Многорукие бандиты

Селиханович Даниил, МФТИ ФУПМ, 572 группа

4 апреля 2018 г.

Постановка задачи

Многорукие бандиты. Имеется два «одноруких бандита» (так называют игровые автоматы с ручкой, дергая за которую получаем случайный выигрыш). Вероятность выиграть 1 франк на первом автомате $p_1 > 0$ (с вероятностью $1 - p_1$ выигрыш равен 0), а на втором — $p_2 > 0$. Обе вероятности неизвестны. Игрок может в любом порядке раз дергать за ручки «одноруких бандитов». Стратегией игрока является выбор ручки на каждом шаге, в зависимости от результатов всех предыдущих шагов, так чтобы суммарный выигрыш был бы максимальным. Приведите стратегию игрока. Предложите свое обобщение задачи.

Содержание

1 Динамическое программирование.

1.1 Испытания Бернулли.

Постановка задачи: у нас имеется последовательность подбрасываний монетки, о «честности» которой нам ничего не известно. Требуется определить, какова вероятность получить орёл или решку в следующем подбрасывании.

Обозначим вероятность выпадения орла через q, случайные величины, соответствующие i-му подбрасыванию монетки, через X_i, а их значения через x_i: $x_i = 1$, если выпал орёл и $x_i = 0$ для решки. По формуле Байеса плотность вероятности $f(q) = \mathsf{P}\Big(q \;\Big|\; X_1 = x_1, \ldots, X_n = x_n\Big)$ вычисляется так:

$$\mathsf{P}\Big(q \;\Big|\; X_1 = x_1, \ldots, X_n = x_n\Big) = \frac{\mathsf{P}(q)\,\mathsf{P}\Big(X_1 = x_1, \ldots, X_n = x_n \;\Big|\; q\Big)}{\mathsf{P}\Big(X_1 = x_1, \ldots, X_n = x_n\Big)}. \tag{1}$$

Мы будем считать, что априорное распределение «честности» монетки **<u>равномерное</u>**, то есть пока мы не видели ни одного результата, мы вообще ничего не можем сказать о вероятности выпадения орла и считаем все такие вероятности одинаково возможными: $\mathsf{P}(q) = 1$ при $0 \le q \le 1$ и $\mathsf{P}(q) = 0$ иначе.

$$\mathsf{P}\Big(X_1 = x_1, \ldots, X_n = x_n \;\Big|\; q\Big) = \prod_{i=1}^{n} q^{x_i}(1-q)^{1-x_i} = q^s(1-q)^{n-s}. \tag{2}$$

Здесь s - число выпадений орлов в эксперименте из n подбрасываний.

Дискретная формула полной вероятности в непрерывном случае, как в данной задаче, пример вид:

$$\mathsf{P}\Big(X_1 = x_1, \ldots, X_n = x_n\Big) = \int\limits_0^1 q^s(1-q)^{n-s}dq = B(s+1, n-s+1) =$$
$$= \frac{\Gamma(s+1)\Gamma(n-s+1)}{\Gamma(n+2)} = \frac{s!(n-s)!}{(n+1)!}$$

Получили, что

$$\mathsf{P}\Big(q \;\Big|\; X_1 = x_1, \ldots, X_n = x_n\Big) = \frac{(n+1)!}{s!(n-s)!}q^s(1-q)^{n-s} \tag{3}$$

Вычислим матожидание $\mathsf{E}[q]$ выпадения орла в следующем подбрасывании монетки при условиях эксперимента $X_1 = x_1, \ldots, X_n = x_n$:

$$\mathsf{E}[q] = \int\limits_0^1 q f(q)\,dq = \int\limits_0^1 q\frac{(n+1)!}{s!(n-s)!}q^s(1-q)^{n-s}\,dq = \frac{(n+1)!}{s!(n-s)!}\int\limits_0^1 q^{s+1}(1-q)^{n-s}\,dq =$$
$$= \frac{(n+1)!}{s!(n-s)!}B(s+2, n-s+1) = \frac{(n+1)!}{s!(n-s)!}\frac{\Gamma(s+2)\Gamma(n-s+1)}{\Gamma(n+3)} =$$
$$= \frac{(n+1)!}{s!(n-s)!}\frac{(s+1)!(n-s)!}{(n+2)!} = \frac{s+1}{n+2}. \quad \mathbf{(4)}$$

1.2 Алгоритм поиска ожидаемой максимальной прибыли.

Пусть мы имеем k разных «одноруких бандитов». Будем обозначать состояние игрока после n_1 использований 1-ой ручки, из которых только w_1 были успешными, $\ldots, n_k$ использований k-ой ручки, из которых только w_k были успешными, через $S = \Big\{(n_1, w_1), \ldots, (n_k, w_k)\Big\}$, а $V^*(S)$ через ожидаемый максимальный выигрыш для состояния с данной историей, если всего можно делать h выборов.

Тогда задача сводится к вычислению $V^*\Big((0,0), \ldots, (0,0)\Big)$ для данного числа выборов h. Базой рекурсии, которую мы хотим составить, будет служить очевидное равенство:

$$V^*\Big(n_1, w_1), \ldots, (n_k, w_k)\Big) = 0 \text{ для всех таких } n_1, \ldots, n_k : \sum_{i=1}^{k} n_i = h \tag{5}$$

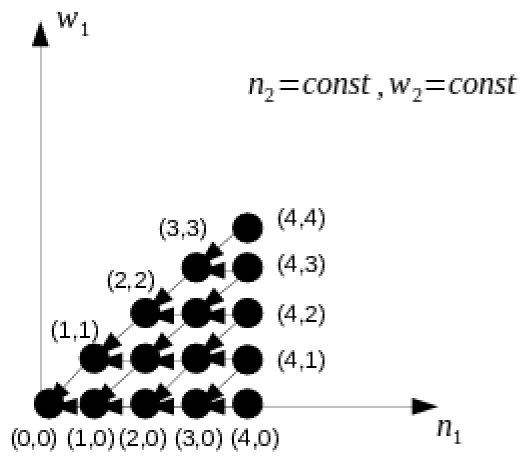
. Если обозначить через ρ_i ожидаемую вероятность того, что после истории испытаний на ручке i (n_i, w_i) игрок в состоянии S получит выигрыш после выбора ручки i, то можем выписать следующую рекурренту:

$$\textbf{(6)}\ V^*\Big((n_1, w_1), \ldots, (n_k, w_k)\Big) =$$
$$= \max_{1 \le i \le k} \Big(\rho_i \Big\{1 + V^*\Big((n_1, w_1), \ldots, (n_i+1, w_i+1), \ldots\Big)\Big\} + (1-\rho_i) V^*\Big((n_1, w_1), \ldots, (n_i+1, w_i), \ldots\Big)\Big)$$

Согласно **(4)** можем считать, что $\rho_i = \frac{w_i+1}{n_i+2}$ в соответствии с условием задачи, так как вероятности выигрыша p_i могут быть произвольными ненулевыми.

