Задание 1

1. [6 баллов] Пусть $X_1, \ldots, X_n \sim F$, где F – функция распределения. Обозначим через $\hat{F}_n$ – эмпирическую функцию распределению, подсчитанную по выборке.

   a) Пусть x, y, $x \neq y$ – произвольные точки на действительной оси. Подсчитать ковариацию между $\hat{F}_n(x)$ и $\hat{F}_n(y)$.

   b) Пусть $a < b$ – фиксированные числа. Определим параметр $\theta = T(F) = F(b) - F(a)$. Пусть $\hat{\theta} = T\left(\hat{F}_n\right) = \hat{F}_n(b) - \hat{F}_n(a)$ используется в качестве оценки этого параметра. Найти значение стандартной ошибки для $\hat{\theta}$. Получить выражение для приближенного доверительного интервала размера $(1 - \alpha)$ для θ.

   c) На странице курса скачайте данные об амплитудах землетрясений вблизи Фиджи. Построить график для $\hat{F}_n$. Подсчитать и построить приближенные 95% доверительный интервал для F. Подсчитать и построить приближенный 95% доверительный интервал для значения $F(4.9) - F(4.3)$.

2. [2 балла] 100 людям давали стандартный антибиотик для лечения инфекции, а другим 100 людям давали новый антибиотик. В первой группе 90 человек выздоровели. Во второй группе выздоровели 85 человек. Пусть p_1 – вероятность выздороветь, принимая стандартное лекарство, а p_2 – вероятность выздороветь, принимая новое лекарство. Необходимо оценить параметр $\theta = p_1 - p_2$. Подсчитать оценку этого параметра, стандартную ошибку оценки, 80% и 95% доверительные интервалы для параметра θ.

3. [3 балла] Провести моделирование, чтобы сравнить различные типы доверительных интервалов, построенные с помощью бутстрепа. Пусть $n = 50$, $T(F) = \frac{\int (x-\mu)^3 dF(x)}{\sigma^3}$ – эксцесс, где F – логнормальное распределение. Выборка из логнормального распределения генерируется следующим образом. Сначала необходимо сгенерировать простую выборку $Y_1, \ldots, Y_n \sim N(0, 1)$, после чего положить $X_i = e^{Y_i}$, $i = 1, 2, \ldots, n$. Построить 95% доверительные интервалы для $T(F)$ (под F понимается распределение элементов выборки $X_1, \ldots, X_n$) по данным $X_1, \ldots, X_n$, используя три подхода на основе бутстрепа.

4. [5 баллов] Пусть $X_1, \ldots, X_n \sim U(0, \theta)$, $\hat{\theta} = X_{(n)} = \max\{X_1, \ldots, X_n\}$. Сгенерировать выборку объема $n = 50$ для случая $\theta = 1$.

   a) Найти распределение $\hat{\theta}$. Сравнить плотность настоящего распределения θ с гистограммой, полученной с помощью бутстрепа.

   b) В рассматриваемом случае бутстреп работает плохо. Выясните причины этого, показав, что $\mathrm{P}\left(\hat{\theta} = \theta\right) = 0$, однако $\mathrm{P}\left(\hat{\theta}^* = \hat{\theta}\right) \approx$

0.632, поскольку $\mathrm{P}\left(\hat{\theta}^* = \hat{\theta}\right) = 1 - \left(1 - \frac{1}{n}\right)^n$, где $\hat{\theta}^*$ – оценка, полученная с помощью бутстрепа.

5. [6 баллов] Пусть $T_n = \overline{X}_n^2$, $\mu = \mathbb{E}\,(X_1)$, $\alpha_k = \int (x-\mu)^k dF(x)$ и $\hat{\alpha}_k = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^k$. Доказать, что оценка дисперсии функционала T с помощью бутстрепа (при подсчете дисперсии функционала T_n используется "бутстрепное усреднение", то есть усреднение по эмпирической функции распределения F_n) равна

$$v_{boot} = \frac{4\overline{X}_n^2 \hat{\alpha}_2}{n} + \frac{4\overline{X}_n \hat{\alpha}_3}{n^2} + \frac{\hat{\alpha}_4}{n^3} + \frac{\hat{\alpha}_2^2(2n-3)}{n^3}.$$