Теория вероятностей и математическая статистика, Коллоквиум IV

Версия от 16.06.2021 14:13

Содержание

Билет 1

Выборка, оценка, статистика. Несмещенность, состоятельность, асимптотическая нормальность и эффективность оценок. Пример отсутствия несмещенной оценки. Отсутствие эффективной оценки в классе всех оценок. Единственность эффективной оценки. Состоятельность асимптотической нормальной оценки.

Выборка, оценка, статистика

Предположим, нам известно, что неизвестное распределение принадлежит какому-то конкретному семейству распределений с функциями распределения F_θ, где $\theta \in \Theta \subseteq \mathbb{R}^k$. Тогда задачей статистики является оценка неизвестного параметра $\theta_0 \in \Theta$, соответствующего нашему неизвестному распределению.

Например, пусть X есть случайная величина и мы знаем распределение этой случайно величины $F_\theta(t)$ с точностью до θ (например, $\mathcal{N}(0, \theta)$). Задача статистики заключается в том, чтобы оценить параметр θ.

Пример. Пусть в ящике N шаров, M из них чёрные. Мы достали из ящика n шаров. Теория вероятностей задается вопросом, с какой вероятностью среди вытянутых шаров есть m чёрных. Математическая статистика задаётся вопросом, сколько всего в ящике чёрных шаров (какое M), если мы достали n шаров и m из них чёрные.

Определение. Вектор $X = (X_1, \dots, X_n)$ с независимыми компонентами, где каждая случайная величина имеет одно и то же распределение, называется **выборкой**.

Определение. Произвольная функция $T_n(X)$, принимающая выборку как аргумент, называется **статистикой**.

Важно то, что статистика зависит от случайных величин $X_1, \dots, X_n$, и не зависит от θ.

Пример. Выборочное среднее $\overline{X_n}$ является статистикой

$$T(X) = \overline{X_n} = \frac{X_1 + \dots + X_n}{n}.$$

Когда проводится серия независимых экспериментов с функцией распределения F_θ (θ неизвестно), мы получаем выборку $X = (X_1, \dots, X_n)$. По выборке хочется определить значение $\widehat{\theta}$, которое в каком-либо смысле близко к реальному θ.

Определение. Статистика $\widehat{\theta}(X_1, \dots, X_n)$ со значением из множества параметров Θ называется **оценкой** неизвестного параметра θ.

Несмещенность, состоятельность, асимптотическая нормальность и эффективность оценок

Определение. Оценка $\widehat{\theta}_n(X)$ является **несмещенной**, если $\mathrm{E}_\theta\left[\widehat{\theta}_n(X)\right] = \theta$ для любого $\theta \in \Theta$.

Напомним определение сходимости случайных величин X_n к случайной величине X *по вероятности*:

$$X_n \xrightarrow[n\to\infty]{\mathrm{P}} X \iff \forall \varepsilon > 0 \lim_{n\to\infty} \mathrm{P}\left[|X_n - X| \geqslant \varepsilon\right] = 0.$$

Определение. Оценка $\widehat{\theta}_n(X)$ является **состоятельной**, если $\widehat{\theta}_n(X) \xrightarrow[n\to\infty]{\mathrm{P}_\theta} \theta$ для любого $\theta \in \Theta$.

Обычно состоятельность оценки является следствием закона больших чисел.

Пример. Пусть $\widehat{\theta}_n = \overline{X_n} = \dfrac{X_1 + \dots + X_n}{n}$ и $\mathrm{E}[X_1] = \theta$. Тогда $\widehat{\theta}_n = \overline{X_n} \xrightarrow[\text{ЗБЧ}]{\mathrm{P}} \mathrm{E}[X_1] = \theta$, откуда следует, что оценка $\widehat{\theta}_n$ состоятельная.

Напомним определение сходимости случайных величин X_n к случайной величине X *почти наверное*:

$$X_n \xrightarrow[n\to\infty]{\text{п.н.}} X \iff \mathrm{P}\left[\lim_{n\to\infty} X_n = X\right] = 1.$$

Определение. Оценка $\widehat{\theta}_n(X)$ является **сильно состоятельной**, если $\widehat{\theta}_n(X) \xrightarrow[n\to\infty]{\text{п.н.}} \theta$ для любого $\theta \in \Theta$.

Напомним определение сходимости случайных величин X_n к случайной величине X *по распределению*:

$$X_n \xrightarrow[n\to\infty]{\mathrm{d}} X \iff \lim_{n\to\infty} F_{X_n}(x) = F_X(x) \text{ в каждой точке } x, \text{ где непрерывна } F_X.$$

Определение. Оценка $\widehat{\theta}_n(X)$ является **асимптотически нормальной** с коэффициентом $\sigma^2(\theta)$, если

$$\sqrt{n}\left(\widehat{\theta}_n(X)-\theta\right) \xrightarrow[n\to\infty]{\mathrm{d}_\theta} \mathcal{N}(0,\sigma^2(\theta)),$$

что эквивалентно $\dfrac{\sqrt{n}\left(\widehat{\theta}_n(X)-\theta\right)}{\sigma(\theta)} \xrightarrow[n\to\infty]{\mathrm{d}_\theta} \mathcal{N}(0,1)$.

Коэффициент $\sigma^2(\theta)$ называется асимптотической дисперсией.

Определение. Пусть K это некое множество (класс) оценок (например, K — несмещенные оценки).

Оценка $\widehat{\theta}_n(X) \in K$ является **эффективной в классе** K, если для каждого $\theta \in \Theta$ и для каждой оценки $\theta_n^* \in K$ выполняется

$$\mathrm{E}_\theta\left[\widehat{\theta}_n(X)-\theta\right]^2 \leqslant \mathrm{E}_\theta\left[\theta_n^*(X)-\theta\right]^2.$$

Если K это класс несмещенных оценок, то условие можно переписать следующим образом:

$$\mathrm{D}_\theta\left[\widehat{\theta}_n(X)\right] \leqslant \mathrm{D}_\theta\left[\theta_n^*(X)\right].$$

Определение. Несмещенную оценку $\widehat{\theta}_n(X)$ в классе всех несмещенных оценок будем называть просто **эффективной**.

Пример отсутствия несмещенной оценки

Пример. Пусть X_i это случайная величина Бернулли (то есть X_i принимает значение 1 с вероятностью $p \in (0;1)$ и 0 с вероятностью $1-p$) и $\theta = \sin(p)$. Запишем математическое ожидание оценки по определению:

$$\mathrm{E}\left[\widehat{\theta}_n(X)\right] = \sum_{(x_1,\ldots,x_n)\in\{0,1\}^n} \widehat{\theta}_n(x_1,\ldots,x_n)\cdot p^{x_1+\cdots+x_n}\cdot(1-p)^{n-x_1-\cdots-x_n}.$$

То есть математическое ожидание оценки это некий полином от p, а полином не может равняться $\sin p$, поэтому оценка не может быть несмещенной. Вместо синуса можно было рассмотреть что-то другое.

Отсутствие эффективной оценки в классе всех оценок

Утверждение. Не существует эффективной оценки в классе всех оценок.

Доказательство. Пусть $\widehat{\theta}_n(X)$ эффективна в классе всех оценок. Тогда для любой $\theta_n^*(X)$ должно выполняться

$$\mathrm{E}_\theta\left[\widehat{\theta}_n(X)-\theta\right]^2 \leqslant \mathrm{E}_\theta\left[\theta_n^*(X)-\theta\right]^2.$$

В том числе, это должно выполняться для любой $\theta_n^*(X) = C = \mathrm{const}$:

$$\mathrm{E}_\theta\left[\widehat{\theta}_n(X)-\theta\right]^2 \leqslant \mathrm{E}_\theta\left[C-\theta\right]^2.$$

Это неравенство также должно выполняться для любого θ. В частности, для $\theta = C$. Тогда

$$\mathrm{E}_\theta\left[\widehat{\theta}_n(X)-\theta\right]^2 \leqslant 0.$$

Это означает, что $\widehat{\theta}_n(X) = \theta = C$ почти наверное. Но мы ведь могли взять и другое $C' \neq C$ и ровно по тем же соображениям получить

$$\widehat{\theta}_n(X) = C \neq C' = \widehat{\theta}_n(X).$$

Пришли к противоречию. ■

Единственность эффективной оценки

Утверждение. Если эффективная оценка существует, то она единственная.

Доказательство. Пусть $\widehat{\theta}_n(X)$ и $\theta_n^*(X)$ — эффективные оценки.

Тогда по определению эффективности:

- $\mathrm{E}_\theta\left[\widehat{\theta}_n(X)\right] = \theta = \mathrm{E}_\theta\left[\theta_n^*(X)\right]$ для любого θ.
- $\mathrm{D}_\theta\left[\widehat{\theta}_n(X)\right] \leqslant \mathrm{D}_\theta\left[\tilde{\theta}(X)\right]$ и $\mathrm{D}_\theta\left[\theta_n^*(X)\right] \leqslant \mathrm{D}_\theta\left[\tilde{\theta}(X)\right]$ для любой несмещенной $\tilde{\theta}_n(X)$.

