Теория вероятностей. Лекция шестнадцатая
Неравенства концентрации меры

Дмитрий Валерьевич Хлопин
glukanat@mail.ru

Институт математики и механики им. Н.Н.Красовского

13.02.2019

Напоминание: вероятностное пространство

- Ω – некоторое множество;
- $\mathcal{F}$ – семейство подмножеств Ω, образующее σ-алгебру; $\mathcal{F}$ называют σ-алгеброй событий;
- $\mathbb{P}$ — функция из $\mathcal{F}$ в $[0,1]$, удовлетворяющая условию σ-аддитивности; саму функцию $\mathbb{P}$ называют вероятностью.

Интерпретация:

- Ω – множество исходов, которые реально могут произойти (полученное число, вектор, последовательность чисел, траектория);
- $\mathcal{F}$ – набор событий, про каждое из которых мы можем сказать, случилось оно или нет (был ли результат меньше некоторого числа, лежит ли результат в круге, имеет ли получившаяся последовательность 0 на k-м месте и т.д.);
- $\mathbb{P}$ – оценка, насколько вероятно то или иное событие из $\mathcal{F}$.

Напоминание: случайные величины

Множество действительных чисел обычно наделяется σ-алгеброй $\mathcal{B}(\mathbb{R})$ борелевских множеств (минимальной σ-алгеброй, содержащей все отрезки). Случайной величиной называется функция $\xi : \Omega \to \mathbb{R}$, обладающая свойством *измеримости*, т.е. для любого множества $A \in \mathcal{B}(\mathbb{R})$

$$\xi^{-1}(A) = \{\omega \in \Omega : \xi(\omega) \in A\} \in \mathcal{F}.$$

Свойство измеримости позволяет считать законным запрос о вероятности события $\xi(\omega) \in A$.
Если $\xi_1, \xi_2, \ldots$ являются случайными величинами, то

$$\xi_1 + \xi_2, \quad \xi_1 \xi_2, \quad \lim_{n\to\infty} \xi_n, \quad \sup_n \xi_n$$

также являются случайными величинами.
Суперпозиция измеримых случайных величин также измерима.

Напоминание: функция распределения

Каждая случайная величина ξ автоматически задает вероятность на $\mathbb{R}$ правилом:

$$(\xi\#\mathbb{P})(A) \triangleq \mathbb{P}(\xi^{-1}(A)).$$

Для большинства приложений при этом достаточно помнить про функцию распределения F_ξ, заданную правилом

$$F_\xi(x) = (\xi\#\mathbb{P})(-\infty, x] = \mathbb{P}(\xi \le x) = \mathbb{P}\{\omega \in \Omega \,|\, \xi(\omega) \le x\}.$$

Функция распределения — ступенчатая тогда и только тогда, когда ξ принимает не более чем счетное число значений;
ξ абсолютно непрерывна тогда и только тогда, когда при всех $t \in \mathbb{R}$ $F_\xi(x) = \int_{-\infty}^{x} f_\xi(t)\,dt$ для некоторой плотности $f_\xi\colon \mathbb{R} \to \mathbb{R}$

Напоминание: матожидание

Математическим ожиданием $\mathbb{E}\xi$ случайной величины $\xi : \Omega \to \mathbb{R}$ назовем значение выражения

$$\int_\Omega \xi(\omega)\,\mathbb{P}(d\omega) = \int_\mathbb{R} x\,(\xi\#\mathbb{P})(dx) = \int_\mathbb{R} x\,dF_\xi(x),$$

если оно существует и конечно.

В простейших случаях, если случайная величина принимает не более чем счетное число значений, то

$$\mathbb{E}\xi = \sum_{i\in\mathbb{N}} p_i\xi_i;$$

если случайная величина имеет плотность, то

$$\mathbb{E}\xi = \int_\mathbb{R} x f_\xi(x)\,dx.$$

При этом можно пользоваться формулой замены переменных

$$E\phi(\xi) = \int_\mathbb{R} \phi(x)(\xi\#\mathbb{P})(dx)\Big(= \sum_{i\in\mathbb{N}} p_i\phi(\xi_i)\Big)\Big(= \int_\mathbb{R} \phi(x) f_\xi(x)\,dx\Big).$$

Напоминание: свойства матожидания попроще

0^0 $\mathbb{E}1_A = \mathbb{P}(A)$;

1^0 $\mathbb{E}\xi \geq 0$, если $\xi \geq 0$ для почти всех $\omega \in \Omega$;

