1. Оценки и доверительные интервалы.

1 (2 балла) Пусть $X_1, \ldots, X_n$ – выборка из распределения $U(-\theta, \theta)$, $\theta > 0$. Построить оценку параметра θ методом максимального правдоподобия. Проверить её на состоятельность (руками).

2 (2 балла) (Метод выбора с отклонением) Пусть X имеет плотность $f(x)$, а мы умеем строить величину с плотностью $g(x)$, причем $f(x)/g(x) \le c$. Сгенерируем величину Y с плотностью $g(x)$ и с вероятностью $f(Y)/(cg(Y))$ возьмем её в качестве X иначе повторим процедуру. Доказать, что данный метод даёт случайную величину с плотностью $f(x)$. Какое c можно взять в данном методе?

3 (2 балла) На примере бесконечной выборки $X = (X_1, X_2, \ldots)$ из произвольного распределения с конечным вторым моментом изучить поведение статистики

$$T_n = \frac{\sum\limits_{i=1}^{n} X_i - nEX_1}{\sqrt{nDX_1}}.$$

Сходится ли T_n на какой-нибудь реализации выборки X? Как вы это объясните? Также проверьте, что центральная предельная теорема всё же выполняется.

4 (2 балла) Пусть выборка $X_1, \ldots, X_{100}$ сделана из распределения Парето с параметром $\alpha > 0$, имеющего плотность $p(x) = \frac{\alpha}{x^{\alpha+1}} I\{x > 1\}$. Для всех $\alpha \in (0, 5)$ (по мелкой решетке) построить по выборке доверительные интервалы уровня доверия γ для параметра α, вывести их на графике в зависимости от α и сделать выводы.

5 (2 балла) Пусть $F(x) = 1 - \exp(-(\beta x)^\alpha),\ x > 0, \alpha, \beta > 0$ – функция распределения Вейбулла. Смоделировать выборку из этого распределения размера 100. Реализовать метод спейсингов и найти этим методом оценки параметров α и β.

6 (2 балла) Пусть $X_1, \ldots, X_n$ – выборка. Определим оценку Пикандса индекса экстремального значения как

$$\gamma_p = \frac{1}{\ln 2} \ln \frac{X_{(n-k)} - X_{(n-2k)}}{X_{(n-2k)} - X_{(n-4k)}}.$$

Известно, что для распределения Парето с параметром α данная оценка, как и оценка Хилла, является состоятельной для параметра $\gamma = \frac{1}{\alpha}$ при $k \to \infty, k/n \to 0, n \to \infty$. Определить, какая оценка лучше при малых значениях n (скажем, $n < 1000$), и обосновать свой выбор. Предложить метод определения оптимального значения k при малых значениях n. Определить значение индекса экстремального значения по датасету Aids2 из пакета MASS языка R.

Замечание 1. В качестве распределения в задаче 3 можно взять распределение $Beta(\alpha, \beta)$, где $\beta = 50 - \alpha$ и α – ваш номер в таблице курса. Те же значения параметров α и β можно выбрать в задаче 5.

Замечание 2. В качестве целевой переменной в задаче 6 брать время между постановкой диагноза и смертью пациента (или концом наблюдения).