Математическая Статистика

4 марта 2014 г.

Глава 1

Основы

1.1 Методы оценок характеристик распределения наблюдаемых случайных величин

$x_1, \ldots, x_n$ — независимые одинаково распределённые случайные величины с неизвестной функцией распределения F.

Цель — найти F или сказать что-то о её свойствах.

Определение 1.1.1. *Эмпирической (выборочной) функцией распределения, построенной по выборке $x_1, \ldots, x_n$ называется функция*

$$F_n(x) = \frac{1}{n} \cdot \sum_{k=1}^{n} \mathbb{1}_{\{x_k \leq x\}}$$

$$F(x) = \mathbb{P}\{x_1 \leq x\} = M\mathbb{1}_{\{x_1 \leq x\}}$$

Теорема 1.1.1.

$$\mathbb{P}\left(F_n(x) \underset{n \to \infty}{\rightrightarrows} F(x)\right) = 1$$

Как можно попытаться отследить плотность распределения? Постараемся найти функцию распределения, а потом и плотность.

Допустим, F имеет хорошую (непрерывную) плотность. Тогда как из F получить p?

Мы знаем, что $F' = p$, но это никому не нужно, так как F_n' — производная ступенчатой функции, которая почти везде будет равна нулю.

Но также мы помним, что

$$F(b) - F(a) = \int\limits_a^b p(x)\,dx$$

Тогда, положив $a = x$, взяв некую Δ, и постановив $b = x + \Delta$, получаем следующее

$$F(x + \Delta) - F(x) = \int\limits_x^{x+\Delta} p(y)\,dy$$

Делим обе части на Δ и при достаточно малых его значениях получаем

$$\frac{1}{\Delta} \cdot \int\limits_{x}^{x+\Delta} p(y)\,dy = \frac{F(x+\Delta) - F(x)}{\Delta} \approx \frac{dF(x)}{dx} = p(x)$$

Значит, можем заменить $p(x)$ не производной, а такой разностью.

$$p(x) \approx \frac{F(x+\Delta) - F(x)}{\Delta}$$

Возьмём выборку из m случайных величин в порядке возрастания $a_1, \ldots, a_m$, обозначим отрезки $I_j = [a_{j-1}, a_j]$ и введём функцию $q(y)$

$$q(y) = \sum_{j=1}^{m} \frac{F(a_j) - F(a_{j-1})}{a_j - a_{j-1}} \cdot \mathbb{1}_{I_j}(y)$$

Теперь введём последовательность функций $q_n(y)$ и видим, что она сходится к $q_n(y)$ почти наверное согласно закону больших чисел, а та в свою очередь имеет сходимость порядка $\dfrac{1}{n}$ к плотности распределения $p(y)$

$$q_n(y) = \sum_{j=1}^{m} \frac{F_n(a_j) - F_n(a_{j-1})}{a_j - a_{j-1}} \cdot \mathbb{1}_{I_j}(y) \xrightarrow[n\to\infty]{a.s.} q(y) \xrightarrow[n\to\infty]{} p(y)$$

q_n — гистограмма. И вот конечная формула

$$q_n(y) = \sum_{j=1}^{m} \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{x_k \in I_j\}} \frac{1}{|I_j|} \cdot \mathbb{1}_{I_j}(y)$$

1.2 Оценка неизвестных параметров

Снова у нас есть $x_1, \ldots, x_n$ — выборка из распределения F_θ, где θ — неизвестный параметр из множества Θ

Пример 1.2.1. *Нормальное распределение с известным СКО $\sigma = 1$ и неизвестным математическим ожиданием, тогда θ — математическое ожидание*

Пример 1.2.2. *Нормальное распределение, в котором неизвестны оба параметра. Тогда θ будет парой (a, σ)*

Главный вопрос — определение основных параметров.

Определение 1.2.1. *Функцию от выборки, значение которой заменяет неизвестный параметр, назвают оценкой*

Пример 1.2.3. *Предположим, что выборка сделана из распределения Бернулли, то есть $\{x_i\}$ — набор одинаково распределённых случайных величин, причём*

$$x_i = \begin{cases} 1, & p \\ 0, & 1-p \end{cases}$$

Тогда неизвестный параметр — величина p (вероятность удачного эксперимента)

$$\theta = p \in [0;1] = \Theta$$

Оглавление