1. Напоминания основных понятий теории вероятностей и простейшая теорема концентрации для линейных функционалов от Бернуллиевских случайных величин.

1.1. **Вероятность.** Пусть $\Omega = \{\omega_1, \ldots, \omega_N\}$ — множество. Функцию $P\colon 2^\Omega \to [0,1]$ со свойствами

1) $P(\Omega) = 1$,

2) $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

называют **вероятностью** или вероятностной мерой. Для $A \subset \Omega$ число $P(A)$ называют вероятностью события A. Задать вероятность P равносильно тому, чтобы задать набор элементарных вероятностей $\{p(\omega_j)\}$ со свойствами 1) $p(\omega_j) \geq 0$, 2) $\sum_{\omega_j \in \Omega} p(\omega_j) = 1$. В этом случае $P(A) := \sum_{\omega_j \in A} p(\omega_j)$. Пара (Ω, P) называется вероятностным пространством.

Простое, но важное свойство вероятностной меры — субаддитивность, а именно,

$$P(A_1 \cup \ldots \cup A_n) \leq \sum_{j=1}^{n} P(A_j).$$

1.2. **Случайные величины.** Любая функция $X\colon \Omega \to \mathbb{R}$ называется **случайной величиной**. Вектор $X = (X_1, \ldots, X_m)$, чьи компоненты являются случайными величинами называют случайным вектором. Каждая случайная величина (и случайный вектор) порождает новое веротяностное пространство $(X(\Omega), \{p_X\})$, где набор элементарных вероятностей $\{p_X\}$, который называется **распределением** случайной величины X, задан по правилу $p_X(x) := P(X = x)$ для каждого $x \in X(\Omega)$. Случайные величины $X_1, \ldots, X_m$ называются **независимыми**, если

$$P(X_1 = x_1, \ldots, X_m = x_m) = P(X_1 = x_1) \cdot \ldots \cdot P(X_m = x_m)$$

для всех возможных наборов точек $x_1, \ldots, x_m \in X_1(\Omega) \times \ldots \times X_m(\Omega)$.

Упражнение 1.1. Проверьте, что $X_1, \ldots, X_m$ независимы тогда и только тогда, когда $P(X_1 \in A_1, \ldots, X_m \in A_m) = P(X_1 \in A_1) \cdot \ldots \cdot P(X_m \in A_m)$ для всех возможных наборов множеств $A_1, \ldots, A_m$.

Определение 1.2. Везде далее $(\varepsilon_1, \ldots, \varepsilon_m)$ обозначает набор независимых случайных величин, принимающих значения ± 1 с вероятностью $1/2$ (случайные величины, принимающие два значения называются Бернуллиевскими).

Таким образом, $(\varepsilon_1, \ldots, \varepsilon_m)$ — вектор, соответсвующий результатам независимых бросков правильной монеты. В случае, когда выпадает орел, пишем 1, если же выпадает решка, то пишем -1.

1.3. **Математическое ожидание и дисперсия случайной величины.** Для случайной величины X ее **математическим ожиданием** $\mathbb{E}[X]$ называется число

$$\mathbb{E}[X] := \sum_{x \in X(\Omega)} xP(X = x) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}).$$

Дисперсией случайной величины X называется число $\mathbb{D}[X] := \mathbb{E}\big[(X - \mathbb{E}[X])^2\big]$.

Свойства:

(i) $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$;

(ii) $X \geq Y \Rightarrow \mathbb{E}X \geq \mathbb{E}Y$;

(iii) X и Y независимые $\Rightarrow \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Упражнение 1.3. Пусть $A_1, \dots, A_k$ — попрано не пересекаются и $A_1 \cup \ldots \cup A_k = \Omega$. Пусть $X\Big|_{A_j} = a_j$. Тогда

$$\mathbb{E}[X] = \sum_{j=1}^{k} a_j P(A_j).$$

Следствие 1.4. *Пусть X — случайный вектор, тогда*

$$\mathbb{E}[f(X)] = \sum_{x \in X(\Omega)} f(x) P(X = x).$$

1.4. **Неравенство Чебышева.** Пусть $X \geq 0$, $t > 0$. Тогда $X \geq t I_{\{X \geq t\}}$, откуда по свойству (ii) математического ожидания $\mathbb{E}[X] \geq t\mathbb{E}[I_{\{X \geq t\}}] = tP(X \geq t)$. Таким образом, имеет место следующее **неравенство Чебышева**:

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

Как следствие, получается неравенство Чебышева с дисперсией

$$P(|X - \mathbb{E}[X]| \geq t) \leq \frac{D[X]}{t^2}.$$

Таким образом, каждая случайная величина в каком-то смысле сосредоточена вокруг своего среднего значения.

Упражнение 1.5. Пусть f — непрерывная функция на $[0, 1]$. Пусть $X_1, \dots, X_n$ — последовательность независимых случайных величин, соответствующих последовательности независимых бросков неправильной монеты с вероятностью успеха p, т.е. $P(X_j = 1) = p$, $P(X_j = 0) = 1 - p$. Пусть $S_n = X_1 + \ldots + X_n$ — количество орлов при n бросках. Докажите, что

$$\sup_{p \in [0,1]} \Big| \mathbb{E}\big[f(\tfrac{S_n}{n})\big] - f(p) \Big| \leq \sup_{p \in [0,1]} \mathbb{E}\Big[\big| f(\tfrac{S_n}{n}) - f(p) \big| \Big] \to 0.$$

Проверьте, что $p \mapsto \mathbb{E}\big[f(\frac{S_n}{n})\big]$ — многочлен и тем самым доказана теорема Вейерштрасса о равномерном приближении непрерывной на отрезке функции многочленами.

(**HINT**: Разбейте математическое ожидание на два слагаемых, в первом все исходы, где $|\frac{S_n}{n} - p| \geq \delta$, и во втором все исходы, где $|\frac{S_n}{n} - p| < \delta$. Тогда в первом слагаемом можно воспользоваться неравенством Чебышева, а во втором равномерной непрерывностью функции f.)

1.5. **Линейные функционалы от Бернуллиевских случайных величин.** Посмотрим, что можно сказать про случайные величины вида $S := \sum_{j=1}^{m} a_j \varepsilon_j$. Заметим, что $\mathbb{E}[S] = 0$ и

$$\mathbb{D}[S] = \mathbb{E}[(a_1\varepsilon_1 + \ldots a_m\varepsilon_m)^2] = \sum_{j=1}^{m} a_j^2 \mathbb{E}[\varepsilon_j^2] + \sum_{j \neq k} a_j a_k \mathbb{E}[\varepsilon_k \varepsilon_k] = \sum_{j=1}^{m} a_j^2,$$

где мы воспользовались тем, что $\varepsilon_j^2 = 1$, $\mathbb{E}[\varepsilon_k \varepsilon_k] = \mathbb{E}[\varepsilon_k]\mathbb{E}[\varepsilon_k] = 0$. По неравенству Чебышева

$$P\Big(\Big| \sum_{j=1}^{m} a_j \varepsilon_j \Big| \geq t \Big) \leq \frac{a_1^2 + \ldots + a_m^2}{t^2}.$$

Оказывается, в реальности, скорость убывания данной вероятности по t гораздо быстрее полиномиальной.

Теорема 1.6. *Имеет место неравенство*

$$P\Big(\Big| \sum_{j=1}^{m} a_j \varepsilon_j \Big| \geq t \Big) \leq 2e^{-\frac{t^2}{2(a_1^2 + \ldots + a_m^2)}}.$$

Доказательство. Основная идея для доказательства этой теоремы, и многих других, которые будут нам встречаться, заключается в переходе к функции $\exp(\lambda \sum\limits_{j=1}^{m} a_j\varepsilon_j)$. Удобство такого перехода легко понять с вероятностной точки зрения: экспонента переведет нам сумму в произведение, а математическое ожидание произведения независимых случайных величин есть произведение математических ожиданий.

