6. Метод Монте-Карло

Введение.

Моделирование независимых случайных величин с заданным законом распределения (например, равномерным) имеет важное приложение: приближённое вычисление интегралов от функций **большого** числа переменных. Пусть $f\colon [0,1]^d \to R$ - «хорошая» (скажем, непрерывная или гладкая) функция, определённая на d – мерном кубе $[0,1]^d$. Требуется вычислить интеграл

$$I(f) = \int\limits_{[0,1]^d} f(x_1, \dots, x_d) dx_1 \dots dx_d.$$

В случае $d = 1$ и *непрерывной* функции f можно применить **метод прямоугольников.** Суммы

$$S_n(f) = \sum_{k=0}^{n-1} \frac{1}{n} f\left(\frac{k}{n}\right) \tag{1}$$

являются аппроксимациями интеграла $I(f)$) , причём можно контролировать качество этих аппроксимаций следующим образом: если $\delta(x)$ модуль непрерывности функции f, то

$$|I(f) - S_n(f)| \leq \delta\left(\frac{1}{n}\right) \tag{2}$$

Напомним определение модуля непрерывности $\delta(\cdot)$:

$$\delta(\varepsilon) = sup_{|x-y|\leq\varepsilon}(|f(x) - f(y)|).$$

В частности, $|f(x) - f(y)| \leq \delta(|x - y|)$, и если $f \in C^1([0,1])$, то для модуля непрерывности этой функции справедлива оценка

$$\delta(\varepsilon) \leq \left(max_{[0,1]}|f'|\right) \cdot \varepsilon.$$

Если f более регулярна, то можно предложить более эффективные процедуры. Так, например, **метод трапеций** состоит в приближении графика функции f непрерывной, кусочно-линейной функцией. Положим

$$S_n = \frac{1}{2n}\left((f_0 + f_1) + (f_1 + f_2) + \cdots + (f_{n-1} + f_n)\right), f_i = f\left(\frac{i}{n}\right).$$

Если f - функция класса $C^2([0,1])$, то

$$|I(f) - S_n(f)| \leq \frac{C}{n^2}. \tag{3}$$

Метод Симпсона состоит в приближении f непрерывной, кусочно-полиномиальной функцией степени 2 (т.е. кусочно-параболической функцией). Если n чётно, положим

$$S_n = \frac{1}{3n}\left((f_0 + 4f_1 + f_2) + (f_2 + 4f_3 + f_4) + \cdots \right.$$
$$\left. + (f_{n-2} + 4f_{n-1} + f_n)\right), f_i = f\left(\frac{i}{n}\right)$$

и если f – функция класса $C^4([0,1])$, то

$$|I(f) - S_n(f)| \leq \frac{C}{n^4}. \tag{4}$$

В высших размерностях $(d \geq 2)$ обобщением суммы (1) является следующая формула

$$S_n(f) = \frac{1}{n^d}\sum_{k_1=0}^{n-1} \dots \sum_{k_d=0}^{n-1} f\left(\frac{k_1}{n}, \dots, \frac{k_d}{n}\right),$$

и ошибка аппроксимации интеграла $I(f)$ по-прежнему имеет вид (если $f \in C^1([0,1]^d)$)

$$|I(f) - S_n(f)| \leq \frac{C}{n}.$$

Поэтому, чтобы получить ошибку аппроксимации порядка ε, т.е. $C/n \approx \varepsilon$, необходимо вычислять значения f в n^d узлах решётки (то есть число точек становится порядка $\approx \varepsilon^{-d}$) . Когда d велико (например, порядка 100), а ε мало, время вычисления становится непомерно большим. С другой стороны, методы, которые мы только что описали, требуют достаточной гладкости фукции f. Возможно, однако, преодолеть эти две трудности, если использовать **вероятностный алгоритм – метод Монте-Карло.**

6.1. Описание метода.

Метод основан на применении ***закона больших чисел и центральной предельной теоремы.*** Если f – функция класса L_1, определённая на кубе $[0,1]^d$ и $X_1, X_2, \dots$ - последовательность независимых случайных векторов, равномерно распределённых на $[0,1]^d$, то последовательность

$$Y_1 = f(X_1), \dots, Y_n = f(X_n), \dots$$

является последовательностью независимых, одинаково распредёленных с.в. с общим законом распределения μ

$$P(f(X_i) \in A) = P(f(X_1) \in A) = \mu(A), \qquad i = 2,3, \dots$$

для любого борелевского множества $A \in R$. Согласно ***усиленному закону больших чисел*** для P – почти всех $\omega \in \Omega$

$$\lim_{n\to\infty} \frac{1}{n}\big(f\big(X_1(\omega)\big) + \cdots + f(X_n(\omega))\big) = Ef\big(X_1(\omega)\big).$$

