1. Пусть элементы выборки $X_1, X_2, \ldots, X_n$ объема $n \geqslant 5$ имеют распределение Пуассона с параметром λ, где $\lambda > 0$. Для какого параметра $\theta = \theta(\lambda)$ оценка $\theta^* = X_1 \cdot X_2 \cdot \ldots \cdot X_5$ будет несмещенной оценкой? Является ли эта оценка состоятельной оценкой для того же параметра θ?

Решение. В силу соответствующего свойства математического ожидания произведения независимых с. в.,

$$\mathsf{E}_\lambda \theta^* = \mathsf{E}_\lambda (X_1 \cdot X_2 \cdot \ldots \cdot X_5) = \mathsf{E}_\lambda X_1 \cdot \mathsf{E}_\lambda X_2 \cdot \ldots \cdot \mathsf{E}_\lambda X_5 = \lambda^5.$$

Итак, оценка θ^* является несмещенной оценкой для параметра $\theta = \theta(\lambda) = \lambda^5$. Состоятельной она не является, поскольку не зависит от n. Соответственно, при $n \to \infty$ оценка θ^* стремится по вероятности разве что к себе самой, не равняясь при этом λ^5.

2. Пусть элементы выборки $X_1, X_2, \ldots, X_n$ имеют распределение Бернулли с параметром p, где $0 < p < 1$. Для какого параметра $\theta = \theta(p)$ оценка $\theta^* = \arcsin\sqrt{\overline{X}}$ будет асимптотически нормальной оценкой? Найти коэффициент асимптотической нормальности.

Решение. Согласно соответствующей теореме, оценка $\theta^* = H\bigl(\overline{g(X)}\bigr) = H\left(\frac{\sum g(X_i)}{n}\right)$ является асимптотически нормальной оценкой параметра $\theta = H\bigl(\mathsf{E}g(X_1)\bigr)$ с коэффициентом

$$\sigma^2 = \bigl(H'\bigl(\mathsf{E}g(X_1)\bigr)\bigr)^2 \cdot \mathsf{D}g(X_1),$$

если существуют и отличны от нуля $H'\bigl(\mathsf{E}g(X_1)\bigr)$ и $\mathsf{D}g(X_1)$, и функция $H'(t)$ непрерывна в точке $t = \mathsf{E}g(X_1)$.

Имеем: $\theta^* = \arcsin\sqrt{\overline{X}} = H\left(\overline{g(X)}\right)$, если в качестве функций g и H взять $g(y) = y$, $H(t) = \arcsin\sqrt{t}$.

Проверим, что выполнены условия теоремы: $\mathsf{E}g(X_1) = \mathsf{E}X_1 = p$; дисперсия $\mathsf{D}g(X_1) = \mathsf{D}X_1 = p(1-p)$ существует и отлична от нуля при $0 < p < 1$; производная

$$H'(t) = (\arcsin\sqrt{t})' = \frac{1}{\sqrt{1-(\sqrt{t})^2}} \cdot \frac{1}{2\sqrt{t}} = \frac{1}{2\sqrt{t(1-t)}}$$

существует и непрерывна в точке $\mathsf{E}g(X_1) = p \in (0,1)$; $H'(\mathsf{E}g(X_1)) = \dfrac{1}{2\sqrt{p(1-p)}}$. Поэтому оценка θ^* — АНО параметра $\theta = \theta(p) = H\left(\mathsf{E}g(X_1)\right) = \arcsin\sqrt{p}$ с коэффициентом

$$\sigma^2(p) = \left(\frac{1}{2\sqrt{p(1-p)}}\right)^2 \cdot p(1-p) = \frac{1}{4p(1-p)}\, p(1-p) = \frac{1}{4}.$$

Выяснить, *для какого параметра* θ^* является асимптотически нормальной оценкой, можно и не прибегая к помощи теоремы. Так как любая асимптотически нормальная оценка состоятельна, то θ^* может быть асимптотически нормальной оценкой ровно того параметра, для которого она является состоятельной оценкой. А поскольку $\overline{X} \xrightarrow{\text{p}} p$, и «арксинус корня» суть непрерывная функция на $(0,1)$, то $\arcsin\sqrt{\overline{X}} \xrightarrow{\text{p}} \arcsin\sqrt{p} = \theta(p)$.

3. Пусть элементы выборки $X_1, X_2, \ldots, X_n$ имеют биномиальное распределение с параметрами 5 и p, где $0 < p < 1$. Проверить, выполнены ли условия неравенства Рао — Крамера и доказать, что оценка $p^* = \dfrac{n+1}{5n}\,\overline{X}$ неизвестного параметра p эффективна в некотором классе. В каком именно?

