \\
&= \exp(-\lambda(1+\delta)\mu) \prod_{i\in[n]} (1 - p_i + p_i e^{\lambda}) \\
&\leq \exp(-\lambda(1+\delta)\mu) \prod_{i\in[n]} \exp(p_i(1 - e^{\lambda})) \\
&= \exp(-\lambda(1+\delta)\mu) \exp(\sum_{i\in[n]} p_i(1 - e^{\lambda})) \\
&= \exp(\mu(-\lambda(1+\delta) + 1 - e^{\lambda})).
\end{aligned}
$$

Now we want to minimize $\exp(\mu(-\lambda(1+\delta)+1-e^{\lambda}))$. By solving the equation that the derivative equals to 0, we can know it is minimized when $\lambda = \ln(1+\delta)$.

Thus, finally we have

$$
P[X \geq (1+\delta)\mu] \leq (\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}})^{\mu}.
$$

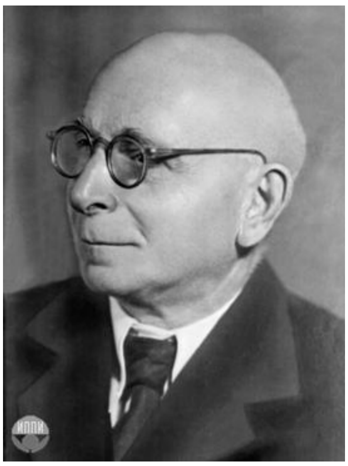
□

1.2 Bernstein's inequality

5: Bernstein's inequality

Let $C = \sup_{i\in[n]}(b_i - a_i)$.

$$
P[|X - E[X]| > t] \leq 2\exp(-\frac{t^2/2}{Var[X] + C \cdot t/3}).
$$



Sergei Natanovich Bernstein (5 March 1880 –26 October 1968) was a Ukrainian and Soviet mathematician of Jewish origin

His father was Natan Osipovich Bernstein (1836-1891), a medical doctor and also an extraordinary professor at the University of Odessa. The family was Jewish, and Natan Bernstein had been an editor of the Odessa magazine Zion: Organ of Russian Jews which had only been published for a year around 1861 before being closed down.

He graduated from high school in 1898. After this, following his mother's wishes, he went with his elder sister to Paris. Bernstein's sister studied biology in Paris and did not return to the Ukraine but worked at the Pasteur Institute. After one year studying mathematics at the Sorbonne, Bernstein decided that he would rather become an engineer and entered the École d'Electrotechnique Supérieure. However, he continued to be interested in mathematics and spent three terms at Göttingen, beginning in the autumn of 1902, where his studies were supervised by David Hilbert.

According to [A.V. Prokhorov, 2001], 5: Bernstein's inequality is proposed by S.N. Bernshtein in "Probability theory", 1911. But, [Wikipedia, 2020a] claimed that there inequalities are proved and published in the 1920s and 1930s.

I think the first date maybe wrong, because it only refers to a later version of this textbook published in 1946. But unfortunately, I cannot find any origin paper or textbook.

As a conclusion, 5: Bernstein's inequality may be proposed in 1920s(far earlier than 6: Hoeffding' inequality, even 4: Chernoff bound!) and it turns out to be one of the strongest bound in some cases.

1.3 Hoeffding' inequality and Azuma's inequality

6: Hoeffding' inequality

For all i, X_i is bounded in $[a_i, b_i]$ almost surely. Then

$$P[|X - E[X]|| > t] < 2\exp(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}).$$

To prove this inequality, we need a lemma first.

