Домашнее задание 1, Новиков Герман, 277

Задача 1. Пусть $\hat{F}_n$ — эмпирическая функция распределения, построенная по простой выборке длины n, с неизвестной функцией распределения F.

- Какое распределение имеет случайная величина $n\hat{F}_n$?
- Используя оценку $\hat{F}_n(x)$ и вариант центральной предельной теоремы, построить приближенный (в виду предельности теоремы) интервал, содержащий значение $F(x)$ с заданной вероятностью α.

Решение:

- По определению:
$$n\hat{F}_n(x) = \sum_{i=1}^{n} I(X_i < x)$$
. Но тогда, для произвольного x: $n\hat{F}_n(x)$ есть сумма n независимых Бернуллевских случайных велечин с $p = F(x)$, а следственно
$$n\hat{F}_n(x) \sim Bin(n, F(x))$$
-
$$\sqrt{n}(\hat{F}_n(x) - F(x)) = \left(\frac{\sum\limits_{i=1}^{n} I(X_i < x) - nF(x)}{\sqrt{n}} \right) =$$
$$= \left(\frac{\sum\limits_{i=1}^{n} I(X_i < x) - n\mathsf{E}(I(X_1 < x))}{\sqrt{n}} \right) \xrightarrow{CLT} N(0, \mathsf{D}(I(X_1 < x)))$$
Где $\mathsf{D}(I(X_1 < x)) = F(x)(1 - F(x))$

Таким образом, получаем ассимптотическую оценку на распределение разности между эмпирической и реальной функцией распределения, взвешенной на $\sqrt{n}$. Положим, что выборка достаточно большая и можно считать, что равенство не ассимптотическое. Тогда мы просто берем двухсторонний квантиль:

$$x^{(1)}_{\frac{\alpha}{\sqrt{n}}} = G^{-1}\left(\frac{\alpha}{2\sqrt{n}}\right), x^{(2)}_{\frac{\alpha}{\sqrt{n}}} = G^{-1}\left(1 - \frac{\alpha}{2\sqrt{n}}\right)$$

Где $G(x)$ - та самая, полученная нами предельная функция распределения.

Теорема 1. *(Неравенство Дворецкого-Кифера-Вольфовитца).*

Пусть $X_1, ..., X_n \sim F$. Тогда для любого $\varepsilon > 0$:

$$\mathsf{P}\left(\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \geqslant \varepsilon\right) \leqslant 2\exp(-2n\varepsilon^2)$$

Задача 2. Построить на основании неравенства Дворецкого-Кифера-Вольфовитца неасимптотический вариант критерия согласия Колмогорова. Указать критическую область при заданном уровне значимости.

Решение: критерий согласия Колмогорова:

$$\mathsf{P}\left(\sqrt{n}\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \leqslant \varepsilon\right) = \mathsf{P}\left(\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \leqslant \frac{\varepsilon}{\sqrt{n}}\right) \leqslant 1 - 2\exp(-2\varepsilon^2)$$

Задаём α - уровень значимости: $\sqrt{n}\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| \geqslant t_\alpha$

В нашем случае $K(\varepsilon) = 1 - 2\exp(-2\varepsilon^2)$

Хотим: $\alpha = 1 - K(\varepsilon)$, следственно выбираем

$$t_\alpha = \varepsilon = K^{-1}(1-\alpha) = \sqrt{\frac{\ln(2) - ln(\alpha)}{2}}$$

Таким образом отвергаем гипотезу H_0 при:

$$\sqrt{n}D_n \geqslant \sqrt{\frac{\ln(2) - ln(\alpha)}{2}}$$

Задача 3. Доказать, что

$$\sup_{x\in\mathbb{R}} |\hat{F}(x) - F(x)| \xrightarrow{a.e.} 0,$$

где предел берется по n - числу элементов выборки.

Доказательство(можно не комментировать, это доказательство из книги, свое я обещаю придумать): Пусть $F(x)$ непрерывна. Фиксируем $\varepsilon = 1/N$. Так как $\forall x F(x) \in [0,1]$ и непрерывна, то существует набор чисел $\{x_i\}_{i=1}^N \subset \mathbb{R} \cup \{-\infty, +\infty\}$ такой, что на каждом из отрезков $[x_i, x_i + 1]$ функция $F(x)$ растёт на $\frac{1}{N}$, где N - произвольное целое. Тогда $\forall x \in [x_i, x_{i+1}]$ выполнено:

$$\hat{F}(x) - F(x) \leqslant \hat{F}(x_{i+1}) - F(x_i) = \hat{F}(x_{i+1}) - F(x_{i+1}) + \varepsilon$$

$$\hat{F}(x) - F(x) \geqslant \hat{F}(x_i) - F(x_i + 1) = \hat{F}(x_i) - F(x_i) - \varepsilon$$

Обозначим A_k - множество элементарных событий ω, на которых $\hat{F}(x_k) \to F(x_k)$. Мера этого множества $P(A_k) = 1$ в силу УЗБЧ. Но тогда $\forall \omega \in A = \cup_{k=0,N} A_k$ существует $n(\omega)$: что $\forall n \geqslant n(\omega)$ выполнено

$$|\hat{F}(x_k) - F(x_k)| < \varepsilon, k = 0, 1, ..., N.$$

Но тогда вместе с предыдущим получаем требуемое утверждение:

$$\sup_z |\hat{F}(x) - F(x)| < 2\varepsilon.$$

Для произвольной функции аналогично. Только нужно более аккуратно поступать с разрывами.

Задача 4. Построить plug-in оценки для математического ожидания и дисперсии. Являются ли эти оценки несмещенными?

Решение:

- Для математического ожидания $\overline{x} = \hat{\mathsf{E}}(X) = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$. Она является несмещенной, т.к. $\mathsf{E}\hat{\mathsf{E}}(X) = \mathsf{E}\frac{1}{n}\sum\limits_{i=1}^{n} X_i = \frac{1}{n}\sum\limits_{i=1}^{n} \mathsf{E}X_i = \mathsf{E}X$, где $X \sim F(x)$

- Для дисперсии: $S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{x})^2$. Она не является несмещенной:

  $S^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{x})^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i^2 - 2X_i\overline{x} + \overline{x}^2) = \frac{1}{n}\sum\limits_{i=1}^{n}X_i^2 - 2\overline{x}(\frac{1}{n}\sum\limits_{i=1}^{n}X_i) + \frac{1}{n}\sum\limits_{i=1}^{n}X_i^2 = \frac{1}{n}\sum\limits_{i=1}^{n}X_i^2 - \overline{x}^2 = \overline{x^2} - \overline{x}^2$

  Тогда получим, что $\mathsf{E}S^2 = \mathsf{E}(\overline{x^2} - \overline{x}^2) = ... = \sigma^2 - \frac{\sigma^2}{n} = \sigma^2(1 - \frac{1}{n})$. То есть не является несмещенной, однако просто нормируется до несмещенной, домножением на $\frac{n}{n-1}$.