3. Основные методы нахождения оценок

Некоторые определения

1. Параметрический бутстреп. Пусть $\widehat{\theta}$ — оценка θ по выборке $X_1, ..., X_N$, которая получена из распределения P_θ. *Бутстрепная выборка размера N в параметрическом бутстрепе* — это выборка из распределения $\mathsf{P}_{\widehat{\theta}}$.

2. Непараметрический бутстреп. Пусть дана выборка $X_1, ..., X_N$ из распределения P, и пусть P^* — эмпирическое распределение, построенное по этой выборке. *Бутстрепная выборка размера N в непараметрическом бутстрепе* — это выборка из распределения P^*. Легко показать (покажите), что если $i_1, ..., i_N \sim R\{1, ..., N\}$ — независимые случайные величины, то $X_{i_1}, ..., X_{i_N}$ — бутстрепная выборка размера N в непараметрическом бутстрепе (построенная по данной выборке $X_1, ..., X_N$ из некоторого распределения P).

3. Бутстрепная оценка дисперсии. Пусть дана выборка $X_1, ..., X_N$ из распределения P_θ и оценка параметра θ, заданная равенством $\widehat{\theta} = \widehat{\theta}(X_1, \ldots, X_N)$. Сгенерировано K бутстрепных выборок $X^1 = (X_1^1, \ldots, X_N^1), \ldots, X^k = (X_1^k, \ldots, X_N^k)$ (при этом все эти выборки можно генерировать как на основе параметрического бутстрепа, так и на основе непараметрического, но эти серии выборок должны быть сгенерированы одним и тем же способом) и для каждой из них посчитана оценка параметра $\widehat{\theta} = \widehat{\theta}(X^i)$, $i \in \{1, \ldots, k\}$. Далее по этой полученной выборке оценок параметра $\widehat{\theta}(X^1), \ldots, \widehat{\theta}(X^k)$ строится выборочная дисперсия $s^2 = s^2(\widehat{\theta}(X^1), \ldots, \widehat{\theta}(X^k))$, которая и носит название *выборочной оценки дисперсии оценки параметра θ.*

Задачи

1. *(К теоретическим задачам 1-3)* Сгенерируйте выборки $X_1, ..., X_N$ из всех распределений из задач 1-3 ($N = 1000$). Для всех $n \leqslant N$ посчитайте значение полученных оценок (по выборке $X_1, \ldots, X_n$) методом моментов и методом максимального правдоподобия. Оцените дисперсию (с помощью выборочной дисперсии) каждой оценки, сгенерировав для каждой из них $K = 1000$ бутстрепных выборок а) с помощью параметрического бутстрепа (у каждого распределения и у каждой оценки своя бутстрепная выборка), б) с помощью непараметрического бутстрепа (у каждого распределения своя бутстрепная выборка). Проведите эксперимент для разных значений θ (рассмотрите не менее трех различных значений).

2. На высоте 1 метр от поверхности Земли закреплено устройство, которое периодически излучает лучи на поверхность Земли (считайте, что поверхность Земли представляет из себя прямую). Пусть l — перпендикуляр к поверхности Земли, опущенный из точки, в которой закреплено устройство. Угол к прямой l (под которым происходит излучение) устройство выбирает случайно из равномерного распределения на отрезке $(-\pi/2, \pi/2)$ (все выборы осуществляются независимо). Можно доказать, что в

этих предположениях точки пересечения с поверхностью имеют распределение Коши (плотность равна $\frac{\theta}{\pi(\theta^2+(x-x_0)^2)}$) с параметром масштаба $\theta = 1$. Неизвестный параметр сдвига x_0 соответствует проекции (вдоль прямой l) точки расположения устройства на поверхность Земли (направление оси и начало координат на поверхности Земли выбраны заранее некоторым образом независимо от расположения устройства). В файле Cauchy.csv находятся координаты точек пересечения лучей с поверхностью Земли. Оцените параметр сдвига методом максимального правдоподобия а) по половине выборки (первые 500 элементов выборки, т.е. выборка состоит из 1000 наблюдений); б) по всей выборке. Оценку произведите по сетке (т.е. возьмите набор точек с некоторым шагом и верните ту, на которой достигается максимум функции правдоподобия). Известно, что параметр масштаба принадлежит интервалу $[-1000, 1000]$. Выберите шаг равным 0.01. Если получается долго или не хватает памяти, то уменьшите интервал поиска и поясните (в комментариях), почему берете именно такой интервал.

3. В банке каждую минуту подсчитывается баланс по сравнению с началом дня (6 часов утра). В полночь работники банка измеряют две величины: X^1 — максимальное значение баланса за день, X^2 — значение баланса в полночь. Считается, что величина $X = X^1 - X^2$ имеет распределение Вейбулла с функцией распределения $1 - e^{-x^\gamma} I(x \geq 0)$, где $\gamma > 0$ — параметр формы. В течение 10 лет каждый день банк проводил измерение величины X, получив, в результате выборку $X_1, \ldots, X_{3652}$. В файле Weibull.csv находятся соответствующие измерения. Оцените параметр формы методом максимального правдоподобия а) по первым 4 годам; б) по всей выборке. Оценку произведите по сетке (в логарифмической шкале). Известно, что $\log_{10} \gamma \in [-2, 2]$. Выберите шаг равным 10^{-3}.