7: Hoeffding's lemma

Random variable X is bounded in $[a, b]$ almost surely and $E[X] = \mu$. Then

$$E[\exp(\lambda X)] \leq \exp(\lambda\mu + \frac{\lambda^2(b-a)^2}{8}).$$

Proof. $\exp(\lambda X)$ is a convex function of X, let $p = \frac{b-\mu}{b-a}$, $u = (b-a)\lambda$,

$$\begin{aligned}
E[\exp(\lambda X)] &\leq \frac{b - E[x]}{b-a}\exp(\lambda a) + \frac{E[x] - a}{b-a}\exp(\lambda b) \\
&= \exp(\lambda a)(\exp(\lambda(b-a))\frac{\mu - a}{b-a} + \frac{b-\mu}{b-a}) \\
&= ((1-p)\exp(u) + p)\exp((p-a)u + \lambda\mu) \\
&= \exp((p-1)u + \lambda\mu + \log((1-p)\exp(u) + p)) := \exp(f(u)).
\end{aligned}$$

We now try to minimize $f(u)$, by Taylor's expansion,

$$f(u) = f(0) + uf'(0) + \frac{u^2}{2}f''(\xi), \xi \in [0, u] \leq \lambda\mu + \frac{1}{8}u^2 = \lambda\mu + \frac{\lambda^2}{8}(b-a)^2.$$

□

Then we can simply come to the main result.

Proof. By 4: Chernoff bound and 7: Hoeffding's lemma,

$$\begin{aligned} P[X - E[X] \geq t] &\leq \exp(-\lambda t) \prod_{i \in [n]} E[\exp(\lambda(X_i - E[X_i]))] \\ &\leq \exp(-\lambda t) \prod_{i \in [n]} \exp(\frac{\lambda^2}{8}(b_i - a_i)^2) \\ &= \exp(-\lambda t + \frac{1}{8}\lambda^2 \sum_{i \in [n]} (b_i - a_i)^2) \\ &\leq \exp(-\frac{2t^2}{\sum_{i \in [n]} (b_i - a_i)^2}). \end{aligned}$$

Here we set $\lambda = \frac{4t}{\sum_{i \in [n]} (b_i - a_i)^2} > 0$. □

Wassily Hoeffding (June 12, 1914 –February 28, 1991) was a Finnish statistician and probabilist.
Hoeffding was born in Mustamäki, Finland, (Gorkovskoye, Russia since 1940), although his place of birth is registered as St. Petersburg on his birth certificate. His father was an economist and a disciple of Peter Struve, the Russian social scientist and public figure. His paternal grandparents were Danish and his father's uncle was the Danish philosopher Harald Høffding. His mother, née Wedensky, had studied medicine. In 1918 the family left Tsarskoye Selo for Ukraine and, after traveling through scenes of civil war, finally left Russia for Denmark in 1920, where Wassily entered school.
In 1924 the family settled in Berlin. Hoeffding obtained his PhD in 1940 at the University of Berlin. He migrated with his mother to the United States in 1946. His younger brother, Oleg, became a military historian in the United States.

Hoeffding publish this result [Hoeffding, 1994] year 1963. He also pointed out in his paper that after minor modifications this should work for certain dependent variables with bounded difference (section 4d in [Hoeffding, 1994])

> Let $S = Y_1 + Y_2 + \cdots + Y_n$ where the sequence of random variables $Y_1, Y_2, \cdots Y_n$ is $(r - l)$-dependent; that is, the random vectors $(Y_1, \cdots, Y_i)$ and $(Y_j, \cdots, Y_n)$ are independent if $j - i > r$, where r is a positive integer. Then the random variable $Y_i, Y_{r+i}, Y_{2r+i}, \cdots$ are independent.
> ……
> Explicitly, if $a \leq Y_j \leq b$, then $P[S - E[S] \geq nt] \leq \exp(-2[n/r]t^2/(b-a)^2)$.

This is somewhat similar to martingale difference version by Azuma below.

8: Azuma's inequality

Let $\{X_0, X_1, \cdots\}$ be a martingale and $|X_k - X_{k-1}| \leq c_k$ almost surely. Then for all positive integers N and all positive reals ε,

$$P[X_n - X_0 \geq \varepsilon] \leq \exp(\frac{-\varepsilon^2}{2\sum_{k=1}^N c_k^2}), P[X_n - X_0 \leq -\varepsilon] \leq \exp(\frac{-\varepsilon^2}{2\sum_{k=1}^N c_k^2}).$$