Рассмотрим равенство параллелограмма для билинейной формы $\operatorname{cov}(\xi, \nu)$:

$$\operatorname{cov}(\xi+\nu, \xi+\nu) + \operatorname{cov}(\xi-\nu, \xi-\nu) = 2 \cdot \operatorname{cov}(\xi, \xi) + 2 \cdot \operatorname{cov}(\nu, \nu).$$

Так как $\operatorname{cov}(\xi, \xi) = \mathrm{D}[\xi]$, поделив обе стороны на 4 получаем равенство

$$\mathrm{D}\left[\frac{\xi+\nu}{2}\right] + \mathrm{D}\left[\frac{\xi-\nu}{2}\right] = \frac{1}{2}\,\mathrm{D}[\xi] + \frac{1}{2}\,\mathrm{D}[\nu].$$

Воспользовавшись этим равенством для оценки $\dfrac{\widehat{\theta}_n(X) + \theta_n^*(X)}{2}$ получаем

$$\mathrm{D}_\theta\left[\frac{\widehat{\theta}_n(X) + \theta_n^*(X)}{2}\right] = \frac{\mathrm{D}[\widehat{\theta}_n(X)] + \mathrm{D}[\theta_n^*(X)]}{2} - \mathrm{D}_\theta\left[\frac{\widehat{\theta}_n(X) - \theta_n^*(X)}{2}\right] \leqslant \frac{\mathrm{D}[\widehat{\theta}_n(X)] + \mathrm{D}[\theta_n^*(X)]}{2}$$

При этом, так как обе оценки эффективные,

$$\frac{\mathrm{D}[\widehat{\theta}_n(X)] + \mathrm{D}[\theta_n^*(X)]}{2} \leqslant \mathrm{D}_\theta\left[\frac{\widehat{\theta}_n(X) + \theta_n^*(X)}{2}\right].$$

Последнее равенство может показаться неочевидным. Так как $\widehat{\theta}_n(X)$ это эффективная оценка (а значит имеет дисперсию не превосходящую дисперсию любой другой несмещенной оценки) и оценка $\dfrac{\widehat{\theta}_n(X) + \theta_n^*(X)}{2}$ несмещенная (по линейности матожидания), то выполняется неравенство

$$\mathrm{D}_\theta[\widehat{\theta}_n(X)] \leqslant \mathrm{D}_\theta\left[\frac{\widehat{\theta}_n(X) + \theta_n^*(X)}{2}\right] \iff \frac{1}{2}\,\mathrm{D}_\theta[\widehat{\theta}_n(X)] \leqslant \frac{1}{2}\,\mathrm{D}_\theta\left[\frac{\widehat{\theta}_n(X) + \theta_n^*(X)}{2}\right].$$

Аналогичное неравенство выполняется и для оценки $\theta_n^*(X)$. Суммируем эти два неравенства и получаем

$$\frac{\mathrm{D}_\theta[\widehat{\theta}_n(X)] + \mathrm{D}_\theta[\theta_n^*(X)]}{2} \leqslant \mathrm{D}_\theta\left[\frac{\widehat{\theta}_n(X) + \theta_n^*(X)}{2}\right].$$

Мы закончили тем же, с чего и начали. Тогда все неравенства это равенство. Значит, $\mathrm{D}_\theta\left[\dfrac{\widehat{\theta}_n(X) - \theta_n^*(X)}{2}\right] = 0.$

Из этого следует, что почти наверное

$$\widehat{\theta}_n(X) - \theta_n^*(X) = \mathrm{E}_\theta\left[\widehat{\theta}_n(X) - \theta_n^*(X)\right] = \mathrm{E}_\theta\left[\widehat{\theta}_n(X)\right] - \mathrm{E}_\theta\left[\theta_n^*(X)\right] = \theta - \theta = 0.$$

То есть, $\widehat{\theta}_n(X) = \theta_n^*(X)$ почти наверное. ■

Состоятельность асимптотической нормальной оценки

Утверждение. Если оценка асимптотически нормальная, то она состоятельная.

Доказательство. Пусть $\widehat{\theta}_n(X)$ — асимптотически нормальная оценка параметра θ.

По определению асимптотической нормальности $\sqrt{n}\left(\widehat{\theta}_n(X) - \theta\right) \xrightarrow{\mathrm{d}_\theta} \mathcal{N}(0, \sigma^2(\theta))$.

Про сходимость по распределению произведения мы знаем, что если один из пределов сходится к константе, то верна арифметика:

$$\frac{1}{\sqrt{n}} \cdot Z \xrightarrow[n\to\infty]{\mathrm{d}} 0 \cdot Z = 0.$$

Тогда

$$\widehat{\theta}_n(X) - \theta = \frac{1}{\sqrt{n}} \cdot \sqrt{n}\left(\widehat{\theta}_n(X) - \theta\right) \xrightarrow[n\to\infty]{\mathrm{d}_\theta} 0 \cdot \mathcal{N}(0, \sigma^2(\theta)) = 0.$$

Так как есть сходимость по распределению к константе, то есть сходимость по вероятности к константе (факт с предыдущего коллоквиума):

$$\widehat{\theta}_n(X) - \theta \xrightarrow[n\to\infty]{\mathrm{d}_\theta} 0 \iff \widehat{\theta}_n(X) \xrightarrow[n\to\infty]{\mathrm{P}_\theta} \theta,$$

а это и есть определение состоятельности. ■

Билет 2

Метод моментов и его состоятельность. Метод максимального правдоподобия. Энтропия и состоятельность оценки максимального правдоподобия.

Метод моментов и его состоятельность

Пусть $(X_1, \ldots, X_n)$ — выборка, где X_i задано распределением F_θ. Хотим найти состоятельную оценку параметра θ. Пусть g — непрерывная функция, причём $\mathbb{E}_\theta|g(X_1)| < \infty$. Посчитаем матожидание $\mathbb{E}_\theta g(X_1) = f(\theta)$. Предположим, что $\exists\, f^{-1}$, и она непрерывна. Так как на практике мы не знаем параметр θ, то мы не можем посчитать такое матожидание. Но мы можем приближённо посчитать $f(\theta)$ воспользовавшись ЗБЧ:

$$\frac{g(X_1) + \ldots + g(X_n)}{n} \xrightarrow{P_\theta} \mathbb{E}_\theta g(X_1) = f(\theta).$$

Теперь в силу обратимости f можно получить сходимость к θ:

$$f^{-1}\left(\frac{g(X_1) + \ldots + g(X_n)}{n}\right) \xrightarrow{P_\theta} f^{-1}(f(\theta)).$$

Оценкой параметра θ назовём функцию

$$\hat{\theta}(X_1, \ldots, X_n) = f^{-1}\left(\frac{g(X_1) + \ldots + g(X_n)}{n}\right).$$

Эта оценка состоятельна, так как $\hat{\theta}(X_1, \ldots, X_n) \xrightarrow{P_\theta} \theta$. Вместо ЗБЧ можно применить УЗБЧ, и получить сильную состоятельность.

Метод максимального правдоподобия

Энтропия и состоятельность оценки максимального правдоподобия

Определение 1. Обобщённой плотностью ρ_X случайной величины X назовём функцию плотности X, если случайная величина является непрерывно, или функцию $\rho_X(t) = P(X = t)$ в случае, если X имеет дискретное распределение.

Определение 2. Пусть $X = (X_1, \ldots, X_n)$ — выборка из распределения с обобщённой плотностью ρ_θ. Обобщённая плотность вектора X называется функцией правдоподобия, и имеет вид

$$p(X, \theta) = \rho_\theta(X_1) \cdot \ldots \cdot \rho_\theta(X_n)$$

Функцию $\ln p(X, \theta)$ называют логарифмической функцией правдоподобия и обозначают $L(X, \theta)$.

Определение 3. Пусть ρ_0, ρ_1 — положительные вероятностные плотности. Выражение

$$\int \ln \frac{\rho_1(x)}{\rho_0(x)} \cdot \rho_1(x) dx$$

называется энтропией распределения с плотностью ρ_1 относительно распределения с плотностью ρ_0.

Замечание. Здесь и далее интергралы без пределов интегрирования обозначают интегрирование по множеству, на котором задано распределение. Они вовсе не означают неопределённый интеграл.

Следующее утверждение показывает, что энтропия в некотором смысле оценивает расстояние между распределениями:

Лемма (Информационное неравенство)**.** Пусть ρ_0, ρ_1 — положительные вероятностные плотности. Тогда

$$\int \ln \frac{\rho_1(x)}{\rho_0(x)} \cdot \rho_1(x) dx \geqslant 0$$

Равенство достигается тогда и только тогда, когда $\rho_0 = \rho_1$.

Доказательство. Домножим обе части неравенства на (-1) и будем выводить оценку сверху:

$$\int \ln \frac{\rho_0(x)}{\rho_1(x)} \cdot \rho_1(x) dx \overset{?}{\leqslant} 0$$

Воспользуемся неравенством $\ln x \leqslant x - 1$ (очевидно, если, например, посмотреть на графики этих функций: у них есть единственное пересечение в точке $x = 1$):

$$\ln \frac{\rho_0(x)}{\rho_1(x)} \rho_1(x) \leqslant \left(\frac{\rho_0}{\rho_1} - 1\right) \rho_1 = \rho_0 - \rho_1$$

Проинтегрируем обе части:
$$\int \ln \frac{\rho_0(x)}{\rho_1(x)} \cdot \rho_1(x)dx \leqslant \int \rho_0(x)dx - \int \rho_1(x)dx = 0$$
Оба интеграла справа равны 1, в силу того, что под интегралами стоят плотности. Таким образом оценку сверху мы доказали, найдём теперь, когда достигается равенство.

Пусть в неравенстве достигается равенство, т.е. известно, что
$$\int \ln \frac{\rho_0(x)}{\rho_1(x)} \cdot \rho_1(x)dx = 0 \qquad \int \rho_0(x)dx - \int \rho_1(x)dx = 0 \iff \int \left(\frac{\rho_0}{\rho_1} - 1\right)\rho_1 dx = 0$$
Тогда
$$\int \left(\left(\frac{\rho_0}{\rho_1} - 1\right) - \ln \frac{\rho_0}{\rho_1}\right)\rho_1 dx = 0$$
Так как $\ln x \leqslant x - 1$, то $0 \leqslant x - 1 - \ln x$ и функция в скобках неотрицательна. Теперь очевидно, что 0 достигается только в случае $\rho_0 = \rho_1$:
$$\left(\frac{\rho_0}{\rho_1} - 1\right) - \ln \frac{\rho_0}{\rho_1} = 0 \iff \left(\frac{\rho_0}{\rho_1} - 1\right) = \ln \frac{\rho_0}{\rho_1} \iff \rho_0 = \rho_1$$
■

Вывели утверждение, которое показывает, что энтропия, в некотором смысле оценивает расстояние между плотностями, т.е. расстояние между распределениями. Теперь будем применять это утверждение для построения оценки.

