2 О параметрическом оценивании, оценках максимального правдоподобия и спейсингов

Мы привыкли рассматривать параметрическую модель $X_i \sim F_\theta$, где F — какое-то заданное семейство функций распределения, причем обычно довольно узкое — например, все нормальные распределения.

Будем обозначать $X_1, \ldots, X_n$ выборку, то есть сами случайные величины, а $x_1, \ldots, x_n$ — реализацию, то есть принятые ими в эксперименте значения.

2.1 Оценки и их свойства

Оценкой мы называем измеримую функцию $\widehat{\theta}(X_1, \ldots, X_n)$, которая приближает параметр θ или функцию $g(\theta)$ от него.

2.1.1 Свойства оценок

Важными свойствами оценок для нас были:

- Несмещенность $\mathbf{E}_\theta \widehat{\theta}(X_1, \ldots, X_n) = \theta$ при всех $\theta \in \Theta$.
- Состоятельность: $\widehat{\theta}(X_1, \ldots, X_n) \xrightarrow{P_\theta} \theta$ при всех $\theta \in \Theta$.
- Асимптотическая нормальность: $\sqrt{n}(\widehat{\theta}(X_1, \ldots, X_n) - \theta) \xrightarrow{d} Z \sim \mathcal{N}(0, \sigma^2(\theta))$. Величину $\sigma^2(\theta)$ называют асимптотической дисперсией. Вместо $\sqrt{n}$ иногда рассматривают другие скорости сходимости.

Если наша цель была оценить не θ, а какую-то функцию $g(\theta)$, то в правой части 1)-3) мы будем ставить $g(\theta)$.

Свойство несмещенности нам важно, если мы повторяем опыт, тогда мы можем гарантировать себе, что у нас не будет систематического занижения или завышения оценки.

Свойство состоятельности необходимо, если мы имеем возможность наращивать количество испытаний.

Наконец, свойство асимптотической нормальности уточняет состоятельность. Непосредственно нормальность предельного распределения дает достаточно распространенный класс предельных распределений и некоторые удобные свойства, в частности функциональную инвариантность, о которой мы сейчас поговорим.

2.1.2 Функциональная инвариантность

Отметим несколько важных для нас замечаний, связанных с так называемой функциональной инвариантностью оценок. Если оценка $\widehat{\theta}$ обладает каким-то свойством как оценка $g(\theta)$, то обладает ли $h(\widehat{\theta})$ тем же свойством как оценка $h(g(\theta))$.

- Если оценка $\widehat{\theta}(X_1, \ldots, X_n)$ несмещенная для функции $g(\theta)$, то $h(\widehat{\theta}(X_1, \ldots, X_n))$ не обязана быть несмещенной оценкой для $h(g(\theta))$. Более того, если, скажем, h строго выпукла вверх или вниз, а $\widehat{\theta}$ не константа, то это заведомо не так.
- Если оценка $\widehat{\theta}(X_1, \ldots, X_n)$ состоятельная для функции $g(\theta)$, а h — непрерывная функция, то $h(\widehat{\theta}(X_1, \ldots, X_n))$ — состоятельная оценка $h(g(\theta))$.
- Если оценка $\widehat{\theta}(X_1, \ldots, X_n)$ асимптотически нормальная для функции $g(\theta)$ с асимптотической дисперсией $\sigma^2(\theta)$, а h — дифференцируемая функция, то $h(\widehat{\theta}(X_1, \ldots, X_n))$ будет асимптотически нормальной оценкой для $h(g(\theta))$ с дисперсией $(\sigma(\theta)h'(g(\theta)))^2$.