Замечаем, что данная рекуррента с учётом начальных условий позволяет определить $V^*\Big((0,0), \ldots, (0,0)\Big)$ для данного числа выборов h, не проводя сам эксперимент. Тем самым возникает наивный алгоритм действий игрока, желающего получить ожидаемый максимальный выигрыш при игре с однорукими бандитами.

Рис. 1: Порядок заполнения таблицы



Алгоритм

1) предрасчётом вычислить таблицу $V^*\Big((n_1, w_1), \ldots, (n_k, w_k)\Big)$ для всех $n_1, \ldots, n_k, w_1, \ldots, w_k$: $0 \le w_i \le n_i, \sum\limits_{i=1}^{k} n_i \le h$. Сложность заполнения таблицы - $O(k|V^*|)$, где $|V^*| = O\Big((h+1)^k \sum\limits_{i=0}^{h} \binom{i+k-1}{i}\Big) = O\Big(h^k \binom{h+k}{k}\Big)$ - размер заполняемой+- таблицы. В общем случае получаем экспоненциальное время заполнения от h, но при небольшом числе ручек k получаем полином от h - так, для двух ручек получим сложность заполнения таблицы $O(h^4)$.

2) на каждом шаге с данным фиксированным состояние опыта $S = \Big\{(n_1, w_1), \ldots, (n_k, w_k)\Big\}$ выбирать действие i, для которого достигается

$$\max_{1 \le i \le k} \Big(\rho_i \Big\{1 + V^*((n_1, w_1), \ldots, (n_i+1, w_i+1), \ldots)\Big\} + (1-\rho_i) V^*((n_1, w_1), \ldots, (n_i+1, w_i), \ldots)\Big).$$

3) переходить в новое состояние S_{new}.

Анализ алгоритма: при значениях k, сравнимых с h, время заполнения экспоненциально от времени игры, так что можем использоваться только в тех случаях, когда нужно планировать игру не слишком далеко. По сути - полный перебор всех возможных событий. Зато таблица $V^*\Big((n_1, w_1), \ldots, (n_k, w_k)\Big)$ может использоваться многократно.