Proof. [Roch, 2015] Applying 4: Chernoff bound to $X_n - X_0$, we have

$$P[X_t - X_0 \geq \varepsilon] \leq \exp(-\varepsilon\lambda)E[\exp(\lambda \sum_{r\in[t]} (X_r - X_{r-1}))].$$

Then condition on the filtration

$$E[E[\exp(\lambda \sum_{r\in[t]} (X_r - X_{r-1})) \mid \mathcal{F}_{t-1}]] = E[\exp(\lambda \sum_{r\in[t]} (X_r - X_{r-1}))E[\exp(\lambda(X_t - X_{t-1})) \mid \mathcal{F}_{t-1}]].$$

By 7: Hoeffding's lemma, it holds almost surely that

$$E[\exp(\lambda(X_t - X_0)) \mid \mathcal{F}_{t-1}] \leq \exp(\frac{c_t^2\lambda^2}{8}).$$

By induction we can have

$$\begin{aligned} E[\exp(\lambda(X_t - X_0))] &\leq \exp(\frac{\lambda^2 \sum_{r\leq t} c_r^2}{8}) \\ &\leq \exp(-\frac{2\varepsilon^2}{\sum_{i\leq t} c_i^2}). \end{aligned}$$

□

There is also a maximal version and its proof in [Roch, 2015] via Doob's martingale inequality.

9: Maximal Azuma-Hoeffding inequality

$$P[\sup_{0\leq i\leq t} (Z_i - Z_0) \geq \beta] \leq \exp(-\frac{2\beta^2}{\sum_{i\leq t} c_i^2}).$$

Of course there can be a symmetric minimal form of 9: Maximal Azuma-Hoeffding inequality.

1.4 Summary

Table 1 gives a brief summary of the date and authors of these inequalities. Note that martingale was introduced by Paul Lévy in 1934.

Table 1: A brief summary for the inequalities

inequality	date	actually by
2: Markov's inequality		Pafnuty Lvovich Chebyshev
3: Chebyshev's inequality	1867	Irénée-Jules Bienaymé
4: Chernoff bound	1952	Herman Rubin
5: Bernstein's inequality	1930s	
6: Hoeffding' inequality	1963	
8: Azuma's inequality	1967	

2 Applications

2.1 Coin flipping

We often use the example of coin flipping to illustrate the lows of large numbers and how the result might be concentrate to its expectation. Here, our coin will fall heads with probability $p(0 \leq p \leq 1)$ scoring 1, and fall tails with probability $1-p$, scoring 0. We will toss it for N times.

Let X_i be the score after i tosses. By the linearity of expectation, the expectation of score after i tosses is $E[X_i] = ip$.

We would like to figure out how X can deviates from $E[X]$. To make things simpler, we seek for the probability that $X - E[X] > aN(a > 0)$.

What is the 'actual' deviation?

Whatever inequality we use, we may come into an upper bound of $P[X - E[X] > aN]$. Besides the comparison of such results, we also need to know how close they are compared to the real value. Of course there cannot be a formula to compute the actual value, but we can still simulate for a large number of times to reach the 'actual' deviation. Figure 1 shows a simulated distribution, in which we sample 10^6 times, 2671 of them turns to be out of range $[0, \frac{3N}{4}]$. Thus we can tell that the 'actual' probability of deviation $P[X - E[X] > \frac{N}{4}] \approx 2.671 \times 10^{-3}$.