Последнее утверждение называют Delta method (в русскоязычной литературе также встречается название "Лемма об асимптотической нормальности"). Этот факт мы сформулировали для одномерной функции h, аналогичный факт верен и в случае векторной функции:

Теорема 1. *Если векторная оценка $\widehat{\theta}(X_1, \dots, X_n)$ асимптотически нормальная для функции $g(\theta)$, $g: \Theta \to \mathbb{R}^k$ с асимптотической ковариацией $\Sigma(\theta)$, то есть*

$$\sqrt{n}\left(\widehat{\theta}(X_1, \dots, X_n) - g(\theta)\right) \xrightarrow{d} Z \sim \mathcal{N}(\vec{0}, \Sigma(\theta)),\ n \to \infty.$$

Пусть отображение $h: \mathbb{R}^k \to \mathbb{R}^m$ дифференцируемо в точке $g(\theta)$ и имеет матрицу Якоби

$$J(t_1, \dots, t_m) = \left(\frac{\partial h_i(t_1, \dots, t_m)}{\partial t_j},\ i \le k,\ j \le m\right),$$

где $h(t_1, \dots, t_m) = (h_1(t_1, \dots, t_m), \dots, t_k(t_1, \dots, t_m))$. Тогда

$$\sqrt{n}\left(h(\widehat{\theta}(X_1, \dots, X_n)) - h(g(\theta))\right) \xrightarrow{d} Z \sim \mathcal{N}(\vec{0}, J(g(\theta))\Sigma(\theta)J^{\mathrm{t}}(g(\theta))),\ n \to \infty,$$

то есть асимптотическая нормальность сохраняется при действии дифференцируемых отображений.

В векторной форме функциональная инвариантность верна и для состоятельности. В частности, отсюда вытекает и инвариантность около базовых операций, например, суммирования, поскольку $x+y$ — дифференцируемая функция двух переменных.

Сумма двух асимптотически нормальных оценок уже необязательно асимптотически нормальная. Для этого нужно, чтобы вектор-оценка была асимптотически нормальной оценкой вектора.

Три упомянутых свойства оценок далеко не единственные: нас может интересовать эффективность или оптимальность оценок (то есть минимальность их дисперсии в классе несмещенных оценок), робастность (то есть устойчивость к выбросам в выборке) и другие свойства.

2.2 Методы построения оценок

Два базовых метода для построения оценок, которые вы рассматривали в курсе статистики— метод моментов и метод максимального правдоподобия. Напомним как они устроены, обсудим их качества и добавим еще один метод, называемый методом спэйсингов.

2.2.1 Метод моментов

Метод моментов базируется на том, что в силу ЗБЧ и ЦПТ выборочные средние $\overline{X^k} = \frac{X_1^k + \dots + X_n^k}{n}$— хорошие оценки для $\mathbf{E}_\theta X_1^k$, а именно состоятельные и при $\mathbf{E}_\theta X_1^{2k} < \infty$ асимптотически нормальные с асимптотической дисперсией $\mathbf{D}_\theta X_1^k$. С помощью функциональной инвариантности мы можем попытаться сделать из них соответствующие оценки для $\vec{\theta}$.

Для оценки параметра $\vec{\theta} = (\theta_1, \dots, \theta_k)$ предлагается рассмотреть $\overline{X^1}, \dots, \overline{X^k}$ как оценки $\mu_1(\theta) = \mathbf{E}_\theta X, \dots, \mu_k(\theta) = \mathbf{E}_\theta X^k$ и применить к ним такое отображение f, которое переводит $(\mu_1(\theta), \dots, \mu_k(\theta))$ в $\vec{\theta}$. Если это отображение окажется непрерывным, то $f(\overline{X}, \dots, \overline{X^k})$ будет состоятельной для $\vec{\theta}$, а если дифференцируемым, то асимптотически нормальным (если, конечно $\mathbf{E}_\theta X^{2k} < \infty$).

Иначе говоря, чтобы найти оценку методом моментов (ОММ), мы должны решить систему уравнений

$$\begin{cases} \mu_1(\theta_1, \dots, \theta_k) = \overline{X^1}, \\ \mu_2(\theta_1, \dots, \theta_k) = \overline{X^2}, \\ \qquad \dots \\ \mu_k(\theta_1, \dots, \theta_k) = \overline{X^k}. \end{cases}$$

Если система оказалась несовместной, то мы можем убрать часть уравнений, заменив их на соотношения на следующие моменты.

