1 Домашнее задание

1.1 Прогерские задачи

Для начала, некоторые пожелания по присылаемым задачам:

1. Присылать не только код, но и какие-то мысли по результатам его работы. Удалось ли что-то подтвердить, опровергнуть, что-то не получилось, результаты сравнения такие-то...

2. Прикладывать к коду результат его работы — таблицы и графики. Программы часто работают долго, у меня нет возможности запустить каждую

3. Если пишете на третьем Питоне, то давайте как-то это понять (например, указывать в sha-bang'е явно нужный интерпретатор)

Теперь по новым задачам. Мы продолжим исследование различных оценок и их сходимости. Немного расширим список анализируемых величин.

Порядок сходимости Много раз звучали слова "порядок сходимости". Возникает вопрос, как именно его оценивать. Для начала, надо четко определиться с тем, что мы хотим оценить. Для теории удобно исследовать MSE (т.е. квадратичную характеристику), а на практике более интерпретируемо RMSE (корень из MSE, т.е. линейная характеристика). Я предлагаю все-таки рассматривать RMSE, потому что оно сразу дает представление о величине погрешности. Как можно оценить асимптотическое поведение RMSE (или MSE)? Обычно в асимптотике RMSE имеет вид: $RMSE(N) = C/N^{\alpha}$, α и отражает порядок сходимости. Типичные значения α: 0.5 — обычное поведение, 1 — "сверхсходимость" (например, при оценивании верхней границы в первой задаче для оценки максимумом имела место сверхсходимость) и $\alpha < 0.5$ — "плохая" сходимость. Другие значения у нас пока не ожидаются, конкретные значения для третьего случая можно не дифференцировать, плохо значит плохо. Под "плохо" также попадает сходимость медленнее, чем степенная. В последнем случае $\alpha \to 0$. Чтобы наглядно оценить α, можно построить график зависимости ЛОГАРИФМА RMSE от ЛОГАРИФМА объема выборки ($\log RMSE(N) \sim \log N$). График будет иметь вид, близкий к прямой. По наклону этой прямой легко определить α. Если график зашумленный, нечеткий, значит, следует увеличить количество выборок M. Если график для маленьких длин выборки N ведет себя не как прямая, это нормально, значит, "асимптотика еще не насытилась", т.е. объемы выборки недостаточны, чтобы асимптотические свойства выполнялись с приемлемой точностью. Оценивать порядок сходимости следует по правой части графика, при больших N.

ARE На паре я произнес слова "относительная асимптотическая эффективность" (Asimptotic Relative Efficiency, ARE). Формально ARE определяется как предельное при росте объема выборки отношение числа наблюдений, необходимых для достижения одинаковой точности двух сравниваемых процедур[1]. Например, при оценивании верхней границы равномерного распределения (первая задача с прошлого раза) для максимума $MSE \approx 2/N^2$, а для подправленного максимума $MSE \approx 1/N^2$. Значит, вторая оценка вдвое эффективнее. Разумеется, эта мера имеет смысл только для оценок с

[1] Т.е. на ARE можно смотреть как на "прикидку" (в асимптотическом смысле) того, во сколько раз уменьшится бюджет эксперимента при переходе к более эффективной оценке

одинаковым порядком сходимости, иначе она равна бесконечности или нулю. Оценить ARE можно разделив RMSE двух оценок для большого N друг на друга и возведя частное в степень $1/\alpha$.

Робастность Звучало также слово "робастность". Робастность следует понимать как устойчивость метода к влиянию выбросов и других отклонений от теоретической модели. Робастность мы пока будем рассматривать как умозрительное свойство; на базовом уровне понятно, какие оценки робастны, а какие нет: робастны оценки типа медианы, выборочных квантилей и подрезанного среднего, неробастны оценки типа среднего, других выборочных моментов и (особенно) типа максимума-минимума. Понятно, что за робастность приходится платить эффективностью (т.е. большим количеством наблюдений) и величина этой платы будет одним из объектов нашего исследования.