Применим описанный подход в нашем случае. Пусть $t, \lambda > 0$, тогда

$$P\Big(\sum_{j=1}^{m} a_j\varepsilon_j \geq t\Big) = P\Big(\exp\Big(\lambda\sum_{j=1}^{m} a_j\varepsilon_j\Big) \geq e^{\lambda t}\Big) \leq e^{-\lambda t}\mathbb{E}\Big[\exp\Big(\lambda\sum_{j=1}^{m} a_j\varepsilon_j\Big)\Big] = e^{-\lambda t}\prod_{j=1}^{m}\mathbb{E}\big[e^{\lambda a_j\varepsilon_j}\big],$$

где во втором переходе мы применили неравенство Чебышева. Вычислим математическое ожидание:

$$\mathbb{E}\big[e^{\lambda a_j\varepsilon_j}\big] = \frac{e^{\lambda a_j} + e^{-\lambda a_j}}{2} = \operatorname{ch}(\lambda a_j).$$

Заметим, что $\operatorname{ch}\alpha \leq e^{\frac{\alpha^2}{2}}$, т.к.

$$\operatorname{ch}\alpha = \frac{1}{2}(e^{\alpha} + e^{-\alpha}) = \sum_{k=0}^{\infty}\frac{1}{(2k)!}\alpha^{2k} \leq \sum_{k=0}^{\infty}\frac{1}{(2k)!!}\alpha^{2k} = \sum_{k=0}^{\infty}\frac{1}{k!}\Big(\frac{\alpha^2}{2}\Big)^k = e^{\frac{\alpha^2}{2}}.$$

Таким образом,

$$P\Big(\sum_{j=1}^{m} a_j\varepsilon_j \geq t\Big) \leq e^{-\lambda t}\prod_{j=1}^{m}\mathbb{E}\big[e^{\lambda a_j\varepsilon_j}\big] \leq e^{-\lambda t + \frac{\lambda^2}{2}(a_1^2+\ldots+a_m^2)}.$$

Остается минимимзировать правую часть по параметру λ, т.е. найти точку минимума функции $\lambda \mapsto -\lambda t + \frac{\lambda^2}{2}(a_1^2 + \ldots + a_m^2)$. Это точка $\lambda = \frac{t}{a_1^2+\ldots+a_m^2}$, которую мы и подставляем в правую часть, чтобы получить оценку

$$P\Big(\sum_{j=1}^{m} a_j\varepsilon_j \geq t\Big) \leq \exp\Big(-\frac{t^2}{2(a_1^2+\ldots+a_m^2)}\Big).$$

Аналогично,

$$P\Big(\sum_{j=1}^{m} a_j\varepsilon_j \leq -t\Big) \leq \exp\Big(-\frac{t^2}{2(a_1^2+\ldots+a_m^2)}\Big)$$

и

$$P\Big(\Big|\sum_{j=1}^{m} a_j\varepsilon_j\Big| \geq t\Big) \leq 2e^{-\frac{t^2}{2(a_1^2+\ldots+a_m^2)}},$$

что и было заявлено. □

v

Упражнение 1.7. Пусть $X_1, \ldots, X_m$ — независимые случайные величины, принимающие значения на отрезке $[-1, 1]$ (т.е. $X(\omega) \in [-1, 1]$). Пусть $\mathbb{E}[X_j] = 0\ \forall j \in \{1, \ldots, m\}$. Докажите, что

$$P\Big(\Big|\sum_{j=1}^{m} a_jX_j\Big| \geq t\Big) \leq 2e^{-\frac{t^2}{2(a_1^2+\ldots+a_m^2)}}.$$

Как следствие получите следующее неравенство Хёфдинга: пусть $Y_1, \ldots, Y_m$ — независимые случайные величины, принимающие значения в отрезках $[\alpha_j, \beta_j]$ соответственно. Тогда

$$P\Big(\Big|\sum_{j=1}^{m} Y_j - \mathbb{E}[Y_j]\Big| \geq t\Big) \leq 2e^{-\frac{t^2}{2(|\beta_1-\alpha_1|^2+\ldots+|\beta_m-\alpha_m|^2)}}.$$

(**HINT**: Заметьте, что из-за выпуклости экспоненты, при каждом $x \in [-1,1]$ имеет место неравенство $\exp(\alpha x) = \exp\left(\frac{1+x}{2}\alpha + \frac{1-x}{2}(-\alpha)\right) \le \frac{1+x}{2}e^{\alpha} + \frac{1-x}{2}e^{-\alpha} = \mathrm{ch}\alpha + x\mathrm{sh}\alpha$.)

2. Лемма Джонсона–Линденштрауса о малом искажении.

Из доказательства последней теоремы в предыдущем разделе мы увидели, что важную роль играет следующая функция.

Определение 2.1. Для случайной величины S и $\lambda \in \mathbb{R}$ пусть $\psi_S(\lambda) := \ln \mathbb{E}[e^{\lambda S}]$.

В частности, при доказательства предыдущей теоремы мы установили, что

$$\psi_S(\lambda) \le \tfrac{\lambda^2}{2} \quad \forall \lambda \in \mathbb{R}$$

в случае, когда $S = \sum\limits_{j=1}^{m} a_j \varepsilon_j$ и $a_1^2 + \ldots + a_m^2 = 1$.

Анализируя проделанные выше рассуждения, легко понять, что имеет место общее **неравенство Чернова**:

$$P(S \ge t) \le \exp\bigl(-\sup_{\lambda>0}(\lambda t - \psi_S(\lambda))\bigr).$$

Отметим одно простое, но полезное свойство функции ψ_S в случае, когда S — есть сумма независимых случайных величин.

Предложение 2.2. *Пусть* $S = \sum\limits_{j=1}^{m}(X_j - \mathbb{E}X_j)$, *где* X_j *— независимые случайные величины. Тогда*

$$\psi_S(\lambda) \le \sum_{j=1}^{m} \mathbb{E}\bigl[e^{\lambda X_j} - \lambda X_j - 1\bigr].$$

Доказательство. В силу независимости X_j имеет место тождество

$$\psi_S(\lambda) = \ln \mathbb{E}\Bigl[\exp\Bigl(\lambda \sum_{j=1}^{m}(X_j - \mathbb{E}X_j)\Bigr)\Bigr] = \ln\Bigl(\prod_{j=1}^{m} \mathbb{E}\bigl[\exp\bigl(\lambda X_j - \mathbb{E}[\lambda X_j]\bigr)\bigr]\Bigr) =$$

$$= \sum_{j=1}^{m} \ln\bigl(e^{-\mathbb{E}[\lambda X_j]}\mathbb{E}\bigl[e^{\lambda X_j}\bigr]\bigr) = \sum_{j=1}^{m}\Bigl(-\mathbb{E}[\lambda X_j] + \ln\bigl(\mathbb{E}\bigl[e^{\lambda X_j}\bigr]\bigr)\Bigr).$$

Воспользовавшись неравенством для логарифма $\ln(x) \le x - 1$, получаем, что

$$\ln\bigl(\mathbb{E}\bigl[e^{\lambda X_j}\bigr]\bigr) \le \mathbb{E}\bigl[e^{\lambda X_j}\bigr] - 1 = \mathbb{E}\bigl[e^{\lambda X_j} - 1\bigr],$$

что завершает доказательство. □

Подумаем теперь над таким вопросом. *Пусть* $V = \{v^1, \ldots, v^N\} \subset \mathbb{R}^m$ *— конечное множество многомерных данных большой размерности (т.е.* m *— очень большое число). Сколь сильно можно уменьшить размерность этих данных (сжать данные) так, чтобы относительная структура множества* V *не сильно изменилась после сжатия?* Сначала придадим этому вопросу более точную математическую формулировку. Мы хотим отобразить исходное множество данных V в пространство меньшей размерности d с помощью некоторого отображения $f\colon \mathbb{R}^m \to \mathbb{R}^d$ таким образом, чтобы попарные расстояния не сильно поменялись, т.е.

$$(1-\varepsilon)\|v' - v''\| \le \|f(v') - f(v'')\| \le (1+\varepsilon)\|v' - v''\| \quad \forall v', v'' \in V, \tag{2.1}$$

где $\|x\| := \bigl(\sum x_j^2\bigr)^{1/2}$. Оказывается, размерность d можно взять независимой от размерности исходных данных m! Формальный ответ на наш вопрос дает следующая теорема Джонсона–Линденштрауса.

