Неравенства концентрации для выборок без возвращений

Толстихин Илья
ВЦ РАН

Декабрь 2013

Тема доклада

Пусть

- $\{X_1, \ldots, X_n\}$ — выборка случайных величин
- $\xi = f(X_1, \ldots, X_n)$ для некоторой функции $f\colon \mathbb{R}^n \to \mathbb{R}$

Задача: для $t \geq 0$ изучить поведение следующих величин:

$$\mathbb{P}\{\xi - \mathbb{E}[\xi] \geq t\}, \quad \mathbb{P}\{\mathbb{E}[\xi] - \xi \geq t\}$$

Известные результаты:

- Суммы независимых случайных величин: $f = \sum_{i=1}^{n} X_i$
  н-ва Хефдинга, Беннетта, Бернштейна
- Другие функции от независимых случайных величин
  н-во МакДиармида, индукция Талаграна, метод энтропии Леду

Вопрос: что делать, если $\{X_1, \ldots, X_n\}$ — выборка без возарвщений из конечного множества действительных чисел?

План доклада

Введение

- Неравенства концентрации: короткий обзор
- Постановка задачи и мотивация
- Метод Чернова

Раздел I: Суммы случайных величин

- Редукция Хефдинга
- Прямой анализ: метод мартингалов

Раздел II: Функции, определенные на разбиениях

- Неравенство МакДиармида
- Субгауссовское Неравенство Бобкова

Раздел III: Супремумы эмпирических процессов

- Субгауссовское неравенство
- Обобщение неравенства Талаграна

План доклада

Введение

- Неравенства концентрации: короткий обзор
- Постановка задачи и мотивация
- Метод Чернова

Раздел I: Суммы случайных величин

- Редукция Хефдинга
- Прямой анализ: метод мартингалов

Раздел II: Функции, определенные на разбиениях

- Неравенство МакДиармида
- Субгауссовское Неравенство Бобкова

Раздел III: Супремумы эмпирических процессов

- Субгауссовское неравенство
- Обобщение неравенства Талаграна

Неравенства концентрации: независимые с.в.

(Boucheron et al., 2013): книга содержит подробные обзоры многих подходов

Для независимых случайных величин $\{X_1, \ldots, X_n\}$ и функции $f\colon \mathbb{R}^n \to \mathbb{R}$ обозначим $\xi = f(X_1, \ldots, X_n)$.
Задача: получить *неасимптотические* оценки:

$$\mathsf{P}\{\xi - \mathbb{E}[\xi] \geq t\} \leq F_1(n) \text{ и } \mathsf{P}\{\mathbb{E}[\xi] - \xi \geq t\} \leq F_2(n).$$

Базовые неравенства:

- Неравенства Маркова и Чебышева
- Метод Чернова

Суммы независимых случайных величин (или мартингалы):

- Неравенства (Азумы)-Хефдинга, Беннетта и Бернштейна

Общие функции независимых случайных величин:

- *Метод мартингалов* и неравенство МакДиармида (McDiarmid, 1989)
- *Индуктивный метод* и неравенства Талаграна (1995, 1996) (Talagrand, 1996)
- *Энтропийный метод* (Ledoux, 1996) и неравенство Буске (2001)

Постановка задачи и мотивация

Рассмотрим конечное множество $\mathcal{C} = \{c_1, \ldots, c_N\}$ ограниченных действительных чисел $c_i \in [0, 1]$. Всюду далее:

- $\{Z_1, \ldots, Z_n\}$ вытянуты равномерно без возвращений из $\mathcal{C}$;
- $\{X_1, \ldots, X_n\}$ вытянуты равномерно с возвращениями из $\mathcal{C}$;

Задача: для $\xi = f(Z_1, \ldots, Z_n)$ и $t \geq 0$ получить верхние оценки для

$$\mathbb{P}\{\xi - \mathbb{E}[\xi] \geq t\}, \quad \mathbb{P}\{\mathbb{E}[\xi] - \xi \geq t\}.$$

Мотивация: Трансдуктивное обучение (Vapnik, 1998), Комбинаторный подход (Vorontsov, 2010):

- Конечная генеральная совокупность $\mathcal{Z} = \{(x_i, y_i)\}_{i=1}^{N}$
- Обучающая выборка $S = \{(x_i, y_i)\}_{i=1}^{n}$ вытянута равномерно без возвращений из $\mathcal{Z}$
- Ограниченная функция потерь $\ell(y', y'') \in [0, 1]$
- **Цель:** поиск отображения $h\colon x \to y$, такого что

$$\sum_{(x,y) \in \mathcal{Z} \setminus S} \ell\bigl(h(x), y\bigr) \to \min_{h \in \mathcal{H}}$$

Метод Чернова

Как оценить $\mathbb{P}\{\xi \geq t\}$ для с.в. ξ?

- Для всех $\lambda \geq 0$:

$$\mathbb{P}\{\xi \geq t\} = \mathbb{P}\left\{e^{\lambda\xi} \geq e^{\lambda t}\right\}.$$

- Для неотрицательной с.в. $e^{\lambda\xi}$ выполнено н-во Маркова:

$$\mathbb{P}\left\{e^{\lambda\xi} \geq e^{\lambda t}\right\} \leq \frac{\mathbb{E}\left[e^{\lambda\xi}\right]}{e^{\lambda t}}.$$

- Оценка сверху для производящей функции моментов (ПФМ):

$$\mathbb{E}\left[e^{\lambda\xi}\right] \leq F(\lambda).$$

- Минимизируем оценку для $\lambda \geq 0$:

$$\mathbb{P}\{\xi \geq t\} \leq \min_{\lambda \geq 0} \frac{F(\lambda)}{e^{\lambda t}}.$$

Метод Чернова

Как оценить $\mathbb{P}\{\xi \geq t\}$ для с.в. ξ?