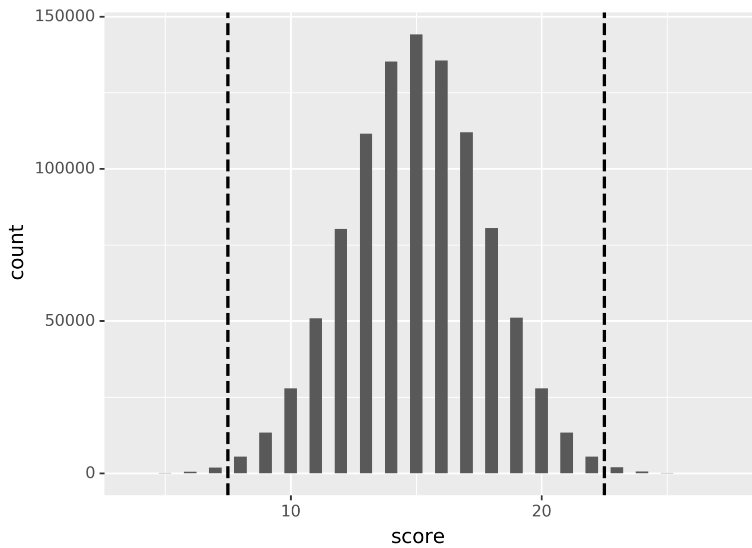


Figure 1: Distribution of scores when tossing a fair coin for 30 times

Analysis

All bounds are straightforward, Table 2 lists the upper bounds we get from section 1.

For example, when $p = 0.5$ and $a = 0.25$, just like Figure 1 shows. Figure 2 are the bound respect to N. In this case, 6: Hoeffding' inequality holds the tightest upper bound.

When p is smaller, we found that bound 4: Chernoff bound becomes better, shown in Figure 3.

Table 2: Upper bounds for coin flipping deviation $P[X - E[X] > aN]$

inequality	bound
2: Markov's inequality	$\leq \frac{E[X]}{aN + E[X]} = \frac{p}{a+p}$
3: Chebyshev's inequality	$\frac{Var[X]}{aN} = \frac{N \cdot (E[X^2] - E^2[X])}{(aN)^2} = \frac{p - p^2}{a^2 N}$
4: Chernoff bound	$(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}})^{\mu} = (\frac{\exp(\frac{a}{p})}{(1+\frac{a}{p})^{(1+\frac{a}{p})}})^{pN}$
5: Bernstein's inequality	$\exp(-\frac{a^2 N^2/2}{N(p-p^2) + aN/3}) = \exp(-\frac{a^2/2}{p - p^2 + a/3} N)$
6: Hoeffding' inequality	$\exp(-\frac{2(aN)^2}{\sum_{i=1}^{N} 1^2}) = \exp(-2a^2 N)$