Рис. 2: Несимметричные вероятности

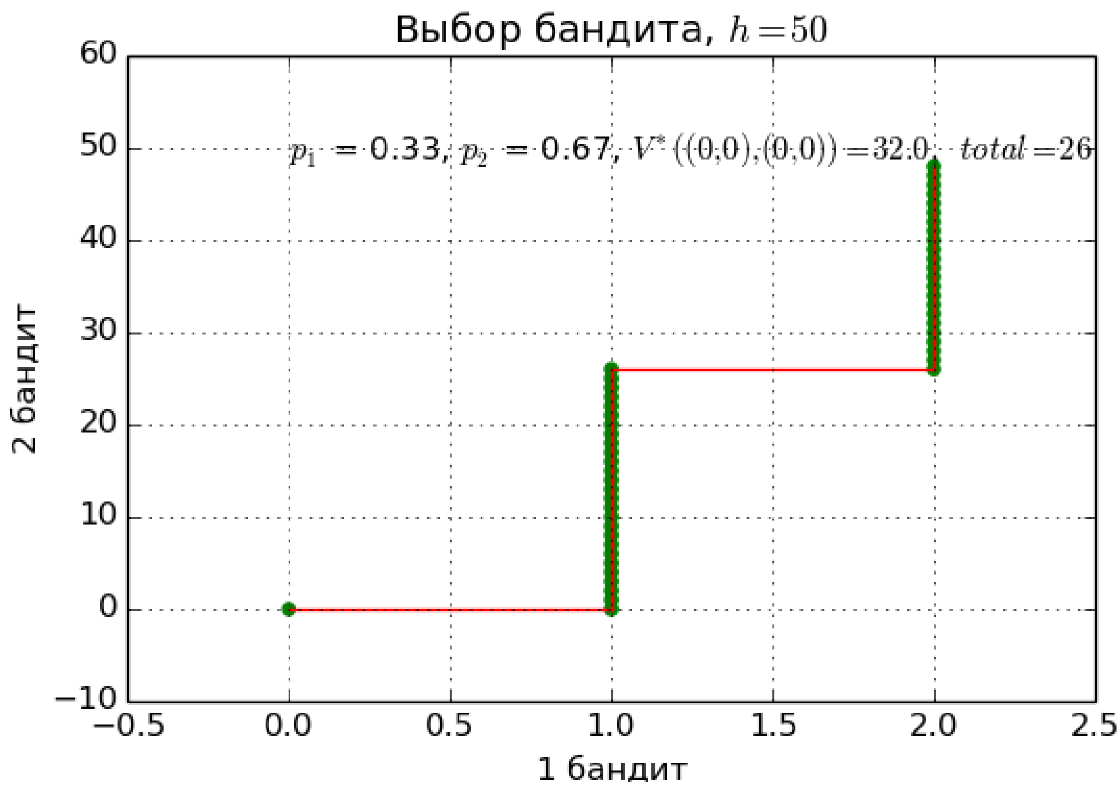


Рис. 3: Несимметричные вероятности

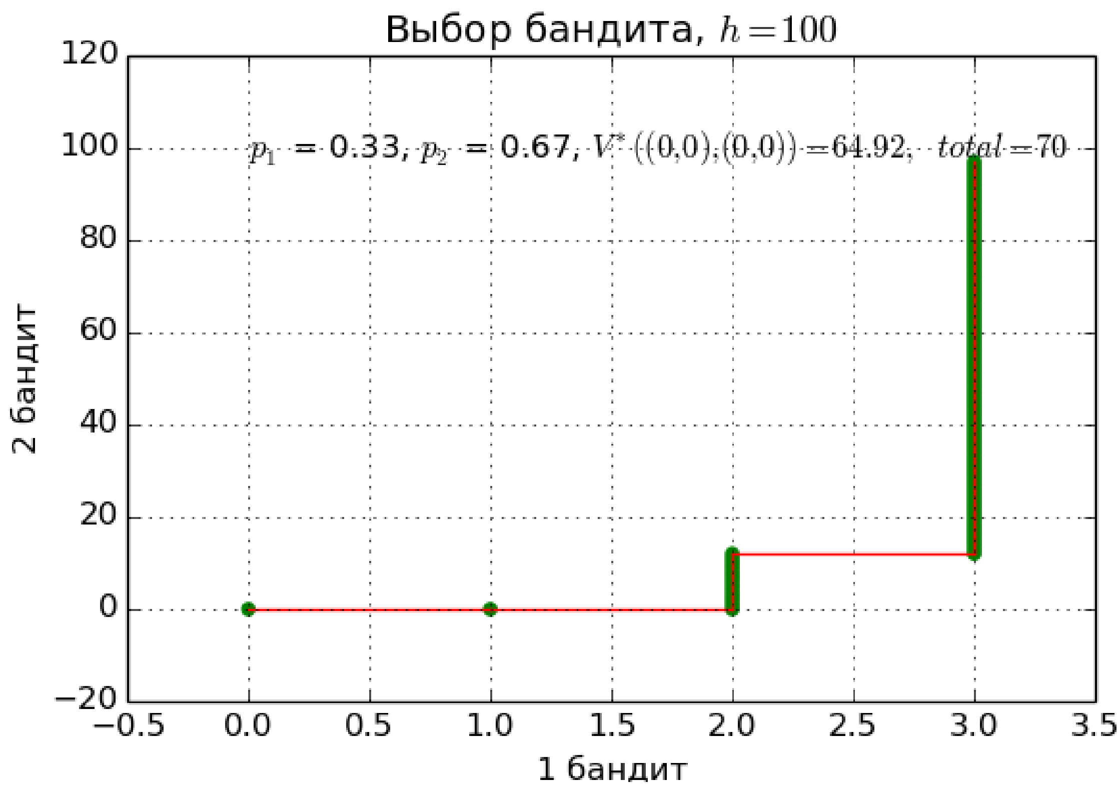


Рис. 4: Несимметричные вероятности

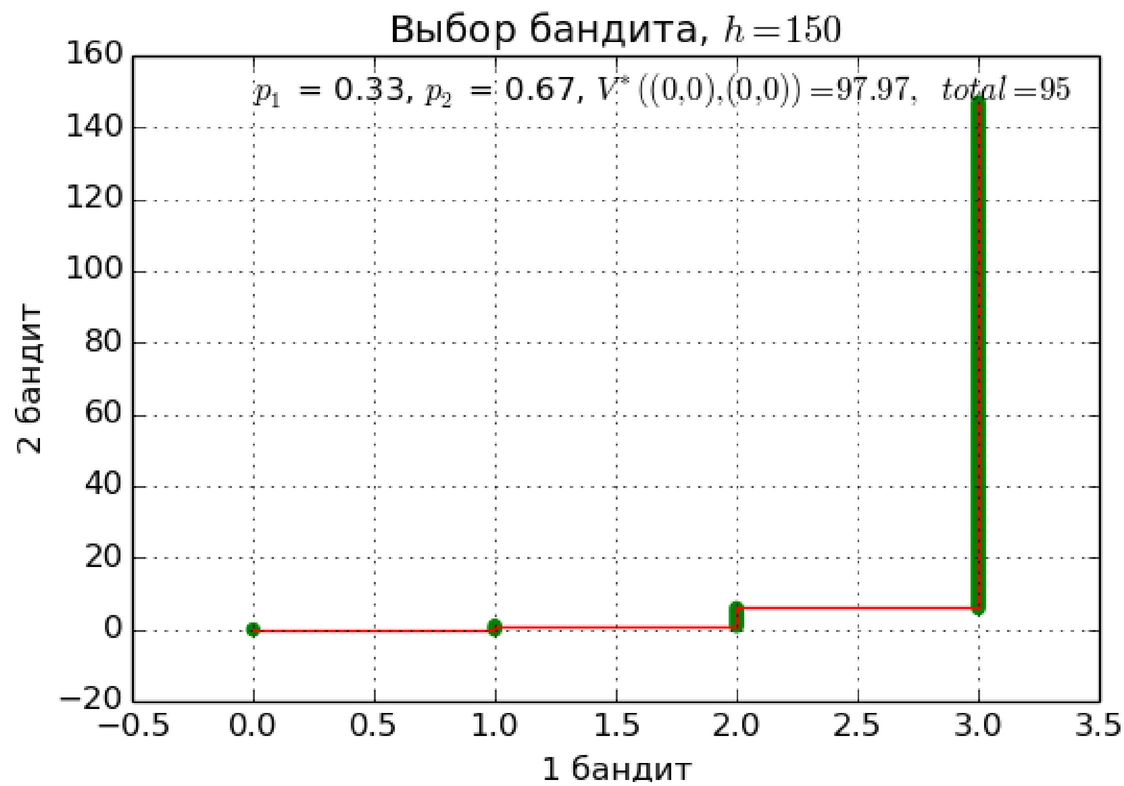


Рис. 5: Несимметричные вероятности