К сожалению, метод моментов не вполне удачен в плане асимптотической дисперсии, она зачастую бывает весьма большой, особенно в случае многомерных параметров. Привязка непосредственно к моментам ограничивает возможности метода. Зачастую используют модификацию, основанную на так называемых *пробных функциях*. В этом случае рассматриваются моменты $\overline{g_i(X)}$, $i = 1, \ldots, k$ и рассматриваются уравнения

$$\mathbf{E}_{\theta_1,\ldots,\theta_k} g_i(X_1) = \overline{g_i(X)},\ i = 1, \ldots, k.$$

Для решения уравнений в R пригодится функция uniroot, в многомерном случае polyroot из библиотеки rootsolve. Для интегрирования можно использовать в R функцию integrate (лучше интегрировать не по всей прямой, а по конечному отрезку).

В Python для решения уравнений есть scipy.optimize.fsolve, для интегрирования scipy.integrate.

Задача 1. Построить алгоритм, численно вычисляющий по однопараметрической функции плотности ОММ.

2.2.2 Метод максимального правдоподобия

Этот метод исходит из простого соображения — мы должны искать такое θ, при котором появление нашей выборки особенно вероятно. Это равносильно максимизации совместного распределения (в случае дискретной выборки) или совместной плотности (в случае абсолютно-непрерывной плотности). Таким образом, метод максимального правдоподобия предписывает искать θ, такие что

$$L(x_1, \ldots, x_n; \theta) = f_\theta(x_1) \cdots f_\theta(x_n) \to \max,$$

где $f_\theta(x)$ — плотность X_i или вероятность $\mathbf{P}_\theta(X = x)$. Практическую зачастую удобнее искать аргумент максимума $\ln L(x_1, \ldots, x_n; \theta)$.

2.2.3 Недостатки ОМП

В некоторых случаях ОМП может быть неединственной или ее не будет совсем. Приведем некоторые случаи, когда у ОМП возникают проблемы:

1. Если распределение X_i устроено так, что при разных параметрах θ величины принимают значения из разных множеств, то есть носитель распределения зависит от параметра.

   Вопрос 1. Что будет ОМП в случае $X_i \sim R[\theta, \theta + 1]$?

2. Если плотность распределения X_i не гладко зависит от параметра.

   Вопрос 2. Что будет ОМП в случае X_i с плотностью $\exp(-|x - \theta|)/2$?

   Вопрос 3. Что будет с ОМП, если X_i бернуллиевские с параметром θ при $\theta \in \mathbb{Q} \cap (0, 1)$ и $1 - \theta$ иначе?

3. Если модель содержит большое число параметров.

   Вопрос 4. Что будет с ОМП, если $(X_{2i}, X_{2i-1}) \sim \mathcal{N}(\theta_i, \sigma^2)$. Будет ли ОМП для σ^2 состоятельной?

4. Если множество изменения параметра не является замкнутым, а плотность может неограниченно расти при определенных значениях аргумента и приближении параметра к границе области изменения.
   Такого рода ситуация может привести к тому, что ОМП будет отсутствовать, поскольку глобальный максимум равен бесконечности. Примеров такого рода достаточно много, рассмотрим такой

   Вопрос 5. Пусть $f(x)$ равна $c/\sqrt{x}$ при $|x| \leq 1$ и равна $d/x^{3/2}$ при $|x| > 1$, где c, d подобраны так, что это плотность, $X_i \sim f(x - \theta)$. Как будет выглядеть ОМП?

   Тот же вопрос для X_i плотностью $\sum_k f(x - \theta - q_k)/2^k$ при $k \geq 1$, q_k — упорядоченные рациональные числа.

   Однако, ситуация может оказаться еще более коварной — скажем, максимум правдоподобия существует, и ОМП даже сходится к какому-то значению, однако вовсе не совпадающему со значением параметра. Суть проблемы в том, что оценка максимального правдоподобия рассчитывает на то, что правдоподобие достигает максимального значения засчет того что основная масса наблюдений при таком значении параметра имеет достаточно большую плотность. Может, однако, оказаться что максимум достигается за счет огромных значений плотности в одном из элементов реализации, хотя во всех остальных точках значения плотности достаточно маленькие. Пример такого рода случая рассмотрен в задаче ниже.