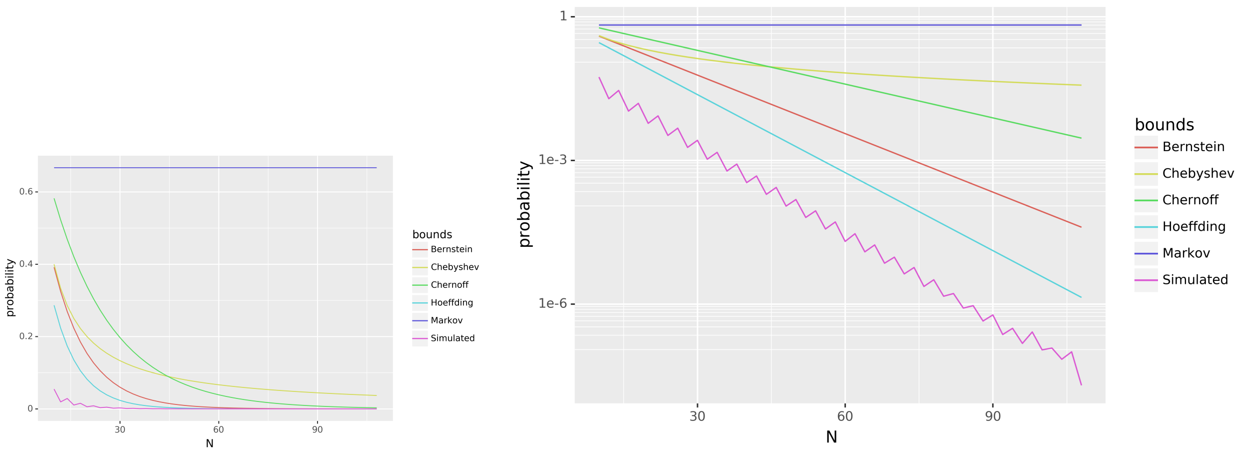


Figure 2: Upper bounds respect to N when $p = 0.5$ and $a = 0.25$

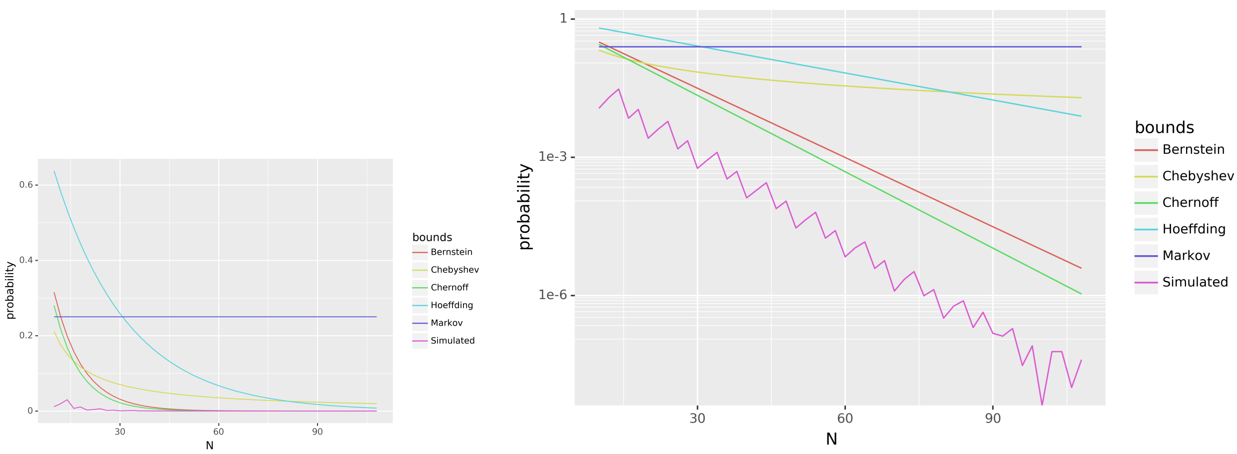


Figure 3: Upper bounds respect to N when $p = 0.05$ and $a = 0.15$

2.2 Mechanism design

In this section, I would like to share an interesting example of concentration inequalities, in mechanism design, a filed in theoretical computer science and economics. In this filed, we design 'games' in which players are rational with private information, and our goal is to find a mechanism to make the best of values.

Note! There are some properties we use it for granted without a proof in order to describe everything more fluently and concentrate on our topic. For example, why the mechanism can prevents agents from cheating. It is natural if a reader find something 'wrong' and I apologize for that.

2.2.1 Background and notations

For example, assume there is an item to be sold with n buyers, namely an auction. The seller thinks it to worth $v_0 \in \mathbb{R}$, while each buyer $i \in [n]$ thinks it's worth $v_i \in \mathbb{R}$. Then we need to design a mechanism for this auction, which takes the valuation $\mathbf{v} = (v_0, v_1, \cdots, v_n)$ as input, and generates the allocation $\mathbf{q} \in [0,1]^{n+1}$ together with the monetary transfer $\mathbf{z} \in \mathbb{R}^{n+1}$, devoting the probability that buyer i can get(or for 0, the seller can sell) the item and how much should they pay(or for 0, how much the seller can receive) in expectation correspondingly. The only information we can used to design such mechanism is that the valuations $\mathbf{v}$ is distributed according to a joint cumulative distribution function $\mathcal{F}(\cdot)$ and bounded in $[0, \overline{v}]$.

The utility ϕ of an agent is how much he earns, which can be described simply by $\phi = vq - z$. vq is the expected earning in this auction, while z is how much he need to pay. The bigger this gap is, the more utility he gets.

The situation here, from [Balseiro et al., 2019], is rather simple. We assume that the auction will be processed for T identical and independent rounds, in round $t \in [T]$, the vector is donated by a superscript t like $\mathbf{v}^t$, and the valuation from seller is constant v_0. With more than one round, our mechanism is able to make use of history information, which is proved later to be quite useful.

