Задача №1

Рассмотрим случайную величину X_i – цена i-го проданного кокошника, где $i \in \{1,2,3,4,5\}$. Рассмотрим множество платьев, стоимостью s. Через s_n обозначим количество таких платьев. Тогда очевидно, что $P(X_1 = s) = \frac{s_n}{5}$. Отсюда следует, что:

$$P(X_2 = s) = P(X_2 = s \mid X_1 = s) * P(X_1 = s) + P(X_2 = s \mid X_1 \neq s) * P(X_1 \neq s) =$$
$$= \frac{s_n - 1}{4} * \frac{s_n}{5} + \frac{s_n}{4} * \left(1 - \frac{s_n}{5}\right) = \frac{s_n}{5} = P(X_1 = s)$$

Таким образом X_2 и X_1 имеют одинаковые маргинальные распределения. По индукции нетрудно показать, что маргинальные распределения всех X_i совпадают. Отметим, что отсюда следует, что $E(X_i) = E(X_j)$ и $Var(X_i) = Var(X_j)$ для любых $i, j \in \{1,2,3,4,5\}$.

Обратим внимание, что, несмотря на равенство маргинальных распределений, случайные величины $X_1,...,X_5$ – зависимы. При этом $\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_1, X_2), \forall i \neq j$. Поскольку маргинальные распределения X_i совпадают, то достаточно показать, что $E(X_i X_j) = \mathrm{E}(X_1 X_2), \forall i \neq j$. Через $S = \{s_1,...,s_m\}$ обозначим множество стоимостей платьев, а через $n = \{n_1,..,n_m\}$ – соответствующих им количеств, где $m \leq 5$. Без потери общности допустим, что $k < i$ и $k < j$. Тогда получаем, что:

$$E(X_k X_i) = \sum_{t=1}^{m} E(X_k X_i \mid X_i = s_t) * P(X_i = s_t) = \sum_{t=1}^{m} s_t E(X_k \mid X_i = s_t) * \frac{n_t}{5} =$$
$$= \sum_{t=1}^{m} s_t E(X_k \mid X_j = s_t) * \frac{n_t}{5} = \sum_{t=1}^{m} E(X_k X_j \mid X_j = s_t) * P(X_j = s_t) = E(X_k X_j)$$

В свете вышесказанного математическое ожидание вырученных Василисой денег принимает вид:

$$E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) = E(X_1) + E(X_1) + E(X_1) = 3500 + 3500 + 3500 = 10500$$

Теперь найдем дисперсию:

$$Var(X_1 + X_2 + X_3) = Var(X_1) + Var(X_2) + Var(X_3) + 2Cov(X_1, X_2) + 2Cov(X_1, X_3) + 2Cov(X_2, X_3) =$$
$$= Var(X_1) + Var(X_1) + Var(X_1) + 2Cov(X_1, X_2) + 2Cov(X_1, X_2) + 2Cov(X_1, X_2) = 3Var(X_1) + 6Cov(X_1, X_2)$$

Из условия известно, что $Var(X_1) = 500^2 = 250000$, поэтому остается найти лишь $Cov(X_1, X_2)$. Во-первых, обратим внимание, что:

$$Cov(X_1, X_1 + X_2 + X_3 + X_4 + X_5) = Cov(X_1, 5 * 3500) = 0$$

Во-вторых, это же выражение можно расписать следующим образом:

$$Cov(X_1, X_1 + X_2 + X_3 + X_4 + X_5) = Var(X_1) + Cov(X_1, X_2) + Cov(X_1, X_3) +$$
$$+ Cov(X_1, X_4) + Cov(X_1, X_5) = Var(X_1) + 4Cov(X_1, X_2) = 250000 + 4Cov(X_1, X_2)$$

Объединяя оба полученных выше результата имеем:

$$250000+4Cov(X_1,X_2)=0$$

Решая получаем, что $Cov(X_1,X_2)=-62500$. Подставляя данный результат в изначальное уравнение имеем:

$$Var(X_1+X_2+X_3)=3Var(X_1)+6Cov(X_1,X_2)=3*250000-6*62500=375000$$

Задача №2

Воспользуемся первым начальным моментом, то есть математическим ожиданием:

$$E(X_1)=\int_0^\theta x\frac{2x}{\theta^2}dx=\frac{2x^3}{3\theta^2}|_0^\theta=\frac{2}{3}\theta$$

Отсюда следует, что $\theta=\frac{3}{2}E(X_1)$. Подставляя вместо математического ожидания его оценку получаем: $\hat{\theta}=\frac{3}{2}\bar{X}$, где $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$.