- Для всех $\lambda \geq 0$:

$$\mathbb{P}\{\xi \geq t\} = \mathbb{P}\left\{e^{\lambda\xi} \geq e^{\lambda t}\right\}.$$

- Для неотрицательной с.в. $e^{\lambda\xi}$ выполнено н-во Маркова:

$$\mathbb{P}\left\{e^{\lambda\xi} \geq e^{\lambda t}\right\} \leq \frac{\mathbb{E}\left[e^{\lambda\xi}\right]}{e^{\lambda t}}.$$

- Оценка сверху для производящей функции моментов (ПФМ):

$$\mathbb{E}\left[e^{\lambda\xi}\right] \leq F(\lambda).$$

- Минимизируем оценку для $\lambda \geq 0$:

$$\mathbb{P}\{\xi \geq t\} \leq \min_{\lambda \geq 0} \frac{F(\lambda)}{e^{\lambda t}}.$$

План доклада

Введение

- Неравенства концентрации: короткий обзор
- Постановка задачи и мотивация
- Метод Чернова

Раздел I: Суммы случайных величин

- Редукция Хефдинга
- Прямой анализ: метод мартингалов

Раздел II: Функции, определенные на разбиениях

- Неравенство МакДиармида
- Субгауссовское Неравенство Бобкова

Раздел III: Супремумы эмпирических процессов

- Субгауссовское неравенство
- Обобщение неравенства Талаграна

Суммы случайных величин: н-во Хефдинга

Лемма ((Hoeffding, 1963))

Пусть ξ *— ограниченная случайная величина* $a \leq \xi \leq b$ *и* $\mathbb{E}[\xi] = 0$*. Тогда для всех* $\lambda \geq 0$*:*

$$\mathbb{E}\left[e^{\lambda\xi}\right] \leq e^{\lambda^2(b-a)^2/8}.$$

Теорема (Хефдинг, 1963)

Пусть $\{X_1, \ldots, X_n\}$ *— независимые с.в. с* $\mathbb{E}[X_i] = 0$ *и* $a_i \leq X_i \leq b_i$*. Тогда для всех* $t > 0$

$$\mathbb{P}\left\{\sum_{i=1}^n X_i \geq t\right\} \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

$$\mathbb{P}\left\{\sum_{i=1}^n X_i \leq -t\right\} \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right)$$

Суммы случайных величин: н-во Хефдинга

Идея доказательства:

Пусть $\{X_1, \dots, X_n\}$ — незавизисимые с.в. с $\mathbb{E}[X_i] = 0$ и $a_i \le X_i \le b_i$.
Тогда для всех $t > 0$ и $\lambda \ge 0$:

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{n} X_i\right)\right] = \prod_{i=1}^{n} \mathbb{E}\left[\exp \lambda X_i\right]$$

$$\le \exp\left(\frac{1}{8}\lambda^2 \sum_{i=1}^{n} (b_i - a_i)^2\right).$$

Применим метод Чернова:

$$\mathbb{P}\left\{\sum_{i=1}^{n} X_i \ge t\right\} \le \frac{\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{n} X_i\right)\right]}{e^{\lambda t}}$$

$$\le \exp\left(\frac{1}{8}\lambda^2 \sum_{i=1}^{n} (b_i - a_i)^2 - \lambda t\right) \to \min_{\lambda \ge 0}.$$

Минимум достигается при $\lambda^* = \frac{4t}{\sum_{i=1}^{n} (b_i - a_i)^2}$

Суммы случайных величин: н-во Беннетта

(Bennett, 1962; Bernstein, 1946)

Лемма (Бернштейн)

Пусть ξ *— ограниченная сверху с.в.* $\xi \leq 1$ *с* $\mathbb{E}[\xi] = 0$. *Обозначим* $\psi(s) = e^s - s - 1$. *Тогда для всех* $\lambda \geq 0$*:*

$$\mathbb{E}\left[e^{\lambda\xi}\right] \leq \exp\left(\psi(\lambda)\mathbb{E}[\xi^2]\right).$$

Теорема (Беннетт, Бернштейн)

Пусть $\{X_1, \ldots, X_n\}$ *— независимые с.в. с* $\mathbb{E}[X_i] = 0$ *и* $X_i \leq 1$. *Обозначим*

$$\sigma^2 = \sum_{i=1}^{n} \mathbb{V}[X_i].$$

Тогда для всех $t > 0$

$$\mathbb{P}\left\{\sum_{i=1}^{n} X_i \geq t\right\} \leq \exp\left(-\sigma^2 h\left(\frac{t}{\sigma^2}\right)\right) \leq \exp\left(-\frac{t^2}{2\sigma^2 + \frac{2}{3}t}\right)$$

где $h(u) = (1+u)\log(1+u) - u \geq u^2/(2 + 2u/3)$ *для* $u \geq 0$.

Суммы случайных величин: н-во Беннетта

Идея доказательства:

Пусть $\{X_1, \ldots, X_n\}$ — независимые с.в. с $\mathbb{E}[X_i] = 0$ и $X_i \leq 1$. Тогда для всех $t > 0$ и $\lambda \geq 0$:

$$\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{n} X_i\right)\right] = \prod_{i=1}^{n} \mathbb{E}\left[\exp \lambda X_i\right] \leq \exp\left((e^{\lambda} - \lambda - 1)\sum_{i=1}^{n} \mathbb{V}[X_i]\right).$$

Применим метод Чернова:

$$\mathbb{P}\left\{\sum_{i=1}^{n} X_i \geq t\right\} \leq \frac{\mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^{n} X_i\right)\right]}{e^{\lambda t}} \leq \exp\left((e^{\lambda} - \lambda - 1)\sum_{i=1}^{n} \mathbb{V}[X_i] - \lambda t\right) \to \min_{\lambda \geq 0}.$$