Пусть есть выборка $X = (X_1, \ldots, X_n)$ из распределения с обобщённой плотностью ρ_θ. Пусть реальное значение параметра θ равно θ_1. Рассмотрим функцию следующего вида:
$$W(\theta) = \mathbb{E}_{\theta_1} \ln \rho_\theta(X_1) = \int \ln \rho_\theta(x)\rho_{\theta_1}(x)dx$$
Можно показать, что $W(\theta) \leqslant W(\theta_1) \,\forall\, \theta$, действительно:
$$W(\theta) - W(\theta_1) = \int \ln \rho_\theta(x)\rho_{\theta_1}(x)dx - \int \ln \rho_{\theta_1}(x)\rho_{\theta_1}(x)dx = \int \ln \frac{\rho_\theta(x)}{\rho_{\theta_1}(x)}\rho_{\theta_1}(x)dx \leqslant 0$$
Причём наибольшее значение $W(\theta)$ достигается при $\theta = \theta_1$. Таким образом можно естественно оценить реальный параметр, если найти точку максимума функции $W(\theta)$. В чём проблема: мы не знаем ρ_{θ_1}, и потому функция $W(\theta)$ нам так же не известна. Решение проблемы: $W(\theta)$ это некоторое матожидание. По ЗБЧ известно, что выборочное среднее по вероятности сходится к матожиданию. Т.е.
$$\frac{\ln \rho_\theta(X_1) + \ldots + \ln \rho_\theta(X_n)}{n} \xrightarrow{P} \mathbb{E}_{\theta_1} \ln \rho_\theta(X_1) = W(\theta)$$
Немного преобразуем левую часть:
$$\frac{\ln \rho_\theta(X_1) + \ldots + \ln \rho_\theta(X_n)}{n} = \frac{1}{n}\sum_{i=1}^{n} \ln \rho_\theta(X_i) = \frac{1}{n}L(X, \theta)$$
Таким образом, вместо того, чтобы искать максимум неизвестной функции, мы будем искаль максимум того, что к ней приближается, и найденное значение и будем называть **оценкой максимального правдоподобия**.

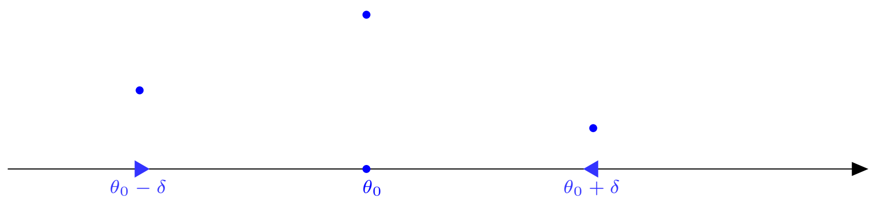
Определение 4. Оценкой максимального правдоподобия параметра θ называется максимум функции $L(X, \theta)$.

Предложение. (Состоятельность оценки максимального правдоподобия.)

Пусть $\theta \in (a, b)$, и на этом отрезке функция $\theta \to L(X, \theta)$ имеет единственную точку локального максимума $\hat{\theta}$. Тогда $\hat{\theta} \xrightarrow{P} \theta_0$.

Доказательство. Будем пользоваться тем, что $\frac{1}{n}L(X, \theta) \xrightarrow{P} W(\theta)$. Хотим доказать, что $P(|\hat{\theta} - \theta_0| \geqslant \delta) \to 0 \,\forall\, \delta > 0$ (просто определение сходимости по вероятности). Рассмотрим точки $\theta_0 - \delta, \theta_0 + \delta$. Про эти точки известно следующее:
$$\begin{cases} W(\theta_0) > W(\theta_0 - \delta) \xleftarrow{P} \frac{1}{n}L(X, \theta_0 - \delta) \\ W(\theta_0) > W(\theta_0 + \delta) \xleftarrow{P} \frac{1}{n}L(X, \theta_0 + \delta) \end{cases}$$
Можно ожидать, что при достаточно большом n, с вероятностью, близкой к 1 будут выполнены неравенства
$$\begin{cases} \frac{1}{n}L(X, \theta_0) > \frac{1}{n}L(X, \theta_0 - \delta) \\ \frac{1}{n}L(X, \theta_0) > \frac{1}{n}L(X, \theta_0 + \delta) \end{cases}$$

Посмотрим теперь на функцию $\theta \to L(X, \theta)$:



Ясно, что на интервале $(\theta_0 - \delta, \theta_0 + \delta)$ сущетвует точка, значение в которой строго больше, чем на концах, а значит, функция имеет на этом отрезке точку локального максимума. Т.е. для точки локального максимума $\hat{\theta}$ выполнено $|\hat{\theta} - \theta_0| < \delta$. Чтобы завершить доказательство, нужно обосновать фразу "при достаточно большом n, с вероятностью, близкой к 1...". Другими словами, хотим доказать, что

$$P\left(\frac{1}{n}L(X, \theta_0 - \delta) < \frac{1}{n}L(X, \theta_0)\right) \xrightarrow[n\to\infty]{} 1$$

Положим $W(\theta_0) - W(\theta_0 - \delta) = \varepsilon > 0$. Из ЗБЧ следует, что

$$\begin{cases} P\left(\left|\frac{1}{n}L(X, \theta_0 - \delta) - W(\theta_0 - \delta)\right| \geqslant \frac{\varepsilon}{4}\right) \to 0 \\ P\left(\left|\frac{1}{n}L(X, \theta_0) - W(\theta_0)\right| \geqslant \frac{\varepsilon}{4}\right) \to 0 \end{cases}$$

Поймём, почему из этого следует, что $P\left(\frac{1}{n}L(X, \theta_0 - \delta) \geqslant \frac{1}{n}L(X, \theta_0)\right) \to 0\,(*)$. Пусть величины $\frac{1}{n}L(X, \theta_0 - \delta)$ и $W(\theta_0 - \delta)$ отличаются менее, чем на $\frac{\varepsilon}{4}$. Аналогично для $\frac{1}{n}L(X, \theta_0)$ и $W(\theta_0)$. Тогда верна следующая цепочка равенств:

$$W(\theta_0) \leqslant \frac{1}{n}L_n(X, \theta_0) + \frac{\varepsilon}{4} \overset{(1)}{\leqslant} \frac{1}{n}L_n(X, \theta_0 - \delta) + \frac{\varepsilon}{4} \leqslant W(\theta_0 - \delta) + \frac{\varepsilon}{2}$$

Переход (1) следует из неравенства $(*)$. Тогда мы получаем, что $W(\theta_0) - W(\theta_0 - \delta) \leqslant \frac{\varepsilon}{2}$. Но $W(\theta_0) - W(\theta_0 - \delta) = \varepsilon$, и получается противоречие. Значит, или $\frac{1}{n}L(X, \theta_0 - \delta)$ и $W(\theta_0 - \delta)$ отличаются более, чем на $\frac{\varepsilon}{4}$, или же величины $\frac{1}{n}L(X, \theta_0)$ и $W(\theta_0)$. Но тогда мы получаем, что исход из события $\{\frac{1}{n}L(X, \theta_0 - \delta) \geqslant \frac{1}{n}L(X, \theta_0)\}$ лежит в объединении

$$\{\left|\frac{1}{n}L(X, \theta_0 - \delta) - W(\theta_0 - \delta)\right| \geqslant \frac{\varepsilon}{4}\} \cup \{\left|\frac{1}{n}L(X, \theta_0) - W(\theta_0)\right| \geqslant \frac{\varepsilon}{4}\}$$

А вероятность таких событий стремится к нулю. Теперь методом пристального взгляда можно заметить, что мы всё доказали. ■

Билет 3

Информация Фишера и неравенство Рао-Крамера. Критерий равенства в неравенстве Рао-Крамера.

