Практикум 2. Состоятельность. Асимптотическая нормальность.

1.1. $X_1, ..., X_n \sim \mathrm{R}[0, \theta]$. Во всех пунктах нужно расмотреть $n = 20$, 50, 100, 500.
1) Рассмотрим две несмещенные состоятельные оценки параметра θ: $((n+1)/n)X_{(n)}$ и $2\overline{X}$. Чтобы определить, какая из них "лучше угадывает"значение θ, смоделируйте 100 выборок (для каждого n) и подсчитайте, в скольких из них $2\overline{X}$ оказывается ближе к θ. Выведите это число, а также среднее (по всем выборкам) абсолютное отклонение каждой из статистик от θ.
2) Оценка $2\overline{X}$ асимптотически нормальна, проиллюстрируем это. Постройте гистограммы (смоделировав по 1000 реализаций) для $\sqrt{n}(2\overline{X} - \theta)$ и изобразите на гистограмме нормальную плотность.
3) Будет ли оценка $((n+1)/n)X_{(n)}$ асимптотически нормальна? Постройте гистограммы (смоделировав по 1000 реализаций) для $X_{(n)}$ при и сравните с нормальной плотностью.
4) Постройте гистограммы (смоделировав по 1000 реализаций) для $n(X_{(n)} - \theta)$, сравните с экспоненциальной плотностью.
5*) В качестве оценки для θ^2 можно использовать $\hat{\theta}_1 = (X_1X_2 + X_2X_3 + \cdots + X_{n-1}X_n)/(n-1)$, $\hat{\theta}_2 = (X_1X_2 + X_3X_4 + \cdots + X_{n-1}X_n)/(n/2)$, $\hat{\theta}_3 = (\overline{X})^2$ и $\hat{\theta}_4 = (((n+1)/n)X_{(n)})^2$. Сравните эти оценки, как в п. 1), для $\sqrt{n}(\hat{\theta}_i - \theta)$, $i = 1, 2, 3$, постройте гистограммы на одном графике, чтобы сравнить асимптотические дисперсии.

2.1. $X_1, \ldots, X_n \sim F_\theta(x) = (1 - e^{-(x-\theta)})\,\mathrm{I}\{x \geq \theta\}$.
Во всех пунктах нужно расмотреть $n = 20$, 50, 100, 500.
1) Рассмотрим две несмещенные состоятельные оценки параметра θ: $X_{(1)} - 1/n$ и $\overline{X} - 1$. Чтобы определить, какая из них "лучше угадывает"значение θ, смоделируйте 100 выборок (для каждого n) и подсчитайте, в скольких из них $2\overline{X}$ оказывается ближе к θ. Выведите это число, а также среднее (по всем выборкам) абсолютное отклонение каждой из статистик от θ.
2) Оценка $\overline{X} - 1$ асимптотически нормальна, проиллюстрируем это. Постройте гистограммы (смоделировав по 1000 реализаций) для $\sqrt{n}(\overline{X} - 1 - \theta)$ и изобразите на гистограмме нормальную плотность.
3) Будет ли оценка $X_{(1)}$ асимптотически нормальна? Постройте гистограммы (смоделировав по 1000 реализаций) для $X_{(n)}$ при и сравните с нормальной плотностью.
4) Постройте гистограммы (смоделировав по 1000 реализаций) для $n(X_{(1)} - \theta)$, сравните с экспоненциальной плотностью.
5*) В качестве оценки для θ^3 можно использовать $\hat{\theta}_1 = (X_1X_2X_3 + X_2X_3X_4 + \cdots + X_{n-2}X_{n-1}X_n)/(n-2)$, $\hat{\theta}_2 = (X_1X_2X_3 + X_4X_5X_6 + \cdots + X_{n-2}X_{n-1}X_n)/(n/2)$, $\hat{\theta}_3 = (\overline{X} - 1)^3$ и $\hat{\theta}_4 = (X_{(1)})^3$. Сравните эти оценки, как в п. 1), для $\sqrt{n}(\hat{\theta}_i - \theta)$, $i = 1, 2, 3$, постройте гистограммы на одном графике, чтобы сравнить асимптотические дисперсии.

3.1. $X_1, \ldots, X_n \sim f_\theta(x) = (\pi(1 + (x-\theta)^2))^{-1}$. Во всех пунктах нужно расмотреть $n = 20$, 50, 100, 500.
1) Распределение Коши имеет тяжелые хвосты, у него нет математического ожидания, поэтому $\overline{X}$ не будет состоятельной оценкой для θ. Постройте гистограммы (смоделировав по 1000 реализаций) для $\overline{X}$, сравните с нормальной плотностью и с плотностью Коши.
2) Посмотрим, можно ли улучшить ситуацию, выкинут самые большие и самые маленькие значения. Рассмотрим $\overline{X}_\alpha$ – усеченное среднее,

$$\overline{X}_\alpha = \frac{1}{n - 2k}(X_{(k+1)} + \cdots + X_{(n-k)}), \quad k = [\alpha n].$$

Постройте гистограммы (смоделировав по 1000 реализаций) для $\sqrt{(}\overline{X}_\alpha - \theta)$, сравните с нормальной плотностью при разных α.
3) В этой модели можно постоить состоятельную асимптотически нормальную оценку MED – выборочная медиана:

$$MED = \begin{cases} X_{(k+1)}, & n = 2k+1, \\ \dfrac{X_{(k+1)} + X_{(k+1)}}{2}, & n = 2k, \end{cases}$$

Постройте гистограммы (смоделировав по 1000 реализаций) для $\sqrt{n}(MED - \theta)$ и сравнить с нормальной плотностью.
4) Чтобы определить, MED или$\overline{X}_\alpha$ "лучше угадывает"значение θ, смоделируйте 100 выборок (для каждого n) и подсчитайте, в скольких из них MED оказывается ближе к θ. Выведите это число, а также среднее (по всем выборкам) абсолютное отклонение каждой из статистик от θ.
5*) Пусть $X_1, \ldots, X_n \sim \mathcal{N}(\theta, 1)$. В качестве оценки для θ^2 можно использовать $\hat{\theta}_1 = (X_1X_2 + X_2X_3 + \cdots + X_{n-1}X_n)/(n-1)$, $\hat{\theta}_2 = (X_1X_2 + X_3X_4 + \cdots + X_{n-1}X_n)/(n/2)$, $\hat{\theta}_3 = (\overline{X})^2$ и $\hat{\theta}_4 = (MED)^2$. Сравните эти оценки, как в п. 4), для $\sqrt{n}(\hat{\theta}_i - \theta)$, $i = 1, 2, 3$, постройте гистограммы на одном графике, чтобы сравнить асимптотические дисперсии.