Минимум достигается при $\lambda^* = \ln\left(\frac{t + \sum_{i=1}^{n} \mathbb{V}[X_i]}{\sum_{i=1}^{n} \mathbb{V}[X_i]}\right)$

План доклада

Введение

- Неравенства концентрации: короткий обзор
- Постановка задачи и мотивация
- Метод Чернова

Раздел I: Суммы случайных величин

- Редукция Хефдинга
- Прямой анализ: метод мартингалов

Раздел II: Функции, определенные на разбиениях

- Неравенство МакДиармида
- Субгауссовское Неравенство Бобкова

Раздел III: Супремумы эмпирических процессов

- Субгауссовское неравенство
- Обобщение неравенства Талаграна

Редукция Хефдинга

Лемма (Редукция Хефдинга, (Hoeffding, 1963))

Пусть $\{X_1, \ldots, X_n\}$ *и* $\{Z_1, \ldots, Z_n\}$ *вытянуты равномерно из конечного множества* $\mathcal{C} = \{c_1, \ldots, c_N\}$ *с и без возвращений соответственно. Тогда для любой выпуклой функции* $g\colon \mathbb{R} \to \mathbb{R}$ *справедливо:*

$$\mathbb{E}\left[g\left(\sum_{i=1}^{n} Z_i\right)\right] \leq \mathbb{E}\left[g\left(\sum_{i=1}^{n} X_i\right)\right]$$

Идея доказательства:

$$\frac{1}{N^n}\sum_{i_1=1}^{N}\cdots\sum_{i_n=1}^{N} g(c_{i_1}, \ldots, c_{i_n}) = \frac{n!}{N!}\sum_{(N,n)} g^*(c_{i_1}, \ldots, c_{i_n}),$$

где $\sum_{(N,n)}$ — сумма по всевозможным n-элементным кортежам $(i_1, \ldots, i_n)$ без повторений, а g^* — взвешенное арифметическое среднее g по нескольким выборкам.

Редукция Хефдинга

Обозначим $\xi = \sum_{i=1}^{n} Z_i$ и $\xi' = \sum_{i=1}^{n} X_i$.

- Воспользуемся методом Чернова для $t \geq 0$ и $\lambda \geq 0$:

$$\begin{aligned} \mathsf{P}\{\xi - \mathbb{E}[\xi] \geq t\} &= \mathsf{P}\{e^{\lambda(\xi - \mathbb{E}[\xi])} \geq e^{\lambda t}\} \\ &\leq \frac{\mathbb{E}\left[e^{\lambda(\xi - \mathbb{E}[\xi])}\right]}{e^{\lambda t}} \\ &= \frac{\mathbb{E}\left[e^{\lambda \xi}\right] e^{-\lambda \mathbb{E}[\xi]}}{e^{\lambda t}} \end{aligned}$$

- Заметим, что для $\lambda \geq 0$ функция $f(s) = e^{\lambda s}$ выпукла
- Воспользуемся редукцией Хефдинга:

$$\frac{\mathbb{E}\left[e^{\lambda \xi}\right]}{e^{\lambda t}} \leq \frac{\mathbb{E}\left[e^{\lambda \xi'}\right]}{e^{\lambda t}}$$

- Заметим, что $\mathbb{E}[\xi] = \mathbb{E}[\xi'] = n \cdot \frac{1}{N} \sum_{i=1}^{N} c_i \triangleq n \cdot \mu$

Редукция Хефдинга

Лемма (Хефдинг, 1963)

Для любых $t \geq 0$ и $\lambda \geq 0$:

$$\mathsf{P}\left\{\sum_{i=1}^{n}(Z_i - \mu) \geq t\right\} \leq \frac{1}{e^{\lambda t}}\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^{n}(X_i - \mu)\right)\right]$$

Отсюда немедленно следует, что для $\sum_{i=1}^{n} Z_i$ справедливы:

- Неравенство Хефдинга
- Неравенства Беннетта и Бернштейна
- kl-неравенство
- Любые другие неравенства, основанные на методе Чернова

Редукция Хефдинга

Заметим, что:

$$\mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} Z_i\right] = \frac{N-n}{N-1}\,\mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \leq \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

Кажется, что вытягивание без возвращений «концентрируется» сильнее, чем вытягивание с возвращениями.

Проблема: Редукция Хефдинга не дает улучшений.

Вопрос: Можно ли получить оценки лучше?

План доклада

Введение

- Неравенства концентрации: короткий обзор
- Постановка задачи и мотивация
- Метод Чернова

Раздел I: Суммы случайных величин

- Редукция Хефдинга
- Прямой анализ: метод мартингалов

Раздел II: Функции, определенные на разбиениях

- Неравенство МакДиармида
- Субгауссовское Неравенство Бобкова

Раздел III: Супремумы эмпирических процессов

- Субгауссовское неравенство
- Обобщение неравенства Талаграна

Метод мартингалов и н-во Серфлинга, 1974

(Serfling, 1974)

Лемма (Серфлинг, 1974)

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^{n}(Z_i-\mu)\right)\right] \le \exp\left(\frac{1}{8}\lambda^2 n\left(1-\frac{n-1}{N}\right)\right)$$

Теорема (Серфлинг, 1974)

Обозначим $\Delta(n,N)=1-\frac{n-1}{N}$. *Тогда для всех* $t>0$*:*

$$\mathbb{P}\left\{\sum_{i=1}^{n}(Z_i-\mu)\ge t\right\} \le \exp\left(-\frac{2t^2}{n\cdot\Delta(n,N)}\right)$$

$$\mathbb{P}\left\{\sum_{i=1}^{n}(\mu-Z_i)\ge t\right\} \le \exp\left(-\frac{2t^2}{n\cdot\Delta(n,N)}\right)$$