Решение. Для (дискретного) биномиального распределения «плотность» (она же вероятность попасть в точку y) имеет вид $f_p(y) = \mathsf{P}(X_1 = y) = C_5^y p^y (1-p)^{5-y}$ при $y = 0, 1, 2, 3, 4, 5$.
Условие (R), очевидно, выполнено: функция $\sqrt{f_p(y)} = \sqrt{C_5^y}\, p^{y/2}\,(1-p)^{(5-y)/2}$ при всех возможных $y = 0, 1, \ldots, 5$ непрерывно дифференцируема по p для любого $0 < p < 1$.
Чтобы проверить выполнение условия (RR), вычислим информацию Фишера:

$$\ln f_p(y) = \ln C_5^y + \frac{y}{2}\ln p + \frac{5-y}{2}\ln(1-p); \quad \frac{\partial}{\partial p}\ln f_p(y) = \frac{y}{p} - \frac{5-y}{(1-p)} = \frac{y-5p}{p(1-p)};$$

Вспомним, что $\mathsf{E}X_1 = 5p$ и $\mathsf{D}X_1 = 5p(1-p)$. Тогда

$$I(p) = \mathsf{E}\left(\frac{\partial}{\partial p}\ln f_p(X_1)\right)^2 = \mathsf{E}\left(\frac{X_1 - 5p}{p(1-p)}\right)^2 = \frac{\mathsf{D}X_1}{p^2(1-p)^2} = \frac{5p(1-p)}{p^2(1-p)^2} = \frac{5}{p(1-p)}.$$

Итак, информация Фишера $I(p) = \dfrac{5}{p(1-p)}$ конечна, положительна и непрерывна по p при всех $0 < p < 1$.
Найдем смещение оценки p^*:

$$b(p) = \mathsf{E}p^* - p = \frac{n+1}{5n}\,\mathsf{E}\,\overline{X} - p = \frac{n+1}{5n}\,\mathsf{E}X_1 - p = \frac{n+1}{5n}\,5p - p = \frac{p}{n}.$$

То есть оценка p^* лежит в классе $K_{p/n}$ оценок со смещением p/n. Вычислим ее дисперсию.

$$\mathsf{D}p^* = \frac{(n+1)^2}{25n^2}\,\mathsf{D}\,\overline{X} = \frac{(n+1)^2}{25n^2}\,\frac{\mathsf{D}X_1}{n} = \frac{(n+1)^2}{25n^2}\,\frac{5p(1-p)}{n} = \frac{(n+1)^2 p(1-p)}{5n^3}.$$

Все условия неравенства Рао — Крамера выполнены. Посмотрим, обращается ли для нашей оценки неравенство Рао — Крамера (в рамочке) в равенство.

$$\frac{(n+1)^2 p(1-p)}{5n^3} = \boxed{\mathsf{D}p^* \geqslant \frac{(1+b'(p))^2}{nI(p)}} = \frac{\left(1+\frac{1}{n}\right)^2}{n\,\frac{5}{p(1-p)}} \equiv \frac{(n+1)^2 p(1-p)}{5n^3}.$$

В неравенстве Рао — Крамера достигается равенство, следовательно, оценка p^* эффективна в классе $K_{p/n}$.

4. Пусть элементы выборки $X_1, X_2, \ldots, X_n$ имеют равномерное распределение на отрезке $[a, b]$, где $a < b$. Проверить, является ли оценка $\theta^* = X_{(n)} - X_{(1)}$ а) несмещенной; б) состоятельной; в) асимптотически нормальной оценкой параметра $\theta = b - a$.

Решение.

а) Функции распределения случайных величин $X_{(n)}$ и $X_{(1)}$ равны, соответственно,

$$F_{X_{(n)}}(y) = \begin{cases} 0, & y < a, \\ \left(\dfrac{y-a}{b-a}\right)^n, & a \leqslant y \leqslant b, \\ 1, & y > b; \end{cases} \qquad F_{X_{(1)}}(y) = \begin{cases} 0, & y < a, \\ 1 - \left(\dfrac{b-y}{b-a}\right)^n, & a \leqslant y \leqslant b, \\ 1, & y > b. \end{cases}$$

Математические ожидания первой и последней порядковых статистик равны (проверить!)

$$\mathsf{E}X_{(n)} = a + \frac{n}{n+1}(b-a), \quad \mathsf{E}X_{(1)} = a + \frac{1}{n+1}(b-a),$$

поэтому

$$\mathsf{E}\theta^* = \mathsf{E}X_{(n)} - \mathsf{E}X_{(1)} = a + \frac{n}{n+1}(b-a) - a - \frac{1}{n+1}(b-a) = \frac{n-1}{n+1}(b-a) \neq b-a,$$

то есть θ^* не является несмещенной оценкой параметра θ.