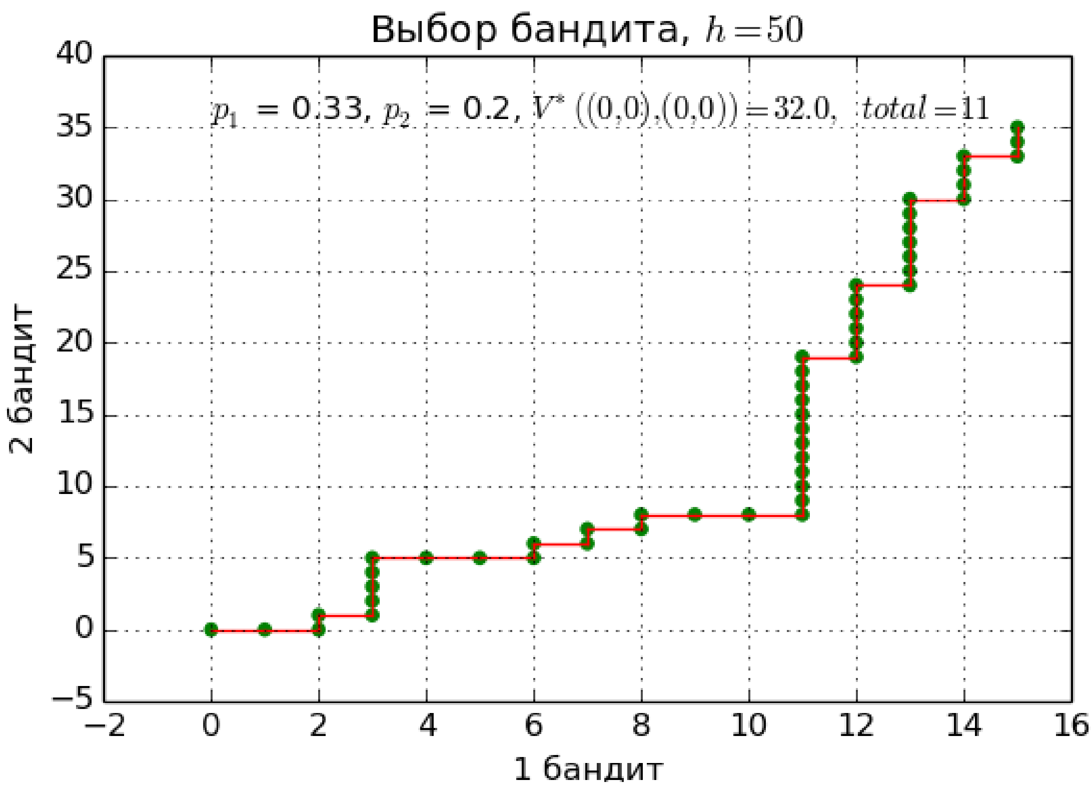


Рис. 6: Несимметричные вероятности

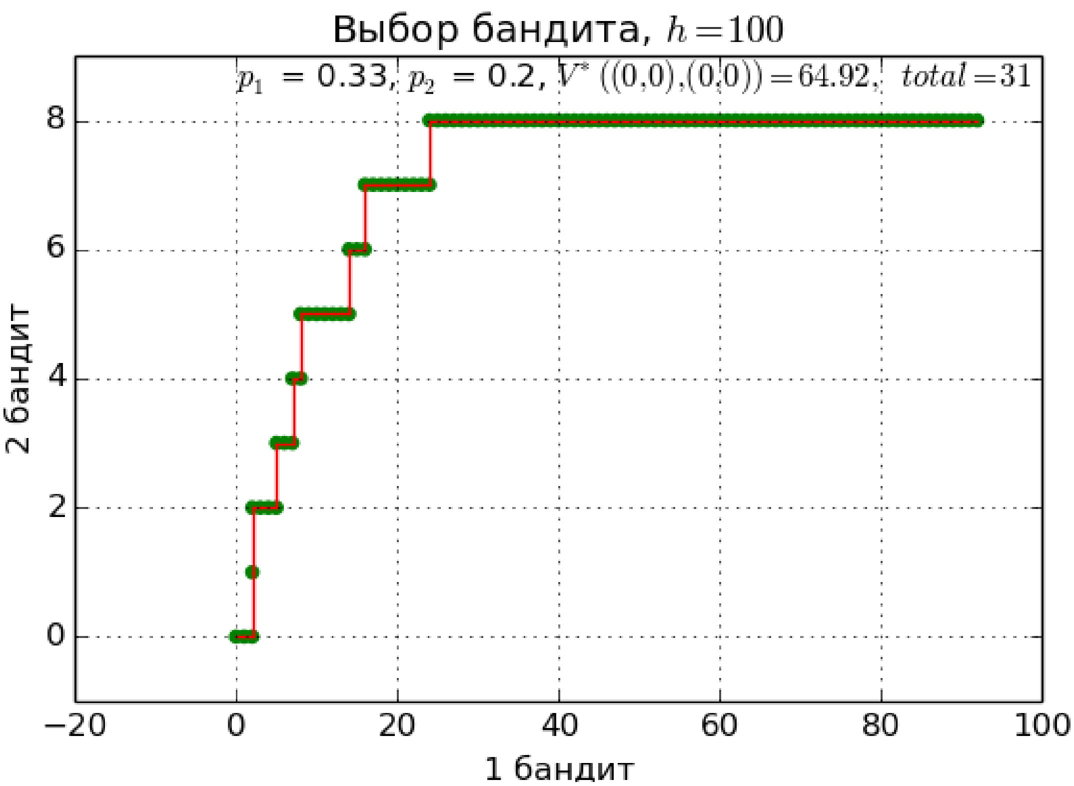


Рис. 7: Несимметричные вероятности

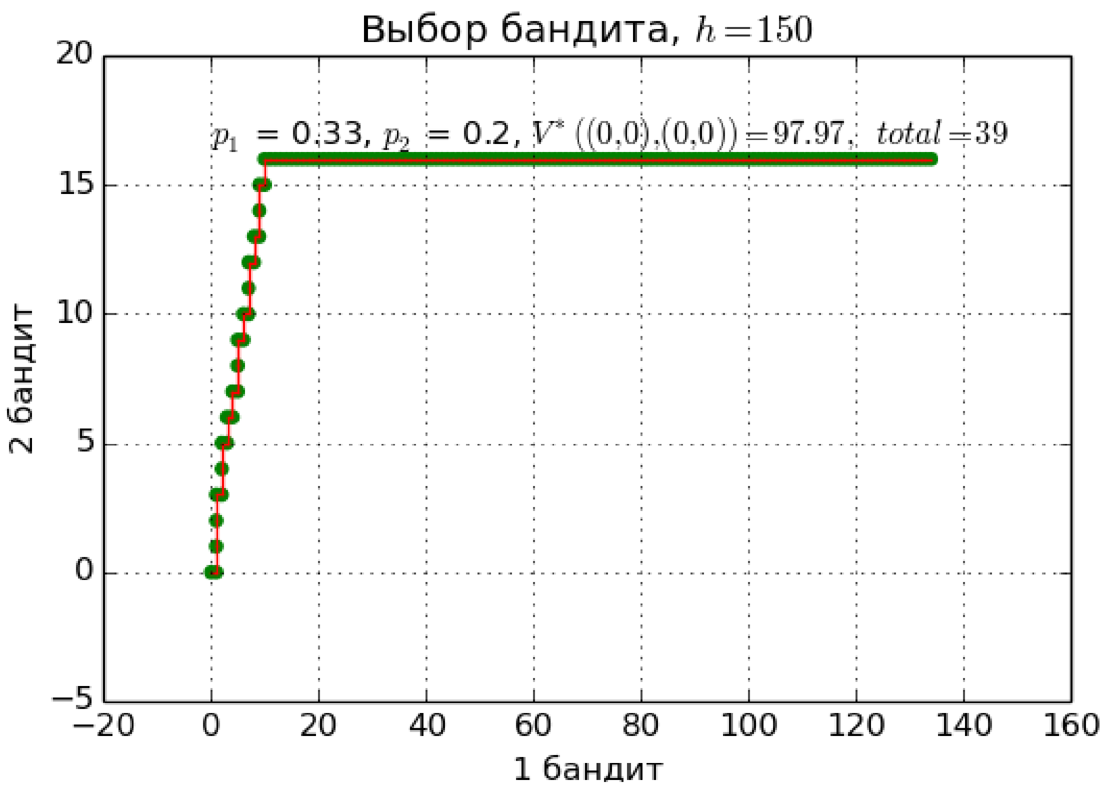


Рис. 8: Симметричные вероятности

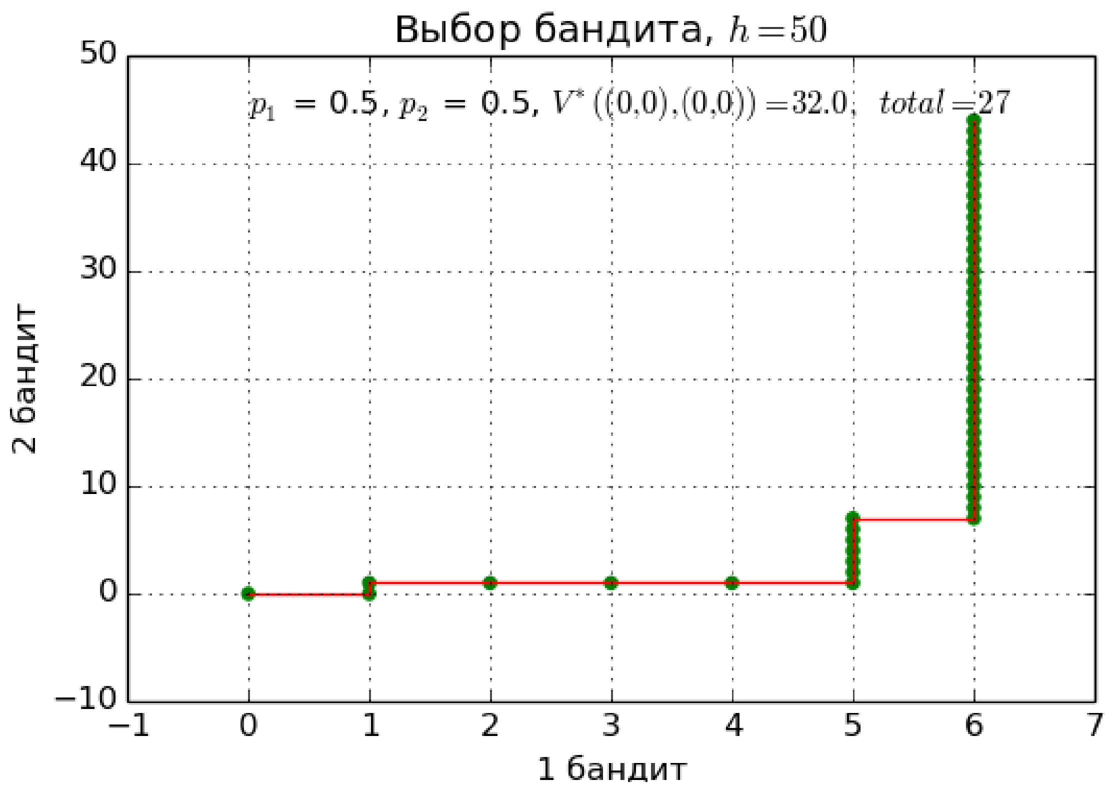