По формуле перехода

$$Ef\big(X_1(\omega)\big) = \int\limits_{R^d} f(x_1, \dots, x_d)\rho\,(x_1, \dots, x_d)dx_1 \dots dx_d,$$

где ρ – плотность равномерного распределения на единичном кубе $[0,1]^d$

$$\rho(x_1, \dots, x_d) = \mathbf{1}_{[0,1]^d}(x_1, \dots, x_d).$$

Подставляя эту плотность, имеем

$$Ef\big(X_1(\omega)\big) = \int\limits_{[0,1]^d} f(x_1, \dots, x_d)\, dx_1 \dots dx_d = I(f),$$

и, следовательно, $\boldsymbol{P}$ – почти наверное

$$\lim_{n\to\infty} \frac{1}{n}\Big(f\big(X_1(\omega)\big) + \cdots + f(X_n(\omega))\Big) = \int\limits_{[0,1]^d} f(x_1, \dots, x_d)\, dx_1 \dots dx_d.$$

Таким образом, мы получили, что ***с вероятностью 1 среднее арифметическое***

$$\frac{1}{n} S_n(f) := \frac{1}{n}\Big(f\big(X_1(\omega)\big) + \cdots + f(X_n(\omega))\Big)$$

сходится к интегралу, который мы хотим вычислить:

$$I(f) = \int\limits_{[0,1]^d} f(x_1, \dots, x_d)\, dx_1 \dots dx_d.$$

Для приложений важно оценить ***скорость сходимости*** к интегралу $I(f)$. Предположим, что f интегрируема с квадратом. Поскольку

$$Y_1 = f(X_1), \dots, Y_n = f(X_n), \dots$$

последовательность независимых одинаково распределённых, интегрируемых с квадратом случайных величин, то, согласно ***центральной предельной теореме,*** случайная величина

$$\frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\Big(f\big(X_1(\omega)\big) + \cdots + f(X_n(\omega))\Big) - \boldsymbol{E}f\big(X_1(\omega)\big)\right), \sigma^2 = Var\big(f(X_1)\big),$$

сходится по распределению к стандартному нормальному закону. Другими словами, если мы обозначим через σ^2 – дисперсию с.в. $f(X_1)$, то

$$\sigma^2 = \boldsymbol{E}f^2\big(X_1(\omega)\big) - \Big(\boldsymbol{E}f\big(X_1(\omega)\big)\Big)^2 = I(f^2) - (I(f))^2,$$

и для любого $a > 0$

$$\lim_{n\to\infty} \boldsymbol{P}\left(\left|\frac{1}{n}S_n(f) - I(f)\right| < \frac{\sigma a}{\sqrt{n}}\right) = \int_{-a}^{a} \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} dx.$$

Интегралы от стандартного нормального распределения табулированы. Например, интеграл

$$\int_{-a}^{a} \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} dx$$

приблизительно равен $0{,}95$ для $a = 1{,}96 \approx 2$ и 0,99 для $a = 2{,}6$. Таким образом, если мы умеем оценивать (хотя бы грубо) величину σ, а $a = 2$ или $a = 2{,}6$, то можно сказать, что с большой вероятностью (0,95 или 0,99 соответственно) $\frac{S_n(f)}{n}$ приближает $I(f)$ с точностью до $\sigma a/\sqrt{n}$, если n достаточно велико. Например, если ε достаточно мало, $a = 2,$ а $n = (2\sigma/\varepsilon)^2$, то $\frac{S_n(f)}{n}$ приближает $I(f)$ с точностью до $\frac{2\sigma}{\sqrt{n}} = \varepsilon$ с вероятностью 0,95. Заметим, что вычисление $S_n(f)$ требует вычисления $n = (2\sigma/\varepsilon)^2$ значений f, и что это число ***не зависит от размерности пространства d.*** Это существенное преимущество по сравнению с методами, описанными во Введении.

К недостаткам метода можно отнести следующие два:

1) Центральная предельная теорема справедлива ***только асимптотически:*** понятия «достаточное малое» (про ε) и «достаточно большое» (про n) не определены *априори.*

2) Метод предполагает ***наличие разумной оценки $\boldsymbol{\sigma}$***, то есть предполагает, что имеются оценки $I(f)$ и $I(f^2)$.

Каждая из этих двух проблем допускает решение, по крайней мере с точки зрения теории. Теоретическое решение первой проблемы основано на следующем результате, дающем оценку скорости сходимости в центральной предельной теореме.