Этот вывод можно получить и не вычисляя математических ожиданий: заметим, что $\theta^* < \theta$ почти наверное, так что $\mathsf{E}\theta^* < \theta$ (см. соотв. свойство математического ожидания).

б) Проверим состоятельность. Докажем, что $X_{(n)} \xrightarrow{\mathrm{p}} b$ и $X_{(1)} \xrightarrow{\mathrm{p}} a$, откуда сразу же будет следовать, что $\theta^* \xrightarrow{\mathrm{p}} \theta$. Пусть $\varepsilon > 0$. Согласно (например!) неравенству Чебышёва,

$$\mathsf{P}(|X_{(n)} - b| > \varepsilon) = \mathsf{P}(-X_{(n)} + b > \varepsilon) \leqslant \frac{\mathsf{E}(-X_{(n)} + b)}{\varepsilon} = \frac{b-a}{(n+1)\varepsilon} \to 0 \text{ при } n \to \infty;$$

$$\mathsf{P}(|X_{(1)} - a| > \varepsilon) = \mathsf{P}(X_{(1)} - a > \varepsilon) \leqslant \frac{\mathsf{E}(X_{(1)} - a)}{\varepsilon} = \frac{b-a}{(n+1)\varepsilon} \to 0 \text{ при } n \to \infty.$$

Исчезновение модулей, помимо потусторонних соображений, объясняется тем, что почти наверное (с вероятностью 1) $X_{(n)} < b$ и $X_{(1)} > a$.

Итак, θ^* — состоятельная оценка параметра $\theta = b - a$.

в) Ни о какой асимптотической нормальности не может быть и речи, хотя бы потому, что $\theta^* = X_{(n)} - X_{(1)} < b - a = \theta$ почти наверное.

Действительно, по определению слабой сходимости, если θ^* — АНО параметра θ, то функция распределения $F_n(y) = \mathsf{P}(\sqrt{n}(\theta^* - \theta) < y)$ сходится к функции распределения $\Phi_{0,\sigma^2(\theta)}(y)$ какого-то нормального закона с нулевым математическим ожиданием, причем эта сходимость имеет место для всех y, поскольку функция распределения нормального закона всюду непрерывна.

Возьмем, например, $y = 0$ и убедимся в том, что в этой точке нет сходимости $F_n(0)$ к $\Phi_{0,\sigma^2(\theta)}(0) = 1/2$. Действительно, $F_n(0) = \mathsf{P}(\sqrt{n}(\theta^* - \theta) < 0) = 1$, не зависит от n и, тем самым, не сходится к $1/2$.

Итак, θ^* не является асимптотически нормальной оценкой параметра θ.

5. Пусть элементы выборки $X_1, X_2, \ldots, X_n$ имеют распределение с плотностью

$$f_\theta(y) = \begin{cases} 2y/\theta^2, & \text{если } 0 \leqslant y \leqslant \theta, \\ 0 & \text{иначе.} \end{cases}$$

Найти оценку максимального правдоподобия $\widehat{\theta}$ параметра θ.

Решение. Функция правдоподобия имеет вид:

$$\Psi(\vec{X}, \theta) = \begin{cases} \dfrac{2^n X_1 \cdot \ldots \cdot X_n}{\theta^{2n}}, & \text{если } 0 \leqslant X_{(1)} \leqslant \ldots \leqslant X_{(n)} \leqslant \theta, \\ 0 & \text{иначе} \end{cases} =$$

$$= \begin{cases} \dfrac{2^n X_1 \cdot \ldots \cdot X_n}{\theta^{2n}}, & \text{если } \theta \geqslant X_{(n)} \text{ и } X_{(1)} \geqslant 0, \\ 0 & \text{иначе} \end{cases}$$

В случае $X_{(1)} \geqslant 0$ функция правдоподобия
- равна $\dfrac{2^n X_1 \cdot \ldots \cdot X_n}{\theta^{2n}}$ в области $\theta \geqslant X_{(n)}$ и монотонно убывает с ростом θ;
- равна нулю для всех прочих $\theta > 0$;
- не определена для $\theta \leqslant 0$.