Рис. 9: Симметричные вероятности

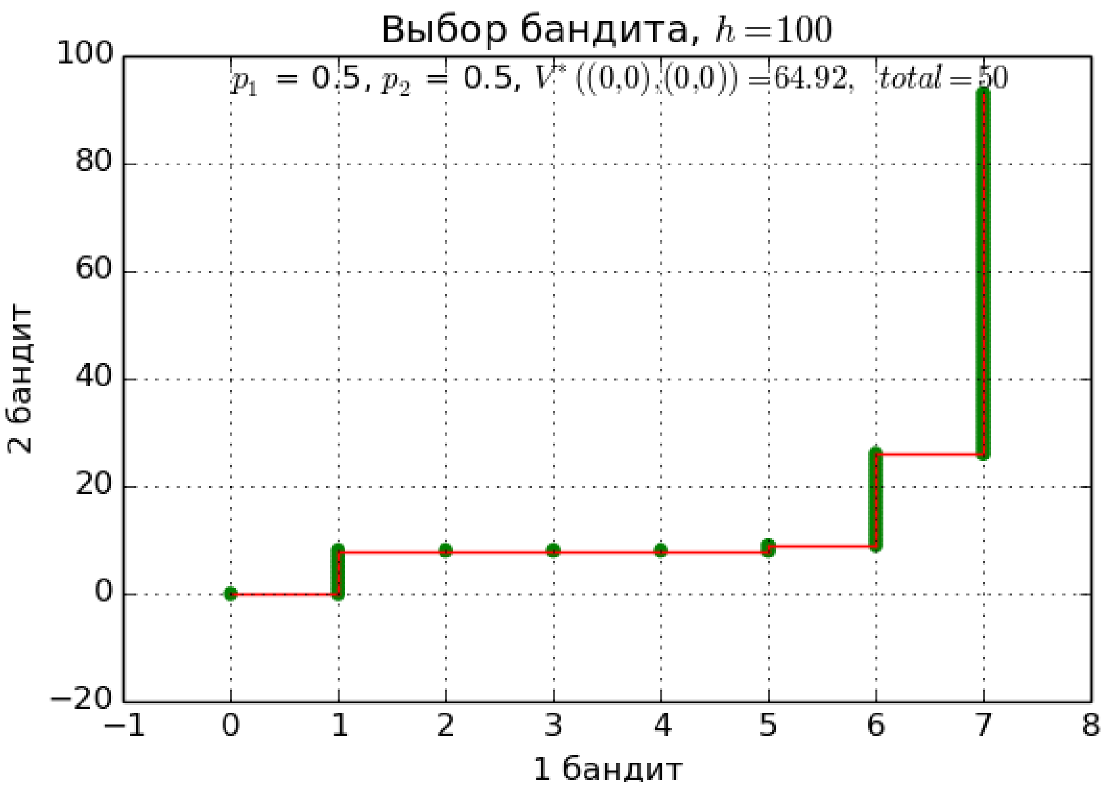
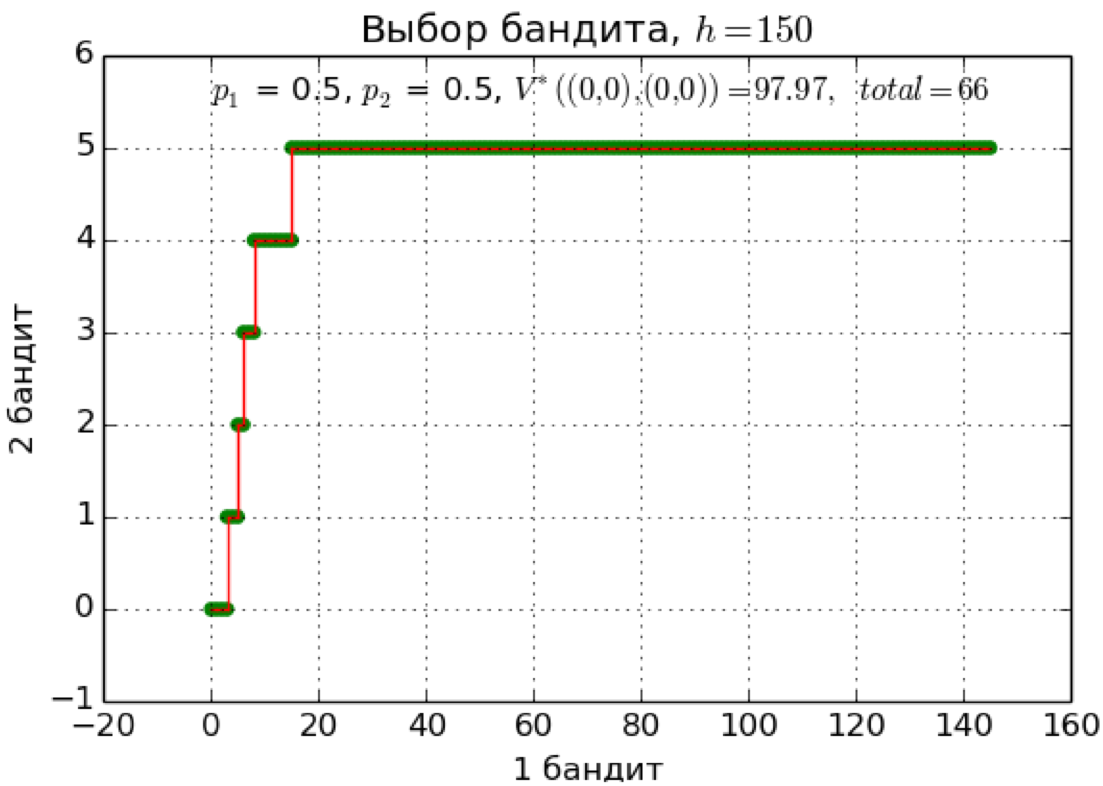


Рис. 10: Симметричные вероятности



2 Заключение

Выражаю благодарность **Гасникову Александру Владимировичу** и **Суворикову Александре** за помощь в исследовании задачи.