Задание 1.1. Рассмотрим интеграл $\int_0^1 (\sin x)^{-2/3} dx$. Точное его значение можно получить, например, с помощью функции `integrate()` в `R` или `scipy.integrate` в Python. Этот интеграл можно оценить с помощью метода Монте-Карло, используя следующие плотности:

1. Равномерное распределение (можно смотреть на равномерное распределение как на Бета-распределение $\mathcal{B}(1,1)$)

2. Бета-распределение $\mathcal{B}(1/3,1)$

3. Бета-распределение $\mathcal{B}(1/2,1/2)$

Проверить состоятельность оценок, определить порядок (степень) сходимости и сравнить оценки в смысле относительной асимптотической эффективности (ARE). Сравнивать желательно с наилучшей, т.е. дать ответ в виде "такая-то оценка лучше, а остальным требуется во столько-то раз большая выборка".

Проверить для данного случая правильность теоретического результата — относительная эффективность равна обратному отношению дисперсий одиночных оценок, которая находится по формуле

$$\mathsf{D}\eta = \int \frac{f(x)^2}{\rho_\xi(x)} dx - \left(\int f(x) dx \right)^2 .$$

Задание 1.2. Рассмотрим распределение $U[a-1,\, a+1]$. Параметр сдвига a можно оценить следующими способами:

1. Обычное выборочное среднее

2. Выборочная медиана

3. Квартильное среднее — среднее между выборочным 25%-м и 75%-м квантилями[2]

4. Подрезанное среднее (trimmed mean) с отбрасыванием 5% наибольших и 5% наименьших значений

[2]Квантили уровня 25%, 50% и 75% принять называть *квартилями*, а квантили уровня 1%, 2%, ..., 99% — *процентилями*

5. Так называемое winsorized mean — считается аналогично trimmed mean, но все, что меньше 5%-го квантиля и все, что больше 95%-го квантиля не откидывается, а заменяется на значение соответствующего квантиля[3]

Проверить состоятельность этих оценок, исследовать порядок сходимости, сравнить оценки в смысле ARE.

Убедиться, что для этой задачи выборочное среднее асимптотически в 3 раза более эффективно, чем выброчная медиана. *Какие из этих оценок являются робастными?*

Задание 1.3. Рассмотрим распределение $\mathcal{N}(a, 1)$, т.е. сдвинутое стандартное нормальное. Для параметра сдвига a возможны те же оценки, что и в задаче 1.2. Проверить состоятельность этих оценок, исследовать порядок сходимости, сравнить оценки в смысле ARE.

Убедиться в справедливости следующего теоретического результата: для этой задачи выборочное среднее в $\pi/2$ раз эффективнее, чем медиана.

1.2 Бумажные задачи

Задание 1.4. Дана выборка $X_1, \dots X_N$ из равномерного распределения $U[a, b]$. Построить для параметров a, b ОМП-оценки и оценки по методу моментов. Аналитически найти MSE для оценок, проверить их состоятельность и несмещенность.

Задание 1.5. Для *сдвинутого экспонциального* распределения с параметрами a, λ, т.е. такого распределения ξ, что $\xi - a \sim Exp(\lambda)$, построить оценки параметров по методу моментов и ОМП-оценки.

Задание 1.6. Для распределения $U[a, b]$ найти значение MAD. MAD — median absolute deviation является медианным аналогом standard deviation и вычисляется как $\mathrm{mad}(\xi) = \mathrm{median}(|\xi - \mathrm{median}(\xi)|)$.

2 Разобранные задачи

Задание 2.1. Четверо друзей из общаги собирались месяц назад и один из них ходил за лимонадом. Тогда они не могли договориться, кто пойдет, но, к счастью, у них была с собой монетка и они просто кинули жребий. Однако, в этот раз их собралось всего трое. Кроме того, они обнаружили, что монета погнута и орлы выпадают теперь немного чаще. Как с помощью такой монетки им провести справедливую жеребьевку?