Идея Доказательства

Обозначим $Z_1^i \triangleq \{Z_1, \dots, Z_i\}$. Рассмотрим величины $S_k = \frac{1}{N-k}\sum_{i=1}^{k}(Z_i - \mu)$.

$$\mathbb{E}[Z_k | Z_1^{k-1}] = \frac{N\mu - \sum_{i=1}^{k-1} Z_i}{N-k+1} = \frac{(N-k+1)\mu + (k-1)\mu - \sum_{i=1}^{k-1} Z_i}{N-k+1} = \mu - S_{k-1}.$$

Кроме того:

$$S_k = S_{k-1} + \frac{Z_k - \mu + S_{k-1}}{N-k}$$

Получаем:

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda S_n\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\exp\left(\lambda S_n\right)\middle| Z_1^{n-1}\right]\right] \\
&= \mathbb{E}\left[\exp\left(\lambda S_{n-1}\right)\mathbb{E}\left[\lambda \frac{Z_n - \mu + S_{n-1}}{N-n}\middle| Z_1^{n-1}\right]\right] \\
&\leq \mathbb{E}\left[\exp\left(\lambda S_{n-1}\right)\right]\exp\left(\frac{\lambda^2}{8(N-n)^2}\right) \\
&\dots \\
&\leq \exp\left(\frac{1}{8}\sum_{i=1}^{n}\frac{\lambda^2}{(N-i)^2}\right)
\end{aligned}$$

Идея Доказательства

$$\mathbb{E}\left[\exp\left(\lambda S_n\right)\right] \le \exp\left(\frac{1}{8}\sum_{i=1}^{n}\frac{\lambda^2}{(N-i)^2}\right) \tag{1}$$

Заметим, что $\sum_{i=1}^{n}(Z_i-\mu) = (N-n)S_n$. Применим (1) с $\lambda' = \lambda(N-n)$:

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^{n}(Z_i-\mu)\right)\right] = \mathbb{E}\left[\exp\left(\lambda(N-n)S_n\right)\right] \le \exp\left(\frac{1}{8}\sum_{i=1}^{n}\frac{\lambda^2(N-n)^2}{(N-i)^2}\right)$$

Можно убедиться, что:

$$\sum_{i=1}^{n}\frac{(N-n)^2}{(N-i)^2} \le n\left(1-\frac{n-1}{N}\right)$$

Окончательно получаем:

$$\mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^{n}(Z_i-\mu)\right)\right] \le \exp\left(\frac{\lambda^2}{8}n\left(1-\frac{n-1}{N}\right)\right)$$

□

Другие неравенства для сумм

- Н-во Серфлинга уточняет н-во Хефдинга при $n \to N$.
- Аналогичное усиление справедливо и для н-ва Бернштейна! (Bardenet and Maillard, 2013)
- А также для эмпирического н-ва Бернштейна! (Bardenet and Maillard, 2013)

Мы убедились, что редукция Хефдинга — простой в применении метод. Но непосредственный анализ ПФМ может приводить к существенно лучшим результатам.

План доклада

Введение

- Неравенства концентрации: короткий обзор
- Постановка задачи и мотивация
- Метод Чернова

Раздел I: Суммы случайных величин

- Редукция Хефдинга
- Прямой анализ: метод мартингалов

Раздел II: Функции, определенные на разбиениях

- Неравенство МакДиармида
- Субгауссовское Неравенство Бобкова

Раздел III: Супремумы эмпирических процессов

- Субгауссовское неравенство
- Обобщение неравенства Талаграна

Функции, определенные на разбиениях

- Пусть $\boldsymbol{\pi}$ — случайная перестановка, равномерная на симметрической группе перестановок $\mathrm{Sym}(\{1, \ldots, N\})$
- Тогда, если $\boldsymbol{\pi}_i$ — i-я координата $\boldsymbol{\pi}$, то

$$\sum_{i=1}^{n} Z_i \equiv \sum_{i=1}^{n} c_{\pi_i}$$

- Вытягивание $\boldsymbol{\pi}$ $\Leftrightarrow$ Вытягивание $Z_1, \ldots, Z_n$ без возвращений
- Пусть $X^n \cup X^{N-n}$ — разбиение множества $\{1, \ldots, N\}$ на два непересекающихся подмножества: $\{1, \ldots, N\} = X^n \cup X^{N-n}$.
- Пусть пара (X^n, X^{N-n}) распределена равномерно на множестве из C_N^n таких разбиений.
- Тогда

$$\sum_{i=1}^{n} Z_i \equiv \sum_{i \in X^n} c_i$$

Функции, определенные на разбиениях

- Сумма допускает определение через (X^n, X^{N-n}), поскольку $f(x_1, \ldots, x_n) = \sum_{i=1}^n x_i$ симметрична относительно перестановки координат.
- Оказывается, для функций $f\colon \mathbb{R}^n \to \mathbb{R}$, обладающих определенными симметриями, справедливы **нетривиальные неравенства концентрации**.

План доклада

Введение

- Неравенства концентрации: короткий обзор
- Постановка задачи и мотивация
- Метод Чернова

Раздел I: Суммы случайных величин

- Редукция Хефдинга
- Прямой анализ: метод мартингалов

Раздел II: Функции, определенные на разбиениях

- Неравенство МакДиармида
- Субгауссовское Неравенство Бобкова

Раздел III: Супремумы эмпирических процессов

- Субгауссовское неравенство
- Обобщение неравенства Талаграна

Неравенство МакДиармида

Функция $f\colon \mathbb{R}^n \to \mathbb{R}$ имеет **ограниченные разности**, если для некоторых вещественных констант $c_1, \ldots, c_n$:

$$\sup_{\substack{x_1,\ldots,x_n,\\ x_i' \in \mathbb{R}}} \left| f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) \right| \leq c_i,\ 1 \leq i \leq n.$$

Пример: $f(x_1, \ldots, x_n) = \sum_{i=1}^n x_i$ для $x_i \in [0, 1]$.