Приведем еще один пример однопараметрического распределения, на котором ОМП работает достаточно неудачно, в частности, несостоятельна, хотя плотность ограничена при любом наборе параметров. Причина в том, что глобальный максимум достигается засчет роста плотности в окрестности одного из наблюдений за счет изменения значений параметра.

Задача 2. Моделировать выборку с плотностью

$$f_\theta(x) = \frac{1}{2\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) + \frac{1}{2\sqrt{2\pi}\sigma(\theta)}\exp\left(-\frac{(x-\theta)^2}{2\sigma^2(\theta)}\right),\ \sigma^2(\theta) = \exp\left(-\frac{1}{\theta^2}\right).$$

Это распределение представляет смесь двух нормальных выборок — величина равновероятно выбирается либо из $\mathcal{N}(0,1)$, либо из $\mathcal{N}(\theta,\sigma^2(\theta))$. Построить график логарифма правдоподобия и посмотрите чему оно равно при параметре, равном минимальной по модулю точке выборки.

2.2.4 Достоинства ОМП

Мы указали столько минусов, что может возникнуть вопрос: "Зачем вообще нужна такая оценка?" Однако, не все так плохо — зависимость носителя от параметра чаще всего не так пагубно влияет на оценку (так в примере из первого пункта ОМП не единственна, но все оценки обладают неплохими свойствами, в частности состоятельны и имеют достаточно маленькую дисперсию), отсутствие гладкости в соответствующем примере с распределением Лапласа также не приводит к качественным проблемам, хотя оценка опять же перестает быть единственной. Пример со схемой Бернулли очевидно достаточно искусственный, а в примере из пункта 3) мы работаем с $n+1$ параметром при $2n$ наблюдениях (к слову, если перейти к $X_{2i} - X_{2i-1}$ и рассмотреть соответствующую модель, то ОМП сразу же станет отличной оценкой). Наконец в ситуации последнего пункта можно рассматривать локальные максимумы, некоторые из которых будут близки к искомому параметру.

Зато в случае достаточно гладко зависящих от θ распределений, замкнутого множества изменения параметра и независимости носителя от параметра ОМП оказываются совершенно выдающимися оценками — состоятельными, асимптотически нормальными, да еще и с наилучшей возможной асимптотической дисперсией среди оценок с непрерывной асимптотической дисперсией.

Вопрос 6. Рассмотрим $\widehat{\theta}_n = \overline{X}(1 + I_{|\overline{X}|>n^{2/3}})/2$, $X_i \sim \mathcal{N}(\theta, 1)$. Показать, что эта оценка асимптотически нормальна и ее асимптотическая дисперсия равна дисперсии ОМП при $\theta \neq 0$ и меньше ее при $\theta = 0$.

Обе рассмотренных оценки являются эквивариантными, то есть если $\widehat{\theta}$ — ОМП или ОММ для θ, то $f(\widehat{\theta})$ — ОМП или ОММ для $f(\theta)$. Это удобное свойство, которое, к сожалению, плохо сочетается с несмещенностью.

Для решения уравнений в R пригодится функция uniroot, а для минимизации — nlm или optimize. Функция nlm подходит и для векторного случая, также минимизацию можно осуществлять optim, а вот uniroot в многомерном случае заменяется на polyroot из библиотеки rootsolve). Реализованная оценка ОМП с помощью optim находится в пакете stats4, эта функция носит название mle, ее аргументом является логарифмическая функция правдоподобия с обратным знаком.

В Python для решения уравнений есть scipy.optimize.fsolve, для максимизации функций – scipy.optimize.

Задача 3. Построить алгоритм, численно вычисляющий по однопараметрической функции плотности ОМП.

2.2.5 Предельное распределение ОМП

Для сильно регулярных моделей (условия сильной регулярности могут иметь различный вид, одна из версий приведена в конце файла) ОМП удовлетворяет соотношению

$$\sqrt{n}(\widehat{\theta}(X_1, \ldots, X_n) - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right),$$

где $I(\theta) = \mathbf{E}\left(\frac{\partial}{\partial\theta} \ln f_\theta(X_1)\right)^2$ — информация Фишера.

Таким образом, можно узнать предельное распределение ОМП, просто найдя информацию Фишера.

Вопрос 7. Какое предельное распределение имеет ОМП для Коши?