2^0 $\mathbb{E}\xi_1 \geq \mathbb{E}\xi_2$, если $\xi_1(\omega) \geq \xi_2(\omega)$ для почти всех $\omega \in \Omega$;

3^0 $\mathbb{E}c = c$ для случайной величины, почти всюду равной константе $c \in \mathbb{R}$;

4^0 $\mathbb{E}(c\xi) = c\mathbb{E}\xi$ для любого $c \in \mathbb{R}$;

5^0 $\mathbb{E}\xi_1 + \mathbb{E}\xi_2 = \mathbb{E}(\xi_1 + \xi_2)$;

6^0 $a\mathbb{E}\xi_1 + b\mathbb{E}\xi_2 = \mathbb{E}(a\xi_1 + b\xi_2)$ при любых $a, b \in \mathbb{R}$;

7^0 $\mathbb{E}(\xi\eta) = \mathbb{E}\xi\mathbb{E}\eta$ для независимых случайных величин ξ, η.

Схема Бернулли. Заблуждения

Пусть эксперимент проводится n раз, вероятность успеха $p = 1/2$, неуспеха — $q = 1/2$.

Факт. Если n четно, то наиболее вероятно число успехов m, равное $n/2$.

Заблуждение. Если выпал орел, то следом выпадет решка, потому что вероятность $50/50$.

Столь же наивное заблуждение. Если достаточно долго подождать, то орлов и решек всегда поровну: $m/n = 1/2$.

Факт. Пусть n четно, $p = 1/2$, то

$$\mathbb{P}_n(\text{число успехов} = n/2) = C_n^{n/2} 2^{-n} \approx \sqrt{\frac{2}{\pi n}} \to 0.$$

Схема Бернулли. Что на самом деле

Следствие из пока не доказанной теоремы Муавра-Лапласа.

$$\frac{\text{число успехов}}{n} \approx p \pm O\left(\frac{1}{\sqrt{n}}\right).$$

$$\mathbb{P}_n\left(\left|\frac{\text{число успехов}}{n} - p\right| \le \frac{3}{\sqrt{npq}}\right) \to 0,9973\ldots$$

$$\mathbb{P}_n\left(\left|\frac{\text{число успехов}}{n} - p\right| \le \frac{4}{\sqrt{npq}}\right) \to 0,999937\ldots$$

Вывод: в точности $m/n = p = 1/2$ бывает чем дальше, тем реже.
Однако,
при конечном числе n повторений заданных условий доля числа m случаев, когда случится успех, то есть так называемая *частота* m/n, как правило, близка к p.

Цель

Вы проводите ряд экспериментов, в каждом находите, вообще говоря, свою случайную величину X_i
(в мире предыдущего слайда: произошло событие или не произошло, 1 или 0 соответственно, но, в общем случае, X_i может означать все что угодно, от средней температуры по больнице до курса акций Apple). Что можно сказать при достаточно больших n о случайной величине

$$\frac{X_1 + \cdots + X_n}{n}?$$

Какие условия надо наложить на эксперимент, чтобы при $n \to \infty$ можно было перейти к пределу?
Чему этот предел равен?
В каком смысле этот предел понимать?
Насколько велико должно быть n, чтобы погрешность была удовлетворительна?

Этих вопросов хватит надолго. Сначала просто оценки.

Неравенство Маркова

Предложение 1. Для всякой случайной величины ξ и любого числа $a > 0$ имеет место неравенство

$$\mathbb{P}(|\xi| \geq a) \leq \frac{\mathbb{E}|\xi|}{a}.$$

Доказательство. Имеем, что

$$\mathbb{E}|\xi| = \mathbb{E}|\xi|\mathbf{1}_{|\xi|<a} + \mathbb{E}|\xi|\mathbf{1}_{|\xi|\geq a}.$$

Поскольку первое слагаемое снизу может быть оценено 0, а второе — величиной $a\mathbb{P}(|\xi| \geq a)$, получаем, что

$$\mathbb{E}|\xi| \geq a\mathbb{P}(|\xi| \geq a).$$

Неравенство Маркова, общий случай

Следствие 1. Для всякой случайной величины ξ и любого числа a, для любой монотонно возрастающей положительной функции ϕ имеет место неравенство

$$\mathbb{P}(\xi \geq a) \leq \frac{\mathbb{E}\phi(\xi)}{\phi(a)}.$$

Доказательство. Действительно,

$$\mathbb{P}(\xi \geq a) = \mathbb{P}(\phi(\xi) \geq \phi(a)) \leq \frac{\mathbb{E}\phi(\xi)}{\phi(a)}.$$

