Элементы теории оценивания

Олег Николаевич Граничин

Санкт-Петербургский государственный университет,
математико-механический факультет

20 марта 2013

Элементы теории оценивания

- Метод эмпирического функционала
- Байесовские оценки
- Метод максимума правдоподобия

- **Метод эмпирического функционала**
- Байесовские оценки
- Метод максимума правдоподобия

Метод эмпирического функционала

Рассмотрим задачу минимизации функционала среднего риска

$$F(\mathbf{x}) = \int F(\mathbf{x}, \mathbf{w}) \mathrm{P}_{\mathbf{w}}(d\mathbf{w}),$$

в котором $\mathrm{P}_{\mathbf{w}}(\cdot)$ функция распределения, заданная на некотором вероятностном пространстве событий. Возможность аналитического решения этой задачи предполагает знание распределения $\mathrm{P}_{\mathbf{w}}(\cdot)$. Вместе с тем, в ряде практических задач распределение $\mathrm{P}_{\mathbf{w}}(\cdot)$ не известно, но в распоряжении экспериментатора может быть определяемая им независимая выборка $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_N$. Одним из естественных подходов к задаче нахождения минимума является восстановление по этой выборке распределения $\mathrm{P}_{\mathbf{w}}(\cdot)$, точнее построение *эмпирического функционала*

$$F_N(x) = \frac{1}{N} \sum_{k=1}^{N} f(\mathbf{x}, \mathbf{w}_k),$$

который соответствует вычислению *методом Монте–Карло* интеграла задающего исходный функционал $F(x)$.

Метод Монте–Карло

Метода статистического моделирования Монте-Карло предложен в Лос-Аламосе Джоном фон Нейманом, Николасом Метрополисом и Станиславом Уламом в ходе работ над Манхэттенским проектом. Первоначально стохастический подход использовали для аппроксимации многомерных интегралов в уравнениях переноса, возникших в связи с задачей о движении нейтрона в изотропной среде. Рассмотрим следующий пример способа вычисления интеграла

$$F(\mathbf{x}) = \int_{\mathbf{w} \in \mathbb{W}} F(\mathbf{x}, \mathbf{w}) d\mathbf{w},$$

которым удобно пользоваться при невозможности аналитического интегрирования. Пусть множество $\mathbb{W}$ является параллелепипедом с объемом $V(\mathbb{W})$ и предположим, что значения функции f ограничены: $0 \leq f(\mathbf{x}, \mathbf{w}) \leq f_{\max}$. Геометрический смысл интеграла — объем под графиком функции f.

Метод Монте–Карло

1. Зафиксируем натуральное число N.
2. Выберем случайно в параллелепипеде $\mathbb{W} \times [0, f_{max}] \subset \mathbb{R}^{r+1}$ с равномерным одинаковым распределением N независимых точек $z_1, z_2, \ldots, z_N$.
3. Подсчитаем среди $z_1, z_2, \ldots, z_N$ число K тех z_i, $i \in 1..N$,, у которых последняя координата не превосходит значения функции f в соответствующей точке, определяемой первыми r координатами.
4. Вычислим оценку

$$\widehat{F}(N) = \frac{K}{N} Vol(\mathbb{W}) f_{\max}.$$

Точность оценивания I

Причем необходимую точность приближения можно гарантировать *априори* при выборе N в силу закона больших чисел и неравенства Чебышева

$$\forall \varepsilon > 0 \ \ \mathrm{P}\left\{\left|\frac{F(\mathbf{x})}{V(\mathbb{W})f_{\max}} - \frac{K}{N}\right| \geq \varepsilon\right\} \leq \frac{f(\theta)(V(\mathbb{W})f_{\max} - F(\mathbf{x}))}{N\varepsilon^2(V(\mathbb{W})f_{\max})^2}.$$