Expected profit and social welfare

As a platform, like Didi or Uber, we need not only to do the allocation job, but also earn some profit $\Pi = E[\sum_{t=1}^{T}(\sum_{i=1}^{n} z_i^t - z_0^t)]$. Also, if we know the utility U that players get, $\Pi + U = S$ is the social welfare created during these auctions, we have

$$S = E[\sum_{t\in[T]} \sum_{i\in[n]} (v_i^t - v_0^t) q_i^t].$$

Intuitively, S describes the expectation utility that the transaction make. For buyer i, if he think the item worth v_i^t but the seller thinks it only worth v_0^t, then the welfare is $v_i^t - v_0^t$.

There is a baseline for profit, the first-best benchmark.

10: Goal

Give a mechanism that is good and makes much profit.

- The mechanism is good if it do not make sellers pay and make buyers receive money, not make a buyer to pay more than his value expectation, not hold an asset such that $q_0 \neq \sum_{i\in[n]} q_i$, also, it should prevent players from cheating, i.e., reporting a fake valuation.
- And the mechanism makes much profit if it can reach first-best benchmark. It is an obvious observation that in a first-price auction the platform can earn most, because seller will never pay more and the buyer will never receive less. This profit

$$\Pi^{FB} = T \cdot E[(\max_{i\in[n]} v_i - v_0)^+] = O(T).$$

Our mechanism should have $\Pi/\Pi^{FB} \to 1$ as $T \to \infty$.

Note that a first-price auction(FPA) is not suitable for it cannot prevent from cheating, nor a second-auction auction (SPA) for its poor profit.

Promised utility

Promised utility is an important concept in this mechanism, donated by w. It means the expected utility each player promised by out mechanism. Thus it must satisfy

$$w_i^t = E_{v_t}[\phi_i^t(v_i, v^t; w^t) + u_i^t(v^t, w^t)], \forall i \in [n] \cup \{0\}, t \leq T, w_i^{T+1} = 0.$$

This seems quite unbelievable, because the platform determines the expected utility in the whole game w^1, even before the auctions started! But it is natural though, because the platform knows the valuation distribution in advance anyway.

2.2.2 Mechanism description

We can now introduce the mechanism which can be described by a tuple (q, z, u, w^1). q and z are the allocation function and monetary transfer function above, and u is a function used to determine promised utility change, i.e.

$$w_i^{t+1} = u_i^t(v^t; w^t)$$

, and finally, w^1 is the promised utility for every agent before the auctions started, and is indeed the expected profit an agent can earn in T rounds. Here we use $\underline{w}$ to donate the expected utility an agent can get in a SPA and $\overline{w}_i = (T - t + 1)\underline{w}$ donating as if $\underline{w}$ for the rest of the game.

The main idea is to allocate differently when the buyer is with different different promised utility. The first three situations are for the winner, i.e., the buyer with highest valuation.

- High promised utility (second-price auction, SPA): $\overline{w}_i^t - \overline{v} \leq w_i^t \leq \overline{w}_i^t$.

  $$q_i^t = 1, \quad z_i^t = \max v_{-i}^t, \quad u_i^t = w_i^t - \underline{w}_i.$$

  When the promised utility is quite high, we need to keep our promise, letting this agent earns much. So we do a SPA to fast decrease his w.

- Medium promised utility (first-price auction, FPA): $\underline{w}_i \leq w_i^t < \overline{w}_i^t - \overline{v}$.

$$q_i^t = 1, \quad z_i^t = v_i^t, \quad u_i^t = w_i^t - \underline{w}_i + (v_i^t - \max v_{-i}^t).$$

  Then, when its not too high but also not too low, we do a FPA to make as much profit for the platform as possible. But in this case, the agent actually earns not enough, so wee need to add $v_i^t - \max v_{-i}^t$ for him, promising this utility can be delivered in later auctions.
- Low promised utility (throttled second-price auction): $\underline{w}_i \leq w_i^t < \overline{w}_i^t - \overline{v}$.

$$q_i^t = w_i^t / \underline{w}_i, \quad z_i^t = q_i^t \cdot \max v_{-i}^t, \quad u_i^t = 0.$$

  When w is too low for a SPA, we would rather do a 'partial' SPA, in which the winner can only be allocation with probability. After this auction, its w can be finally set to zero.
- Losing buyer:

$$q_i^t = 0, \quad z_i^t = 0, \quad u_i^t = (w_i^t - \underline{w}_i)^+.$$

  All losers can also be considered to attend an SPA, decreasing $\overline{w}$ promised utility.

