Домашнее задание

В этот раз прогерских задач будет две, похожих. Нужно будет сравнить в среднеквадратичном различные оценки для параметров некоторых распределений с помощью моделирования.

Действовать лучше по следующей схеме: фиксируем параметры распределения (любые разумные) некоторую длину выборки N (100–1000) и число серий M. Моделируем M выборок по N случайных величин, для каждой выборки считаем все статистики[1], получаем выборку длины M для для каждой статистики. Для каждой статистики находим среднеквадратичное отклонение от известного значения параметра (т.е. $\mathrm{RMSE}_m = \sqrt{\sum_{j=1}^{M}(t_j^{(m)} - \theta)^2/M}$, где $t_j^{(m)}$ — оценка по выборке j, полученная m-м методом). M следует выбирать так, чтобы считалось разумное время, но при этом результаты с.к.о. были устойчивы и устойчиво сравнивались. Далее все это стоит обернуть в цикл по N, чтобы показать состоятельность оценок. Результаты выводим в виде уже привычного графика $\mathrm{RMSE}_m(N) \sim N$ для всех m в общих осях в логарифмах.

5.1 Uniform

Рассмотрим распределение $U[0, \theta]$. Пусть $X_1, \ldots, X_N$ — выборка из этого распределения. Рассмотрим следующие оценки параметра θ:

1. $\widehat{\theta} = 2\overline{X}$;
2. $\widehat{\theta} = \max_i X_i$;
3. $\widehat{\theta} = \frac{N+1}{N} \max_i X_i$;
4. $\widehat{\theta} = 2 \operatorname{median}(X)$;
5. $\widehat{\theta} = 4q_{.25}(X)$;
6. $\widehat{\theta} = 4/3q_{.75}(X)$,

где $\operatorname{median}(\cdot)$ обозначает выборочную медиану, а $q_p(X)$ — выборочный p-квантиль.

Собственно, задание. Во-первых, нужно убедиться, что они состоятельны (теоретически и с помощью моделирования). Во-вторых, нужно построить зависимость RMSE от объема выборки для всех оценок и упорядочить оценки от лучших к худшим.

Опционально:

1. Для трех последних оценок провести анализ, аналогично анализу, который мы провели для первых трех на паре. Распределение порядковой статистики (т.е. выборочного квантиля) для равномерного распределения известно. Удобно аналогично считать истинное значение $\theta = 1$, все остальные случаи получаются простым шкалированием. Значения N можно при этом считать, например, всегда нечетными (более точно, $N \mod 4 = 1$), чтобы медиана и квартили были определены однозначно. Нас интересует асимтотическое поведение, поэтому такое упрощение возможно.

[1] На самом деле, можно считать разные оценки по разным выборкам, если это окажется проще закодить, в данном (и только в данном!) случае это непринципиально

2. Вспомнить вид байесовских оценок для параметра θ в аналогичной задаче и провести параллели. Во что превращается байесовская оценка в случае неинформативного априорного распределения? Во что превращается байсовская оценка при больших объемах выборки?

5.2 Cauchy

Рассмотрим распределение Коши со сдвигом, т.е. распределение с плотностью

$$f(x) = \frac{1}{\pi\left(1 + (x-a)^2\right)}$$

Пусть мы имеем выборку $X_1, \ldots, X_N$ из этого распределения и хотим оценить параметр *сдвига* a.

Для "хороших" распределений параметр сдвига обычно оценивается с помощью выборочного среднего, но в этом случае это не пройдет (ЗБЧ работает только в том случае, если существует матожидание, а здесь его нет). Убедитесь в этом, для начала.

А после проверьте являются ли следующие оценки состоятельными и сравните их в смысле среднеквадратичного отклонения от истинного значения сдвига:

1. Выборочная медиана;

2. Среднее, посчитанное после отбрасывания наибольшего и наименьшего значения выборки;

3. Среднее, посчитанное после отбрасывания 5% самых больших и 5% самых маленьких значений выборки (такая оценка известна как "trimmed mean" — подрезанное среднее);

Распределение Коши (без сдвига) можно промоделировать как отношение двух независимых стандартных нормальных случайных величин или воспользоваться готовой функцией.

Конспект (примеры + доказательство некоторых результатов)

Задача 1. Пусть $X_1, \ldots, X_N$ — 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 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 $N-i$.

Задача 2. 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 $x_{[1]} \leq x_{[2]}$.

Задача 3. Рассмотрим распределение $U[0, \theta]$. Пусть $X_1, \ldots, X_N$ — 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. $\widehat{\theta}_4 = \frac{N+1}{N-k} X_{[N-k]}$;

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:

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

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 [2].

[2] *Асимтотическая относительная эффективность, ARE* — это характеристика пары оценок $\widehat{\theta}_1, \widehat{\theta}_2$, имеющих одинаковый порядок сходимости, асимтотика отношения N_1/N_2 — объемов выборок, необходимых для достижения одинаковой точности в смысле RMSE. Можно смотреть на относительную эффективность как на коэффициент увеличения объема выборки (т.е. бюджета эксперимента) при замене более эффективной оценки на менее эффективную при сохранении той же требуемой точности оценок

Задача 4. Убедиться, что метод Монте-Карло для многомерных интегралов может давать лучшие результаты, чем классические детерминированные методы.

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

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

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
```