8 Теория больших уклонений в многомерном случае-2

8.1 Критерий обобщенного отношения правдоподобий и критерий хи-квадрат

Применим теорему Санова к задаче проверки гипотезы согласия $H_0 : p_1 = p_1^0, ..., p_k = p_k^0$, где p_i^0 заданы, с помощью критериев отношения правдоподобия и хи-квадрат. Критерий отношения правдоподобий основан на статистике

$$T_1(X_1, ..., X_n) = 2\sum_{i=1}^{k} N_i \ln \frac{N_i}{np_i^0},$$

критерий хи-квадрат основан на статистике

$$T_2(X_1, ..., X_n) = \sum_{i=1}^{k} \frac{(N_i - np_i^0)^2}{np_i^0},$$

обе из которых имеют асимптотическое распределение χ^2_{k-1} при выполнении гипотезы. Тем самым вероятность $\mathbf{P}_0(T_1(X_1, ..., X_n) > c)$ и $\mathbf{P}_0(T_2(X_1, ..., X_n) > c)$ аппроксимируются вероятностью $1 - F_{\chi^2_{k-1}}(c)$.

Можно сказать, что критерий хи-квадрат получается из критерия отношения правдоподобий аппроксимацией логарифма первыми членами разложения, а критерий хи-квадрат вытекает из нормальной аппроксимации для вектора частот $(\nu_1, \ldots, \nu_k)$.

Эта аппроксимация довольно точна при c порядка константы, но, скажем, при c порядка n она является неудовлетворительной.

Фактический уровень значимости критериев в этом случае можно оценить с помощью теоремы Санова, используя представления

$$\mathbf{P}(T_1(X_1, ..., X_n) > na) = \mathbf{P}\left(2\sum_{i=1}^{k} \nu_i \ln\left(\frac{\nu_i}{p_i^0}\right) > a\right) = \mathbf{P}((\nu_1, \ldots, \nu_{k-1}) \in A),$$

$$\mathbf{P}(T_2(X_1, ..., X_n) > na) = \mathbf{P}\left(\sum_{i=1}^{k} \frac{(\nu_i - p_i^0)^2}{p_i^0} > a\right) = \mathbf{P}((\nu_1, \ldots, \nu_{k-1}) \in B),$$

где

$$A = \{\theta_i : 2\sum_{i=1}^{k} \theta_i \ln\left(\frac{\theta_i}{p_i^0}\right) > a\}, \quad B = \{\theta_i : \sum_{i=1}^{k} \theta_i = 1,\ \sum_{i=1}^{k} \frac{(\theta_i - p_i^0)^2}{p_i^0} > a\}.$$

При этом
$$\Lambda(\vec{\theta}) = \sum_{i=1}^{k} \theta_i \ln\left(\frac{\theta_i}{p_i^0}\right)$$
выпуклая функция, достигающая минимума на границе рассматриваемых областей. При этом $\Lambda(\vec{\theta}) = a/2$ при всех $\vec{\theta} \in \partial A$, поэтому наша асимптотика вероятности ошибки первого рода будет порядка $\exp(-na/2)$. Хи-квадрат аппроксимация дала бы вероятность
$$1 - F_{\chi^2_{k-1}}(na) \sim \frac{2}{2^{(k-1)/2}\Gamma((k-1)/2)}(na)^{(k-1)/2-1}e^{-na/2}$$
где последнее соотношение выводится с помощью правила Лопиталя из формулы гамма-плотности. Тем самым, хи-квадрат аппроксимация продолжает работать даже в зоне маленьких ошибок.

А вот B — внешность эллипсоида
$$\sum_{i=1}^{k} \frac{(\theta_i - p_i^0)^2}{p_i^0} > c,$$
пересеченная с плоскостью $\sum_{i=1}^{k} \theta_i = 1$. Здесь минимум Λ в B также достигается на границе, но лишь в некоторых точках $(\theta_1^{(j)}, ..., \theta_{k-1}^{(j)})$ с
$$\sum_{i=1}^{k} \frac{(\theta_i^{(j)} - p_i^0)^2}{p_i^0} = c,$$
где $\theta_k^{(0)}$ дополняет остальные до единичной суммы.

Раскладывая Λ в окрестности точек $\theta^{(j)}$ и, оценивая слагаемые вне этой окрестности, мы получим асимптотику типа (мы смотрим только на экспоненциальную часть)
$$\exp\left(-\sum_{i=1}^{k} \theta_i^{(0)} \ln \frac{\theta_i}{p_i^0} n\right)$$
при некоторых C, $l < k$. При этом, поскольку каждая из точек дает нам вероятность порядка
$$\exp(-\Lambda(\theta_1^0, \ldots, \theta_k^0)n),$$
итог с точностью до степенного множителя получится таким же. При этом хи-квадрат аппроксимация даст все тот же ответ $\exp(-na/2)$, что и раньше. Области A и B имеют одну и ту же вероятность с точки зрения распределения

χ^2_{k-1}, поэтому можно считать, что это вероятность попасть в A. Но на границе A функция Λ будет иметь меньший минимум, чем на границе B (просто потому, что A не содержится внутри B). Значит хи-квадрат аппроксимация даст нам заниженный ответ (на лекции ошибочно было сказано, что завышенный). Тем самым, и без того маленькая вероятность будет еще сильнее уменьшена погрешностью метода.

8.2 Теорема Стоуна

В нерешетчатом случае для векторов мы также можем сформулировать интегро-локальную теорему.