Информация Фишера и неравенство Рао-Крамера

Определение. Информация Фишера $I(\theta) = \mathbb{E}_\theta \left(\dfrac{\partial}{\partial\theta} L(x,\theta) \right)^2$

Выведение и альтернативные варианты (здесь θ_0 — реальный параметр):

$$L(x,\theta) \overset{\text{ф-ла Тейлора}}{\simeq} L(x,\hat\theta_n(x)) + \frac{1}{2} \underbrace{\frac{\partial^2}{\partial\theta^2} L(x,\hat\theta_n(x))}_{\leqslant 0 \text{ тк точка максимума}} \left(\theta - \hat\theta_n(x)\right)^2$$

Точка максимума близка к параметру, посмотрим на вторую производную в реальном параметре:

$$-\mathbb{E}_{\theta_0} \frac{\partial^2}{\partial\theta^2} L(x,\theta_0) = -\mathbb{E}_{\theta_0} \frac{\partial}{\partial\theta} \left(\frac{\frac{\partial}{\partial\theta} P(x,\theta_0)}{P(x,\theta_0)} \right) = -\mathbb{E}_{\theta_0} \left(\frac{\frac{\partial^2}{\partial\theta^2} P(x,\theta_0)}{P(x,\theta_0)} - \frac{\left(\frac{\partial}{\partial\theta} P(x,\theta_0)\right)^2}{P(x,\theta_0)^2} \right) \ominus$$

$$\left[\text{трюки с производными так как } \left(\frac{a}{b}\right)' = \frac{a'}{b} - \frac{ab'}{b^2}, \text{ но здесь } a = b' \implies \left(\frac{b'}{b}\right)' = \frac{b''}{b} - \frac{(b')^2}{b^2} \right]$$

$$\ominus - \int\limits_{\mathbb{R}^n} \left(\frac{\frac{\partial^2}{\partial\theta^2} P(x,\theta_0)}{P(x,\theta_0)} - \frac{\left(\frac{\partial}{\partial\theta} P(x,\theta_0)\right)^2}{P(x,\theta_0)^2} \right) P(x,\theta_0)dx = -\int\limits_{\mathbb{R}^n} \frac{\partial^2}{\partial\theta^2} P(x,\theta_0)dx + \int\limits_{\mathbb{R}^n} \left(\frac{\frac{\partial}{\partial\theta} P(x,\theta_0)}{P(x,\theta_0)} \right)^2 P(x,\theta_0)dx$$

Предположим, что выполнены условия регулярности

- $P(x,\theta)$ дважды непрерывно дифференцируема по θ
- $P(x,\theta) > 0$ на каком-то множесте иксов (прямая, отрезок, точки в дискретном случае) $\forall\theta$
- Производную и интеграл можно переставить

Из этого следует

$$\int\limits_{\mathbb{R}^n} P(x,\theta)dx = 1 \implies \int\limits_{\mathbb{R}^n} \frac{\partial}{\partial\theta} P(x,\theta)dx = \frac{\partial 1}{\partial\theta} = 0, \qquad \int\limits_{\mathbb{R}^n} \frac{\partial^2}{\partial\theta^2} P(x,\theta)dx = 0$$

Итак

$$-\mathbb{E}_{\theta_0} \frac{\partial^2}{\partial\theta^2} L(x,\theta_0) = \underbrace{-\int\limits_{\mathbb{R}^n} \frac{\partial^2}{\partial\theta^2} P(x,\theta_0)dx}_{=0} + \int\limits_{\mathbb{R}^n} \left(\frac{\frac{\partial}{\partial\theta} P(x,\theta_0)}{P(x,\theta_0)} \right)^2 P(x,\theta_0)dx = \mathbb{E}_{\theta_0} \left(\frac{\partial}{\partial\theta} L(x,\theta_0) \right)^2 = I(\theta)$$

В дискретном случае меняем интегралы на суммы
Предположения про $P(x,\theta)$ очень натуральны, поэтому их никто не проверяет

Утверждение. Пусть выполнены условия регулярности. Тогда

$$I(\theta) = \mathbb{D}_\theta \left(\frac{\partial}{\partial\theta} L(x,\theta) \right) = n \cdot i(\theta), \text{ где } i(\theta) = \mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} \ln \rho_\theta(X_1) \right)^2.$$

Здесь $i(\theta)$ — информация Фишера выборки из одного элемента

Доказательство.

$$\mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} L(x,\theta) \right) = \int\limits_{\mathbb{R}^n} \frac{\frac{\partial}{\partial\theta} P(x,\theta)}{P(x,\theta)} P(x,\theta)dx = \int\limits_{\mathbb{R}^n} \frac{\partial}{\partial\theta} P(x,\theta)dx = 0 \implies \mathbb{D}_\theta \left(\frac{\partial}{\partial\theta} L(x,\theta) \right) = \mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} L(x,\theta) \right)^2$$

$$\frac{\partial}{\partial\theta} L(x,\theta) = \sum_i \underbrace{\frac{\partial}{\partial\theta} \ln \rho_\theta(X_i)}_{\substack{\text{независимые} \\ \text{одинаково распределенные}}} \implies \mathbb{D}_\theta \left(\frac{\partial}{\partial\theta} L(x,\theta) \right) = n\mathbb{D}_\theta \left(\frac{\partial}{\partial\theta} \ln \rho_\theta(X_1) \right) = n\mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} \ln \rho_\theta(X_1) \right)^2 = n \cdot i(\theta)$$

∎

Теорема. Неравенство Рао-Крамера Пусть выполняются условия регулярности, а $\theta_n(x)$ — несмещенная оценка функции $\tau(\theta)$ (как правило $\tau(\theta) = \theta$, но иногда мы пытаемся оценить не саму θ, а какую-то функцию θ), тогда

$$\mathbb{D}_\theta(\theta_n(x)) \geqslant \frac{(\tau'(\theta))^2}{I(\theta)}$$

Неравенство нужно чтобы находить эффективные оценки: там где достигается равенство, там и оценка эффективна.

Доказательство.

$$\tau(\theta) = \mathbb{E}_\theta(\theta_n(x)) = \int\limits_{\mathbb{R}_n} \theta_n(x) P(x,\theta) dx.$$

$$\tau'(\theta) = \int\limits_{\mathbb{R}^n} \theta_n(x) \frac{\partial}{\partial\theta} P(x,\theta) dx = \int\limits_{\mathbb{R}^n} \theta_n(x) \frac{\partial}{\partial\theta} P(x,\theta) dx - \tau(\theta) \overbrace{\int\limits_{\mathbb{R}^n} \frac{\partial}{\partial\theta} P(x,\theta) dx}^{=0} = \int\limits_{\mathbb{R}^n} (\theta_n(x) - \tau(\theta)) \frac{\partial}{\partial\theta} P(x,\theta) dx =$$

$$= \int\limits_{\mathbb{R}^n} (\theta_n(x) - \tau(\theta)) \underbrace{\frac{\frac{\partial}{\partial\theta} P(x,\theta)}{P(x,\theta)}}_{= \frac{\partial}{\partial\theta} \ln P(x,\theta) = \frac{\partial}{\partial\theta} L(x,\theta)} P(x,\theta) dx = \mathbb{E}_\theta \left((\theta_n(x) - \tau(\theta)) \frac{\partial}{\partial\theta} L(x,\theta) \right) \overset{\text{Коши-Буняковский}}{\overset{\downarrow}{\leqslant}}$$

$$\leqslant \sqrt{\mathbb{E}_\theta \left(\theta_n(x) - \tau(\theta)\right)^2} \sqrt{\mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} L(x,\theta) \right)^2} = \sqrt{\mathbb{D}_\theta \left(\theta_n(x)\right)} \sqrt{I(\theta)}$$

$$\tau'(\theta) \leqslant \sqrt{\mathbb{D}_\theta \left(\theta_n(x)\right)} \sqrt{I(\theta)} \implies (\tau'(\theta))^2 \leqslant \mathbb{D}_\theta \left(\theta_n(x)\right) I(\theta).$$

■

Критерий равенства в неравенстве Рао-Крамера

Равенство достигается когда достигается равенство в Коши-Буняковском, то есть

$$\theta_n(x) - \tau(\theta) = \underbrace{c_n(\theta)}_{\text{const}} \frac{\partial}{\partial\theta} L(x,\theta)$$

Пример в ситуации бернулли:

$$\frac{\partial}{\partial\theta} L(x,\theta) = \frac{\sum_i X_i - n\theta}{\theta(1-\theta)} = \underbrace{\frac{n}{\theta(1-\theta)}}_{\text{const}} (\overline{X}_n - \theta) \implies \text{ в Рао-Крамере достигается равенство } \implies \bar{X}_n \text{ эфф оценка}$$

Если (в случае оценки θ, то есть $\tau(\theta) = \theta$) существует несмещенная оценка $\hat{\theta}_n$, на которой достигается равенство в Рао-Крамере, то это оценка максимального правдоподобия

$$\hat{\theta}_n(x) - \theta = c_n(\theta) \frac{\partial}{\partial\theta} L(x,\theta) \quad \text{Возьмем } \theta = \theta^*(x) \text{ оценка макс правдоподобия} \implies \frac{\partial}{\partial\theta} L(x,\theta^*(x)) = 0 \implies \hat{\theta}_n(x) = \theta^*(x)$$

Билет 4

Доверительные интервалы. Различимые методы построения доверительных интервалов (с помощью неравенств на вероятность больших уклонений, с помощью центральной статистики, с помощью асимптотически нормальной оценки). Примеры.

Доверительные интервалы

Знать, что оценка $\hat{\theta}_n(X)$ состоятельна (сходится по вероятности к θ) это, конечно, круто, но особо много информации о ней нам не даёт. Нам хотелось бы знать как быстро она куда-то там сходится – хотим для фиксированного $\alpha \in (0,1)$ и фиксированного $\varepsilon > 0$ знать такой номер n, что $P_\theta(|\hat{\theta}_n(X) - \theta| < \varepsilon) > 1 - \alpha$.

Определение. $(\hat{\theta}_1(X), \hat{\theta}_2(X))$ — доверительный интервал уровня доверия $1-\alpha$, если

$$P_\theta(\theta \in (\hat{\theta}_1(X), \hat{\theta}_2(X))) \geqslant 1 - \alpha$$

$$P_\theta(\hat{\theta}_1(X) \leqslant \theta \leqslant \hat{\theta}_2(X)) \geqslant 1 - \alpha$$

Определение. Последовательность оценок $\hat{\theta}_1^n(X), \hat{\theta}_2^n(X)$ образует асимтотический доверительный интервал, если $\liminf\limits_{n\to\infty} P_\theta(\hat{\theta}_1^n(X) \leqslant \theta \leqslant \hat{\theta}_2^n(X)) \geqslant 1 - \alpha$

Пример. Пусть есть выборка из случайных величин с нормальным распределением $X_j \sim \mathcal{N}(\theta, 1)$.

Знаем, что $\overline{X}_n = \dfrac{X_1 + \dots + X_n}{n} \xrightarrow{P_\theta} \theta$ (ЗБЧ) — среднее хорошо приближает θ.