Теорема 2.3 (лемма Джонсона–Линденштрауса о малом искажении)**.** *Пусть* $\varepsilon \in (0,1)$. *Пусть* $V = \{v^1, \dots, v^N\} \subset \mathbb{R}^m$. *Тогда для каждого* $d \geq c\varepsilon^{-2}\ln(2N)$ *(*c *— некотороя числовая постоянна) существует такое (линейное) отображение* $f\colon \mathbb{R}^m \to \mathbb{R}^d$, *что*

$$(1-\varepsilon)\|v'-v''\|^2 \leq \|f(v')-f(v'')\|^2 \leq (1+\varepsilon)\|v'-v''\|^2 \quad \forall v', v'' \in V.$$

Доказательство.

1) Мы будем искать отображение f в виде случайного линейного отображения. Пусть $\{\varepsilon_{i,j}\}$, $i \in \{1,\dots,d\}$, $j \in \{1,\dots,m\}$, — набор независимых случайных величин со значениями ± 1, и принимающими эти значения с вероятностью $1/2$. Пусть

$$f_i(v) := \sum_{j=1}^m \varepsilon_{i,j} v_j, \quad f(v) := \bigl(\tfrac{1}{\sqrt{d}} f_i(v)\bigr)_{i=1}^d. \tag{2.2}$$

Посмотрим на то, что происходит в среднем. Заметим, что

$$\mathbb{E}\bigl[f_i(v)^2\bigr] = \sum_{j=1}^m v_j^2 \mathbb{E}[\varepsilon_{i,j}^2] + \sum_{j \neq j'} v_j v_{j'} \mathbb{E}[\varepsilon_{i,j}]\mathbb{E}[\varepsilon_{i,j'}] = \sum_{j=1}^m v_j^2 = \|v\|^2.$$

Таким образом,

$$\mathbb{E}\bigl[\|f(v)\|^2\bigr] = \frac{1}{d}\sum_{i=1}^d \mathbb{E}\bigl[(f_i(v))^2\bigr] = \|v\|^2.$$

Пусть теперь

$$T := \left\{ \frac{v'-v''}{\|v'-v''\|} : v', v'' \in V, i \neq j \right\} \subset \mathbb{R}^m.$$

Сразу заметим, что $\#T \leq \frac{N(N-1)}{2}$ и что $\|v\| = 1$ $\forall v \in T$. В силу линейности отображения f, нам достаточно показать, что

$$\sup_{v \in T} \bigl| \|f(v)\|^2 - 1 \bigr| \leq \varepsilon.$$

Заметим, что

$$\|f(v)\|^2 - 1 = \frac{1}{d}\sum_{i=1}^d (f_i(v)^2 - 1).$$

2) Попробуем опять понять, как ведет себя вероятность

$$P\Bigl(\Bigl|\sum_{i=1}^d (f_i(v)^2 - 1)\Bigr| \geq t\Bigr)$$

при фиксированном $v \in \mathbb{R}^m$, $\|v\| = 1$. Для этого исследуем функцию $\psi_{S_1}(\lambda)$, где

$$S_1 := \sum_{i=1}^d (f_i(v)^2 - 1) = \sum_{i=1}^d \bigl(f_i(v)^2 - \mathbb{E}[f_i(v)^2]\bigr).$$

По наблюдению выше

$$\psi_{S_1}(\lambda) \leq \sum_{i=1}^d \mathbb{E}\bigl[e^{\lambda f_i(v)^2} - \lambda f_i(v)^2 - 1\bigr].$$

Заметим, что

$$e^{\lambda f_i(v)^2} - \lambda f_i(v)^2 - 1 = \sum_{k=2}^\infty \frac{1}{k!}\lambda^k f_i(v)^{2k},$$

откуда

$$\mathbb{E}\bigl[e^{\lambda f_i(v)^2} - \lambda f_i(v)^2 - 1\bigr] \leq \sum_{k=2}^\infty \frac{1}{k!}\lambda^k \mathbb{E}\bigl[f_i(v)^{2k}\bigr].$$

Мы знаем, что $P(|f_i(v)| \geq \tau) \leq 2e^{-\frac{\tau^2}{2}}$ $\forall \tau > 0$. Поэтому

$$\mathbb{E}\big[f_i(v)^{2k}\big] = \sum_{s=0}^{\infty} \mathbb{E}\big[f_i(v)^{2k} I_{\{s \leq |f_i(v)| < s+1\}}\big] \leq \sum_{s=0}^{\infty} (s+1)^{2k} P(s \leq |f_i(v)| < s+1) =$$

$$= \sum_{s=0}^{\infty} (s+1)^{2k} \big(P(|f_i(v)| \geq s) - P(|f_i(v)| \geq s+1)\big) =$$

$$= \sum_{s=0}^{\infty} (s+1)^{2k} P(|f_i(v)| \geq s) - \sum_{s=1}^{\infty} s^{2k} P(|f_i(v)| \geq s) = 1 + \sum_{s=1}^{\infty} \big[(s+1)^{2k} - s^{2k}\big] P(|f_i(v)| \geq s) \leq$$

$$\leq 1 + 2\sum_{s=1}^{\infty} 2k(s+1)^{2k-1} e^{-s^2/2}.$$

Воспользовавшись тем, что $e^{s^2/2} \geq \frac{1}{(k+1)!}(\frac{s^2}{2})^{k+1}$, получаем оценку

$$1 + 2\sum_{s=1}^{\infty} 2k(s+1)^{2k-1} e^{-s^2/2} \leq 1 + (4k)(k+1)! 2^{k+1} \sum_{s=1}^{\infty} (s+1)^{2k-1} s^{-2k-2} \leq k! C^k,$$

где $C > 0$ некоторая числовая постоянна. Таким образом,

$$\mathbb{E}\big[e^{\lambda f_i(v)^2} - \lambda f_i(v)^2 - 1\big] \leq \sum_{k=2}^{\infty} (C\lambda)^k = \frac{C^2\lambda^2}{1 - C\lambda}$$

при $\lambda \in (0, 1/C)$. Т.е.

$$\psi_{S_1}(\lambda) \leq \frac{C^2 d\lambda^2}{1 - C\lambda}$$

при $\lambda \in (0, 1/C)$, откуда

$$\sup_{\lambda>0}(\lambda t - \psi_S(\lambda)) \geq \sup_{\lambda\in(0,1/C)} \Big(\lambda t - \frac{C^2 d\lambda^2}{1 - C\lambda}\Big) = d \sup_{\lambda\in(0,1/C)} \Big(C\lambda \frac{t}{Cd} - \frac{(C\lambda)^2}{1 - C\lambda}\Big) =$$

$$= d \sup_{s\in(0,1)} \Big(s\frac{t}{Cd} - \frac{s^2}{1-s}\Big) \geq d\Big(\frac{t}{2Cd+t} \cdot \frac{t}{Cd} - \frac{\frac{t^2}{(2Cd+t)^2}}{\frac{2Cd}{2Cd+t}}\Big) = \frac{t^2}{2C(2Cd+t)},$$

где в предпоследнем неравенстве мы оценили супремум подстановкой $s = \frac{t}{2Cd+t}$. По неравенству Чернова

$$P\Big(\sum_{i=1}^{d} (f_i(v)^2 - 1) \geq t\Big) \leq \exp\Big(-\frac{t^2}{2C(2Cd+t)}\Big).$$

Посмотрим теперь на $S_2 = -\sum_{i=1}^{d}(f_i(v)^2 - 1) = \sum_{i=1}^{d}\big(-f_i(v)^2 - \mathbb{E}[-f_i(v)^2]\big)$ Как и в случае S_1, при $\lambda > 0$, имеет место неравенство

$$\psi_{S_2}(\lambda) \leq \sum_{i=1}^{d} \mathbb{E}\big[e^{-\lambda f_i(v)^2} - (-\lambda f_i(v)^2) - 1\big].$$

Но в данном случае все дальнейшее рассмотрение гораздо проще, т.к. имеет место следующее неравенство.

Упражнение 2.4. Проверьте, что $e^{-x} + x - 1 \leq \frac{x^2}{2}$ при $x > 0$.