Задача №3

а) Оценка является несмещенной, поскольку:

$$E(\hat{\theta}_n)=E\left(\frac{3}{2}\bar{X}\right)=\frac{3}{2}E(X_1)=\frac{3}{2}*\frac{2}{3}\theta=\theta$$

б) Найдем дисперсию оценки:

$$Var(\hat{\theta}_n)=Var(\frac{3}{2}\bar{X})=\frac{9}{4n}Var(X_1)=\frac{9}{4n}(E(X_1^2)-E(X_1)^2)=\frac{9}{4n}(E(X_1^2)-\frac{4}{9}\theta^2)$$

Необходимо посчитать второй начальный момент:

$$E(X_1^2)=\int_0^\theta x^2\frac{2x}{\theta^2}dx=\frac{x^4}{2\theta^2}|_0^\theta=\frac{1}{2}\theta^2$$

Подставляя полученный результат в выражение для дисперсии оценки получаем:

$$Var(\hat{\theta}_n)=\frac{9}{4n}\left(\frac{1}{2}\theta^2-\frac{4}{9}\theta^2\right)=\frac{1}{8n}\theta^2$$

в) Оценка является состоятельной, так как соблюдены оба достаточных условия:

1. $\lim\limits_{n\to\infty}E(\hat{\theta}_n)=\lim\limits_{n\to\infty}\theta=\theta$

2. $\lim\limits_{n\to\infty}Var(\hat{\theta}_n)=\lim\limits_{n\to\infty}\frac{1}{8n}\theta^2=0$

г) Сперва найдем распределение оценки $T_n=\max(X_1,...,X_n)$. Для этого найдем функцию распределения на интервале $x\in[0,\theta]$, пользуясь свойствами независимости и одинаковой распределенности элементов выборки:

$$F_{T_n}(x) = P(T_n \le x) = P(\max(X_1,...,X_n) \le x) = P(X_1 \le x \cap ... \cap X_n \le x) =$$
$$= P(X_1 \le x)*...*P(X_n \le x) = F_{X_1}(x)*...*F_{X_n}(x) = F_{X_1}(x)*...*F_{X_1}(x) =$$
$$= \left(F_{X_1}(x)\right)^n = \left(\int_0^x \frac{2t}{\theta^2}dt\right)^n = \frac{x^{2n}}{\theta^{2n}}, x \in [0,\theta]$$

Дифференцируя функцию распределения получаем функцию плотности:

$$f_{T_n}(x) = \begin{cases} \frac{2nx^{2n-1}}{\theta^{2n}}, \text{ при } x \in [0,\theta] \\ 0, \text{ при } x \notin [0,\theta] \end{cases}$$

Теперь нетрудно показать, что оценка является смещенной, поскольку:

$$E(T_n) = \int_0^\theta x\frac{2nx^{2n-1}}{\theta^{2n}}dx = \int_0^\theta \frac{2nx^{2n}}{\theta^{2n}}dx = \frac{2n\theta^{2n+1}}{(2n+1)\theta^{2n}} = \frac{2n}{(2n+1)}\theta \ne \theta$$

Величина смещения составляет:

$$|E(T_n) - \theta| = |\frac{2n}{(2n+1)}\theta - \theta| = \frac{1}{(2n+1)}\theta$$

д) Поскольку оценка $\hat{\theta}$ является несмещенной, то:

$$MSE(\hat{\theta}_n) = Var(\hat{\theta}_n) = \frac{1}{8n}\theta^2$$

Для оценки T_n получаем:

$$MSE(T_n) = E((T_n - \theta)^2) = E(T_n^2 + \theta^2 - 2T_n\theta) = E(T_n^2) + \theta^2 - 2\theta\mathrm{E}(T_n)$$

Найдем второй начальный момент этой оценки:

$$E(T_n^2) = \int_0^\theta x^2\frac{2nx^{2n-1}}{\theta^{2n}}dx = \int_0^\theta \frac{2nx^{2n+1}}{\theta^{2n}}dx = \frac{2n\theta^{2n+2}}{(2n+2)\theta^{2n}} = \frac{2n}{(2n+2)}\theta^2$$

Подставляя моменты в выражение для MSE имеем:

$$MSE(T_n) = \frac{2n}{(2n+2)}\theta^2 + \theta^2 - 2\theta*\frac{2n}{(2n+1)}\theta = \frac{1}{(n+1)(2n+1)}\theta^2$$

Рассмотрим разницу в MSE:

$$MSE(T_n) - MSE(\hat{\theta}_n) = \frac{1}{(n+1)(2n+1)}\theta^2 - \frac{1}{8n}\theta^2 = \frac{-2n^2+5n-1}{8n(n+1)(2n+1)}\theta^2$$

Знак этой разницы зависит от знака $-2n^2+5n-1$, $n \in N$. Поэтому, следует рассмотреть три случая.