Теорема (Неравенство МакДиармида)

Пусть с.в. $\{X_1, \ldots, X_n\}$ *непрерывны и у функции* f *ограниченные разности с константами* $c_1, \ldots, c_n$. *Обозначим* $\xi = f(X_1, \ldots, X_n)$. *Тогда:*

$$\mathbb{P}\big\{\xi - \mathbb{E}[\xi] \geq t\big\} \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$$

$$\mathbb{P}\big\{\mathbb{E}[\xi] - \xi \geq t\big\} \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)$$

Аналог для выборок без возвращений

(El-Yaniv and Pechyony, 2009)

Функция $f\colon \boldsymbol{\pi} \to \mathbb{R}$, заданная на симметрической группе перестановок $\mathrm{Sym}(\{1,\ldots,N\})$, называется (n,u)-**симметричной**, если она не меняет своих значений при замене порядка первых n и/или последних $N-n$ координат $\boldsymbol{\pi}$.

Пусть $\boldsymbol{\pi}^{i,j}$ — перестановка, полученная из π транспозицией ее i-й и j-й координат.

Теорема (El-Yaniv and Pechyony, 2009)

Пусть $f(\boldsymbol{\pi})$ — (n,u)-симметричная функция, для которой существует константа $\beta > 0$, такая что $|f(\boldsymbol{\pi}) - f(\boldsymbol{\pi}^{i,j})| \leq \beta$ для всех $\boldsymbol{\pi}$, $i \in \{1,...,n\}$ и $j \in \{n+1,\ldots,N\}$. Тогда для любого $t \geq 0$:

$$\mathsf{P}\left\{f(\boldsymbol{\pi}) - \mathbb{E}[f(\boldsymbol{\pi})] \geq t\right\} \leq \exp\left\{-\frac{2t^2}{m\beta^2}\left(\frac{N-1/2}{N-n}\right)\left(1-\frac{1}{2\max(n,N-n)}\right)\right\}.$$

Аналогичное неравенство справедливо и для $\mathsf{P}\left\{\mathbb{E}[f(\boldsymbol{\pi})] - f(\boldsymbol{\pi}) \geq \epsilon\right\}$.

Аналог для выборок без возвращений

- Доказательство основано на *методе мартингалов*
- Сумма $\sum_{i=1}^{n}(Z_i - \mu)$ удовлетворяет условиям теоремы с $\beta = 1$: получаем н-во Серфлинга с немного худшими константами.
- Предположения теоремы непосредственно связаны с условием ограниченных разностей

Проблема: неравенство учитывает только длину интервала $f(\boldsymbol{\pi})$

Вопрос: можно ли учитывать дисперсию?

План доклада

Введение

- Неравенства концентрации: короткий обзор
- Постановка задачи и мотивация
- Метод Чернова

Раздел I: Суммы случайных величин

- Редукция Хефдинга
- Прямой анализ: метод мартингалов

Раздел II: Функции, определенные на разбиениях

- Неравенство МакДиармида
- Субгауссовское Неравенство Бобкова

Раздел III: Супремумы эмпирических процессов

- Субгауссовское неравенство
- Обобщение неравенства Талаграна

Субгауссовское неравенство Бобкова
(Bobkov, 2004)

- Любая перестановка $\boldsymbol{\pi}$ разбивает множество $\mathcal{C} = \{c_1, \ldots, c_N\}$ на два подмножества $\{c_{\pi_1}, \ldots, c_{\pi_n}\}$ и $\{c_{\pi_{n+1}}, \ldots, c_{\pi_N}\}$
- Обозначим $I = \{1, \ldots, n\}$ и $J = \{n+1, \ldots, N\}$
- Рассмотрим квадрат длины дискр. град. функции f в точке $\boldsymbol{\pi}$:

$$V^f(\boldsymbol{\pi}) = |\nabla f(\boldsymbol{\pi})|^2 = \sum_{i \in I} \sum_{j \in J} \left(f(\boldsymbol{\pi}) - f(\boldsymbol{\pi}^{i,j})\right)^2$$

Теорема (Bobkov, 2004)

Пусть $f(\boldsymbol{\pi})$ — (n, u)-симметричная функция, для которой существует константа $\Sigma^2 \geq 0$, такая что $V^f(\boldsymbol{\pi}) \leq \Sigma^2$ для всех $\boldsymbol{\pi}$. Тогда для любого $t \geq 0$ справедливо:

$$\mathsf{P}\left\{f(\boldsymbol{\pi}) - \mathbb{E}[f(\boldsymbol{\pi})] \geq \epsilon\right\} \leq \exp\left\{-\frac{(N+2)\epsilon^2}{4\Sigma^2}\right\}.$$

Аналогичное неравенство справедливо для $\mathsf{P}\left\{\mathbb{E}[f(\boldsymbol{\pi})] - f(\boldsymbol{\pi}) \geq \epsilon\right\}$.