Список литературы

[1] *Баяндина А. С., Гасников А.В., Гулиев Ф.Ш., Лагуновская А.А.* Безградиентные двухточечные методы решения задач стохастической негладкой выпуклой оптимизации при наличии малых шумов не случайной природы: https://arxiv.org/ftp/arxiv/papers/1701/1701.03821.pdf.

[2] *Гасников А. В., Нестеров Ю.Е., Спокойный В.Г.* Об эффективности одного метода рандомизации зеркального спуска в задачах онлайн оптимизации: https://arxiv.org/pdf/1410.3118.pdf.

[3] *Гасников А. В., Крымова Е.А., Лагуновская А.А., Усманова И.Н., Федоренко Ф.А.* Стохастическая онлайн оптимизация. Одноточечные и двухточечные нелинейные многорукие бандиты. Выпуклый и сильно выпуклый случаи: https://arxiv.org/ftp/arxiv/papers/1509/1509.01679.pdf.

[4] *Гасников А. В., Крымова Е.А., Лагуновская А.А., Усманова И.Н., Федоренко Ф.А.* Безградиентные прокс-методы с неточным оракулом для негладких задач выпуклой стохастической оптимизации на симплексе: https://arxiv.org/ftp/arxiv/papers/1412/1412.3890.pdf.

[5] *Аникин А. С., Гасников А.В., Двуреченский П.Е., Тюрин А.И., Чернов А.В.* Двойственные подходы к задачам минимизации сильно выпуклых функционалов простой структуры при аффинных ограничениях: https://arxiv.org/ftp/arxiv/papers/1602/1602.01686.pdf.

[6] Лекция Зорича В.А про концентрацию меры: $http://www.mathnet.ru/php/seminars.phtml?option_lang = rus\&presentid =$ 14604.

[7] *Николенко С. И., Тулупьев А.Л.* Самообучающиеся системы. Российская академия наук. Санкт-Петербургский институт информатики и автоматизации РАН. Москва, Издательство МЦНМО, 2009.

[8] *Bubeck S., Cesa-Bianchi N.* Regret analysis of stochastic and nonstochastic multi-armed bandit problems, Foundation and Trends in Machine Learning, 5:1 (2012) 1–122; Lugosi G., Cesa-Bianchi N. Prediction, learning and games. New York: Cambridge University Press, 2006.

[9] *Lev Bogolubsky, Pavel Dvurechensky, Alexander Gasnikov, Gleb Gusev, Yurii Nesterov, Andrey Raigorodskii, Aleksey Tikhonov, Maksim Zhukovskii* Learning Supervised PageRank with Gradient-Free Optimization Methods: https://arxiv.org/pdf/1411.4282.pdf.

Спасибо за внимание!

3 Дополнение. Метод зеркального спуска в задачах стохастической выпуклой оптимизации.

3.1 Постановка задачи в терминах оптимизации.

Рассмотрим смежную задачу о нелинейных бандитах с шумом, в которой выбор ручки приносит нам случайные потери на шаге k ручки $i(k)$ $(1 \leq i \leq n)$, равные $f^k_{i(k)}$, зависящие от номера шага, номера ручки и от того, какой стратегии мы придерживались до шага k включительно. Наша стратегия на шаге k описывается вектором распределения вероятностей $x^k \in S_n(1) = \{x \geq 0 \mid \sum\limits_{i=1}^{n} x_i = 1\}$, согласно которому мы независимо ни от чего выбираем ручку, которую будем дёргать. Все, чем мы располагаем на шаге k, это вектором

$$\Big((x^1, i(1), f^1_{i(1)}); \ldots ; (x^{k-1}, i(k-1), f^{k-1}_{i(k-1)})\Big).$$

Мы считаем, что потери на k-ом шаге f^k зависят от x^k (но не от результата разыгрывания из распределения x^k), зависят от $(x^1, \ldots, x^{k-1})$ и от результатов соответствующих разыгрываний, а также зависят от $(f^1, \ldots, f^{k-1})$. Таким образом, целью является организовать процедуру выбора ручек, чтобы ожидаемые суммарные потери были бы минимальны.

Предполагаем, что функция $f(x^k)$ выпуклая на $Q = S_n(1)$ и равна потерям после выбора на k-ом шаге вектора вероятностей x^k. Однако игрок не может наблюдать значение самой функции $f(x)$, а видит лишь зашумленную реализацию значений этой функции:

$$\widetilde{f}(x^k, \xi^k) = f(x^k, \xi^k) + \delta(x^k, \xi^k), \tag{7}$$

$$\mathsf{E}_{\xi^k}(f(x^k, \xi^k)) = f(x^k), |\delta(x^k, \xi^k)| < \delta. \tag{8}$$

Предполагается, что $\{\xi^k\}_k$ - независимые одинаково распределённые случайные величины (их реализации одинаковы для всех игроков), функция $f(x, \xi)$ как функция x в μ_0-окрестности выпуклого множества Q является:

- **(9)** выпуклой;
- удовлетворяющей условию **(10)** $|f(x,\xi) - f(y,\xi)| \leq M\big\|y - x\big\|_2 \ \forall x, y, \xi$.

Успешность стратегии игрока, сыгравшего $N >> 1$ раз, характеризуется величиной:

$$Regret_f\Big(\{x^k\}_{k=0}^{N-1}\Big) = \mathsf{E}\Big\{\frac{1}{N}\sum_{k=0}^{N-1} f(x^k)\Big\} - \min_{x \in Q} f(x) \tag{11}$$

Пусть $x^k = x_*$ - оптимальная стратегия игрока и игрок начинает в состоянии x^0. Введём прокс-функцию $d(x) \geq 0$ $(d(x^0) = 0)$, которая предполагается сильно выпуклой относительно евклидовой нормы, и определим $R^2 = V(x_*, x^0)$, где прокс-расстояние (расстояние Брэгмана) определяется формулой:

$$V(x, z) = d(x) - d(z) - \langle \nabla d(z), x - z \rangle. \tag{12}$$

На каждой итерации можно один раз обратиться к оракулу за зашумленным значением стохастического градиента $\nabla_x \widetilde{f}(x^k, \eta^k)$. Пусть для любых $\widetilde{N} \le N$ выполнено (Θ^{k-1} - сигма алгебра, порождённая случайными величинами $\eta^1, \dots, \eta^{k-1}$):

$$\sup_{\{x^k = x^k(\xi^1,\dots,\xi^{k-1})\}_{k=1}^{\widetilde{N}} \in Q} \mathsf{E}\Big\{ \frac{1}{\widetilde{N}} \sum_{k=1}^{\widetilde{N}} \Big\langle \mathsf{E}_{\eta^k}\Big\{ \nabla_x f(x^k, \eta^k) - \nabla_x \widetilde{f}(x^k, \eta^k) \mid \Theta^{k-1} \Big\}, x^k - x_* \Big\rangle \Big\} \le \sigma; \tag{13}$$

$$\mathsf{E}_{\eta^k}\Big\{ \nabla_x f(x^k, \eta^k) \Big\} = \nabla f(x^k), \mathsf{E}_{\eta^k}\Big\{ \big\| \nabla_x \widetilde{f}(x^k, \eta^k) \big\|_2^2 \Big\} \le \widetilde{M}^2, \tag{14}$$

где $\{\eta^k\}_{k=0}^{N-1}$ - независимые одинаково распределённые случайные величины.

