Примеры возможных задач по Лекции 2 и их решения

1. Пусть выборка $Z_n = \{X_1, \ldots, X_n\}$ соответствует распределению $F(x)$. Доказать, что порядковые статистики $X_{(1)}$ и $X_{(n)}$ имеют функции распределения, соответственно,

$$F_{(1)}(x) = 1 - (1 - F(x))^n \quad \text{и} \quad F_{(n)}(x) = F^n(x).$$

Решение. Для $X_{(1)} = \min\{X_1, \ldots, X_n\}$:

$$\mathrm{P}(X_{(1)} < x) = 1 - \mathrm{P}(X_{(1)} \geq x) = 1 - \mathrm{P}(X_1 \geq x, \ldots, X_n \geq x).$$

Поскольку $\mathrm{P}(X_i \geq x) = 1 - F(x)$, где $F(x) = \mathrm{P}(X < x)$, то

$$\mathrm{P}(X_{(1)} < x) = 1 - (1 - F(x))^n.$$

Следовательно, функция распределения $X_{(1)}$:

$$F_{(1)}(x) = \mathrm{P}(X_{(1)} < x) = 1 - (1 - F(x))^n.$$

Для $X_{(n)} = \max\{X_1, \ldots, X_n\}$:

$$F_{(n)}(x) = \mathrm{P}(X_{(n)} < x) = \mathrm{P}(X_1 < x, \ldots, X_n < x) = F(x)^n.$$

2. (Обобщение задачи 1) Найти функцию распределения k-й порядковой статистики $X_{(k)}$, $k = 1, 2, \ldots, n$.

Решение. Событие $\{X_{(k)} < x\}$ означает, что не менее k элементов выборки $< x$. Число элементов выборки, меньших x, имеет биномиальное распределение с параметрами n и $p = F(x)$. Следовательно,

$$F_{(k)}(x) = \mathrm{P}(X_{(k)} < x) = \sum_{m=k}^{n} \mathrm{C}_n^m F(x)^m (1 - F(x))^{n-m}.$$

Это выражение может быть записано через **неполную бета-функцию**:

$$F_{(k)}(x) = I_{F(x)}(k, n - k + 1),$$

где неполная бета-функция определяется как

$$I_x(a, b) = \frac{1}{B(a, b)} \int_0^x t^{a-1} (1 - t)^{b-1} \, dt,$$

а $B(a, b) = \int_0^1 t^{a-1} (1 - t)^{b-1} \, dt$ — полная бета-функция.

3. Выборка соответствует распределению $R[0,1]$ — непрерывному равномерному распределению на $[0,1]$. Найти $\mathrm{M}\{X_{(m)}\}$ и $\mathrm{D}\{X_{(m)}\}$ — математическое ожидание и дисперсию статистики $X_{(m)}$.

   Решение. Для равномерного распределения $R[0,1]$ плотность $f(x)=1$, функция распределения $F(x)=x$, $0\le x\le 1$. Плотность m-й порядковой статистики $X_{(m)}$:

$$f_{(m)}(x)=\frac{n!}{(m-1)!(n-m)!}F(x)^{m-1}(1-F(x))^{n-m}f(x)=$$

$$=\frac{n!}{(m-1)!(n-m)!}x^{m-1}(1-x)^{n-m},\quad 0\le x\le 1.$$

   Это бета-распределение с параметрами m и $n-m+1$. Его моменты:

$$\mathrm{M}X_{(m)}=\frac{m}{n+1},\quad \mathrm{D}X_{(m)}=\frac{m(n-m+1)}{(n+1)^2(n+2)}.$$

   В частности, при $m=n$ получаем:

$$\mathrm{M}X_{(n)}=\frac{n}{n+1},\quad \mathrm{D}X_{(n)}=\frac{n}{(n+1)^2(n+2)}.$$

4. Пусть выборка Z_n порождена случайной величиной X с конечным моментом ν_r. Доказать, что при любом n выборочный начальный момент $\bar{\nu}_r(n)$ обладает по отношению к ν_r свойством несмещённости, т.е.

$$\mathrm{M}\{\bar{\nu}_r(n)\}=\nu_r.$$

   Решение. Пусть $\nu_r=\mathrm{M}X^r$ — теоретический начальный момент порядка r. Выборочный начальный момент:

$$\bar{\nu}_r(n)=\frac{1}{n}\sum_{i=1}^{n}X_i^r.$$

   Тогда

$$\mathrm{M}\bar{\nu}_r(n)=\frac{1}{n}\sum_{i=1}^{n}\mathrm{M}X_i^r=\frac{1}{n}\cdot n\nu_r=\nu_r.$$

   Таким образом, $\bar{\nu}_r(n)$ — несмещённая оценка для ν_r.

5. Пусть выборка Z_n порождена случайной величиной X, имеющей распределение $R[0,1]$. Для любого $\varepsilon > 0$ оценить $\mathrm{P}\{|F_n^*(x) - x| \le \varepsilon\}$ при $n \gg 1$. В частности, получить оценку этой вероятности для $x = \frac{1}{2}$, если $\varepsilon = 0.1$, $n = 100$.

   Решение. Для $X \sim R[0,1]$ имеем $F(x) = x$, $0 \le x \le 1$. Эмпирическая функция распределения:

$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{X_i < x\}}.$$

   Тогда $nF_n^*(x) \sim \mathrm{Bin}(n, x)$. По неравенству Чебышёва:

$$\mathrm{P}\left(|F_n^*(x) - x| > \varepsilon\right) \le \frac{\mathrm{D}F_n^*(x)}{\varepsilon^2} = \frac{x(1-x)}{n\varepsilon^2} \le \frac{1}{4n\varepsilon^2}.$$

   Следовательно,

$$\mathrm{P}\left(|F_n^*(x) - x| \le \varepsilon\right) \ge 1 - \frac{1}{4n\varepsilon^2}.$$

   Для $x = \frac{1}{2}$, $\varepsilon = 0.1$, $n = 100$:

$$\mathrm{P}\left(\left|F_{100}^*\left(\frac{1}{2}\right) - \frac{1}{2}\right| \le 0.1\right) \ge 1 - \frac{1}{4 \cdot 100 \cdot 0.01} = 1 - \frac{1}{4} = 0.75.$$

   Более точную оценку можно получить, используя центральную предельную теорему и функцию Лапласа Φ_0:

$$\sqrt{n}\frac{F_n^*(x) - x}{\sqrt{x(1-x)}} \xrightarrow{d} N(0,1).$$

   Функция Лапласа: $\Phi_0(z) = \frac{1}{\sqrt{2\pi}}\int_0^z e^{-t^2/2}dt$. Тогда при больших n

$$\mathrm{P}\left(|F_n^*(x) - x| \le \varepsilon\right) \approx 2\Phi_0\left(\varepsilon\sqrt{\frac{n}{x(1-x)}}\right).$$

   Для данных значений:

$$\varepsilon\sqrt{\frac{n}{x(1-x)}} = 0.1 \cdot \sqrt{\frac{100}{0.25}} = 0.1 \cdot 20 = 2,$$

$$\mathrm{P} \approx 2\Phi_0(2) \approx 2 \cdot 0.4772 = 0.9544.$$