Eventually, all promised utilities will be set to zero, which means our mechanism keeps its promise.

2.2.3 Profit approximation

11: Theorem

If the initial promised utilities are set to $w_i^1 = \overline{v}\sqrt{8T\log T} + \underline{w}_i$, then the expected profit of the mechanism, denoted by Π, satisfies:

$$\Pi \geq \Pi^{FB} - O(\sqrt{T\log T}).$$

whenever $w_i^1 \leq \overline{w}_i^1 - \overline{v}$. In particular, the mechanism is asymptotically optimal, i.e., $\Pi/\Pi^{FB} \to 1$ as $T \to \infty$.

Proof. As described above, let S be the expected social welfare of the mechanism.

$$S = \Pi + \sum_{i=0}^{n} w_i = E_v[\sum_{t=1}^{T}\sum_{i\in[n]}(v_i^t - v_0^t)q_i^t].$$

Let $\tau_i = \inf\{t \geq 1 : w_i^t < \underline{w}_i\}$ be the stopping time measuring the first time in which w_i^t falls into the throttled SPA region.

Let $\hat{q}_i^t$ be the efficient allocation, i.e., a first price auction for the FB benchmark.

$$\begin{aligned}
\Pi^{FB} - S &= E_v[\sum_{t=1}^{T}\sum_{i\in[n]}(v_i^t - v_0^t)\cdot\hat{q}_i^t - \sum_{t=1}^{T}\sum_{i\in[n]}(v_i^t - v_0^t)q_i^t] \\
&= \sum_{i\in[n]} E_v[\sum_{t=\tau_i}^{T}(v_i^t - v_0^t)(\hat{q}_i^t - q_i^t)] \\
&\leq \overline{v}\sum_{i\in[n]} E_v[(T - \tau_i + 1)^+].
\end{aligned}$$

Now we define

$$\mu_i^t = w_i^t + \sum_{\tau=1}^{t-1}(v_i^\tau \cdot q_i^\tau - z_i^\tau) \quad \text{(A process from expectation to reality)}.$$

Let $\mathcal{H}_t = \sigma(v_1, v_2, \cdots, v_{t-1})$ be the natural filtration. We have

$$E[\mu_i^{t+1} \mid \mathcal{H}_t] = \mu_i^t + E[v_i^t \cdot q_i^t - z_i^t + w_i^{t+1} \mid \mathcal{H}_t] - w_i^t = \mu_i^t.$$

Therefore, $\{\mu_i^t\}_{t=1}^T$ is a martingale with respect to the natural filtration.

Note that the SPA and throttled SPA region are absorbing. Let $T_i^* = T - \overline{v}/\underline{w}_i \geq 1$. By 9: Maximal Azuma-Hoeffding inequality,

$$P[\tau_i \leq t] = P[\min_{s \leq t} \mu_i^s \leq \underline{w}_i] \leq \exp(-\frac{(w_i^1 - \underline{w}_i)^2}{8\sum_{k=1}^t \overline{v}^2}) = \exp(-\frac{(w_i^1 - \underline{w}_i)^2}{8t\overline{v}^2}).$$

This implies that

$$\begin{aligned} E_v[(T - \tau_i + 1)^+] = \sum_{t=1}^T P[\tau_i \leq t] &\leq \sum_{t=1}^{T_i^*} P[\tau_i \leq T_i^*] + T - T_i^* \\ &\leq T_i^* \exp(-\frac{(w_i^1 - \underline{w}_i)^2}{8T_i^*\overline{v}^2}) + T - T_i^*. \end{aligned}$$