3.2 Оценка регрета.

Метод зеркального спуска состоит в выборе на очередном шаге:

$$\textbf{(15)}\ x^{k+1} = Mirr_{x^k}\Big(h \nabla_x \widetilde{f}(x^k, \eta^k) \Big), Mirr_{x^k}(v) = \arg\min_{x \in Q} \Big\{ \Big\langle v, x - x^k \Big\rangle + V(x, x^k) \Big\},$$

где шаг спуска h будет впоследствии выбран.

Основное свойство МЗС:

(16) $2V(x, x^{k+1}) \le 2V(x, x^k) + 2h \Big\langle \nabla_x \widetilde{f}(x^k, \eta^k), x - x^k \Big\rangle + h^2 \Big\| \nabla_x \widetilde{f}(x^k, \eta^k) \Big\|_2^2$. Тогда можем получить:

$$\textbf{(17)}\ f(x^k) - f(x) \le_{\textbf{(9)}} \Big\langle \nabla f(x^k), x^k - x \Big\rangle =_{\textbf{(14)}} \Big\langle \mathsf{E}_{\eta^k}\Big\{ \nabla_x f(x^k, \eta^k) \mid \Theta^{k-1} \Big\}, x^k - x \Big\rangle \le_{\textbf{(8)}}$$

$$\le \Big\langle \mathsf{E}_{\eta^k}\Big\{ \nabla_x f(x^k, \eta^k) \mid \Theta^{k-1} \Big\} - \nabla_x \widetilde{f}(x^k, \eta^k), x^k - x \Big\rangle + \frac{1}{h}\Big(V(x, x^k) - V(x, x^{k+1}) \Big) + \frac{h}{2} \big\| \nabla_x \widetilde{f}(x^k, \eta^k) \big\|_2^2.$$

Берём условное матожидание $E_{\eta^k}\Big\{ \cdot \mid \Theta^{k-1} \Big\}$ от обеих частей неравенства **(17)** и воспользуемся неравенством (14):

$$\textbf{(18)}\ f(x^k) - f(x) \le \Big\langle \mathsf{E}_{\eta^k}\Big\{ \nabla_x f(x^k, \eta^k) - \nabla_x \widetilde{f}(x^k, \eta^k) \mid \Theta^{k-1} \Big\}, x^k - x \Big\rangle +$$
$$+ \frac{1}{h}\Big(V(x, x^k) - E_{\eta^k}\Big\{ V(x, x^{k+1}) \mid \Theta^{k-1} \Big\} \Big) + \frac{h}{2} \widetilde{M}^2.$$

Суммируем последнее неравенство по $k = 0, \dots, N-1$, делим на N обе части, а затем берём полное математическое ожидание от обеих частей и полагаем $x = x_* = \min_{x \in Q} f(x)$ и если такое x^* не единственно, тогда выбираем то, которое доставляет минимум $R^2 = V^*(x_*, x^0)$:

$$\textbf{(19)}\ \mathsf{E}\Big\{ \frac{1}{N} \sum_{k=0}^{N-1} f(x^k) \Big\} - f(x_*) \le \mathsf{E}\Big\{ \frac{1}{N} \sum_{k=0}^{N-1} \Big\langle \mathsf{E}_{\eta^k}\Big\{ \nabla_x f(x^k, \eta^k) - \nabla_x \widetilde{f}(x^k, \eta^k) \mid \Theta^{k-1} \Big\}, x^k - x_* \Big\rangle \Big\} +$$
$$+ \frac{1}{hN}\Big(V(x_*, x^0) - V(x_*, x^N) \Big) + \frac{h}{2} \widetilde{M}^2.$$

В силу того, что расстояние Брэгмана всегда неотрицательно, то с учётом (13) получаем оценку на регрет в зависимости от шага h в МЗС:

$$Regret_f\Big(\{x^k\}_{k=0}^{N-1}\Big) \le \sigma + \frac{R^2}{hN} + \frac{h}{2}\widetilde{M}^2. \tag{20}$$

Здесь $R^2 = V(x_*, x^0)$. Желаем минимизировать оценку по h, значит выбираем

$$h = \sqrt{\frac{2R^2}{N\widetilde{M}^2}} \tag{21}$$

При такой оценке получаем

$$Regret_f\Big(\{x^k\}_{k=0}^{N-1}\Big) \le \sigma + \sqrt{\frac{2R^2\widetilde{M}^2}{N}} \tag{22}$$

Значит, после

$$N = \frac{2R^2\widetilde{M}^2}{\varepsilon^2} \tag{23}$$

испытаний мы получим оценку

$$Regret_f\Big(\{x^k\}_{k=0}^{N-1}\Big) \le \sigma + \varepsilon. \tag{24}$$

3.3 Сглаживание задачи.

Пусть $\widetilde{e} \in RB_2^n(1)$ - случайный вектор, равномерно распределённый на шаре единичного радиуса в норме $\|\cdot\|_2$ в R^n. Сгладим исходную функцию $f(x,\widetilde{\eta})$ с помощью локального усреднения по шару радиуса $\mu > 0$ ($\mu < \mu_0$), который будет выбран позже, а $\widetilde{e}$ и $\widetilde{\eta}$ предполагаются независимыми случайными величинами:

$$f^\mu(x,\widetilde{\eta}) = \mathsf{E}_{\widetilde{e}}\Big\{f(x+\mu\widetilde{e},\widetilde{\eta})\Big\} = \frac{1}{V_{RB_2^n(1)}} \int\limits_{RB_2^n(1)} f(x+\mu t,\widetilde{\eta})dt, \tag{25}$$

$$f^\mu(x) = \mathsf{E}_{\widetilde{\eta}}\Big\{f^\mu(x,\widetilde{\eta})\Big\}. \tag{26}$$

Из условия **(10)** получаем:

$$f^\mu(x,\widetilde{\eta}) - f(x,\widetilde{\eta}) = \frac{1}{V_{RB_2^n(1)}} \int\limits_{RB_2^n(1)} \Big(f(x+\mu t,\widetilde{\eta}) - f(x,\widetilde{\eta})\Big)dt \le \frac{1}{V_{RB_2^n(1)}} \int\limits_{RB_2^n(1)} M\mu\|t\|_2 dt \le M\mu$$