Наибольшее значение функция правдоподобия достигает при самом маленьком значении θ в области $\theta \geqslant X_{(n)}$, то есть при $\theta = X_{(n)}$. Итак, $\hat{\theta} = X_{(n)}$.

6. Пусть элементы выборки $X_1, X_2, \ldots, X_n$ имеют распределение Пуассона с параметром λ, где $\lambda > 0$. Для какого параметра $\theta = \theta(\lambda)$ оценка а) $\theta_1^* = \overline{X}e^{-\overline{X}}$; б) $\theta_2^* = \dfrac{1}{n}\sum\limits_{i=1}^{n} \mathrm{I}\{X_i = 1\}$ будет состоятельной оценкой? Проверить, будет ли оценка θ_1^* несмещенной оценкой того же параметра.

Указание. Для нахождения распределения $n\overline{X}$ воспользоваться свойством устойчивости распределения Пуассона относительно суммирования.

Решение. Поскольку $\overline{X} \xrightarrow{\mathrm{p}} \lambda$ при $n \to \infty$, то

$$\theta_1^* = \overline{X}e^{-\overline{X}} \xrightarrow{\mathrm{p}} \lambda e^{-\lambda} = \theta(\lambda).$$

Далее, $\mathbf{I}\{X_1 = 1\}, \mathbf{I}\{X_2 = 1\}, \ldots$ — независимые и одинаково распределенные с. в. с конечным первым (и не только) моментом (имеющие распределение Бернулли), поэтому

$$\theta_2^* = \frac{1}{n}\sum_{i=1}^{n} \mathbf{I}\{X_i = 1\} \xrightarrow{\mathrm{p}} \mathsf{E}\,\mathbf{I}\{X_1 = 1\} = \mathsf{P}\{X_1 = 1\} = \lambda e^{-\lambda} = \theta(\lambda).$$

Вычислим математическое ожидание оценки θ_1^*. В силу устойчивости распределения Пуассона по суммированию, случайная величина $\xi = n\overline{X} = \sum\limits_{i=1}^{n} X_i$ имеет распределение Пуассона $\Pi_{n\lambda}$, то есть $\mathsf{P}(\xi = k) = \dfrac{(n\lambda)^k}{k!}\,e^{-n\lambda}$ для $k = 0, 1, 2, \ldots$. Поэтому

$$\mathsf{E}\,\theta_1^* = \mathsf{E}\,\frac{\xi}{n}\,e^{-\frac{\xi}{n}} = \sum_{k=0}^{\infty} \frac{k}{n}\,e^{-\frac{k}{n}}\,\mathsf{P}(\xi = k) = \sum_{k=0}^{\infty} \frac{k}{n}\,e^{-\frac{k}{n}}\,\frac{(n\lambda)^k}{k!}\,e^{-n\lambda} =$$

$$= \frac{e^{-n\lambda}}{n}\sum_{k=1}^{\infty} \frac{\left(n\lambda e^{-\frac{1}{n}}\right)^k}{(k-1)!} = \left(n\lambda e^{-\frac{1}{n}}\right)\frac{e^{-n\lambda}}{n}\underbrace{\sum_{k-1=0}^{\infty} \frac{\left(n\lambda e^{-\frac{1}{n}}\right)^{k-1}}{(k-1)!}}_{e^{n\lambda\, e^{-1/n}}} = \lambda e^{n\lambda(e^{-1/n}-1)-1/n}$$

Итак, $\mathsf{E}\theta_1^* = \lambda e^{n\lambda(e^{-1/n}-1)-1/n} \neq \theta(\lambda) = \lambda e^{-\lambda}$, то есть оценка θ_1^* не является несмещенной оценкой параметра $\theta = \theta(\lambda) = \lambda e^{-\lambda}$.

Но легко убедиться (разложив в ряд Тейлора экспоненту, стоящую в показателе степени), что оценка θ_1^* является *асимптотически несмещенной* оценкой параметра $\theta = \theta(\lambda)$ (этого делать не требовалось).

Полезно также убедиться в том, что оценка θ_2^* не только является оценкой метода моментов для параметра θ, полученной с помощью замены истинного математического ожидания с.в. $g(X_1) = \mathbf{I}\{X_1 = 1\}$ на соответствующее выборочное среднее $\overline{g(X)}$, но и асимптотически нормальна, а также является единственной эффективной оценкой в классе K_0 несмещенных оценок параметра θ.