2.2.6 Метод спейсингов

Метод спейсингов (method of maximal spacing) предлагает для оценивания параметра θ семейства распределений с плотностями, сосредоточенными на отрезке $[a, b]$, рассмотреть

$$S(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} D_i(\theta),$$

где $D_i(\theta) = \mathbf{P}_\theta(X \in [x_{(i)}, x_{(i+1)}]) = F_\theta(x_{(i+1)}) - F_\theta(x_{(i)})$, $x_{(0)} = a$, $x_{(n+1)} = b$, где $x_{(i)}$ — вариационный ряд. Тогда θ, максимизирующее S, называют оценкой методом спейсингов.

Логика довольно проста — при правильном θ $Y_i = F_\theta(X_{(i)})$ есть вариационный ряд равномерного распределения. Тогда $Y_i - Y_{i-1}$ одинаково распределены. Но $\max_{y_1+\cdots+y_{n+1}}(y_1, \ldots, y_{n+1})$ достигается при $y_i = 1/(n+1)$. Следовательно, максимизация S будет связана с выравниванием $D_i(\theta)$, что соответствует искомому θ.

Эта оценка состоятельна, а в случае регулярных оценок асимптотически эффективна. В регулярных моделях она достаточно близка к ОМП, но при этом зачастую избегает ее проблем в нерегулярных. Конечно, она неприятна в вычислительном плане, но численно ее поиск не представляет таких уж проблем.

Вопрос 8. Какая оценка методом спейсингов для равномерного распределения $R[\theta_1, \theta_2]$?

Задача 4. Пусть $F_{\theta_1,\theta_2}(x) = 1 - e^{-(x-\theta_1)^{\theta_2}}$, $x > \theta_1$. Построить численно оценку методом спейсингов. Сравнить ее с ОМП.

3 Ответы на вопросы

Ответ 1. Что будет ОМП в случае $X_i \sim R[\theta, \theta+1]$?
Если $X_i \sim R[\theta, \theta+1]$, то функция правдоподобия будет иметь вид

$$L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} \frac{1}{\theta + 1 - \theta} I_{x_i \in [\theta, \theta+1]} = I_{x_i \in [\theta, \theta+1], i \le n}.$$

Эта функция равняется 1, если $\theta \le \min(x_i) \le \max(x_i) \le \theta + 1$ и 0 иначе. Поэтому ОМП будет целый отрезок $[\max(x_i)-1, \min(x_i)]$. Это, впрочем, не мешает всем этим ОМП быть состоятельными, поскольку этот отрезок стягивается с ростом n в точку θ.

Ответ 2. Что будет ОМП в случае X_i с плотностью $\exp(-|x - \theta|)/2$?
В этом случае функция правдоподобия имеет вид

$$L(x_1, \ldots, x_n; \theta) = 2^{-n} \exp\left(-\sum_{i=1}^{n} |x_i - \theta|\right)$$

и его максимизация равносильна минимизации $\sum_{i=1}^n |x_i - \theta|$. Эта функция линейна на каждом отрезке $[x_i, x_{i+1}]$, $i = 1, \ldots, n-1$. Давайте предположим, что a — (n+1)/2-й по возрастанию из x_i, если выборка имеет нечетный размер. Тогда при $\theta < a$ прямые имеют отрицательный наклон, а при $\theta > a$ положительный и минимум будет достигаться при $\theta = a$. А если выборка имеет четный размер и a, b — $n/2$ и $n/2+1$ из наблюдений по возрастанию, то при $\theta \in [a, b]$ график будет горизонтальный и ОМП будет весь отрезок $[a, b]$.

Ответ 3. Что будет с ОМП, если X_i бернуллиевские с параметром θ при $\theta \in \mathbb{Q} \cap (0,1)$ и $1-\theta$ иначе? В этом случае правдоподобие имеет вид

$$L(x_1, \ldots, x_n; \theta) = \theta^{x_1+\cdots+x_n}(1-\theta)^{n-x_1-\cdots-x_n},\ \theta \in \mathbb{Q},\ L(x_1, \ldots, x_n; \theta) = (1-\theta)^{x_1+\cdots+x_n}\theta^{n-x_1-\cdots-x_n},\ \theta \notin \mathbb{Q}.$$

Тогда ОМП будет $\overline{x}$, если $\overline{x} \in \mathbb{Q}$ и $1-\overline{x}$ иначе. Но $\overline{x}$ рациональна по определению, поэтому оценка всегда будет $\overline{x}$. При иррациональных θ она не состоятельна.