Теорема 6.1. (неравенство Берри-Эссеена). *Пусть $X_1, \dots, X_n, \dots$ последовательность независимых одинаково распределённых случайных величин, $E(X_1) = 0, E(X_1^2) = \sigma^2, E(|X_1|^3) = \rho < \infty$. Если $F_n(x)$ - функция распределения с.в. $S_n/\sigma\sqrt{n}$, то для всех $x \in R$*

$$\left|F_n(x) - \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt\right| \le \frac{3\rho}{\sigma^3\sqrt{n}}.$$

Отсюда немедленно следует оценка скорости сходимости, дающая ответ на первый из поставленных вопросов:

Теорема 6.2. *Для всех* a

$$\left| \boldsymbol{P}\left(\left| \frac{1}{n} S_n(f) - I(f) \right| < \frac{\sigma a}{\sqrt{n}} \right) - \int_{-a}^{a} \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} dx \right| \le 6 \left(\frac{|f|_{C^0}}{\sigma} \right)^3 \frac{1}{\sqrt{n}}.$$

Здесь $|f|_{C^0} = \max_{[0,1]^d} |f(x)|$. Для доказательства достаточно применить Теорему 6.1 к случайным величинам $Y_i(\omega)$

$$Y_i(\omega) = f\big(X_i(\omega)\big) - Ef\big(X_i(\omega)\big).$$

Чтобы ответить на второй вопрос, естественно заменить неизвестную дисперсию σ^2 на её оценку Σ_n^2

$$\Sigma_n^2 = \frac{1}{n} S_n(f^2) - \left(\frac{1}{n} S_n(f) \right)^2.$$

Следующая теорема показывает, что при такой замене вид предельного распределения сохраняется.

Теорема 6.3.

$$\lim_{n\to\infty} \boldsymbol{P}\left(\left| \frac{1}{n} S_n(f) - I(f) \right| < \frac{\Sigma_n a}{\sqrt{n}} \right) = \int_{-a}^{a} \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} dx.$$

Доказательство. Обозначим

$$Z_n = \frac{\sqrt{n}}{\sigma}\left(\frac{S_n}{n} - I(f) \right),\ \tilde{Z}_n = \frac{\sqrt{n}}{\Sigma_n}\left(\frac{S_n}{n} - I(f) \right), Y_n = \frac{\sigma}{\Sigma_n}.$$

то есть $\tilde{Z}_n = Y_n Z_n$. Поскольку по усиленному закону больших чисел Σ_n сходится почти наверное к σ (т.е. Y_n сходится почти наверное к 1), а, как мы показали, Z_n сходится по распределению к стандартному нормальному закону Φ, то достаточно воспользоваться следующей леммой, которую мы приводим без доказательства.

Лемма 6.1.. *Если* U_n *сходится по распределению к* U *и если* $\frac{V_n}{U_n}$ *сходится почти наверое к 1, то и* V_n *сходится по расределению к* U.

Заметим, что в нашем случае $U_n = Z_n,\ U = \Phi, V_n = \tilde{Z}_n$.

6.2. Варианты метода.

Пусть теперь требуется вычислить интегралы по R^d (а не по параллелепипедам) . Тогда можно использовать следующую процедуру. Пусть $X_1, \dots, X_n$ ... последовательность независимых, одинаково распределённых случайных векторов с плотностью ρ, ***которую умеем моделировать***. Требуется вычислить

$$I(f) = \int\limits_{R^d} f(x_1, \dots, x_d)\, dx_1 \dots dx_d.$$

Имеем (поделим и умножим на плотность ρ)

$$I(f) = \int\limits_{R^d} f(x_1, \dots, x_d)\, dx_1 \dots dx_d =$$

$$I(f) = \int\limits_{R^d} \frac{f(x_1, \dots, x_d)}{\rho(x_1, \dots, x_d)} \rho(x_1, \dots, x_d) dx_1 \dots dx_d = E\big(\phi(X)\big),$$

где математическое ожидание берётся уже по плотности ρ и

$$\phi(x_1, \dots, x_d) = \frac{f(x_1, \dots, x_d)}{\rho(x_1, \dots, x_d)}.$$

Будем предполагать, что ***ρ не обращается в ноль на носителе функции f .*** Согласно закону больших чисел, случайная величина

$$S_n(\phi) = \frac{1}{n}\Big(\phi\big(X_1(\omega)\big) + \dots + \phi(X_n(\omega))\Big)$$

сходится почти наверное к $E\big(\phi(X_1)\big) = I(f)$. Рассуждая так же, как в предыдущем разделе, можно доказать следующую теорему:

Теорема 6.4. Положим

$$\sigma_n^2 = \frac{1}{n} S_n(\phi^2) - \left(\frac{1}{n} S_n(\phi)\right)^2$$

Тогда

$$\lim_{n\to\infty} \boldsymbol{P}\left(I(f) \in \left[\frac{1}{n} S_n(\phi) - \frac{\sigma_n a}{\sqrt{n}}, \frac{1}{n} S_n(\phi) + \frac{\sigma_n a}{\sqrt{n}}\right]\right) = \int_{-a}^{a} \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2} dx.$$

Выбор плотности ρ произволен, однако, на практике ищут плотность, которую достаточно легко моделировать.