Задание 2.2. Пусть $X_1, \dots, X_N$ — выборка из известного непрерывного распределения (с функцией распределения F и плотностью f). В терминах функции распределения и плотности выразить функцию распределения и плотность максимума и минимума выборки ($\xi = \max_i X_i, \eta = \min_i X_i$)

Задание 2.3. Пусть $X_1, \dots, X_N$ — выборка из известного непрерывного распределения (с функцией распределения F и плотностью f). В терминах функции распределения и плотности выразить функцию распределения и плотность порядковых статистик $X_{[i]}$). *Порядковая статистика — это статистика, занимающая строго определенное место в ранжированной совокупности, т.е. $X_{[i]}$ это i-й элемент в упорядоченном массиве*

[3]Напомню, что значение “5%” не несет сакрального смысла, а просто принятое по умолчанию значение параметра процедуры. Также распространены варианты: 10%, 25%, 2.5% и другие

X-ов. Примеры порядковых статистик — минимум, максимум, медиана, выборочные квантили

Решение. На паре я сообщил решение для общего вида этой задачи, плотность для любой порядковой статистики:

$$\rho_{x_{[i]}}(t) = \frac{F(t)^{i-1}(1-F(t))^{N-i}f(t)}{B(i, N-i+1)}. \tag{1}$$

В знаменателе Бета-функция. В случае, когда распределение — стандартное равномерное, распределение порядковых статистик — Бета-распределение с параметрами $(i, N-i+1)$.

Доказательство: Выпишем функцию распределения для порядковой статистики:

$$F_{x_{[i]}}(t) = P(x_{[i]} < t) = P(\text{“хотя бы } i \text{ среди } x_k \text{ меньше } t\text{”}) = \\ = \sum_{j=i}^{N} P(\text{“ровно } j \text{ среди } x_k \text{ меньше } t\text{”}) = \sum_{j=i}^{N} \binom{N}{j} F(t)^j (1-F(t))^{N-j}.$$

Мы получили выражение для ФР, но нам нужно показать, что его производная в самом деле равна выражению (1).

Я предлагаю следующее доказательство (мне кажется, что вам должна понравится идея). Очевидно, что выражение для ФР — это полином N-й степени от $F(t)$. Обозначим этот полином за $\mathcal{P}(x)$. Тогда плотность $f_{x_{[i]}}(t) = (\mathcal{P}(F(t)))' = \mathcal{P}'(F(t))f(t)$, что похоже на нужное нам выражение. Осталось только лишь показать, что $\mathcal{P}(x)' = C \cdot x^{i-1}(1-x)^{N-i}$. По обыкновению я не буду следить за константой — она естественно получится из нормировки.

Заметим, что полином $\mathcal{P}$ имеет корень 0 кратности не меньше i (так как является суммой полиномов, имеющих корень 0 с кратностями $i, \ldots, N$). Что означает, что его производная имеет корень 0 с кратностью не меньше $i-1$. Если мы сумеем показать, что эта же производная имеет корень 1 с кратностью не меньше $N-i$, то задача будет решена, так как полином определяется своими корнями с точностью до константы, а производная полинома степени N – полином степени $N-1$.

Чтобы доказать последний нужный нам факт, рассмотрим $1 - F_{x_{[i]}}(t)$:

$$1 - F_{x_{[i]}}(t) = P(x_{[i]} > t) = P(\text{“не больше, чем } i-1 \text{ среди } x_k \text{ меньше } t\text{”}) = \\ = \sum_{j=0}^{i-1} \binom{N}{j} F(t)^j (1-F(t))^{N-j} = \mathcal{Q}(F(t)).$$

Полином $\mathcal{Q}$ имеет корень 1 не менее чем $N-i+1$-й кратности. При этом, $\mathcal{Q}(x)' = -\mathcal{P}(x)'$, следовательно, интересующий нас полином $\mathcal{P}(x)'$ имеет корень 1 кратности не менее чем $N-i$.

Задание 2.4. Пусть $\xi_1, \ldots, \xi_n \sim U[0, 1]$, независимы. Пусть $\xi_{[1]}, \ldots, \xi_{[n]}$ — *порядковые статистики* выборки ξ_i, т.е. отсортированные по возрастанию значения ξ_i. Найти распределение для каждой порядковой статистики $\xi_{[i]}$ и проверить их независимость.