Ответ 4. Что будет с ОМП, если $(X_{2i}, X_{2i-1}) \sim \mathcal{N}(\theta_i, \sigma^2)$. Будет ли ОМП для σ^2 состоятельной? Логарифм правдоподобия для этой задачи имеет вид

$$\ln L(x_1, \ldots, x_{2n}; \theta_1, \ldots, \theta_n, \sigma) = -2n\ln\sigma - n\ln(2\pi) - \frac{1}{2\sigma^2}\left(\sum_{i=1}^n (x_{2i-1}-\theta_i)^2 - \sum_{i=1}^n (x_{2i}-\theta_i)^2\right).$$

ОМП $\widehat{\theta}_i$ для θ_i получаются равными $(x_{2i-1}+x_{2i})/2$, а для σ определяется из условия равенства нулю производной

$$-\frac{2n}{\sigma} + \frac{1}{\sigma^3}\left(\sum_{i=1}^n (x_{2i-1}-\widehat{\theta}_i)^2 - \sum_{i=1}^n (x_{2i}-\widehat{\theta}_i)^2\right).$$

Поскольку

$$(x_{2i-1}-\widehat{\theta}_i)^2 = (x_{2i}-\widehat{\theta}_i)^2 = (x_{2i}-x_{2i-1})^2/4,$$

то

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^n (x_{2i}-x_{2i-1})^2}{4n}$$

Величины под знаком квадрата имеют $\mathcal{N}(0, 2\sigma^2)$ распределение, откуда

$$\widehat{\sigma}^2 \xrightarrow{d} \sigma^2/2,\ n \to \infty.$$

Следовательно, оценка не состоятельна, а сходится к половине настоящего значения параметра.

Ответ 5. Пусть $f(x)$ равна $c/\sqrt{x}$ при $|x| \le 1$ и равна $d/x^{3/2}$ при $|x| > 1$, где c, d подобраны так, что это плотность, $X_i \sim f(x-\theta)$. Как будет выглядеть ОМП?
Тот же вопрос для X_i плотностью $\sum_k f(x-\theta-q_k)/2^k$ при $k \ge 1$, q_k — упорядоченные рациональные числа.

Плотность может быть увеличена неограничено в ситуации, когда x лежится близко к θ. Получаем, что при выборе $\widehat{\theta}$ близко к любому из X_i правдоподобие уходит в бесконечность, поэтому ОМП является каждое из наблюдений.

Во втором случае множество ОМП будет всюду плотным множеством.

Ответ 6. Рассмотрим $\widehat{\theta}_n = \overline{X}(1 + I_{|\overline{X}|>n^{2/3}})/2$, $X_i \sim \mathcal{N}(\theta, 1)$. Показать, что эта оценка асимптотически нормальна и ее асимптотическая дисперсия равна дисперсии ОМП при $\theta \ne 0$ и меньше ее при $\theta = 0$. При $\theta = 0$ величина $\sqrt{n}\overline{X}I_{\overline{X}>n^{2/3}}$ сходится к 0 по вероятности, поскольку

$$\frac{\overline{X}}{n^{2/3}} \xrightarrow{P} 0$$

из ЦПТ. Следовательно, при $\theta = 0$

$$\sqrt{n}(\widehat{\theta}_n - \theta) = \frac{1}{2}\sqrt{n}\overline{X} + \frac{1}{2}\sqrt{n}\left(\overline{X}I_{\overline{X}>n^{2/3}}\right) \xrightarrow{d} Z \sim \mathcal{N}\left(0, \frac{1}{4}\right).$$