Следствие 2. Для всякой случайной величины ξ имеют место неравенства

$$\begin{aligned}
\mathbb{P}(\xi \geq a) &\leq \inf_{t>0} e^{ta}\mathbb{E}e^{-t\xi} \quad \forall a \in \mathbb{R} \text{ (граница Чернова)} \\
\mathbb{P}(|\xi| \geq |a|) &\leq \inf_{t>0} a^{-t}\mathbb{E}|\xi|^t \quad \forall a > 0 \text{ (моментная граница)}.
\end{aligned}$$

Неравенство Чебышева

Предложение 2. Для всякой имеющей дисперсию случайной величины ξ и любого числа $a > 0$ имеет место неравенство

$$\mathbb{P}(|\xi - \mathbb{E}\xi| \geq a) \leq \frac{D\xi}{a^2}.$$

Доказательство.
Заметим, что

$$\mathbb{P}(|\xi - \mathbb{E}\xi| \geq a) = \mathbb{P}(|\xi - \mathbb{E}\xi|^2 \geq a^2).$$

Отсюда, применяя неравенство Маркова к случайной величине $\eta = |\xi - \mathbb{E}\xi|^2$, математическое ожидание которой равно $D\xi$, мы получаем неравенство Чебышева.

Следствие 3 [Правило трех сигм]. Для всякой случайной величины ξ с $D\xi = \sigma^2$ имеет место неравенство

$$\mathbb{P}(|\xi - \mathbb{E}\xi| > 3\sigma) \leq \frac{1}{9}.$$

Закон больших чисел

Следствие 4. Для последовательности попарно независимых, одинаково распределенных случайных величин X_i с конечной дисперсией, при всех положительных a имеет место неравенство

$$\mathbb{P}\left(\left|\frac{X_1+\dots+X_n}{n}-\mathbb{E}X_1\right|\geq a\right)\leq\frac{DX_1}{na^2}.$$

Доказательство.

Заметим, что $\mathbb{E}\frac{X_1+\dots+X_n}{n}=\mathbb{E}X_1$, $D\frac{X_1+\dots+X_n}{n}=\frac{n}{n^2}DX_1=DX_1/n$. Теперь,

$$\mathbb{P}\left(\left|\frac{X_1+\dots+X_n}{n}-\mathbb{E}X_1\right|\geq a\right)\leq\frac{D\frac{X_1+\dots+X_n}{n}}{a^2}=\frac{DX_1}{na^2}.$$

Подумать: можно ослабить требование конечной дисперсии, рассматривать слабо коррелированные случайные величины, избавиться от общего распределения у X_i...

Мы позже увидим более точные оценки, но они и требуют большего.

Пример задачки

В поселке $N = 23025$ жителей, каждый из которых случайно и независимо от остальных ездит в город в среднем два раза в месяц. Поезд из поселка в город идет один раз в сутки. Найдите распределение числа K пассажиров, их среднее число и дисперсию. Какова должна быть вместимость w поезда, чтобы он переполнился с вероятностью, не превышающей $0,01$?

Неравенство Чебышева дает $w \sim 1914$, граница Чернова — $w \sim 1652$. Поскольку здесь медиана (как и среднее) — $N/15 = 23025/15 = 1535$ пассажиров, откуда $w > 1535$, то оценка 1652 кажется много лучше 1914.

Экспоненциальная оценка

Сейчас наша цель вывести

Следствие 5. Пусть дана последовательность независимых в совокупности, одинаково распределенных случайных величин X_i, принимающих значения из отрезка $[0,1]$. Тогда, для всех натуральных n имеем

$$\mathbb{P}\left(\left|\frac{X_1+\dots+X_n}{n}-\mathbb{E}X_1\right|\geq\delta\mathbb{E}X_1\right)\leq 2e^{-n\delta^2\mathbb{E}X_1/2}\qquad\forall\delta\in(0,1).$$

Подумать: что дает такой вариант закона больших чисел, по сравнению с предыдущим, полиномиальным вариантом? Чем приходится жертвовать?

Само следствие выводится сразу из замечаний к следующему предложению.