Посмотрим на разность эмпирического среднего и реальной θ: $\dfrac{X_1 + \dots + X_n}{n} - \theta = \dfrac{\overbrace{(X_1 - \theta)}^{\sim\mathcal{N}(0,1)} + \dots + \overbrace{(X_n - \theta)}^{\sim\mathcal{N}(0,1)}}{n} \sim \mathcal{N}(0, \dfrac{1}{n})$

$\Longrightarrow \sqrt{n}(\dfrac{X_1 + \dots + X_n}{n} - \theta) \sim \mathcal{N}(0,1)$

Теперь по таблице значений функции распределения нормального закона найдём квантили $z_{\frac{\alpha}{2}}$ и $z_{1-\frac{\alpha}{2}}$: $\Phi(z_{\frac{\alpha}{2}}) = \dfrac{\alpha}{2}, \Phi(z_{1-\frac{\alpha}{2}}) = 1 - \dfrac{\alpha}{2}$.
$\Longrightarrow$

$$P_\theta(z_{\frac{\alpha}{2}} \leqslant \sqrt{n}(\frac{X_1 + \dots + X_n}{n} - \theta) \leqslant z_{1-\frac{\alpha}{2}}) = \Phi(z_{1-\frac{\alpha}{2}}) - \Phi(z_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$$
$$P(\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}} - \overline{X_n} \leqslant -\theta \leqslant \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} - \overline{X_n}) = 1 - \alpha$$
$$P(\overline{X_n} - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leqslant \theta \leqslant \overline{X_n} - \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}) = 1 - \alpha$$

Заметим, что мы взяли симметричный интервал: $z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$. В таком случае наш интервал принимает вид: $(\overline{X_n} - \dfrac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \overline{X_n} + \dfrac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}})$. В таком случае длина этого интервала равна $O(\dfrac{1}{\sqrt{n}})$

Но зачем мы решили взять симметричный интервал? Вспомним, что мы от него хотим: минимальной длины. А какой интервал на графике нормального распределения будет захватывать нужную площадь и при этом быть самым коротким среди всех? Правильно, симметричный с центром в пике колокола нормального распределения.

Различимые методы построения доверительных интервалов (с помощью неравенств на вероятность больших уклонений, с помощью цен-тральной статистики, с помощью асимптотически нормальной оценки). Примеры

1. Неравенства Чебышёва или Чернова

   $X_1, \dots, X_n \sim \mathrm{Bern}(\theta), P(X_j = 1) = \theta$

   Чебышёв:

   $P_\theta(|\overline{X_n} - \theta| \geqslant \varepsilon) \leqslant \dfrac{\mathbb{D}X_1}{n\varepsilon^2} = \dfrac{\theta(1-\theta)}{n\varepsilon^2} \leqslant \dfrac{1}{4n\varepsilon^2} = \alpha \Longrightarrow \varepsilon = \dfrac{1}{\sqrt{2n\alpha}} \Longrightarrow P_\theta(\overline{X_n} - \dfrac{1}{\sqrt{2n\alpha}} < \theta < \overline{X_n} + \dfrac{1}{\sqrt{2n\alpha}}) \geqslant 1 - \alpha.$

   Чернов: $P_\theta(|\overline{X_n} - \theta| \geqslant \varepsilon) \leqslant 2e^{-\frac{n\varepsilon^2}{4}} = \alpha$
   $-\dfrac{n\varepsilon^2}{4} = \ln\dfrac{\alpha}{2}$

$\varepsilon = 2\sqrt{-\frac{\ln\frac{\alpha}{2}}{n}}$

$\implies P_\theta(\overline{X_n} - 2\sqrt{-\frac{\ln\frac{\alpha}{2}}{n}} < \theta < \overline{X_n} + 2\sqrt{-\frac{\ln\frac{\alpha}{2}}{n}}) \geqslant 1-\alpha$

Заметим, что в обоих оценках мы получили, что длина интервала равна $\mathrm{O}(\frac{1}{\sqrt{n}})$, но несложно заметить, что константа Чернова значительно лучше, чем у Чебышёва.

2. Метод центральной статистики

   Определение. $\mathrm{V}(X,\theta)$ называется центральной статистикой, если:

   (a) её распределение не зависит от θ: $P_\theta(\mathrm{V}(X,\theta) \leqslant t) = F(t)$

   (b) $\forall X : \theta \mapsto \mathrm{V}(X,\theta)$ — монотонная

   Пусть у нас есть такая статистика. Вопрос: как с её помощью строить доверительные интервалы? Предельно просто: подберём числа t_1 и t_2 таким образом, чтобы $P_\theta(t_1 \leqslant \mathrm{V}(X,\theta) \leqslant t_2) \geqslant 1-\alpha$. Мы можем так сделать, потому что распределение V не зависит от θ. Теперь поскольку при любом X наша функция монотонна, то данная оценка равносильна тому, что $P_\theta(\hat\theta_1(X) \leqslant \theta \leqslant \hat\theta_2(X)) \geqslant 1-\alpha$ — чисто из-за монотонности по θ.

   Пример. $X_j \sim \mathcal{U}(0,\theta) \implies \theta^{-1}X_j \sim \mathcal{U}(0,1)$. Это уже центральная статистика, однако она зависит всего от одного элемента выборки. Рассмотрим $X_{(n)} = \max\limits_{1\leqslant j\leqslant n} X_j$: $P_\theta(\theta^{-1}X_{(n)} \leqslant t) = P_\theta(\max\limits_{1\leqslant j\leqslant n}\theta^{-1}X_j \leqslant t) = \prod\limits_{j=1}^{n} P_\theta(\underbrace{\theta^{-1}X_j}_{\sim\mathcal{U}(0,1)} \leqslant t) = t^n$

   Теперь грубо попробуем оценить, куда там наша статистика попадает:

   $P_\theta(\underbrace{t}_{t_1} \leqslant \theta^{-1}X_{(n)} \leqslant \underbrace{1}_{t_2}) = 1 - t^n = 1-\alpha \implies t = \alpha^{\frac{1}{n}}$

   Теперь попробуем вытащить отсюда θ:

   $P_\theta(\alpha^{\frac{1}{n}} \leqslant \theta^{-1}X_{(n)} \leqslant 1) = 1-\alpha$

   $P_\theta(\frac{\alpha^{\frac{1}{n}}}{X_{(n)}} \leqslant \theta^{-1} \leqslant \frac{1}{X_{(n)}}) = 1-\alpha$

   $P_\theta(\underbrace{X_{(n)}}_{\hat\theta_1(X)} \leqslant \theta \leqslant \underbrace{\frac{X_{(n)}}{\alpha^{\frac{1}{n}}}}_{\hat\theta_2(X)}) = 1-\alpha$

   Теперь посмотрим на длину полученного доверительного интервала:

   $(\alpha^{-\frac{1}{n}} - 1)X_{(n)}$

   Что мы можем сказать про $\alpha^{-\frac{1}{n}} - 1$? Разложим это дело по Тейлору:

   $\alpha^{-\frac{1}{n}} - 1 \sim e^{-\frac{\ln\alpha}{n}} - 1 \sim \frac{-\ln\alpha}{n} = \underline{\mathrm{O}}(\frac{1}{n}) \to \infty$

   Получается длина доверительного интервала с ростом количества элементов выборки стремится к нулю. Получается мы построили что-то более менее разумное.

   Часто в роли центральной статистики можно взять следующую лабуду: $\mathrm{V}(X,\theta) = -\sum\limits_{j=1}^{n}\ln F_\theta(X_j)$ — это сумма независимых распределений, поэтому достаточно показать что одно не зависит от θ — тогда в силу независимости сумма тоже будет не зависеть от θ:

   $P_\theta(-\ln F_\theta(X_j) \leqslant) = P_\theta(F_\theta(X_j) \geqslant e^{-t}) = P_\theta(X_j \geqslant F_\theta^{-1}(e^{-t})) = 1 - F_\theta(F_\theta^{-1}(e^{-t})) = 1 - e^{-t}$, а это экспоненциальное распределение. Сумма экспоненциальных распределений это Гамма распределение $\implies \mathrm{V}(X,\theta) = \Gamma(n,1)$

3. Построение асимптотических доверительных интервалов

   Пусть у нас есть $\hat\theta_n(X)$ — асимптотически нормальная оценка θ с асимптотической дисперсией $\sigma^2(\theta)$. Это значит, что

   $\frac{\sqrt{n}(\hat\theta_n(X) - \theta)}{\sigma(\theta)} \xrightarrow{d_\theta} Z \sim \mathcal{N}(0,1)$

   Теперь мы хотим получить доверительный интервал. Если бы у нас $\sigma(\theta)$ была константой, то мы могли бы уже привычно взять там квантили нормального распределения, туды сюды и получить интервал:

   $P_\theta(t_1 \leqslant \frac{\sqrt{n}(\hat\theta_n(X) - \theta)}{\sigma(\theta)} \leqslant t_2) \to \Phi(t_2) - \Phi(t_1) = 1-\alpha.$

Тогда мы могли бы просто взять такие $\Phi(t_2) = 1 - \dfrac{\alpha}{2}$ и $\Phi(t_1) = \dfrac{\alpha}{2}$ и получить, что

$P_\theta(\hat{\theta}_n(X) - \dfrac{t_{1-\frac{\alpha}{2}}\sigma(\theta)}{\sqrt{n}} \leqslant \theta \leqslant \hat{\theta}_n(X) + \dfrac{t_{1-\frac{\alpha}{2}}\sigma(\theta)}{\sqrt{n}}) \to 1 - \alpha$