Субгауссовское неравенство Бобкова

Идея доказательства:

Доказательство использует энтропийный метод Леду:

- Для случайной перестановки $\boldsymbol{\pi}$ и $f\colon \pi^n \to \mathbb{R}^+$ определим
$$\mathrm{Ent}(f) = \mathbb{E}[g(\boldsymbol{\pi}) \ln g(\boldsymbol{\pi})] - \mathbb{E}[g(\boldsymbol{\pi})] \ln \mathbb{E}[g(\boldsymbol{\pi})]$$
- Автор показывает, что для любого $\lambda \in \mathbb{R}$:
$$(N+2)\mathrm{Ent}\left[e^{\lambda f}\right] \le \Sigma^2 \lambda^2 \mathbb{E}\left[e^{\lambda f(\boldsymbol{\pi})}\right]$$
- Получаем дифференциальное неравенство на
$$F(\lambda) = \ln \mathbb{E}\left[e^{\lambda f(\boldsymbol{\pi})}\right]$$
- Решая его, получаем:
$$\mathbb{E}\Big[\exp\Big(\lambda\big(f(\boldsymbol{\pi}) - \mathbb{E}[f(\boldsymbol{\pi})]\big)\Big)\Big] \le e^{\Sigma^2 \lambda^2/(N+2)}$$
- Метод Чернова завершает доказательство $\square$

Субгауссовское неравенство Бобкова

Теорема (Bobkov, 2004)

Пусть $f(\boldsymbol{\pi})$ *—* (n,u)*-симметричная функция, для которой существует константа* $\Sigma^2 \geq 0$*, такая что* $V^f(\boldsymbol{\pi}) \leq \Sigma^2$ *для всех* $\boldsymbol{\pi}$*. Тогда для любого* $t \geq 0$ *справедливо:*

$$\mathsf{P}\left\{f(\boldsymbol{\pi}) - \mathbb{E}[f(\boldsymbol{\pi})] \geq t\right\} \leq \exp\left\{-\frac{(N+2)t^2}{4\Sigma^2}\right\}.$$

- Пусть (n,u)-симметричная функция f удовлетворяет условиям Теоремы (El-Yaniv and Pechyony, 2009). Тогда:

$$V^f(\boldsymbol{\pi}) = \sum_{i\in I}\sum_{j\in J}\big(f(\boldsymbol{\pi}) - f(\boldsymbol{\pi}^{i,j})\big)^2 \leq \sum_{i\in I}\sum_{j\in J}\beta^2 = n(N-n)\beta^2.$$

- Откуда получаем:

$$\mathsf{P}\left\{f(\boldsymbol{\pi}) - \mathbb{E}[f(\boldsymbol{\pi})] \geq t\right\} \leq \exp\left\{-\frac{t^2}{4n\beta^2}\left(\frac{N+2}{N-n}\right)\right\}.$$

Субгауссовское неравенство Бобкова

- Однако, условие

$$\sum_{i\in I}\sum_{j\in J}\big(f(\boldsymbol{\pi})-f(\boldsymbol{\pi}^{i,j})\big)^2\le\Sigma^2 \qquad (1)$$

является менее строгим, чем

$$|f(\boldsymbol{\pi})-f(\boldsymbol{\pi}^{i,j})|\le\beta,\quad\forall\boldsymbol{\pi},i\in\{1,...,n\},j\in\{n+1,\dots,N\}. \qquad (2)$$

- В случае независимых случайных величин известны примеры, когда условия типа (1) ведут к существенным улучшениям по сравнению с условиями типа (2).
- Пример: наибольшее собственное значение $\lambda_{\max}$ случайной симметр. матрицы $\mathbb{R}^{n\times n}$ с независимыми элементами $X_{i,j}$, такими что $|X_{i,j}|\le 1$. См. (Boucheron et al., 2013)
- Также, Теорема Бобкова позволяет получать н-ва для $\sum_{i=1}^n Z_i$, учитывающие дисперсию.

План доклада

Введение

- Неравенства концентрации: короткий обзор
- Постановка задачи и мотивация
- Метод Чернова

Раздел I: Суммы случайных величин

- Редукция Хефдинга
- Прямой анализ: метод мартингалов

Раздел II: Функции, определенные на разбиениях

- Неравенство МакДиармида
- Субгауссовское Неравенство Бобкова

Раздел III: Супремумы эмпирических процессов

- Субгауссовское неравенство
- Обобщение неравенства Талаграна

Супремумы эмпирических процессов

- Рассмотрим конечное множество $\mathcal{C} = \{c_1, \dots, c_N\}$ действительных чисел $c_i \in \mathbb{R}$.
- $\{Z_1, \dots, Z_n\}$ вытянуты равномерно без возвращений из $\mathcal{C}$;
- $\{X_1, \dots, X_n\}$ вытянуты равномерно с возвращениями из $\mathcal{C}$;
- $\mathcal{F} = \{f : \mathcal{C} \to \mathbb{R}\}$ — счетное множество отображений, таких что $\mathbb{E}[f(X_1)] = 0$ и $f(x) \in [-1, 1]$ для всех $f \in \mathcal{F}$, $x \in \mathcal{C}$.

Рассмотрим случайные величины:

$$Q_n = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(X_i), \quad Q'_n = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(Z_i).$$

- Q_n — супремум эмпирического процесса
- Q_n хорошо изучена в теории эмпирических процессов
- Один из центральных результатов: н-во Талаграна (1996)

Супремумы эмпирических процессов

$$Q_n = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(X_i), \quad Q'_n = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(Z_i).$$

Теорема ((Bousquet, 2002), Talagrand, 1996)

Положим $\sigma^2 = \sup_{f \in \mathcal{F}} \mathbb{V}[f(X_1)]$, $v = n\sigma^2 + 2\mathbb{E}[Q_n]$ *и для* $u \geq -1$ *определим* $\phi(u) = e^u - u - 1$, $h(u) = (1+u)\log(1+u) - u$. *Тогда для* $\lambda \geq 0$ *справедливо:*

$$E\left[e^{\lambda(Q_n - \mathbb{E}[Q_n])}\right] \leq e^{v\phi(\lambda)}.$$

Кроме того, для $t \geq 0$ *справедливо:*

$$\mathsf{P}\left\{Q_n - \mathbb{E}[Q_n] \geq t\right\} \leq e^{-vh(t/v)} \leq \exp\left(-\frac{t^2}{2v + \frac{2}{3}t}\right).$$

Заметим: если $\mathcal{F} = \{f_0\}$, то Теорема в точности воспроизводит н-ва Беннетта и Бернштейна.