By letting $w_i^1 = \overline{v}\sqrt{8T\log T} + \underline{w}_i$ and using that $T_i^* \leq T$,

$$\begin{aligned} E_v[(T - \tau_i + 1)^+] &\leq T_i^* \exp(-\frac{(w_i^1 - \underline{w}_i)^2}{8T_i^*\overline{v}^2}) + T - T_i^* \\ &\leq T \cdot \exp(-\log T) + T - T_i^* = 1 + T - T_i^* = \overline{v}/\underline{w}_i + 1. \end{aligned}$$

Hence,

$$\Pi_{FB} - S \leq \overline{v} \sum_{i \in [n]} (\overline{v}/\underline{w}_i + 1) = O(1).$$

Finally,

$$\Pi = S - \sum_i w_i^1 = \Pi^{FB} - O(1) - O(\sqrt{T\log T}) = \Pi^{FB} - O(\sqrt{T\log T}).$$

□

2.2.4 Profit concentration

12: Theorem

The probability that the aggregate budget balance constraint is violated goes to zero as time goes to infinity.

$$Pr[\sum_{t=1}^T (\sum_{i \in [n]} z_i^t - z_0^t) \geq 0] \to 1 \text{ as } T \to \infty.$$

By 11: Theorem we know that the expectation for profit is very delightful. But it is not satisfactory if the profit often deviates from expectation much. So here we focus on how concentrate our profit is. Here, we only focus on the probability for this mechanism to make positive profit, i.e., $P[\Pi > 0]$.

Of course by the definitions above, our mechanism will never loss money. So finding upper bound for $P[\Pi > 0]$ makes no sense here. But later in this paper we started to focus on a more general case, in which we subsidize trade sometimes. And this property also holds for these cases, preventing the platform from losing money too often.

Proof. Let $\sigma(v^t) = (\max v_{-0}^t - v_0^t)^+$ be the social welfare for period t. Note that Π here is a function calculating platform profit with given value vector v instead of the expected profit.

We argue that $\frac{1}{T}E[|\Pi^{FB} - \Pi(v)]$ converges to zero:

$$\begin{aligned} & E[|\Pi^{FB} - \Pi(v)|] \\ \text{(Minkowski's inequality)} \ & \leq E[|\Pi^{FB} - \Pi^{FB}(v)|] + E[|\Pi^{FB}(v) - \Pi(v)|] \\ \text{(Jensen's inequality)} \ & \leq \sqrt{Var(\Pi^{FB}(v))} + \Pi^{FB} - \Pi = \sqrt{T}\bar{v}^2/4 + \tilde{O}(\sqrt{T}) = \tilde{O}(\sqrt{T}) \end{aligned}$$

Then we have

$$\begin{aligned} Pr[\Pi(v) < 0] = Pr[\Pi^{FB} - \Pi(v) > \Pi^{FB}] \leq Pr[|\Pi^{FB} - \Pi(v)| \geq \Pi^{FB}] \\ \text{(2: Markov's inequality)} \ \leq \frac{\tilde{O}(\sqrt{T})}{TE[\sigma(v^1)]} \to 0. \end{aligned}$$

□

References

[A.V. Prokhorov, 2001] A.V. Prokhorov, N.P. Korneichuk, V. M. (2001). Bernstein inequality.

[Balseiro et al., 2019] Balseiro, S. R., Mirrokni, V., Leme, R. P., and Zuo, S. (2019). Dynamic double auctions: Towards first best. In Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 157–172. SIAM.

[Chernoff, 2014] Chernoff, H. (2014). A career in statistics. Past, Present, and Future of Statistical Science, 29.

[Chernoff et al., 1952] Chernoff, H. et al. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. The Annals of Mathematical Statistics, 23(4):493–507.

[Hoeffding, 1994] Hoeffding, W. (1994). Probability inequalities for sums of bounded random variables. In The Collected Works of Wassily Hoeffding, pages 409–426. Springer.

[Roch, 2015] Roch, S. (2015). Modern discrete probability: An essential toolkit.

[Wikipedia, 2020a] Wikipedia (2020a). Bernstein inequalities (probability theory) — Wikipedia, the free encyclopedia.

[Wikipedia, 2020b] Wikipedia (2020b). Chebyshev's inequality — Wikipedia, the free encyclopedia.