Но тут есть проблема — у нас слева и справа есть $\sigma(\theta)$ в числителе, что совершенно ломает корректность статистики, мы ведь хотим чтобы штуки слева и спрва от θ в неравенстве не зависили от θ. Как это решать? Очень просто, перейти от $\sigma(\theta)$ к состоятельной оценке $\sigma(\theta)$. Возможны следующие случаи:

(a) σ — непрерывная функция

Тогда $\sigma(\hat{\theta}_n(X)) \xrightarrow{P_\theta} \sigma(\theta)$ и мы можем везде в наших рассуждениях заменить $\sigma(\theta)$ на $\sigma(\hat{\theta}_n(X))$ и сходимость сохранится:

$P_\theta(\hat{\theta}_n(X) - \dfrac{t_{1-\frac{\alpha}{2}}\sigma(\hat{\theta}_n(X))}{\sqrt{n}} \leqslant \theta \leqslant \hat{\theta}_n(X) + \dfrac{t_{1-\frac{\alpha}{2}}\sigma(\hat{\theta}_n(X))}{\sqrt{n}}) \to 1 - \alpha$

(b) Изначально было ЦПТ

$\hat{\theta}_n(X) = \overline{X_n}, \sigma^2(\theta) = \mathbb{D}_\theta X$

В таком случае мы можем использовать выборочную дисперсию в качестве состоятельной оценки дисперсии:

$s^2 = \dfrac{1}{n-1}\sum\limits_{j=1}^{n}(X_j - \overline{X_n})^2$ — выборочная дисперсия

(c) Можно поправить нашу асимптотическую дисперсию: подобрать такую функцию φ, что

$\sqrt{n}(\varphi(\hat{\theta}_n(X)) - \varphi(\theta)) \to \underbrace{\mathcal{N}(0,1)}_{=\varphi'(\theta)\cdot\mathcal{N}(0,\sigma^2(\theta))} \implies \varphi'^2(\theta)\sigma^2(\theta) = 1$

Билет 5

Построение точных доверительных интервалов для параметров нормального распределения.

Матожидание.

Оцениваем параметры случайной величины $\sim \mathcal{N}(a, \sigma^2)$ по выборке $X_1, \ldots, X_n$:

1. σ — известно. Оцениваем матожидание a:

   Центрируем и нормируем случайную величину разности оценки и параметра:

   $$\sqrt{n}\,\frac{\overline{X_n} - a}{\sigma} \sim \mathcal{N}(0,1) \text{ — центральная статистика}$$

   Пусть для $t_{1-\frac{\alpha}{2}}$ верно, что $\Phi\left(t_{1-\frac{\alpha}{2}}\right) = 1 - \dfrac{\alpha}{2}$, тогда

   $$P_\theta\left(-t_{1-\frac{\alpha}{2}} \le \frac{\sqrt{n}(\overline{X_n} - a)}{\sigma} \le t_{1-\frac{\alpha}{2}}\right) = 1-\alpha \Leftrightarrow P_\theta\left(\underbrace{\overline{X_n} - \frac{t_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}}_{\hat{\Theta}_1(x)} \le a \le \underbrace{\overline{X_n} + \frac{t_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}}_{\hat{\Theta}_2(x)}\right) = 1-\alpha$$

2. σ не известно.

   Лемма. Пусть $X = (X_1, \ldots, X_n) \sim \mathcal{N}\left(a, \operatorname{diag}(\sigma^2)\right)$, $\operatorname{diag}(\sigma^2) = \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix}$, $\{X_j\}$ независимы, $\mathbb{D}X_j = \sigma^2$, $\mathbb{E}X_j = a$.

   Тогда $\overline{X_n}$ и $\sum\limits_{j=1}^{n}(X_j - \overline{X_n})^2$ независимы.

   Доказательство. Не ограничивая общности, считаем, что $\mathbb{E}X_j = a = 0$.

   Рассмотрим $U = \begin{pmatrix} \frac{1}{\sqrt{n}} & \ldots & \frac{1}{\sqrt{n}} \\ \star & \ldots & \star \\ \star & \ldots & \star \end{pmatrix}$ — ортогональную, и случайную величину $u = UX \sim \mathcal{N}(0, \operatorname{diag}(\sigma^2))$,

   $$\left(U^* C_x U = U^* \operatorname{diag}(\sigma^2) U = \sigma^2 U^* I U = \sigma^2 U^* U = \sigma^2 I = \operatorname{diag}(\sigma^2)\right)$$

   $\{u_\sigma\}$ — незав., $\mathbb{E}u_j = 0$, $\mathbb{D}u_j = \sigma^2$.

   Далее $|u|$ — вторая векторная норма, т.е. $|u| = \sqrt{u_1^2 + \ldots + u_n^2}$. К слову, вторая норма — унитарно-инвариантна, поэтому $|u| = |UX| = |X|$ в силу ортогональности U.

   $$\sum_{j=2}^{n} u_j^2 = |u|^2 - u_1^2 = |u|^2 - \left(\frac{\sum X_j}{\sqrt{n}}\right)^2 = |X|^2 - n\overline{X_n}^2 \text{ (т.к. } U \text{ — орт.)}$$

   $$\sum_{j=1}^{n}(X_j - \overline{X_n})^2 = \sum_{j=1}^{n} X_j^2 - 2n\overline{X_n}\frac{1}{n}\sum_{j=1}^{n} X_j + n\overline{X_n}^2 = \sum_{j=1}^{n} X_j^2 - n\overline{X_n}^2$$

   $$\sum_{j=2}^{n} u_j^2 = \sum_{j=1}^{n}(X_j - \overline{X_n})^2 \text{ — независимо с } u_1 = \frac{X_1 + \ldots + X_n}{\sqrt{n}} = \sqrt{n}\,\overline{X_n}$$

   При этом $\sum\limits_{j=2}^{n} u_j^2 = \sigma^2 \sum\limits_{j=2}^{n}(\sigma^{-1}u_j)^2$, и $\sigma^{-1}u_j \sim \mathcal{N}(0,1)$, то есть

   $$\chi_n^2 = \sum_{j=1}^{n}(X_j - \overline{X_n}s)^2 = \sum_{j=1}^{n}\xi_k^2,\ \{\xi_k\} \text{ — незав.}, \xi_k \sim \mathcal{N}(0,1),$$

   где χ_n^2 — распределение χ-квадрат с n степенями свободы, распределение величины $\sum\limits_{j=1}^{n}\eta_j^2$, η_j — независимые нормально распределенные с параметрами 0 и 1 величины.

   В итоге $\overline{X_n}$ и $s^2 = \dfrac{1}{n-1}\sum\limits_{j=1}^{n}(x_j - \overline{X_n})^2$ независимы и $\sigma^{-2}s^2 \sim \dfrac{1}{n-1}\,\chi_{n-1^2}$. ■

Так как σ неизвестна, заменим ее на $\sqrt{s^2}$ и получим статистику:

$$T_{n-1}(X) = \frac{\sqrt{n}(X_n - a)}{\sqrt{s^2}} = \frac{\frac{\sqrt{n}(X_n - a)}{\sigma}}{\sqrt{\sigma^2 s^2}} \sim \frac{\xi}{\sqrt{\frac{1}{n-1}\,\chi^2_{n-1}}},$$

ξ и χ независимы.

$T_n = \dfrac{\xi}{\sqrt{\frac{1}{n}\,\chi^2_n}}$ — распределение Стьюдента с $(n-1)$ степенями свободы.

Его плотность:

$$\varrho(x) = C_n \left(1 + \frac{x^2}{n-1}\right)^{-n/2}$$

Т.к. плотность симметрична, можем выбрать 1 квантиль:

$$F_{T_{n-1}}(t_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2},\; F_{T_{n-1}}(-t_{1-\frac{\alpha}{2}}) = \frac{\alpha}{2}$$

Попадаем в случай 1 с известной дисперсией:

$$P_\theta\left(-t_{1-\frac{\alpha}{2}} \le \frac{\sqrt{n}(\overline{X_n} - a)}{\sqrt{s^2}} \le t_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \;\Leftrightarrow\; P_\theta\left(\underbrace{\overline{X_n} - \frac{t_{1-\frac{\alpha}{2}}\sqrt{s^2}}{\sqrt{n}}}_{\hat{\Theta}_1(x)} \le a \le \underbrace{\overline{X_n} + \frac{t_{1-\frac{\alpha}{2}}\sqrt{s^2}}{\sqrt{n}}}_{\hat{\Theta}_2(x)}\right) = 1 - \alpha$$

Дисперсия.

Из доказательства леммы выше:

$$\sigma^{-2}(n-1)S_n^2 \sim \chi^2_{n-1}$$

Выберем $x_{\alpha/2}$ и $x_{1-\alpha/2}$ такое, что $F_{\chi^2_{n-1}}(x_{\alpha/2}) = \alpha/2$ и $F_{\chi^2_{n-1}}(1 - x_{\alpha/2}) = 1 - \alpha/2$ ($F_{\chi^2_{n-1}}$ — функция распределения случайной величины χ_{n-1^2}). Тогда

$$P\left(x_{\alpha/2} \le \frac{(n-1)S_n^2}{\sigma^2} \le x_{1-\alpha/2}\right) = 1 - \alpha,$$

и интервал уровня $1 - \alpha$ имеет вид

$$\left(\frac{\sqrt{n-1}\,S_n}{\sqrt{x_{1-\alpha/2}}},\; \frac{\sqrt{n-1}\,S_n}{\sqrt{x_{\alpha/2}}}\right).$$

Билет 6

Проверка гипотез. Ошибки 1-го и 2-го рода. Уровень значимости и мощность статистического критерия. Пример построения критерия с помощью доверительного интеграла. Нижняя оценка суммы вероятностей ошибок 1-го и 2-го рода.

Проверка гипотез

Пускай есть выборка $X_1,\ldots,X_n$ с распределением P_θ.