Граница Чернова для ограниченных сл.в. I

Предложение 3. Пусть дана последовательность независимых в совокупности, одинаково распределенных случайных величин X_i, принимающих значения из отрезка $[0,1]$. Тогда для всех натуральных n и положительных δ имеем

$$\mathbb{P}\Big(X_1 + \cdots + X_n \geq (1+\delta)M\Big) \leq e^{M(\delta-(1+\delta)\ln(1+\delta))},$$

где $M \triangleq \mathbb{E}(X_1 + \cdots + X_n)$.

Доказательство. Примем $t = \ln(1+\delta) > 0$. Пусть $X = X_1 + \cdots + X_n$, $Y_i = e^{tX_i}$, $Y = e^{tX}$.

Заметим, что $e^{tx} \leq 1 - x + e^t x$ для $x \in [0,1]$ в силу

$$e^{xt+(1-x)0} = \exp(tx) \leq x\exp(t) + (1-x)\exp(0),$$

Граница Чернова для ограниченных сл.в. II

откуда,

$$\mathbb{E}[e^{tX_i}] \le \mathbb{E}[1 - X_i + e^t X_i] = 1 + (e^t - 1)\mathbb{E}X_i \le e^{(e^t-1)\mathbb{E}X_i}.$$

Теперь из $\mathbb{E}Y = \prod_{i=1}^n \mathbb{E}Y_i \le \prod_{i=1}^n e^{(e^t-1)\mathbb{E}X_i} \le e^{(e^t-1)\sum_{i=1}^n \mathbb{E}X_i} \le e^{(e^t-1)M}$ имеем

$$\begin{aligned}
\mathbb{P}\Big(X_1 + \cdots + X_n \ge (1+\delta)M\Big) &= \mathbb{P}(e^{t(X_1+\cdots+X_n)} \ge e^{t(1+\delta)M}) \\
&= \mathbb{P}(Y \ge e^{t(1+\delta)M}) \le \frac{\mathbb{E}Y}{e^{t(1+\delta)M}} \\
&\le \frac{e^{(e^t-1)M}}{e^{t(1+\delta)M}} = e^{M(e^t-1-t(1+\delta))}.
\end{aligned}$$

Подставляя $t = \ln(1+\delta)$, получаем

$$\mathbb{P}\Big(X_1 + \cdots + X_n \ge (1+\delta)M\Big) \le e^{M(\delta-(1+\delta)\ln(1+\delta))}.$$

Замечания к предложению 3

Из $\frac{1}{2}\ln(1+\delta) = \frac{t}{2} \geq \tanh\frac{t}{2} = \frac{e^t-1}{e^t+1} = \frac{\delta}{2+\delta}$ следует
$(1+\delta)\ln(1+\delta) \geq \frac{2(1+\delta)\delta}{2+\delta} = \delta + \frac{\delta^2}{2+\delta}$, откуда имеем

Замечание 1. В условиях предложения 3 выполнено

$$\begin{aligned}
\mathbb{P}\Big(X_1+\dots+X_n \geq (1+\delta)M\Big) &\leq e^{-M\delta^2/(2+\delta)} \qquad \forall\delta>0,\\
\mathbb{P}\Big(X_1+\dots+X_n \geq (1+\delta)M\Big) &\leq e^{-M\delta^2/3} \qquad \forall\delta\in[0,1],\\
\mathbb{P}\Big(X_1+\dots+X_n \geq (1+\delta)M\Big) &\leq e^{-M\delta/3} \qquad \forall\delta>1.
\end{aligned}$$

Проведя это же доказательство для $t=\ln(1-\delta)$, получим

Замечание 2. В условиях предложения 3 выполнено

$$\mathbb{P}\Big(X_1+\dots+X_n \leq M(1-\delta)\Big) \leq e^{-M(\delta+(1-\delta)\ln(1-\delta))} \leq e^{-M\delta^2/2} \quad \forall\delta\in(0,1).$$

Подумать: как изменятся эти оценки, если значения X_i из отрезка $[-K,K]$.

Что разобрали:

- Неравенства концентрации меры
- Сходимости случайных величин
- Усиленный закон больших чисел
- Слабая сходимость
- Характеристические функции
- Центральная предельная теорема

На самом деле этих неравенств концентрации меры до чертиков, потому как, используя минимальную априорную информацию (знание дисперсии или знание априорной ограниченности), можно получать разные оценки. Не столько даже для того, чтобы оценить погрешность среднего выборочного от матожидания, это все же попроще, но и, например, чтобы оценить "тяжелые хвосты", вероятность "черных лебедей" и прочая.