Мы получаем, что

$$\psi_{S_2}(\lambda) \leq \sum_{i=1}^{d} \frac{\lambda^2}{2} \mathbb{E}\big[f_i(v)^4\big] \leq \sum_{i=1}^{d} \sum_{k=2}^{\infty} \frac{1}{k!} \lambda^k \mathbb{E}\big[f_i(v)^{2k}\big].$$

Эту сумму мы уже оценили выше, поэтому

$$P\Big(\Big|\sum_{i=1}^{d}(f_i(v)^2-1)\Big|\ge t\Big)\le 2\exp\Big(-\frac{t^2}{2C(2Cd+t)}\Big).$$

3) Возвращаясь к нашей исходной задаче, получаем, что

$$P\big(\sup_{v\in T}\big|\|f(v)\|^2-1\big|>\varepsilon\big)=P\Big(\sup_{v\in T}\Big|\sum_{i=1}^{d}(f_i(v)^2-1)\Big|\ge d\varepsilon\Big)=$$

$$=P\Big(\bigcup_{v\in T}\Big\{\Big|\sum_{i=1}^{d}(f_i(v)^2-1)\Big|\ge d\varepsilon\Big\}\Big)\le\sum_{v\in T}P\Big(\Big|\sum_{i=1}^{d}(f_i(v)^2-1)\Big|\ge d\varepsilon\Big)\le$$

$$\le \#T\cdot 2\exp\Big(-\frac{d^2\varepsilon^2}{2C(2Cd+d\varepsilon)}\Big)\le N^2\exp\Big(-\frac{d\varepsilon^2}{2C(2C+\varepsilon)}\Big)\le N^2\exp\Big(-\frac{d\varepsilon^2}{C_1}\Big),$$

где $C_1:=2C(2C+1)$. Т.к. мы хотим, чтобы существовало отображение f, для которого $\sup_{v\in T}\big|\|f(v)\|^2-1\big|\le\varepsilon$, то давайте попросим даже большего. Скажем, посмотрим, когда такое отображение (указанного выше вида (2.2)) существует с вероятностью не менее $1/2$. Для этого посмотрим, при каких d

$$N^2\exp\Big(-\frac{d\varepsilon^2}{C_1}\Big)\le 1/2.$$

Такое неравенство выполнено при всех $d\ge C_1\varepsilon^{-2}\ln(2N^2)$, Ясно, что тогда оно выполнено и при всех $d\ge C_2\varepsilon^{-2}\ln(2N)$, где $C_2=2C_1$. Теорема доказана. $\Box$

Упражнение 2.5. Пусть $\delta\in(0,1)$, $d\ge c\varepsilon^{-2}\ln(\frac{2N}{\sqrt{\delta}})$. Проверьте, что с вероятностью не менее, чем $1-\delta$, случайное отображение вида (2.2) будет отображением со свойством (2.1). Т.е. искомое в лемме Джонсона–Линденштрауса отображение можно просто строить (с большой веротяностью) с помощью случайного выбора знаков.

3. Энтропия.

Определение 3.1. Пусть X — некоторый случайный вектор, задающий распределение вероятностей $P(X=x)=p_X(x)$ на множестве $Q=X(\Omega)$. **Энтропией (по Шеннону)** случайной величины X называется число

$$H(X):=\mathbb{E}[-\ln p_X(X)]=-\sum_{x\in X(\Omega)}p_X(x)\log p_X(x).$$

Замечание 3.2. Заметим, что $H(X)\ge 0$.

Определение 3.3. Пусть p и q два вероятностных распределения на множестве Q. **Расстоянием Кульбака–Лейблера** (или относительной энтропией) распределения p относительно распределения q называется величина

$$D(p\|q):=\sum_{x\in Q}p(x)\ln\frac{p(x)}{q(x)}$$

(предполагаем, что $p(x)=0$, если $q(x)=0$).

Упражнение 3.4. Проверьте, что $\ln s\ge 1-\frac{1}{s}$ (причем равенство достигается только в точке $s=1$).

Замечание 3.5. Заметми, что

$$D(p\|q) := \sum_{x\in Q} p(x)\ln\frac{p(x)}{q(x)} \geq \sum_{x\in Q} p(x)\Big(1-\frac{q(x)}{p(x)}\Big) = 0,$$

причем равенство достигается только в случае $p(x)=q(x)\ \forall x\in Q$. Таким образом, расстояние Кульбака–Лейблера измеряет, отклонение распределения p от распределения q.

Как следствие,

$$\sum_{x\in X(\Omega)} p_X(x)\ln\frac{p_X(x)}{\frac{1}{\#X(\Omega)}} = D(p_X\|p_U)\geq 0,$$

где U — случайный вектор с равномерным распределением на множестве $Q=X(\Omega)$, т.е. $P(U=x)=\frac{1}{\#X(\Omega)}$. Таким образом, $H(X)\leq \ln\#X(\Omega)=H(U)$ и энтропия (по Шеннону) измеряет уклонение распределения от равномерного распределения.

Определение 3.6. Пусть $X=(X_1,\ldots,X_m)$ — случайный вектор. Введем следующее обозначение: $X^{(j)}=(X_1,\ldots,X_{j-1},X_{j+1},\ldots,X_m)$.

Теорема 3.7 (Неравенство Хана)**.** *Пусть $X=(X_1,\ldots,X_m)$ — случайный вектор. Тогда*

$$H(X)\leq\frac{1}{m-1}\sum_{j=1}^m H(X^{(j)}).$$

Доказательство. Зафиксируем j. Пусть $(X_1,\ldots,X_{j-1})=U$, $(X_{j+1},\ldots,X_m)=V$, $(x_1,\ldots,x_{j-1})=u$, $(x_{j+1},\ldots,x_m)=v$. Заметим, что

$$H(X)-H(X^{(j)}) = -\sum_x p_X(x)\ln p_X(x)+\sum_{u,v}p_{(U,V)}(u,v)\ln p_{(U,V)}(u,v) = -\sum_x p_X(x)\ln\frac{p_X(x)}{p_{(U,V)}(u,v)},$$

где мы использовали тождество

$$\begin{aligned}p_{(U,V)}(u,v) &= P\big((U,V)=(u,v)\big) = P(X_1=x_1,\ldots,X_{j-1}=x_{j-1},X_{j+1}=x_{j+1},\ldots,X_m=x_m) = \\ &= \sum_{x_j}P(X_1=x_1,\ldots,X_{j-1}=x_{j-1},X_j=x_j,X_{j+1}=x_{j+1},\ldots,X_m=x_m) = \sum_{x_j}p_X(x).\end{aligned}$$

Заметим, что

$$-\sum_x p_X(x)\ln\frac{p_X(x)}{p_{(U,V)}(u,v)} = -\sum_x p_X(x)\ln\frac{p_{(U,X_j)}(u,x_j)}{p_U(u)} - \sum_x p_X(x)\ln\frac{p_U(u)p_X(x)}{p_{(U,X_j)}(u,x_j)p_{(U,V)}(u,v)}$$

и что

$$\begin{aligned}\sum_x p_X(x)\ln\frac{p_U(u)p_X(x)}{p_{(U,X_j)}(u,x_j)p_{(U,V)}(u,v)} &\geq \sum_x p_X(x)\Big(1-\frac{p_{(U,X_j)}(u,x_j)p_{(U,V)}(u,v)}{p_U(u)p_X(x)}\Big) = \\ &= 1-\sum_{u,x_j,v}\frac{p_{(U,X_j)}(u,x_j)p_{(U,V)}(u,v)}{p_U(u)} = 1-\sum_{u,x_j}p_{(U,X_j)}(u,x_j)=0,\end{aligned}$$

где мы воспользовались тем, что $\sum\limits_v p_{(U,V)}(u,v)=p_U(u)$. Таким образом,

$$H(X)-H(X^{(j)})\leq -\sum_x p_X(x)\ln\frac{p_{(U,X_j)}(u,x_j)}{p_U(u)} = -\sum_x p_X(x)\ln\frac{p_{(X_1,\ldots,X_j)}(x_1,\ldots,x_j)}{p_{(X_1,\ldots,X_{j-1})}(x_1,\ldots,x_{j-1})}.$$

Суммируя по j, получаем оценку

$$\sum_{j=1}^{m}\big(H(X)-H(X^{(j)})\big) \le -\sum_{j=1}^{m}\sum_{x} p_X(x)\ln\frac{p_{(X_1,\dots,X_j)}(x_1,\dots,x_j)}{p_{(X_1,\dots,X_{j-1})}(x_1,\dots,x_{j-1})} =$$

$$= -\sum_{x} p_X(x)\ln\Big[\prod_{j=1}^{m}\frac{p_{(X_1,\dots,X_j)}(x_1,\dots,x_j)}{p_{(X_1,\dots,X_{j-1})}(x_1,\dots,x_{j-1})}\Big] = -\sum_{x} p_X(x)\ln p_X(x) = H(X).$$