Супремумы эмпирических процессов

- Случайная величина Q_n и н-во Талаграна играют ключевую роль в современных подходах теории статистического обучения
- Локальный радемахеровский анализ, который ведет к быстрым скоростям сходимости порядка $o(1/\sqrt{n})$, основан на н-ве Талаграна.
- Однако, в трансдуктивной постановке случайные величины вытягиваются без возвращений, а значит зависимы
- Для применения радемахеровского анализа к трансдуктивному подходу необходим аналог н-ва Талаграна

Н-во для Q'_n типа МакДиармида

- с.в. Q'_n может быть представлена с помощью перестановок:

$$Q'_n = Q'_n(\boldsymbol{\pi}) = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(c_{\pi_i})$$

- Очевидно, Q'_n— (n, u)-симметричная функция
- Для $i \in \{1, \ldots, n\}$, $j \in \{n+1, \ldots, N\}$ рассмотрим:

$$\begin{aligned}
\left|Q'_n(\boldsymbol{\pi}) - Q'_n(\boldsymbol{\pi}^{i,j})\right| &= \left|\sup_{f \in \mathcal{F}} \sum_{k=1}^{n} f(c_{\pi_k}) - \sup_{f \in \mathcal{F}} \sum_{k=1}^{n} f(c_{\pi_k^{i,j}})\right| \\
= [Q'_n(\boldsymbol{\pi}) > Q'_n(\boldsymbol{\pi}^{i,j})] &\left(\sup_{f \in \mathcal{F}} \sum_{k=1}^{n} f(c_{\pi_k}) - \sup_{f \in \mathcal{F}} \sum_{k=1}^{n} f(c_{\pi_k^{i,j}})\right) \\
+[Q'_n(\boldsymbol{\pi}) \le Q'_n(\boldsymbol{\pi}^{i,j})] &\left(\sup_{f \in \mathcal{F}} \sum_{k=1}^{n} f(c_{\pi_k^{i,j}}) - \sup_{f \in \mathcal{F}} \sum_{k=1}^{n} f(c_{\pi_k})\right) \le 2
\end{aligned}$$

- Справедлива Теорема (El-Yaniv and Pechyony) с $\beta = 2$

Н-во для Q'_n типа МакДиармида

Теорема (El-Yaniv, Pechyony, 2009)

Для любого $t \geq 0$ справедливо:

$$\mathsf{P}\left\{Q'_n - \mathbb{E}[Q'_n] \geq t\right\} \leq \exp\left\{-\frac{t^2}{2n}\left(\frac{N-1/2}{N-n}\right)\left(1-\frac{1}{2\max(n, N-n)}\right)\right\}$$

Аналогичное неравенство справедливо и для $\mathsf{P}\left\{\mathbb{E}[Q'_n] - Q'_n \geq t\right\}$.

- \+ Глобальный трансдуктивном радемахеровский анализ (*El-Yaniv and Pechyony, 2009*)
- — Локальный трансдуктивном радемахеровский анализ: нужен учет дисперсии

План доклада

Субгауссовское неравенство для Q'_n

(Толстихин, 2012)

Теорема (Толстихин, 2012)

Введем обозначение $\sigma^2 = \sup_{f \in \mathcal{F}} \mathbb{V}[f(X_1)]$. *Тогда для любого* $t \geq 0$ *справедливо:*

$$\mathsf{P}\left\{Q'_n - \mathbb{E}[Q'_n] \geq t\right\} \leq \exp\left(-\frac{(N+2)t^2}{8N^2\sigma^2}\right) \leq \exp\left(-\frac{t^2}{8n}\frac{n}{N\sigma^2}\right)$$

Аналогичное неравенство справедливо и для $\mathsf{P}\left\{\mathbb{E}[Q'_n] - Q'_n \geq t\right\}$.

- Сравнение с (El-Yaniv and Pechyony, 2009):

$$\left(\frac{N - 1/2}{N - n}\right)\left(1 - \frac{1}{2\max(n, N-n)}\right) \to \frac{n}{N}\frac{1}{\sigma^2}$$

- Теперь учитываем дисперсию $\sup_{f \in \mathcal{F}} \mathbb{V}[f(X_1)]$
- Но плохое поведение по n.

План доклада

Введение

- Неравенства концентрации: короткий обзор
- Постановка задачи и мотивация
- Метод Чернова

Раздел I: Суммы случайных величин

- Редукция Хефдинга
- Прямой анализ: метод мартингалов

Раздел II: Функции, определенные на разбиениях

- Неравенство МакДиармида
- Субгауссовское Неравенство Бобкова

Раздел III: Супремумы эмпирических процессов

- Субгауссовское неравенство
- Обобщение неравенства Талаграна

Редукция Хефдинга для Q'_n

Толстихин И. О., 2012

Лемма (Редукция Хефдинга, 1963)

Пусть $\{U_1, \dots, U_n\}$ *и* $\{W_1, \dots, W_n\}$ *вытянуты равномерно из конечного множества* d*-мерных векторов* $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_N\} \subset \mathbb{R}^d$ *с и без возвращений соответственно. Тогда для любой выпуклой функции* $f \colon \mathbb{R}^d \to \mathbb{R}$ *справедливо:*

$$\mathbb{E}\left[f\left(\sum_{i=1}^{n} W_i\right)\right] \leq \mathbb{E}\left[f\left(\sum_{i=1}^{n} U_i\right)\right]$$

Вопрос: можно ли применить редукцию Хефдинга для Q'_m?