Решение. Ответ: $\mathcal{B}(i, n-i+1)$. При этом порядковые статистики одной выборки между собой, очевидно, зависимы.

Доказательство: См. предыдущую задачу, это частный случай. Зависимость следует, очевидно, из упорядоченности — Бета-распределение распределено на всем отрезке $[0, 1]$, в то время как $x_{[1]} \leq x_{[2]}$.

Задание 2.5. Рассмотрим распределение $U[0, \theta]$. Пусть $X_1, \dots, X_N$ — выборка из этого распределения. Найти аналитически MSE следующих оценок:

1. $\widehat{\theta}_1 = 2\overline{X}$
2. $\widehat{\theta}_2 = \max_i X_i$
3. $\widehat{\theta}_3 = \frac{N+1}{N} \max_i X_i$
4. $\widehat{\theta}_4 = \frac{N+1}{N-k} X_{[N-k]}$
5. $\widehat{\theta}_5 = \frac{N+1}{N-k_N} X_{[N-k_N]}$, где $k_N = \lfloor \log N \rfloor$

Решение. Применим небольшой трюк: рассмотрим частный случай $\theta = 1$. Это не сильно нас ограничит, ведь все три оценки линейны по θ, т.е. чтобы получить MSE для общего случая, нам будет достаточно домножить MSE для стандартного равномерного $U[0,\ 1]$ на θ^2. Также мы будем пользоваться "теоремой Пифагора":

$$MSE = \mathsf{D}\widehat{\theta} + \mathrm{bias}(\widehat{\theta})^2.$$

Первая оценка. Найдем отдельно смещение и дисперсию. Смещение нулевое, т.к. $\mathsf{E}\overline{X} = \mathsf{E}X_1 = 1/2$. Дисперсия: $\mathsf{D}\overline{X} = \mathsf{D}X_1/N = \frac{1}{12N}$. $\mathsf{D}\widehat{\theta} = 4\mathsf{D}\overline{X} = \frac{1}{3N}$. Итого:

$$MSE_1 = \mathsf{D}\widehat{\theta}_1 + \mathrm{bias}(\widehat{\theta}_1)^2 = \frac{1}{3N} + 0 = \frac{1}{3N}.$$

Вторая оценка. Максимум — порядковая статистика, значит, имеет распределение Бета: $\mathcal{B}(N, 1)$. Матожидание (для Бета-распределения известно): $\mathsf{E}\widehat{\theta} = \frac{N}{1+N}$. Т.о. оценка смещенная, смещение имеет вид: $1/(N+1)$. Дисперсия для Бета-распределения тоже известна, имеем: $\mathsf{D}\widehat{\theta}_2 = \frac{N \cdot 1}{(N+1)^2(N+1+1)} = \frac{N}{(N+1)^2(N+2)}$. Итого:

$$MSE_2 = \mathsf{D}\widehat{\theta} + \mathrm{bias}(\widehat{\theta})^2 = \frac{N}{(N+1)^2(N+2)} + \frac{1}{(N+1)^2} = \\ = \frac{N+N+2}{(N+1)^2(N+2)} = \frac{2}{(N+1)(N+2)} \approx \frac{2}{N^2}.$$

Опа, сверхсходимость!

Третья оценка. Смещение нулевое, потому что мы так ее построили. А дисперсия вот увеличивается в $\left(\frac{N+1}{N}\right)^2$ раз. Итого:

$$MSE_3 = \mathsf{D}\widehat{\theta}_3 + 0 = \frac{(N+1)^2}{N^2}\mathsf{D}\widehat{\theta}_2 = \frac{(N+1)^2}{N^2} \cdot \frac{N}{(N+1)^2(N+2)} = \frac{1}{N(N+2)} \approx \frac{1}{N^2}.$$

Четвертая оценка. Смещение нулевое, потому что так строили (вспоминаем распределение порядковых статистик и матожидание для Бета-распределения). Остается только дисперсия (смотрим или вспоминаем дисперсию для Бета-распределения). Итого:

$$MSE_4 = \mathsf{D}\widehat{\theta}_4 + 0 = \frac{(N+k)^2}{N^2} \cdot \frac{(N-k)(k+1)}{(N+1)^2(N+2)} = \frac{(N+k)^2(N-k)(k+1)}{N^2(N+1)^2(N+2)} \approx \frac{k+1}{N^2}$$

Пятая оценка. Аналогично четвертой, но формально асимптотика сходимости будет хуже: $MSE_5 \approx \frac{\log N}{N^2}$.