Это равносильно заявленной в условии оценке. Теорема доказана. □

Замечание 3.8. Посмотрим на неравенство Хана в конкретном случае $Q = \{-1,1\}^m$, $U = (\varepsilon_1,\dots,\varepsilon_m)$, $X = (X_1,\dots,X_m)$ — некоторый случайный вектор со значениями в Q. Пусть $f(x) := \frac{p_X(x)}{p_U(x)} = \frac{p_X(x)}{2^{-m}}$. Тогда, по неравенству Хана

$$\mathbb{E}[f(U)\ln f(U)] = \sum_{x\in Q} p_X(x)\ln\frac{p_X(x)}{p_U(x)} = D(p_X\|p_U) = \sum_{x\in Q} p_X(x)\ln p_X(x) - \sum_{x\in Q} p_X(x)\ln 2^{-m} =$$

$$= -H(X) - \ln 2^{-m} \ge -\frac{1}{m-1}\sum_{j=1}^{m} H(X^{(j)}) - \ln 2^{-m} = -\frac{1}{m-1}\sum_{j=1}^{m} H(X^{(j)}) - \frac{m}{m-1}\ln 2^{-(m-1)} =$$

$$= \frac{1}{m-1}\sum_{j=1}^{m}\big(-H(X^{(j)}) - \ln 2^{-(m-1)}\big) = \frac{1}{m-1}\sum_{j=1}^{m} D(p_{X^{(j)}}\|p_{U^{(j)}}).$$

Заметим, что

$$D(p_{X^{(j)}}\|p_{U^{(j)}}) = \mathbb{E}[f_j(U^{(j)})\ln f_j(U^{(j)})],$$

где

$$f_j(x^{(j)}) = \frac{p_{X^{(j)}}(x^{(j)})}{2^{-(m-1)}} = \sum_{x_j\in\{-1,1\}} \frac{p_X(x)}{2^{-m}}\cdot\frac{1}{2} = \mathbb{E}\big[f(x_1,\dots,x_{j-1},\varepsilon_j,x_{j+1},\dots,x_m)\big].$$

Усреднение только по $j-$ координате обозначим $\mathbb{E}^{(j)}$. Тогда

$$\mathbb{E}[f(U)\ln f(U)] \ge \frac{1}{m-1}\sum_{j=1}^{m}\mathbb{E}\big[\mathbb{E}^{(j)}[f(U)]\ln\mathbb{E}^{(j)}[f(U)]\big],$$

что, после арифметических преобразований и тождества $\mathbb{E}[f(U)\ln f(U)] = \mathbb{E}\big[\mathbb{E}^{(j)}[f(U)\ln f(U)]\big]$ принимает вид

$$\mathbb{E}[f(U)\ln f(U)] \le \mathbb{E}\Big[\sum_{j=1}^{m}\mathbb{E}^{(j)}[f(U)\ln f(U)] - \mathbb{E}^{(j)}[f(U)]\ln\mathbb{E}^{(j)}[f(U)]\Big].$$

Заметим, что на функцию f (помимо условия неотрицательности) наложено условие $\mathbb{E}[f(U)] = 1$. Пусть теперь f — произвольная неотрицательная функция на Q. Тогда для функции $\widetilde{f} := \frac{f}{\mathbb{E}[f(U)]}$ выполнено неравенство

$$\mathbb{E}[\widetilde{f}(U)\ln\widetilde{f}(U)] \le \mathbb{E}\Big[\sum_{j=1}^{m}\mathbb{E}^{(j)}[\widetilde{f}(U)\ln\widetilde{f}(U)] - \mathbb{E}^{(j)}[\widetilde{f}(U)]\ln\mathbb{E}^{(j)}[\widetilde{f}(U)]\Big],$$

которое после умножения на $\mathbb{E}[f(U)]$ принимает вид (**проверьте это!**)
(3.1)

$$\mathbb{E}[f(U)\ln f(U)] - \mathbb{E}[f(U)]\ln\mathbb{E}[f(U)] \le \mathbb{E}\Big[\sum_{j=1}^{m}\mathbb{E}^{(j)}[f(U)\ln f(U)] - \mathbb{E}^{(j)}[f(U)]\ln\mathbb{E}^{(j)}[f(U)]\Big].$$

Определение 3.9. Пусть X — произвольный случайный вектор, $f \geq 0$. **Энтропией** функции f называется величина

$$\mathrm{Ent}(f) := \mathbb{E}[f(X)\ln f(X)] - \mathbb{E}[f(X)]\ln \mathbb{E}[f(X)].$$

Замечание 3.10. Напомним, что в нашем примере с булевым кубом, в случае $\mathbb{E}[f] = 1$,

$$\mathrm{Ent}(f) = D(f \cdot p_U \| p_U).$$

Т.е. для функций с единичным средним, энтрипия показывает степень уклонения функции от константы 1, т.к. расстояние Кульбака–Лейблера показывает степень отклонения $f \cdot p_U$ от p_U. Таким образом, в общем случае, для неотрицательной функции f энтропия функции измеряет близость функции к своему среднему, т.е. к (некоторой) константе.

Свойство (3.1) называется свойством **субаддитивности энтропии**. На самом деле, неравенство (3.1) справедливо не только для равномерного распределения на булевом кубе, но и в более общем случае.

Теорема 3.11. *Пусть $X = (X_1, \ldots, X_m)$ — случайный вектор с независимыми компонентами. Тогда для произвольной $f \geq 0$ имеет место неравенство*

$$\mathrm{Ent}(f) \leq \mathbb{E}\Big[\sum_{j=1}^{m} \mathrm{Ent}^{(j)}(f)\Big],$$

где $\mathrm{Ent}^{(j)}(f) := \mathbb{E}^{(j)}[f(X)\ln f(X)] - \mathbb{E}^{(j)}[f(X)]\ln \mathbb{E}^{(j)}[f(X)]$.

4. Логарифмическое неравенство Соболева на булевом кубе. Рассуждение Хербста. Концентрация для липшицевых функций.

Как мы уже поняли, энтропия функции является способом измерения отличия неотрицательной функции от постоянной. Другим естественным способом померить отличие функции от постоянной является производная, а в многомерном случае градиент $\nabla f(x) := (\partial_1 f(x), \ldots, \partial_m f(x))$, где $\partial_j f(x)$ — производная по j-й переменной при фиксированных остальных. На булевом кубе роль частной производной играет следующая величина. Для точки $x = (x_1, \ldots, x_m) \in Q := \{-1, 1\}^m$ пусть $\overline{x}^j := (x_1, \ldots x_{j-1}, -x_j, x_{j+1}, \ldots, x_m)$.

Определение 4.1. Пусть $f \colon \{-1, 1\}^m \to \mathbb{R}$. Тогда

$$\partial_j f(x) := \frac{f(\overline{x}^j) - f(x)}{2}$$

и $\nabla f(x) := (\partial_1 f(x), \ldots, \partial_m f(x))$.

Теорема 4.2 (Логарифмическое неравенство Соболева)**.** *Пусть $f \colon \{-1, 1\}^m \to \mathbb{R}$. Тогда*

$$\mathrm{Ent}(f^2) \leq 2\mathbb{E}\big[\|\nabla f(U)\|^2\big] = 2\mathbb{E}\Big[\sum_{j=1}^{m} |\partial_j f(U)|^2\Big].$$

Доказательство. Из-за свойства субаддитивности энтропии (3.1) нам достаточно проверить утверждение только для функций одной переменной на множестве $\{-1, 1\}$. Пусть $f(1) = a$, $f(-1) = b$. Нам нужно проверить, что

$$a^2 \ln a^2 \frac{1}{2} + b^2 \ln b^2 \frac{1}{2} - \frac{a^2 + b^2}{2} \ln \frac{a^2 + b^2}{2} \leq \frac{(a-b)^2}{2}.$$

Т.к. $(|a| - |b|)^2 \leq (a-b)^2$, то без ограничения общности можно считать, что $a, b > 0$. Из-за симметрии можем считать, что $a \geq b$. Пусть

$$h(a) := (a-b)^2 + (a^2 + b^2)\ln \frac{a^2 + b^2}{2} - a^2 \ln a^2 - b^2 \ln b^2.$$

