Основные методы поиска оценок.

Необходимым условием получения положительного балла за задачу является присутствие комментариев в тексте, а также наличие окончательного вывода. Не забывайте об этом!

Необходимые определения.

1. **Параметрический бутстреп.** Пусть $\hat{\theta}$ – оценка параметра θ по выборке $X_1, \ldots, X_n$, которая получена из распределения P_θ. Бутстрепная выборка размера N в параметрическом бутстрепе – это выборка из распределения $P_{\hat{\theta}}$.

2. **Непараметрический бутстреп.** Пусть дана выборка $X_1, \ldots, X_n$ 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 видеть, что если $i_1, \ldots, i_N \sim R\{1, \ldots, N\}$ – независимые случайные величины, то $X_{i_1}, \ldots, X_{i_N}$ – бутстрепная выборка размера N в непараметрическом бутстрепе (построенная по выборке $X_1, \ldots, X_n$ из некоторого распределения P).

3. **Бутстрепная оценка дисперсии.** Пусть дана выборка $X_1, \ldots, X_n$ из распределения P_θ, а $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ – оценка параметра θ. 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 $\hat{\theta}_i = \hat{\theta}(X^i)$, $i \in \{1, \ldots, k\}$. Далее по выборке $\{\hat{\theta}(X^i)\}_{i \geq 1}$ строится выборочная дисперсия $s^2(\hat{\theta}) = s^2(\hat{\theta}(X^1), \ldots, \hat{\theta}(X^k))$, которая и называется бутстрепной оценкой дисперсии оценки $\hat{\theta}$.

Задачи.

1. (К теоретической задаче 3.1) Сгенерируйте выборки $X_1, \ldots, X_N$ из всех распределений из задачи 3.1 $(N = 1000)$. Для всех $n \leq N$ посчитайте значения полученных оценок (по выборке $X_1, \ldots X_n$) методом моментов. Оцените дисперсию каждой оценки, сгенерировав для каждой из них $K = 1000$ бутстрепных выборок а) с помощью параметрического бутстрепа, б) с помощью непараметрического бутстрепа. Проведите эксперимент для разных значений параметров распределений (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). В этих предположениях точки пересечения с

поверхностью имеют распределение Коши с плотностью $p(x) = \frac{1}{\pi(1+(x-x_0)^2}$. Неизвестный параметр сдвига x_0 соответствует проекции 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 распреде ления $F(x) = 1 - e^{-x^\gamma} (x > 0)$, 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}.