Во-первых, предположим, что $-2n^2+5n-1=0$, а значит $n=0.22$ или $n=2.28$, чего быть не может, следовательно, рассматриваемые оценки никогда не будут одинаково эффективными.

Во-вторых, допустим, что $-2n^2+5n-1<0$, тогда $n \ge 3$ и при таком объеме выборки оценка T_n оказывается более эффективной.

В-третьих, если $-2n^2+5n-1>0$, то $n=1$ или $n=2$ и лишь при таком объеме выборки оценка $\hat{\theta}_n$ будет более эффективной согласно рассматриваемому критерию.

Задача №4

а) Функция правдоподобия имеет вид:

$$L(\theta;x)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x_i^2}{2\theta}}$$

Тогда логарифм функции правдоподобия принимает форму:

$$lnL(\theta;x)=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln(\theta)-\sum_{i=1}^{n}\frac{x_i^2}{2\theta}$$

Условия первого порядка принимают вид:

$$\frac{dlnL(\theta;x)}{d\theta}=-\frac{n}{2\theta}+\frac{\sum_{i=1}^{n}x_i^2}{2\theta^2}=0$$

Решая данное равенство получаем точку экстремума:

$$\theta^*=\frac{\sum_{i=1}^{n}x_i^2}{n}$$

Проверим условия второго порядка:

$$\frac{d^2lnL(\theta;x)}{d^2\theta}|_{\theta=\theta^*}=\frac{n}{2\theta^{*2}}-\frac{\sum_{i=1}^{n}x_i^2}{\theta^{*3}}=\frac{1}{2n\left(\sum_{i=1}^{n}x_i^2\right)^2}-\frac{n^3}{\left(\sum_{i=1}^{n}x_i^2\right)^2}<0$$

Полученный результат свидетельствует в пользу того, что θ^* является точкой глобального максимума функции правдоподобия, а значит оценка максимального правдоподобия $\hat{\theta}^{ML}$ параметра θ принимает вид:

$$\hat{\theta}^{ML}=\frac{\sum_{i=1}^{n}X_i^2}{n}$$

б) Полученная оценка является несмещенной, поскольку:

$$E(\hat{\theta}^{ML}) = E\left(\frac{\sum_{i=1}^{n} X_i^2}{n}\right) = E(X_1^2) = \theta$$

в) Вычислим информацию Фишера следующим образом:

$$I(\theta) = -E\left(\frac{d^2 lnL(\theta;X)}{d^2\theta}\right) = -E\left(\frac{n}{2\theta^2} - \frac{\sum_{i=1}^{n} X_i^2}{\theta^3}\right) = \frac{n}{2\theta^2}$$

г) Рассмотрим произвольную несмещенную оценку $\hat{\theta}$ параметра θ. Тогда, согласно неравенству Рао-Крамера, получаем:

$$Var(\hat{\theta}) \geq \frac{2\theta^2}{n}$$

Покажем, что для оценки $\hat{\theta}^{ML}$ данное равенство соблюдается строго, что докажет её эффективность:

$$Var(\hat{\theta}^{ML}) = Var\left(\frac{\sum_{i=1}^{n} X_i^2}{n}\right) = \frac{1}{n} Var(X_1^2) = \frac{1}{n}\left(E(X_1^4) - E(X_1^2)^2\right) =$$

$$= \frac{1}{n}\left(\theta^2 E\left(\left(\frac{X_1}{\sqrt{\theta}}\right)^4\right) - E(X_1^2)^2\right) = \frac{1}{n}\left(3\theta^2 - \theta^2\right) = \frac{2\theta^2}{n}$$

Задача №5

Введем обозначения для реализаций экстремальных статистик: $x_{(1)} = \min(x_1,...,x_n)$ и $x_{(n)} = \max(x_1,...,x_n)$. Обратим пристальное внимание на то, что функция плотности $f(x;\theta)$ принимает значение $\frac{1}{\theta}$ только при $x \in [0,\theta]$, а в противном случае – принимает нулевое значение. Учитывая это обстоятельство, аккуратно запишем функцию правдоподобия:

$$L(\theta;x) = \prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} \prod_{i=1}^{n} \frac{1}{\theta}, \text{ если } x_{(n)} \leq \theta \\ 0, \text{ в противном случае} \end{cases} = \begin{cases} \frac{1}{\theta^n}, \text{ если } x_{(n)} \leq \theta \\ 0, \text{ в противном случае} \end{cases}$$

Поскольку функция правдоподобия строго убывает по θ при $x_{(n)} \leq \theta$ и равна нулю в противном случае, то точкой максимума будет $\theta^* = x_{(n)}$, а значит оценка максимального правдоподобия принимает вид: $\hat{\theta}^{ML} = X_{(n)}$, где $X_{(n)} = \max(X_1,...,X_n)$.