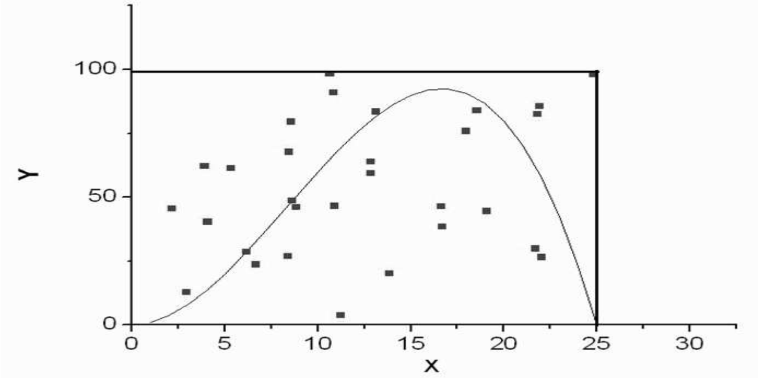
При небольшой размерности r множества $\mathbb{W}$ во многих случаях можно при вычислении интегралов пользоваться и традиционными детерминированными приближенными методами, но при $r \gg 1$ на практике фактически ничем нельзя воспользоваться кроме метода Монте-Карло, сложность которого не зависит r. Метод Монте-Карло может оказаться более предпочтительным и при небольших r, когда функция задана неявно, а необходимо определить область, заданную в виде сложных неравенств.

Точность оценивания II

По неравенству Хеффдинга

$$\forall \varepsilon > 0 \ \ \mathrm{Prob}\left\{ \left| \frac{K}{N} - \frac{F}{Vol(\mathbb{W}) f_{\max}} \right| \geq \varepsilon \right\} \leq 2e^{-2N\varepsilon^2}.$$

Пример. Интегрирование одномерной функции методом Монте-Карло



Из 30 случайных точек только $K = 16$ под графиком:
$f \approx 25 \cdot 100 \cdot 14/30 = 1166\frac{2}{3}$.
Вероятность относительной ошибки более 5% примерно равна $0,11$.

Сходимость значений эмпирических функционалов

Пусть величины $f(\mathbf{x}, \mathbf{w}_k)$, $k = 1, 2, \ldots, N$ — случайные, независимые, одинаково распределенные, и задано некоторое множество $\mathbb{X} \ni \mathbf{x}$. Если подынтегральная функция $f(\mathbf{x}, \mathbf{w})$ равномерно ограничена, или удовлетворяет более слабому условию

$$\sup_{\mathbf{x} \in \mathbb{X}} \int \|f(\mathbf{x}, \mathbf{w})\|^2 \mathrm{P}_{\mathbf{w}}(d\mathbf{w}) < \infty,$$

то в соответствии с законом больших чисел с вероятностью единица и в среднеквадратичном при $N \to \infty$ значения эмпирического функционала сходятся к соответствующим значениям исходного

$$\lim_{N \to \infty} F_N(\mathbf{x}) = F(\mathbf{x}),$$

причем сходимость в среднеквадратичном смысле имеет место равномерно по $\mathbf{x} \in \mathbb{X}$.

Оценки метода эмпирического функционала

Оценкой *метода эмпирического функционала* называется

$$\widehat{\mathbf{x}}_n = \arg\min_{\mathbf{x}\in\mathbb{X}} F_N(\mathbf{x})$$

любая из точек множества $\mathbb{X}$, соответствующая наименьшему значению функционала $F_N(\mathbf{x})$. При сделанных предположениях последовательность оценок $\{\widehat{\mathbf{x}}_n\}$ в среднеквадратичном смысле и с вероятностью единица сходится к множеству точек $\overline{\mathbb{X}}$, минимизирующих функционал $F(\mathbf{x})$. В частности, если это множество $\overline{\mathbb{X}}$ состоит всего из одной точки, то последовательность $\{\widehat{\mathbf{x}}_n\}$ состоятельная.

Метод эмпирического функционала в задачах идентификации динамических объектов

В задачах идентификации динамических объектов и адаптивного управления появляются более сложные эмпирические функционалы вида

$$F_N(\mathbf{x}) = \frac{1}{N} \sum_{k=1}^{N} f(k, \mathbf{x}, \mathbf{w}_k),$$

в котором штрафная функция $f(\cdot, \cdot, \cdot)$ зависит явно от времени, а случайные величины $\{\mathbf{w}_k\}$ не являются независимыми.