Мы хотим показать, что $h(a) \geq 0$ при $a \geq b$. Заметим, что $h(b) = 0$. Посмотрим на производную

$$h'(a) = 2(a-b) + 2a \ln \frac{a^2+b^2}{2a^2}.$$

Заметим, что $h'(b) = 0$. Посмотрим на вторую производную

$$h''(a) = 2+2\ln\frac{a^2+b^2}{2a^2} - \frac{4b^2}{a^2+b^2} = 2-2\ln\frac{2a^2}{a^2+b^2} - 4 + \frac{4a^2}{a^2+b^2} = 2\Big(\frac{2a^2}{a^2+b^2} - 1 - \ln\frac{2a^2}{a^2+b^2}\Big) \geq 0,$$

где мы воспользовались оценкой $\ln x \leq x - 1$. Таким образом, h' не убывает при $a \geq b$, значит $h'(a) \geq h'(b) = 0$. Значит h не убывает при $a \geq b$, значит $h(a) \geq h(b) = 0$. Таким образом, теорема доказана. $\square$

Вернемся к нашему исходному вопросу о концентрации распределения. Как мы уже знаем, для исследования вопросов концентрации случайной величины $S = f(\varepsilon_1, \ldots, \varepsilon_m)$ важную роль играет функция $\psi_S(\lambda) := \ln \mathbb{E}[e^{\lambda S}]$. Попробуем использовать для оценки этой функции логарифмическое неравенство Соболева. Пусть $g(x) = e^{\lambda f(x)/2}$, $\lambda \in \mathbb{R}$. Тогда

$$\mathrm{Ent}(g^2) = \mathrm{Ent}(e^{\lambda f}) = \lambda \mathbb{E}\big[S e^{\lambda S}\big] - \mathbb{E}\big[e^{\lambda S}\big] \ln \mathbb{E}\big[e^{\lambda S}\big].$$

Таким образом,

$$\mathrm{Ent}(g^2) = \lambda (e^{\psi_S(\lambda)})' - e^{\psi_S(\lambda)} \psi_S(\lambda) = \big(\lambda \psi_S'(\lambda) - \psi_S(\lambda)\big) e^{\psi_S(\lambda)}.$$

Теперь оценим правую часть логарифмического неравенства Соболева.

Упражнение 4.3. Проверьте, что

$$(e^{\alpha/2} - e^{\beta/2})^2 \leq \frac{(\alpha-\beta)^2}{4}(e^\alpha + e^\beta).$$

Используя упражнение, получаем, что

$$\begin{aligned}
2\mathbb{E}\Big[\sum_{j=1}^m |\partial_j g(U)|^2\Big] &= \frac{1}{2}\mathbb{E}\Big[\sum_{j=1}^m \big(e^{\lambda f(U)/2} - e^{\lambda f(\overline{U}^{(j)})/2}\big)^2\Big] \leq \\
&\leq \frac{\lambda^2}{8}\mathbb{E}\Big[\sum_{j=1}^m \big(f(U) - f(\overline{U}^{(j)})\big)^2 \big(e^{\lambda f(U)} + e^{\lambda f(\overline{U}^{(j)})}\big)\Big] = \\
&= \frac{\lambda^2}{8}\mathbb{E}\Big[e^{\lambda f(U)} \sum_{j=1}^m \big(f(U) - f(\overline{U}^{(j)})\big)^2\Big] + \frac{\lambda^2}{8}\sum_{j=1}^m \mathbb{E}\Big[e^{\lambda f(\overline{U}^{(j)})}\big(f(U) - f(\overline{U}^{(j)})\big)^2\Big] = \\
&= \frac{\lambda^2}{8}\mathbb{E}\Big[e^{\lambda f(U)} \sum_{j=1}^m \big(f(U) - f(\overline{U}^{(j)})\big)^2\Big] + \frac{\lambda^2}{8}\sum_{j=1}^m \mathbb{E}\Big[e^{\lambda f(U)}\big(f(\overline{U}^{(j)}) - f(U)\big)^2\Big] = \\
&= \frac{\lambda^2}{4}\mathbb{E}\Big[e^{\lambda f(U)} \sum_{j=1}^m \big(f(U) - f(\overline{U}^{(j)})\big)^2\Big],
\end{aligned}$$

где мы воспользовались тем, что вектор $\overline{U}^{(j)}$ также распределен, как и U. Предположим, что

$$\sum_{j=1}^m \big(f(x) - f(\overline{x}^{(j)})\big)^2 \leq \theta^2 \quad \forall x \in \{-1,1\}^m. \tag{4.1}$$

Упражнение 4.4. Проверьте, что указанное условие влечет свойство липшицивости функции f, т.е.

$$|f(x) - f(y)| \leq \frac{\theta}{2}\|x - y\| \quad \forall x, y \in \{-1,1\}^m.$$

Тогда, применив логарифмическое неравенство Соболева, получаем, что

$$\big(\lambda\psi_S'(\lambda)-\psi_S(\lambda)\big)e^{\psi_S(\lambda)}=\mathrm{Ent}(g^2)\le\frac{\lambda^2\theta^2}{4}\mathbb{E}\big[e^{\lambda f(U)}\big]=\frac{\lambda^2\theta^2}{4}e^{\psi_S(\lambda)}.$$

Попробуем извелечь из этого дифференциального неравенства какую-нибудь информацию про функцию ψ_S. Поделив неравенство на положительное число $\lambda^2 e^{\psi_S(\lambda)}$, получаем, что

$$\frac{\psi_S'(\lambda)}{\lambda}-\frac{\psi_S(\lambda)}{\lambda^2}\le\frac{\theta^2}{4}.$$

Заметим, что левая часть есть производная функции $F(\lambda):=\frac{\psi_S(\lambda)}{\lambda}$, причем по правилу Лопиталя

$$F(0):=\lim_{\lambda\to0}\frac{\psi_S(\lambda)}{\lambda}=\lim_{\lambda\to0}\frac{\mathbb{E}\big[Se^{\lambda S}\big]}{\mathbb{E}\big[e^{\lambda S}\big]}=\mathbb{E}[S].$$

Таким образом, при $\lambda>0$, получаем, что

$$\frac{\psi_S(\lambda)}{\lambda}=F(\lambda)-F(0)+\mathbb{E}[S]=\int_0^\lambda F'(s)\,ds+\mathbb{E}[S]\le\int_0^\lambda\frac{\theta^2}{4}\,ds+\mathbb{E}[S]=\frac{\theta^2\lambda}{4}+\mathbb{E}[S].$$

Таким образом, $\psi_S(\lambda)\le\frac{\theta^2\lambda^2}{4}+\lambda\mathbb{E}[S]$ при $\lambda>0$. По неравенству Чернова,

$$P(S\ge\mathbb{E}[S]+t)\le\exp\big(-\sup_{\lambda>0}(\lambda\mathbb{E}[S]+\lambda t-\psi_S(\lambda)\big)=\exp\big(-\sup_{\lambda>0}(\lambda t-\tfrac{\theta^2\lambda^2}{4})\big)=e^{-t^2/\theta^2}.$$

Применив неравенство к случайной величине $-S$, получаем, что

$$P(S\le\mathbb{E}[S]-t)\le e^{-t^2/\theta^2}$$

и, суммарно,

$$P(|S-\mathbb{E}[S]|\ge t)\le 2e^{-t^2/\theta^2}.$$

Таким образом, нами доказана следующая теорема.

Теорема 4.5. *Пусть для функции f на $\{-1,1\}^m$ выполнено условие* (4.1) *и пусть $S=f(\varepsilon_1,\dots,\varepsilon_m)$. Тогда*

$$\psi_S(\lambda)\le\frac{\theta^2\lambda^2}{4}+\lambda\mathbb{E}[S]\quad\forall\lambda>0$$

и

$$P(|S-\mathbb{E}[S]|\ge t)\le 2e^{-t^2/\theta^2}.$$

5. Гауссовская концентрация.

Определение 5.1. Случайный вектор $G=(G_1,\dots,G_m)$ называется **стандартным гауссовским** (или нормальным) случайным вектором, если

$$\mathbb{E}[f(G)]=\int_{\mathbb{R}^m}f(x)\tfrac{1}{(2\pi)^{m/2}}e^{-\frac{\|x\|^2}{2}}\,dx.$$

Отметим, что $G=(G_1,\dots,G_m)$ — стандартный гауссовский вектор тогда и только тогда, когда $\{G_j\}_{j=1}^m$ независимы и G_j — стандартные гауссовские случайные величины.