Как видите, робастность обходится дорого. Относительная эффективность робастной оценки будет иметь вид $1/\sqrt{k+1}$, т.е. ради того, чтобы добавить возможность выдержать три выброса нужно увеличивать число наблюдений вдвое.

Задание 2.6. Убедиться, что метод Монте-Карло для многомерных интегралов может давать лучшие результаты, чем классические детерминированные методы.

Решение. Предлагается рассмотреть следующий интеграл: $\int \cdots \int_{[0,\,1]^r} \prod_{i=1}^r \sin x_i dx_1 \dots dx_r$. Он легко вычисляется аналитически:

$$\int \cdots \int_{[0,\,1]^r} \prod_{i=1}^r \sin x_i dx_i \dots dx_r = \left(\int_0^1 \sin x dx \right)^r = (\cos 0 - \cos 1)^r.$$

Рассмотрим метод Монте-Карло с равномерной интегрирующей плотностью (будем моделировать равномерное распределение на гиперкубе $[0,\,1]^r$, что сводится к моделированию r обычных равномерных $U[0,\,1]$) и “наивный” метод прямоугольников: разобьем гиперкуб $[0,\,1]^r$ на n^r кубиков, по n вдоль каждого ребра и на каждом кубике приблизим интегрируемую функцию константой — значением в центре каждого кубика.

Уже для $r = 7$ метод М-К показывает существенно лучшие результаты. Для проверки предлагается следующий код:

```
rowProds <- function(X) {
  exp(rowSums(log(X)))
}

f <- function(X) {
  rowProds(sin(X))
}

uniform.grid <- function(n, d) {
  edge <- seq(from = 1/(2 * n), by = 1/n, length.out = n)
  as.matrix(expand.grid(rep(list(edge), d)))
}

uniform.random <- function(n, d) {
  matrix(runif(n^d * d), n^d, d)
}

set.seed(1)

d <- 7
exact.value <- (cos(0) - cos(1))^d

n <- 10
x.grid <- uniform.grid(n, d)
x.random <- uniform.random(n, d)

d.grid <- mean(f(x.grid)) - exact.value
```

```
d.random <- mean(f(x.random)) - exact.value

d.grid
d.random

# MSE:
M <- 100
ests <- replicate(M, mean(f(uniform.random(n, d))))
RMSE <- sqrt(mean((ests - exact.value)^2))

RMSE
d.grid / RMSE
```

Результаты:

```
> d.grid
[1] 1.267252e-05
> d.random
[1] -2.66254e-06

> RMSE
[1] 3.102308e-06
> d.grid / RMSE
[1] 4.084867
```

3 Дополнительные задачи

Задание 3.1. Для задачи 1.2 придумать оценку со сверхсходимостью, разработать ее робастный аналог и сравнить их в смысле ARE. *А получится ли также придумать оценку со сверхсходимостью для оценки сдвига в нормальной модели (задача 1.3)*

Задание 3.2. Для задачи 1.4 ($U[a, b]$) построить несмещенные аналоги ОМП-оценок и найти их MSE.

Задание 3.3. В терминах функции квантилей (т.е. обратной функции распределения) для стандартного нормального распределения (в `R` она называется `qnorm()`) выразить значение MAD для нормального распределения $\mathcal{N}(a, \sigma^2)$. Напомню, что MAD — median absolute deviation является медианным аналогом standard deviation и вычисляется как $\mathrm{mad}(\xi) = \mathrm{median}(|\xi - \mathrm{median}(\xi)|)$.