Редукция Хефдинга для Q'_n

Толстихин И. О., 2012

- $Q'_n = \sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(Z_i)$
- Рассмотрим $\mathcal{V} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\} \in \mathbb{R}^M$, где

$$\boldsymbol{v}_j = \big(f_1(c_j), \ldots, f_M(c_j)\big)$$

- $\{W_1, \ldots, W_n\}$ вытянуты без возвращений из $\mathcal{V}$
- $\{U_1, \ldots, U_n\}$ вытянуты с возвращениями из $\mathcal{V}$

Тогда:

$$\mathbb{E}\left[e^{\lambda Q'_n}\right] = \mathbb{E}\left[\exp\left(\lambda \sup_{j=1,\ldots,M}\left(\sum_{i=1}^{n} W_i\right)_j\right)\right]$$

Лемма

Пусть $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Тогда для всех $\lambda > 0$ функция

$$F(\boldsymbol{x}) = \exp\left(\lambda \sup_{i=1,\ldots,d} x_i\right) \text{ выпукла.}$$

Редукция Хефдинга для Q'_n

Толстихин И. О., 2012

Применяя Редукцию Хефдинга получаем:

$$\begin{aligned}\mathbb{E}\left[e^{\lambda Q'_n}\right] &= \mathbb{E}\left[\exp\left(\lambda \sup_{j=1,\dots,M}\left(\sum_{i=1}^{n} W_i\right)_j\right)\right] \\ &\le \mathbb{E}\left[\exp\left(\lambda \sup_{j=1,\dots,M}\left(\sum_{i=1}^{n} U_i\right)_j\right)\right] \\ &= \mathbb{E}\left[e^{\lambda Q_n}\right],\end{aligned}$$

Неравенство Буске дает нам:

$$E\left[e^{\lambda Q_n}\right] \le e^{v\phi(\lambda)+\lambda\mathbb{E}[Q_n]}$$

Мы получаем:

$$E\left[e^{\lambda(Q'_n-\mathbb{E}[Q'_n])}\right] \le e^{v\phi(\lambda)+\lambda(\mathbb{E}[Q_n]-\mathbb{E}[Q'_n])}$$

Редукция Хефдинга для Q'_n

Толстихин И. О., 2012

$$E\left[e^{\lambda(Q'_n - \mathbb{E}[Q'_n])}\right] \le e^{v\phi(\lambda) + \lambda(\mathbb{E}[Q_n] - \mathbb{E}[Q'_n])}$$

Метод Чернова дает нам следующий результат:

Теорема (Толстихин, 2013)

Рассмотрим Q'_n и Q_n для конечного множества $\mathcal{F} = \{f_1, \dots, f_M\}$. Положим $\sigma^2 = \sup_{f \in \mathcal{F}} \mathbb{V}[f(X_1)]$, $v = n\sigma^2 + 2\mathbb{E}[Q_n]$ и для $u \ge -1$ определим $\phi(u) = e^u - u - 1$, $h(u) = (1+u)\log(1+u) - u$. Тогда для $t \ge \mathbb{E}[Q_n] - \mathbb{E}[Q'_n] \ge 0$ справедливо:

$$\begin{aligned} \mathsf{P}\left\{Q'_n - \mathbb{E}[Q'_n] \ge t\right\} &\le \exp\left(-vh\left(\frac{t - \mathbb{E}[Q_n] + \mathbb{E}[Q'_n]}{v}\right)\right) \\ &\le \exp\left(-\frac{(t - \mathbb{E}[Q_n] + \mathbb{E}[Q'_n])^2}{2v + \frac{2}{3}(t - \mathbb{E}[Q_n] + \mathbb{E}[Q'_n])}\right). \end{aligned}$$

Редукция Хефдинга для Q'_n

Толстихин И. О., 2012

Лемма (Толстихин, 2013)

Справедливо следующее неравенство:

$$0 \le \mathbb{E}[Q_n] - \mathbb{E}[Q'_n] \le 2\frac{n^3}{N}.$$

Открытые вопросы

- Локальный трансдуктивный анализ на основе полученных неравенств
- Возможно ли улучшить множитель $\frac{n}{N}$ до $\frac{N}{N-n}$ при применении н-ва Бобкова?
- Возможно ли получить равномерно более точный при $n \to N$ аналог н-ва Буске для выборок без возвращений?

Rémi Bardenet and Odalric-Ambrym Maillard. Concentration inequalities for sampling without replacement. http://arxiv.org/abs/1309.4029, 2013.

G Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57:33–45, 1962.

Sergei N. Bernstein. *Probability Theory*. Moscow-Leningrad, 4^{th} edition, 1946. In Russian.

S. Bobkov. Concentration of normalized sums and a central limit theorem for noncorrelated random variables. *Annals of Probability*, 32, 2004.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

O Bousquet. A bennett concentration inequality and its application to suprema of empirical processes. *C. R. Acad. Sci. Paris, Ser. I*, 334:495–500, 2002.

Ran El-Yaniv and Dmitry Pechyony. Transductive rademacher complexity and its applications. *Journal of Artificial Intelligence Research*, 2009.

W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58 (301):13–30, 1963.

M. Ledoux. On Talagrand's deviation inequalities for product measures. *ESAIM: Probability and Stattistics*, 1, 1996.

C. McDiarmid. On the method of bounded differences. In *Surveys in Combinatorics*, pages 148–188, Cambridge, 1989. Cambridge University Press.

R. J. Serfling. Probability inequalities for the sum in sampling without replacement. *The Annuals of Statistics*, 2(1):39–48, 1974.

Michel Talagrand. New concentration inequalities in product spaces. *Inventiones Mathematicae*, 126, 1996.

Vladimir N. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, 1998.

Konstantin V. Vorontsov. Exact combinatorial bounds on the probability of overfitting for empirical risk minimization. *Pattern Recognition and Image Analysis*, 20(3):269–285, 2010.

И. О. Толстихин. Локализация оценок избыточного риска в комбинаторной теории переобучения. In *Интеллектуализация обработки информации*, pages 54–57, 2012.