с одной стороны. С другой стороны из выпуклости $f(x,\widetilde{\eta})$ как функции аргумента x неравенство Йенсена даёт:

$$f^\mu(x,\widetilde{\eta}) = \frac{1}{V_{RB_2^n(1)}} \int\limits_{RB_2^n(1)} f(x+\mu t,\widetilde{\eta})dt \ge f\Big(x + \int\limits_{RB_2^n(1)} \mu t dt, \widetilde{\eta}\Big) = f(x,\widetilde{\eta})$$

Таким образом, имеют место оценки:

$$0 \le f^\mu(x,\widetilde{\eta}) - f(x,\widetilde{\eta}) \le M\mu. \tag{27}$$

Возьмём матожидание $\mathsf{E}_{\widetilde{\eta}}$ от всех частей данного неравенства:

$$0 \leq f^{\mu}(x) - f(x) \leq M\mu. \tag{28}$$

Тогда если предположить, что:

$$M\mu \leq \frac{\varepsilon}{2} \tag{29}$$

и

$$Regret_{f^{\mu}}\Big(\{x^k\}_{k=0}^{N-1}\Big) \leq \frac{\varepsilon}{2} \tag{30}$$

то можем получить оценку изначального регрета:

$$\textbf{(31)}\ Regret_f\Big(\{x^k\}_{k=0}^{N-1}\Big) = \mathsf{E}\Big\{\frac{1}{N}\sum_{k=0}^{N-1} f(x^k)\Big\} - \min_{x\in Q} f(x) \leq$$
$$\leq \underbrace{\mathsf{E}\Big\{\frac{1}{N}\sum_{k=0}^{N-1} f^{\mu}(x^k)\Big\} - \min_{x\in Q} f^{\mu}(x)}_{Regret_{f^{\mu}}\left(\{x^k\}_{k=0}^{N-1}\right)\leq\frac{\varepsilon}{2}} + \frac{\varepsilon}{2} \leq \varepsilon.$$

Итак, при выполнении (29) оценка регрета для функции f^{μ} с точностью $\frac{\varepsilon}{2}$ позволяет оценить регрет для функции f с точностью ε.

3.4 Корректность оценок для сглаживаемой задачи.

Введём аналог зашумлённого стохастического градиента $\nabla_x \widetilde{f}(x^k, \eta^k)$ из предыдущего пункта:

$$\textbf{(31)}\ \nabla_x \widetilde{f}(x,\eta,e) = \frac{Vol\Big(RS_2^n(\mu)\Big)}{Vol\Big(RB_2^n(\mu)\Big)}\Big(\widetilde{f}(x+\mu e,\eta) - \widetilde{f}(x,\eta)\Big)e =$$
$$= \frac{n\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}\mu^{n-1}}{\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}\mu^{n}}\Big(\widetilde{f}(x+\mu e,\eta) - \widetilde{f}(x,\eta)\Big)e = \frac{n}{\mu}\Big(\widetilde{f}(x+\mu e,\eta) - \widetilde{f}(x,\eta)\Big)e,$$

где $e \in RS_2^n(1)$, т. е. случайный вектор e равномерно распределён на сфере радиуса 1 в евклидовой норме. Считаем, что разыгрывание e происходит независимо ни от чего. Аналогично можно определить незашумленную оценку стохастического градиента $\nabla_x f(x,\eta)$:

$$\nabla_x f(x,\eta,e) = \frac{n}{\mu}\Big(f(x+\mu e,\eta) - f(x,\eta)\Big)e. \tag{32}$$

Основное свойство такого градиента состоит в справедливости первой оценки в (14) и оно следует из векторного варианта формулы Стокса:

$$\mathsf{E}_e\Big\{\nabla_x f(x,\widetilde{\eta},e)\Big\} = \nabla_x f^{\mu}(x,\widetilde{\eta}) \tag{33}$$

Тогда оцениваем

$$\mathsf{E}\Big\{\frac{1}{N}\sum_{k=1}^{N}\Big\langle\nabla_x f(x^k,\eta^k)-\nabla_x\widetilde{f}(x^k,\eta^k),x^k-x_*\Big\rangle\Big\}=$$

$$=\mathsf{E}\Big\{\frac{1}{N}\sum_{k=1}^{N}\Big\langle\frac{n}{\mu}\Big(f(x^k+\mu e_k,\eta^k)-f(x^k,\eta^k)\Big)e_k-\frac{n}{\mu}\Big(\widetilde{f}(x^k+\mu e_k,\eta^k)-\widetilde{f}(x^k,\eta^k)\Big)e_k,x^k-x_*\Big\rangle\Big\}=$$

$$=\mathsf{E}\Big\{\frac{1}{N}\sum_{k=1}^{N}\Big\langle\frac{n}{\mu}\Big(f(x^k+\mu e_k,\eta^k)-\widetilde{f}(x^k+\mu e_k,\eta^k)\Big)e_k+\frac{n}{\mu}\Big(\widetilde{f}(x^k,\eta^k)-f(x^k,\eta^k)\Big)e_k,x^k-x_*\Big\rangle\Big\}.$$

Можем воспользоваться оценкой типа:

$$\langle x+y,z\rangle\le\Big|\langle x+y,z\rangle\Big|=\Big|\langle x,z\rangle+\langle y,z\rangle\Big|\le\Big|\langle x,z\rangle\Big|+\Big|\langle y,z\rangle\Big|$$

и в силу (7) получить:

$$\mathsf{E}\Big\{\frac{1}{N}\sum_{k=1}^{N}\Big\langle\nabla_x f(x^k,\eta^k)-\nabla_x\widetilde{f}(x^k,\eta^k),x^k-x_*\Big\rangle\Big\}\le\frac{2\delta n}{\mu}\mathsf{E}\Big\{\frac{1}{N}\sum_{k=1}^{N}\Big|\langle e_k,x^k-x_*\rangle\Big|\Big\}. \quad (34)$$

Если записать неравенство **(19)** с учётом неотрицательности регрета для произвольного k, то можно получить оценку при выбранном шаге h из (21):

$$\mathsf{E}\Big(\|x^k-x_*\|_2^2\Big)\le V(x_*,x^0)+\frac{Nh^2\widetilde{M}^2}{2}=R^2+R^2=2R^2 \quad (35)$$

Оказывается, явление концентрации равномерной меры на многомерной сфере позволяет из этой оценки и независимости $e_k\in RS_2^n(1)$ от x^k-x_* сделать вывод:

$$\mathsf{E}\Big\{\frac{1}{N}\sum_{k=1}^{N}\Big|\langle e_k,x^k-x_*\rangle\Big|\Big\}\le\frac{2R}{\sqrt{n}},$$

что позволяет выбрать

$$\sigma=\frac{4\delta R\sqrt{n}}{\mu} \quad (36)$$

с учётом (34).