Одно из ключевых свойств гауссовского распределения заключается в том, что оно появляется, как предельное распределение при большом числе испытаний.

Теорема 5.2 (Центральная Предельная Теорема)**.** *Пусть $\{X_n\}_{n=1}^\infty$ последовательность независимых одинаково распределнных случайных величин. Тогда, для каждой непрерывной и ограниченной функции φ,*

$$\lim_{n\to\infty}\mathbb{E}\left[\varphi\Big(\frac{\sum_{j=1}^n(X_j-\mathbb{E}[X_j])}{(\mathbb{D}[\sum_{j=1}^nX_j])^{1/2}}\Big)\right]=\mathbb{E}[\varphi(G)],$$

где G — одномерная гауссовская случайная величина.

Аналогичная теорема верна и в многомерном случае, в частности,

$$\lim_{n\to\infty} \mathbb{E}\Big[\varphi\Big(\frac{\varepsilon_{1,1}+\ldots+\varepsilon_{1,n}}{\sqrt{n}},\ldots,\frac{\varepsilon_{m,1}+\ldots+\varepsilon_{m,n}}{\sqrt{n}}\Big)\Big] = \mathbb{E}[\varphi(G_1,\ldots,G_m)].$$

Например, для функции одной переменной получаем, что

$$\sum_{j=1}^{n}\Big(\varphi\Big(\frac{\varepsilon_{1,1}+\ldots+\varepsilon_{1,j}+\ldots+\varepsilon_{1,n}}{\sqrt{n}}\Big)-\varphi\Big(\frac{\varepsilon_{1,1}+\ldots-\varepsilon_{1,j}+\ldots+\varepsilon_{1,n}}{\sqrt{n}}\Big)\Big)^2 =$$

$$= \sum_{j=1}^{n}\Big(\varphi\Big(\frac{\varepsilon_{1,1}+\ldots+\varepsilon_{1,j}+\ldots+\varepsilon_{1,n}}{\sqrt{n}}\Big)-\varphi\Big(\frac{\varepsilon_{1,1}+\ldots+\varepsilon_{1,j}+\ldots+\varepsilon_{1,n}}{\sqrt{n}}-\frac{2}{\sqrt{n}}\Big)\Big)^2 =$$

$$= \sum_{j=1}^{n}|\varphi'(\xi_j)|^2\cdot\frac{4}{n} \le 4\sup_{x\in\mathbb{R}}|\varphi'(x)|^2.$$

Упражнение 5.3. Проверьте, что и в многомерном случае сохранится подобная оценка, т.е.

$$\sum_{k=1}^{n}\sum_{j=1}^{m}\big(f(x)-f(\overline{x}^{(j,k)})\big)^2 \le 4\sup_{y\in\mathbb{R}^m}\|\nabla\varphi(y)\|^2,$$

где $f(x)=\varphi\big(\frac{x_{1,1}+\ldots+x_{1,n}}{\sqrt{n}},\ldots,\frac{x_{m,1}+\ldots+x_{m,n}}{\sqrt{n}}\big)$.

Теорема 5.4 (концентрация для гауссовского случайного вектора)**.** *Пусть φ — липшицева функция, т.е. $|\varphi(x)-\varphi(y)|\le L\|x-y\|$ $\forall x,y\in\mathbb{R}^m$. Тогда*

$$P(|\varphi(G)-\mathbb{E}[\varphi(G)]|\ge t)\le 2e^{-t^2/4L^2}.$$

Доказательство. Не ограничивая общности, считаем, что φ — достаточно гладкая функция. Тогда из общего курса анализа известно, что $\|\nabla\varphi\|\le L$. Тогда для функции $f(x)=\varphi\big(\frac{x_{1,1}+\ldots+x_{1,n}}{\sqrt{n}},\ldots,\frac{x_{m,1}+\ldots+x_{m,n}}{\sqrt{n}}\big)$ выполнено условие (4.1) с $\theta=2L$. По доказанному ранее

$$\mathbb{E}[e^{\lambda f(\varepsilon_{1,1},\ldots,\varepsilon_{m,n})}]\le\exp(L^2\lambda^2+\lambda\mathbb{E}[f(\varepsilon_{1,1},\ldots,\varepsilon_{m,n})]).$$

Осуществляя предельный переход, получаем, что

$$\mathbb{E}[e^{\lambda\varphi(G)}]\le\exp(L^2\lambda^2+\lambda\mathbb{E}[\varphi(G)]).$$

Дальнейшие рассуждения с применением неравенства Чернова и рассмотрением функции $-\varphi$ для нас уже стандартны. Теорема доказана. ☐

Упражнение 5.5. Докажите следующее логарифмическое неравенство Соболева для гауссовской меры. Пусть

$$\frac{1}{(2\pi)^{m/2}}\int_{\mathbb{R}^m} f^2(x)\,e^{-\frac{\|x\|^2}{2}}\,dx=1,$$

тогда

$$\tfrac{1}{(2\pi)^{m/2}}\int_{\mathbb{R}^m} f^2(x)\ln f^2(x)\,e^{-\frac{\|x\|^2}{2}}\,dx\le\tfrac{2}{(2\pi)^{m/2}}\int_{\mathbb{R}^m}\|\nabla f(x)\|^2\,e^{-\frac{\|x\|^2}{2}}\,dx.$$

6. Теорема Дворецкого.

Зададимся следующим вопросом. Пусть K — выпуклое центрально симметричное тело в $\mathbb{R}^N$ (слово «тело» означает, что K не содержится ни в каком подпространстве меньшей размерности, что оно ограничено и замкнуто, т.е. компактно). Какой размерности существуют сечение тела K «похожие» на эллипсоиды? Если этот вопрос несколько формализовать, то нам бы хотелось понять, какой размерности бывают гиперплоскости $H\subset\mathbb{R}^m$, что в них есть эллипсоид $E\subset H$ со свойством $(1-\delta)E\subset K\cap H\subset(1+\delta)E$. Пока задача все еще кажется несколько эфемерной. Давайте попробуем сформулировать ее на более аналитичском языке.

Определение 6.1. Нормой на $\mathbb{R}^m$ называется такая функция $\|\cdot\|\colon \mathbb{R}^m \to [0,+\infty)$, что
(i) $\|x\| = 0 \Leftrightarrow x = 0$;
(ii) $\|\alpha x\| = |\alpha|\cdot\|x\|$ $\forall x \in \mathbb{R}^m$;
(iii) $\|x+y\| \le \|x\| + \|y\|$ $\forall x, y \in \mathbb{R}^m$.

Более того, с каждым выпуклым центрально симметричным телом K можно связать норму $\|\cdot\|_K$ по правилу

$$\|x\|_K := \inf\{\lambda > 0\colon \lambda^{-1}x \in K\}.$$

Отметим, что $\|\cdot\|_K$ называется **функционалом Минковского**, построенным по множеству K.

Упражнение 6.2. Проверьте, что для каждого выпуклого центрально симметричного тела K функция $\|\cdot\|_K$ действительно обладает всеми свойствами нормы. Кроме того, $K := \{x\colon \|x\|_K \le 1\}$.