Определение. Предположения о значениях θ и называются статистическими гипотезами.

Пример. $H_0\colon \theta \in \Theta_0 -$ статистическая гипотеза

Определение. Простая гипотеза (одноточечная гипотеза) $-$ гипотеза вида $H_0\colon \theta = \Theta_0$, где $\Theta = \{\theta_0\}$

Определение. Гипотеза $H_1\colon \theta = \Theta_1 -$ альтернативная гипотеза

Пример. $H_1\colon \theta = \overline{\Theta_0} -$ альтернативная гипотеза

Для проверки гипотез, строят критерий на основе критического множества как правило $\subset \mathbb{R}^n$, то есть действуют по такому принципу:

Выделяют в области значения параметров критического множества K, так, что $\forall\theta \in \Theta_0$, $P_\theta((X_1\ldots X_n) \in K)$ — «маленькая», тогда $X_1\ldots X_n \in K$ свидетельствует против гипотезы H_0, то есть если $X = (X_1\ldots X_n) \in K \Rightarrow H_0$ отклоняется, иначе принимается.

Далее $X = (X_1\ldots X_n)$

Ошибки 1-го и 2-го рода

Пусть у нас есть критическое множество K. При проверке гипотез мы могли совершить две ошибки:

Определение. Ошибка первого рода: отклонение верной гипотезы H_0, то есть это $\underset{\theta\in\Theta_0}{P_\theta}(X \in K)$. В случае простой гипотезы $P_{\theta_0}(X \in K)$

Определение. Ошибка второго рода: принятие ложной гипотезы H_0, то есть это $\underset{\theta\in\Theta_1}{P_\theta}(X \notin K)$. В случае простой гипотезы $P_{\theta_1}(X \notin K)$

Уровень значимости и мощность статистического критерия

Определение. Критерий K имеет уровень значимости α, если вероятность ошибки первого рода меньше либо равна α, то есть $\underset{\theta\in\Theta_0}{P_\theta}(X \in K) \leqslant \alpha$.

Определение. Мощность критерия K это величина, равная 1 - вероятность ошибки второго рода, то есть $1 - \underset{\theta\in\Theta_1}{P_\theta}(X \notin K) = \underset{\theta\in\Theta_1}{P_\theta}(X \in K)$. В случае простой гипотезы величина $\beta = P_{\theta_1}(X \in 1) -$ мощность

Если имеются два критерия K,S уровня значимости α, то K более мощный, чем S если

$$\forall\theta \in \Theta_1\colon P_\theta(X \in K) \geqslant P_\theta(X \in S)$$

Пример построения критерия с помощью доверительного интеграла

Пример. Пускай $X_1,\ldots,X_n \sim \mathcal{N}(\theta,1)$
$H_0 : \theta = \theta_0$
$H_1 : \theta = \theta_1$

Ранее при данных условиях мы получили следующий доверительный интервал:

$$P_{\theta_0}\left(\overline{X_n} - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leqslant \theta_0 \leqslant \overline{X_n} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) = P_{\theta_0}\left(\theta_0 - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leqslant \overline{X_n} \leqslant \theta_0 + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) = 1-\alpha$$

Выберем критическое множество $K\colon \left\{X : \overline{X_n} > \theta_0 + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right\} \cup \left\{X : \overline{X_n} < \theta_0 - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right\}$

Тогда $P_{\theta_0}(X \in K) = 1 - (1-\alpha) = \alpha$

Найдем ошибку второго рода:

$$P_{\theta_1}(X \notin K) = P_{\theta_1}\left(\theta_0 - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \leqslant \overline{X_n} \leqslant \theta_0 + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\right) =$$

$$= P_{\theta_1}\left(\sqrt{n}\,(\theta_0 - \theta_1) - Z_{1-\frac{\alpha}{2}} \leqslant \underbrace{\sqrt{n}\left(\overline{X_n} - \theta_1\right)}_{\sim\mathcal{N}(0,1)} \leqslant \sqrt{n}\,(\theta_0 - \theta_1) + Z_{1-\frac{\alpha}{2}}\right) =$$

$$= \Phi\left(\sqrt{n}\,(\theta_0 - \theta_1) + Z_{1-\frac{\alpha}{2}}\right) - \Phi\left(\sqrt{n}\,(\theta_0 - \theta_1) - Z_{1-\frac{\alpha}{2}}\right)$$

Посмотрим, что происходит при $n \to \infty$:

1. $\theta_0 > \theta_1 \implies P_{\theta_1}(X \notin K) \to 0$
2. $\theta_0 < \theta_1 \implies P_{\theta_1}(X \notin K) \to 0$

Таким образом, мы получили состоятельный критерий.

Нижняя оценка суммы вероятностей ошибок 1-го и 2-го рода

Теорема. Пусть у нас есть две гипотезы:

1. $H_0 : \rho = f_0$
2. $H_1 : \rho = f_1$

Сумма ошибки первого рода и ошибки второго рода больше либо равна $1 - \frac{1}{2}\int\limits_{\mathbb{R}^n} |f_0 - f_1| dx$

Доказательство. Найдем сумму ошибок первого и второго рода:

$$P_0(X \in K) + P_1(X \notin K) = \int\limits_K f_0 dx + \underbrace{\int\limits_{\mathbb{R}^n \setminus K} f_1 dx}_{1 - \int\limits_K f_1 dx} = 1 + \int\limits_K (f_0 - f_1) dx \geqslant$$

Введем множество $S = \{f_0 \leqslant f_1\}$

$$\geqslant 1 + \int\limits_{K \cap S} (f_0 - f_1) dx \geqslant 1 + \int\limits_S (f_0 - f_1) dx$$

Рассмотрим отдельно интеграл $\int\limits_S (f_0 - f_1) dx$:

$$\int\limits_S (f_0 - f_1) dx = \int\limits_S f_0 dx - \int\limits_S f_1 dx = 1 - \int\limits_{\mathbb{R}^n \setminus S} f_0 \; dx - 1 + \int\limits_{\mathbb{R}^n \setminus S} f_1 dx = \int\limits_{\mathbb{R}^n \setminus S} (f_1 - f_0) dx$$

В силу того, как мы выбрали множество S, можно увидеть, что

1. $\int\limits_S (f_0 - f_1) dx = -\int\limits_S |f_0 - f_1| dx$
2. $\int\limits_{\mathbb{R}^n \setminus S} (f_1 - f_0) dx = -\int\limits_{\mathbb{R}^n \setminus S} |f_0 - f_1| dx$

Тогда мы получаем, что

$$-\int\limits_{\mathbb{R}^n \setminus S} |f_0 - f_1| dx = -\int\limits_S |f_0 - f_1| dx = -\frac{1}{2}\int\limits_{\mathbb{R}^n} |f_0 - f_1| dx \implies P_0(X \in K) + P_1(X \notin K) \geqslant 1 - \frac{1}{2}\int\limits_{\mathbb{R}^n} |f_0 - f_1| dx$$

■

Билет 7

Теорема Неймана-Пирсона и пример её применения.

Теорема Неймана-Пирсона

Пусть гипотеза H_0 утверждает, что плотность выборки – это f_0, а альтернативная гипотеза H_1 утверждает, что плотность выборки – это f_1.
Предположим, что $\forall \alpha \in [0,1]\ \exists t := t(\alpha) : P_0(f_1(x) \geqslant t f_0(x)) = \alpha$.

Теорема (Неймана-Пирсона)**.** В такой постановке наиболее мощный критерий уровня значимости α имеет вид $K_{t(\alpha)} := \{f_1(x) \geqslant t(\alpha) f_0(x)\}$.

Доказательство. Пусть S – тоже критерий уровня значимости α: $P_0(X \in S) \leqslant \alpha = P_0(X \in K_{t(\alpha)})$. Хотим сравнить $P_1(X \in K_{t(\alpha)}) - P_1(X \in S)$. Хотим, чтобы это было больше либо равно нуля. Это и будет означать, что у нас критерий наиболее мощный.
$P_1(X \in K_{t(\alpha)}) - P_1(X \in S) = \int\limits_{K_{t(\alpha)}} f_1 dx - \int\limits_S f_1 dx =$ [можем выкинуть пересечение, так как на пересечении эти интегралы просто сократятся] $= \int\limits_{K_{t(\alpha)} \setminus S} f_1 dx - \int\limits_{S \setminus K_{t(\alpha)}} f_1 dx$.
Заметим, что на $S \setminus K_{t(\alpha)}$ выполнено $f_1 < t(\alpha) f_0$, так как это взято из дополнения к $K_{t(\alpha)}$, где по условию выполняется $f_1(x) \geqslant t(\alpha) f_0(x)$. Поэтому имеем: $\int\limits_{K_{t(\alpha)} \setminus S} f_1 dx - \int\limits_{S \setminus K_{t(\alpha)}} f_1 dx \geqslant t(\alpha) \int\limits_{K_{t(\alpha)} \setminus S} f_0 dx - t(\alpha) \int\limits_{S \setminus K_{t(\alpha)}} f_0 dx =$ [снова добавим пересечение и вынесем $t(\alpha)$] $= t(\alpha) \cdot (\int\limits_{K_{t(\alpha)}} f_0 dx - \int\limits_S f_0 dx) = t(\alpha) \cdot (P_0(X \in K_{t(\alpha)}) - P_0(X \in S)) \geqslant 0$ из построения критерия S ($P_0(X \in S) \leqslant \alpha = P_0(X \in K_{t(\alpha)})$).

