Второе задание по «Математической статистике»

1. На плоскости на расстоянии $a > 0$ (неизвестный параметр) от детектирующей прямой располагается радиоактивный источник, который излучает вспышками равновероятно по любому направлению в этой плоскости. Пусть $\vec{x} = (x_1, ..., x_n)$ – вектор координат вспышек, регистрируемых детектором. Предложите оценку координаты проекции источника на детектирующую прямую.

2. Пусть имеется простая выборка $\vec{x} = (x_1, ..., x_n)$ из распределения с неизвестной гладкой плотностью $p(x) > 0$, сосредоточенной на некотором отрезке. Покажите, что для оценки плотности распределения методом полигона частот (при котором прямыми линиями соединяются середины прямоугольных ступенек и затем неизвестная плотность приближается получающимися трапециями), то относительная погрешность такой аппроксимации оценивается, как $\sqrt{\delta^2_{\min}} \sim n^{-2/5}$.

3. Пусть $\vec{X} = (X_1, X_2, ..., X_n)$ – простая выборка объема n из распределения $F(x)$. Покажите, что в случае когда $F(x)$ – строго возрастающая (на самом деле это не обязательно требовать) непрерывная функция, распределение статистики $D_n(\vec{X}) = \sup\limits_{x \in \mathbb{R}} \left| F_n(x; \vec{X}) - F(x) \right|$ – не зависит от того, какая именно $F(x)$. Здесь $F_n(x; \vec{x})$ – эмпирическая функция распределения:

$F_n(x; \vec{X}) = \dfrac{\mu_n(x)}{n} = \dfrac{1}{n}\sum\limits_{k=1}^{n} \theta(x - X_k)$, где $\mu_n(x) = \left| \{ j : X_j < x \} \right|$, $\theta(z) = \begin{cases} 1, & z > 0, \\ 0, & z \le 0; \end{cases}$.

Указание: Положим $u = F(x)$, тогда u пробегает как минимум все точки интервала $(0, 1)$ (а как максимум – отрезка $[0,1]$), когда x пробегает $\mathbb{R}$. С использованием такой замены переменных имеем $D_n(\bar{X}) = \max\limits_{u \in [0,1]} \left| F_n(F^{-1}(u); \vec{X}) - u \right|$. Далее $F_n(F^{-1}(u)) = \dfrac{1}{n}\sum\limits_{k=1}^{n} \theta(F^{-1}(u) - X_k)$.

Остается показать эквивалентность событий $\{F^{-1}(u) - X_k \le 0\} \sim \{u - F(X_k) \le 0\}$.

4. Закон Хаббла в астрономии гласит: “скорость удаления галактики V прямо пропорциональна (с коэффициентом пропорциональности H – постоянная Хаббла) расстоянию до неё R”. Будем считать, что ошибки измерений некоррелированы, не имеют систематической ошибки и одинаково распределены по нормальному закону, т.е. имеет место “нормальная регрессия”:

$V_i = HR_i + \varepsilon_i$, $\varepsilon_i \in N(0, \sigma^2)$ (V_i и R_i – известны), $i = 1, ..., n$.

а) Предложите формулу для оценивания неизвестного параметра H (ответ аргументируйте).

б) Постройте γ-доверительный интервал для параметра H (σ^2 известно).

Т.е. нужно найти такие статистики $H_1 = H_1(R,V)$ и $H_2 = H_2(R,V)$, (функции от эмпирических данных), что выполняется: $P\{H_1(R,V) \le H \le H_2(R,V)\} \ge \gamma$.

5. Предположим, что вы стоите в очереди в школьной столовой. Пусть время T, за которое школьник оплатит свой обед на кассе, есть случайная величина, имеющая показательное распределение с параметром $\lambda > 0$, то есть $P(T > t) = e^{-\lambda t}$, $t \ge 0$.

а) Покажите, что показательное распределение обладает свойством отсутствия последействия: $P(T > t + s \,|\, T > s) = P(T > t)$, $t, s > 0$.

б) Покажите, что случайная величина $K_t = \max\left\{ n \in \mathbb{N} : \sum\limits_{i=1}^{n} T_i < t \right\}$, равная числу школьников, обслуженных на кассе к фиксированному моменту времени t, имеет пуассоновское распределение, то есть $P\{K_t = l\} = \dfrac{(\lambda t)^l e^{-\lambda t}}{l!}$, $l = 0, 1, 2, \ldots$.

в) Зная K_t ($t \gg 1$), предложите оценку неизвестного параметра λ, постройте γ-доверительный интервал для λ.