Теперь заметим, что любой эллипсоид есть линейный образ шара при невырожденном линейном отображании, поэтому любой эллипсоид E в гиперплоскости $H \subset \mathbb{R}^N$ размерности $\dim H = d$ можно считать линейным образом стандартного шара $B_d := \{x \in \mathbb{R}^d\colon \|x\| \le 1\}$, т.е. $E = T(B_d)$, где $T\colon \mathbb{R}^d \to \mathbb{R}^N$ — невырожденное линейное отображаение, $T(\mathbb{R}^d) = H$. Тогда включение $(1-\delta)E \subset K$ значит, что для $x \in B_d$ выполнено $Tx \in \frac{1}{1-\delta}K$, т.е. из оценки $\|x\| \le 1$ следует $\|Tx\|_K \le \frac{1}{1-\delta}$, что равносильно неравенству

$$\|Tx\|_K \le \frac{1}{1-\delta}\|x\| \quad \forall x \in B_d.$$

С другой стороны, включение $K \cap H \subset (1+\delta)E$ значит, что если $\|Tx\|_K \le 1$, то $\|x\| \le 1+\delta$, что опять равносильно неравенству

$$\|x\| \le (1+\delta)\|Tx\|_K \quad \forall x \in B_d.$$

Во-превых, заметим, что оба неравенства однородны относительно растяжений, а значит можно требовать их выполнения только на сфере $S^{d-1} := \{x \in \mathbb{R}^d\colon \|x\| = 1\}$. Во-вторых, Вместо, оценок с множителями $\frac{1}{1+\delta}$ и $\frac{1}{1-\delta}$ будем требовать оценки с множителями $1-\varepsilon$ и $1+\varepsilon$ для удобства записи. Т.е. нам надо искать отображения T из $\mathbb{R}^d$, с как можно большей размерностью d, и со свойством

$$(1-\varepsilon)\|x\| \le \|Tx\|_K \le (1+\varepsilon)\|x\| \quad x \in S^{d-1}.$$

На самом деле, можно требовать свойство

$$(1-\varepsilon)M\|x\| \le \|Tx\|_K \le (1+\varepsilon)M\|x\| \quad x \in S^{d-1}.$$

и тогда в качестве искомого отображения взять отображение $T' := \frac{1}{M}T$. Также, как и при доказательстве леммы Джоносона–Линденштрауса, будем искать отображение T в виде случайной функции. Пусть $\{g_{k,j}\}_{\substack{1\le k\le N\\ 1\le j\le d}}$ — набор независимых гауссовских случайных величин. Пусть

$$Tx := \Big(\sum_{j=1}^{d} g_{1,j}x_j, \ldots, \sum_{j=1}^{d} g_{N,j}x_j\Big)$$

Заметим, что Tx — случайный вектор с независимыми компонентами, причем, при $x \in S^{d-1}$, каждая компонента также будет стандартной гауссовской (это одно из свойств гауссовских случайных величин). Предположим, что $\|y\|_K \le L\|y\|$ $\forall y \in \mathbb{R}^N$, тогда $\|\cdot\|_K$ — липшицева функция, т.к.

$$\big|\|y_1\|_K - \|y_2\|_K\big| \le \|y_1 - y_2\|_K \le L\|y_1 - y_2\|.$$

В таком случае, мы можем воспользоваться концентрацией для стандартного гауссовского случайного вектора Tx при фиксированном $x \in S^{d-1}$. Таким образом,

$$P\Big(\big|\|Tx\|_K - \mathbb{E}[\|G\|_K]\big| \geq t\Big) \leq 2e^{-t^2/4L^2},$$

где G — стандартный нормальный вектор в $\mathbb{R}^N$. Пусть теперь $W \subset S^{n-1}$ — произвольное подмножество сферы. Тогда, так же, как и при доказательстве леммы Джоносона–Линденштрауса

$$P\Big(\max_{x\in W}\big|\|Tx\|_K - M\big| \geq \varepsilon M\Big) \leq 2 \cdot \#We^{-\varepsilon^2 M^2/4L^2},$$

где $M = \mathbb{E}[\|G\|_K]$.

Определение 6.3. Множество W называется ε-сетью множества $A \subset \mathbb{R}^d$, если

$$A \subset \bigcup_{x\in W} x + tB_d.$$

Пусть

$$N_t(A) := \min\Big\{n \in \mathbb{N}\colon \exists x_1, \ldots, x_n \in A\colon A \subset \bigcup_{j=1}^{n} x_j + tB_d\Big\}.$$

Функция $\log N_t(A)$ называется функцией **метрической энтропии** множества A.

Лемма 6.4. *Пусть, A — компакт в $\mathbb{R}^d$. Тогда*

$$N_t(A) \leq \frac{Vol_n(A + \frac{t}{2}B)}{Vol_n(B)} \frac{2^n}{t^n}.$$

Доказательство. Пусть $n(t)$ — максимальное количество точек в A с попарными расстояниями не менее t и пусть $\{x_1, \ldots, x_{n(t)}\}$ — набор таких точек. Ясно, что множества $x_j + \frac{t}{2} \cdot B_d$ имеют попарно непересекающиеся внутренности,

$$\bigcup_{j=1}^{n(t)}(x_j + \tfrac{t}{2} \cdot B_d) \subset A + \tfrac{t}{2} \cdot B$$

и

$$A \subset \bigcup_{j=1}^{n(t)}(x_j + t \cdot B_d).$$

Таким образом,

$$n(t) \cdot Vol_n(\tfrac{t}{2} \cdot B_d) \leq \lambda_n(A + \tfrac{t}{2} \cdot B_d)$$

и $n \leq n(t) \leq \frac{\lambda_n(A+\frac{t}{2}\cdot B_d)}{\lambda_n(\frac{t}{2}\cdot B_d)}$. Лемма доказана. □

Как следствие, для $A = S^{d-1}$ получаем, что

$$N_t(S^{d-1}) \leq \frac{(1 + t/2)^2}{(t/2)^d} = (1 + 2/t)^d \leq (3/t)^d$$

при $t \in (0, 1)$.

Упражнение 6.5. Пусть $|||\cdot|||$ — некоторая норма на $\mathbb{R}^d$ (не обязательно евклидова). Предположим, что для каждой точки x из δ-сети на S^{d-1} выполнено

$$(1 - \varepsilon)M \leq |||x||| \leq (1 + \varepsilon)M.$$

Проверьте, что

$$\frac{1 - \varepsilon - 2\delta}{1 - \delta}M \leq |||x||| \leq \frac{1 + \varepsilon}{1 - \delta}M \quad \forall x \in S^{d-1}.$$

Пусть теперь W — ε - сеть размера не более $(3/\varepsilon)^d$ на сфере S^{d-1}, $\varepsilon \in (0, \frac{1}{100})$. Тогда, при $(3/\varepsilon)^d e^{-\varepsilon^2 M^2/4L^2} \le \frac{1}{4}$ будет существовать такая реализация отображения T, что

$$(1-\varepsilon)M \le \|Tx\|_K \le (1+\varepsilon)M$$

для каждой точки x из сети W.

Упражнение 6.6. Проверьте, что существуют такие числа $c_1, c_2 > 0$, что неравенство $(3/\varepsilon)^d e^{-\varepsilon^2 M^2/4L^2} \le \frac{1}{4}$ верно при $d \le c_1 \frac{\varepsilon^2 M^2/L^2}{\ln(c_2/\varepsilon)}$.

Таким образом, мы доказали следующее утверждение.

Теорема 6.7 (В. Мильман)**.** *Пусть $\|y\|_K \le L\|y\|$ $\forall y \in \mathbb{R}^N$ и пусть $M = \mathbb{E}[\|G\|_K]$ Тогда существует гиперплоскость H размерности $d \ge C_1 \frac{\delta^2 M^2/L^2}{\ln(C_2/\delta)}$ и эллипсоид E в этой гиперплоскости, что $(1-\delta)E \subset K \cap H \subset (1+\delta)E$ для достаточно мальнького δ.*

Имеет место следующий факт.

Теорема 6.8 (Лемма Дворецкого–Роджерса)**.** *Предположим, что эллипсоид максимального объема, вписанный в K есть LB_N. Тогда*

$$\mathbb{E}[\|G\|_K] \ge C_0 L\sqrt{\ln N}.$$

Теперь переводя любое выпуклое тело линейным преобразованием в выпуклое тело, у которого вписанный эллипсоид максимального объема есть шар, получаем следующую теорему Дворецкого.

Теорема 6.9 (Теорема Дворецкого)**.** *Пусть K — центрально симметричное выпуклое тело в $\mathbb{R}^N$, $\delta \in (0,1)$. Тогда существует гиперплоскость H размерности $d \ge C\frac{\delta^2 \ln N}{\ln(c/\delta)}$ и эллипсоид E в этой гиперплоскости, что $(1-\delta)E \subset K \cap H \subset (1+\delta)E$.*

Список литературы

[1] S. Boucheron, G. Lugosi, P. Massart, Concentration Inequalities: A Nonasymptotic Theory of Independence (2013).
[2] D. Dadush, J. Briet, Geometric Functional Analysis, Lecture Notes, 2019
[3] K. Ball An elementary introduction to modern convex geometry. Flavors of geometry, 31(1–58), (1997).
[4] S. Artstein-Avidan, A. Giannopoulos, V. D. Milman, Asymptotic geometric analysis, Part I and Part II. American Mathematical Society (2015, 2021)