Получили: $P_1(X \in K_{t(\alpha)}) - P_1(X \in S) \geqslant 0$, что и требовалось доказать. ■

Пример применения теоремы Неймана-Пирсона

Пример. Пусть у нас выборка из нормального закона $N(\theta, 1)$. Пусть наша гипотеза H_0 говорит, что $\theta = \theta_0$, а альтернативная гипотеза H_1 говорит, что $\theta = \theta_1 > \theta_0$.

$$f_1(X) = \frac{1}{\sqrt{2\pi}^n} \cdot \exp\left(-\frac{1}{2}\sum_{j=1}^n (X_j - \theta_1)^2\right)$$
$$f_0(X) = \frac{1}{\sqrt{2\pi}^n} \cdot \exp\left(-\frac{1}{2}\sum_{j=1}^n (X_j - \theta_0)^2\right)$$

Зададим критерий K_t из теоремы Неймана-Пирсона (ничего в 0 не обращается – сразу можем поделить):
$K_t = \left\{\frac{f_1}{f_0} \geqslant t\right\} = \left\{exp\left(\frac{1}{2}\sum_{j=1}^n [(X_j - \theta_0)^2 - (X_j - \theta_1)^2]\right) \geqslant t\right\} =$ [логарифмируем, расскрываем скобки, умножаем на два] $= \left\{\sum_{j=1}^n [2X_j(\theta_1 - \theta_0) + n(\theta_0^2 + \theta_1^2)] \geqslant 2\ln t\right\} = \left\{(\theta_1 - \theta_0)\overline{X_n} \geqslant \frac{\ln t}{n} - \frac{(\theta_0^2 + \theta_1^2)}{2}\right\} =$ [по условию $\theta_1 > \theta_0 \Rightarrow$ поделим] $= \left\{\overline{X_n} \geqslant \frac{\frac{\ln t}{n} - \frac{(\theta_0^2 + \theta_1^2)}{2}}{\theta_1 - \theta_0}\right\}$

Таким образом пришли к тому, что $K_t = \left\{\frac{f_1}{f_0} \geqslant t\right\}$ равносильно множеству $\widetilde{K}_s = \{\overline{X_n} \geqslant s\}$. Равносильно в том смысле, что для каждого t мы можем подобрать $s(t)$, что множество K_t совпадает с $\widetilde{K}_{s(t)}$. Теперь будем искать критические множества именно в таком виде (для удобства).

Должно выполняться: $P_0(X \in K_t) = \alpha \Leftrightarrow P_0(X \in \widetilde{K}_{s(t)}) = \alpha$. А что это за вероятности? Это вероятность $P_{\theta_0}(\overline{X_n} \geqslant s) = \alpha$
То есть, $P_{\theta_0}(\sqrt{n}(\overline{X_n} - \theta_0) \geqslant \sqrt{n}(s - \theta_0)) = \alpha$, где $\sqrt{n}(\overline{X_n} - \theta_0) \sim N(0,1)$, поэтому тут просто написано, что $1 - \Phi(\sqrt{n}(s - \theta_0)) = \alpha$.

Значит, выбираем квантиль нормального закона уровня $1-\alpha$: $Z_{1-\alpha} = \sqrt{n}(s-\theta_0) \Rightarrow$
$s = \theta_0 + \dfrac{Z_{1-\alpha}}{\sqrt{n}}$. Выразили s.

Таким образом, наше критическое множество $\left\{\overline{X_n} \geqslant \theta_0 + \dfrac{Z_{1-\alpha}}{\sqrt{n}}\right\}$. Это критерий уровня значимости α.

Теперь посчитаем мощность (это же самый мощный критерий):
$P_{\theta_1}\left(\overline{X_n} \geqslant \theta_0 + \dfrac{Z_{1-\alpha}}{\sqrt{n}}\right) = P_{\theta_1}\big(\sqrt{n}(\overline{X_n}-\theta_1) \geqslant \sqrt{n}(\theta_0-\theta_1) + Z_{1-\alpha}\big) = 1 - \Phi(\sqrt{n}(\theta_0-\theta_1) + Z_{1-\alpha})$.

Заметим, что если объём выборки n устремить к бесконечности, то точка, в которой мы берём Φ стремится к минус бесконечности (так как $(\theta_0-\theta_1) < 0$ по условию), поэтому мощность стремится к 1.

По теореме Неймана-Пирсона выписанная мощность максимальна.

Билет 8

Эмпирическая функция распределения. Теорема Гливенко-Кантелли.

Эмпирическая функция распределения

Говоря об эмпирическом распредлениии и эмпирической функции распределения, мы в качестве параметра θ рассматриваем как бы само распределение, $''\theta = P''$.

Определение. Пусть $X_1, \ldots, X_n$ - выборка из распредления с функцией распределения F. Эмпирическим распределением будем называть $P_n^*(B) = \dfrac{\#\{X_j \in B\}}{n} = \dfrac{1}{n}\sum\limits_{j=1}^{n} I_{\{X_j \in B\}}$, где B - множество, $\#$ - обозначает $''$количество$''$.

Сразу заметим, что:

1. $\mathbb{E}\big[P_n^*(B)\big] = \mathbb{E}\big[I_{X_1 \in B}\big] = P(X_1 \in B)$
2. По УЗБЧ $P_n^*(B) = \dfrac{1}{n}\sum\limits_{j=1}^{n} I_{\{X_j \in B\}} \xrightarrow{\text{п.н.}} \mathbb{E}\big[I_{X_1 \in B}\big] = P(X_1 \in B)$

Рассмотрев в определении эмпирического распределения луч вместо множества, получим эмпирическую функцию распределения.

Определение. Пусть $X_1, \ldots, X_n$ - выборка из распредления с функцией распределения F. Эмпирической функцией распределения будем называть случайную величину $F_n^*(t) = \dfrac{1}{n}\sum\limits_{j=1}^{n} I_{\{X_j \leqslant t\}}$

Заметим:

1. $\mathbb{E}\big[F_n^*(t)\big] = \dfrac{1}{n}\sum\limits_{j=1}^{n} P(X_j \leqslant t) = P(X_1 \leqslant t) = F(t)$
2. По УЗБЧ $F_n^*(t) \xrightarrow{\text{п.н.}} F(t)$

Теорема Гливенко-Кантелли

Теорема (Гливенко-Кантелли)**.** Пусть $X_1, \ldots, X_n$ - выборка из распредления с функцией распределения F. Тогда:

$$\sup_t |F_n^*(t) - F(t)| \xrightarrow{\text{п.н.}} 0$$

Доказательство.

Докажем только для непрерывной F.

Зафиксируем $N \in \mathbb{N}$ и выберем точки $t_0, t_1, \ldots, t_n$ такие, что

$$F(t_0) = 0, F(t_1) = \frac{1}{N}, \ldots, F(t_k) = \frac{k}{N}, \ldots, F(t_N) = 1$$

Для определенности $t_0 = -\infty, t_N = +\infty$.

Посмотрим, что происходит с разностью эмпирической и реальной функций распределения на промежутках вида $t \in [t_k, t_{k+1}]$.

$$F_n^*(t) - F(t) \leqslant \Big[\text{Оцениваем сверху. Хотим получить в } F_n^* \text{ значение побольше, а в } F \text{ - поменьше.}$$

$$\text{Поэтому, используя монотонность, подставляем соответствующие точки из промежутка}\Big] \leqslant F_n^*(t_{k+1}) - F(t_k) =$$

$$= F_n^*(t_{k+1}) - F(t_{k+1}) + \frac{1}{N}$$

Теперь оценим снизу.

$$F_n^*(t) - F(t) \geqslant F_n^*(t_k) - F(t_{k+1}) = F_n^*(t_k) - F(t_k) - \frac{1}{N}$$

Получаем:

$$F_n^*(t_k) - F(t_k) - \frac{1}{N} \leqslant F_n^*(t) - F(t) \leqslant F_n^*(t_{k+1}) - F(t_{k+1}) + \frac{1}{N},$$

$$|F_n^*(t) - F(t)| \leqslant \max_{0 \leqslant k \leqslant N-1} |F_n^*(t_k) - F(t_k)| + \frac{1}{N},$$

$$\sup_t |F_n^*(t) - F(t)| \leqslant \max_{0 \leqslant k \leqslant N-1} |F_n^*(t_k) - F(t_k)| + \frac{1}{N}$$

(Навесили слева супремум по t, т.к. правая часть от t не зависит)

Введем множество $A_N = \left\{\omega : \forall k \in \{0, \ldots, N\}\ F_n^*(t_k) \xrightarrow[n\to\infty]{(\omega)} F(t_k)\right\}$ - пересечение конечного числа событий вероятности 1 (т.к. по УЗБЧ $\forall t_k\ F_n^*(t_k) \xrightarrow{\text{п.н.}} F(t_k)$), поэтому $P(A_N) = 1$.
Если $\omega \in A_N$, то

$$\overline{\lim}_{n\to\infty} \sup_t |F_n^*(t) - F(t)| \leqslant \frac{1}{N}$$

Теперь пересечем множества: $A = \bigcap\limits_{N=1}^{\infty} A_N \Rightarrow P(A) = 1$ - пересекли счетное число событий вероятности 1.

Если $\omega \in A$, то

$$\overline{\lim}_{n\to\infty} \sup_t |F_n^*(t) - F(t)| \leqslant \frac{1}{N}\ \forall N \in \mathbb{N},$$

$$\left[\lim_{N\to\infty}\right]$$

$$\overline{\lim}_{n\to\infty} \sup_t |F_n^*(t) - F(t)| \leqslant 0,$$

$$0 \leqslant \underline{\lim}_{n\to\infty} \sup_t |F_n^*(t) - F(t)| \leqslant \overline{\lim}_{n\to\infty} \sup_t |F_n^*(t) - F(t)| \leqslant 0,$$

$$\forall \omega \in A\ \ \lim_{n\to\infty} \sup_t |F_n^*(t) - F(t)| = 0$$

■