При $\theta \neq 0$ величина $\sqrt{n}\overline{X}I_{\overline{X}\leq n^{2/3}}$ сходится к 0 по вероятности, откуда

$$\sqrt{n}(\widehat{\theta}_n - \theta) = \sqrt{n}(\overline{X} - \theta) - \frac{1}{2}\sqrt{n}\left(\overline{X}I_{\overline{X}\leq n^{2/3}}\right) \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

Таким образом, у нашей оценки асимптотическая дисперсия будет $\frac{1}{4}$ при $\theta = 0$ и 1 иначе. Между тем, ОМП $\overline{X}$ имеет асимптотическую дисперсию 1. Приведенная оценка имеет ту же дисперсию при $\theta \neq 0$ и лучшую при $\theta = 0$. Это не противоречит асимптотической эффективности ОМП, поскольку асимптотическая дисперсия в данном случае разрывна.

Ответ 7. Саму ОМП для Коши найти затруднительно, для этого нужно найти корень многочлена степени $2n-1$. Зато распределение ОМП можно найти с помощью функции вклада U:

$$U(x,\theta) = \frac{\partial}{\partial\theta}\ln f_\theta(x) = \frac{2(x-\theta)}{1+(x-\theta)^2},$$

Информация Фишера есть математическое ожидание квадрата этой величины:

$$I(\theta) = 4\mathbf{E}_\theta\frac{(X-\theta)^2}{(1+(X-\theta)^2)^2} = \frac{4}{\pi}\int_{\mathbb{R}}\frac{x^2}{(1+x^2)^3}dx = \frac{1}{2}.$$

Значит, ОМП в нашем случаче имеет распределение, близкое к

$$\mathcal{N}\left(0, \frac{n}{2}\right).$$

Ответ 8. Какая оценка методом спейсингов для равномерного распределения $R[\theta_1, \theta_2]$?
Для нахождения оценки мы должны максимизировать

$$S(x_1,\ldots,x_n;\theta_1,\theta_2) = \frac{(x_1-\theta_1)(x_2-x_1)\cdots(x_n-x_{n-1})(\theta_2-x_n)}{(\theta_2-\theta_1)^{n+1}},$$

где $\theta_1 \leq x_1$, $\theta_2 \geq x_n$, $x_1 < x_2 < \ldots < x_n$ — упорядоченная по возрастанию выборка. Для максимизации будем рассматривать функцию S без $(x_2-x_1)\cdots(x_n-x_{n-1})$ (назовем это выражение $\widetilde{S}$). Продифференцируем функцию $\widetilde{S}$ по θ_1

$$\frac{\partial}{\partial\theta_1}\widetilde{S} = \frac{-(\theta_2-x_n)}{(\theta_2-\theta_1)^{n+1}} + \frac{(n+1)(x_1-\theta_1)(\theta_2-x_n)}{(\theta_2-\theta_1)^{n+2}} = 0,$$

откуда получаем два возможных варианты: $\theta_2 = x_n$, $\theta_2 - \theta_1 = (n+1)(x_1-\theta_1)$. Аналогично два корня будет у уравнения, полученного из производной по θ_2 — $\theta_1 = x_1$, $\theta_2 - \theta_1 = (n+1)(\theta_2 - x_n)$. При $x_n = \theta_2$ или $x_1 = \theta_1$ функция S принимает нулевое значение, поэтому единственный кандидат на максимум — $\theta_2, \theta_1 : x_1 - \theta_1 = x_2 - \theta_2 = (\theta_2-\theta_1)/(n+1)$. При этом

$$\theta_2 - \theta_1 = \theta_2 - x_n + x_n - x_1 + x_1 - \theta_1 = \frac{2(\theta_2-\theta_1)}{n+1} + x_n - x_1,$$

откуда $\widehat{\theta}_2 - \widehat{\theta}_1 = (n+1)(x_n-x_1)/(n-1)$, $\widehat{\theta}_1 = x_1 - (x_n-x_1)/(n-1)$, $\widehat{\theta}_2 = x_n + (x_n-x_1)/(n-1)$. Убедимся, что данная критическая точка является максимумом. Можно сделать это с помощью матрицы вторых производных, а можно сослаться на то, что функция неотрицательна, стремится к нулю на бесконечности и ноль на границе нашей области, откуда единственная критическая точка обязана быть максимумом.