Теорема 1. *Пусть X_i — нерешетчатые векторы. Тогда*

$$\mathbf{P}(S_n \in \Delta_n[\vec{x}]) = \frac{\Delta_n^d}{(2\pi n)^{d/2}(\det(\Sigma))^{1/2}} \exp\left(-\frac{1}{2n}(\vec{x}-\vec{\mu}n)^t\Sigma^{-1}(\vec{x}-\vec{\mu})\right) + o\left(\frac{\Delta_n^d}{n^{d/2}}\right),$$

при всех Δ_n достаточно медленно стремящихся к нулю, где $o()$ равномерно мало по $\vec{x}$.

Как и прежде, мы можем утверждать, что в крамеровском случае

$$\mathbf{E}\exp(\langle h, X\rangle) < +\infty, \vec{h} \in D$$

теорема будет равномерна по всем сопряженным распределениям $X^{(h)}$ при h из любого компакта в $\operatorname{int} D$.

Теорема 2. *Пусть $\vec{X}_i$ — н.о.р. нерешетчатые векторы с конечной дисперсией, $R(\vec{h}) < \infty$ при $\vec{h} \in D$, G — образ D при отображении $m(\vec{h})$.*
Тогда

$$\mathbf{P}(S_n \in \Delta_n[\vec{x}]) = \frac{\Delta_n^d(1+o(1))}{(2\pi n)^{d/2}\det(\Sigma(\vec{h}_{\vec{x}/n}))^{1/2}} \exp\left(-\Lambda\left(\frac{x}{n}\right)n\right),$$

при всех Δ_n достаточно медленно стремящихся к нулю, где $o()$ равномерно мало по $\vec{x}/n \in G^*$ для любого компакта G^*, содержащегося в G.

Как и прежде, из интегро-локальной теоремы можно получить интегральную, однако, в данной ситуации степенной множитель в асимптотике будет существенно зависить от геометрии линии уровня $\Lambda(\theta)$ и границы множества A, вероятность попадания в которое я изучаю.

8.3 Принцип больших уклонений

Остаток курса мы посвятим грубой асимптотике больших уклонений. Иначе говоря, мы будем говорить только о функции Λ, рассматривая логарифмическую асимптотику рассматриваемых вероятностей.

Определение 1. В общем случае, назовем функцию $f(x)$ *полунепрерывной снизу (сверху)* в точке, если

$$\liminf_{x \to x_0} f(x) \geq f(x_0), \ (\limsup_{x \to x_0} f(x) \leq f(x_0)).$$

Соответственно, полунепрерывность на множестве есть полунепрерывность в каждой точке множества. Это условие равносильно тому, что $f^{-1}(-\infty, a]$ — замкнутое множество при любом a.
Примером полунепрерывной сверху функции является $[x]$, полунепрерывной снизу $\{x\}$.

Определение 2. Назовем *функцией роста* $\Lambda(x)$ неотрицательную полунепрерывную снизу функцию.

Определение 3. Будем говорить, что последовательность мер $\mathbf{P}_n$ удовлетворяет ПБУ (*принципу больших уклонений*) с функцией роста Λ, если

$$-\inf_{x \in A_{int}} \Lambda(x) \leq \liminf \frac{\ln \mathbf{P}_n(A)}{n} \leq \limsup \frac{\ln \mathbf{P}_n(A)}{n} \leq -\inf_{x \in A_{out}} \Lambda(x).$$

Что же это означает? Фактически, мы получаем оценки

$$\exp(-(1+\delta) \inf_{x \in A_{int}} \Lambda(x) n) \leq \mathbf{P}_n(A) \leq \exp(-(1-\delta) \inf_{x \in A_{out}} \Lambda(x) n)$$

при любом наперед взятом δ и достаточно больших n.

Нам приходится рассматривать разные множества в левой и правой частях, как и прежде это необходимая плата за то, что множество A достаточно общего вида.

Нетрудно заметить, что справедливо следующая эквивалентная формулировка:
1) Для любого замкнутого F

$$\limsup_{n \to \infty} \frac{1}{n} \ln \mathbf{P}_n(F) \leq -\inf_{x \in F} \Lambda(x).$$

2) Для любого открытого G

$$\liminf_{n\to\infty} \frac{1}{n} \ln \mathbf{P}_n(G) \geq -\inf_{x\in G} \Lambda(x).$$

Действительно, тогда ПБУ будет следовать из этих утверждения для A_{int} и A_{out}.

8.4 Теорема Крамера

На следующей лекции мы докажем следующую теорему, называемую теоремой Крамера:

Теорема 3 (Крамера)**.** *Пусть S_n — случайное блуждание с $R(h) = \mathbf{E}e^{hX}$ (это матожидание всегда существует, возможно являясь бесконечным). Тогда меры $\mathbf{P}(S_n/n \in \cdot)$ удовлетворяют ПБУ с $\Lambda(x) = \sup_h(hx - \ln R(h))$.*

Отметим, что мы не накладываем никаких условий на конечность $R(h)$ или даже конечность хоть каких-нибудь моментов X. Посмотрим на то, что произойдет с формулировкой в некоторых специфических случаях:

Пример 1. 1) Пусть X_i имеют распределение Коши. Тогда

$$R(0) = 0, \quad R(h) = +\infty,\ h \neq 0.$$

Значит,

$$\Lambda(\theta) = 0 \cdot \theta - \ln R(0) = 0.$$

Таким образом, теорема говорит о том, что

$$e^{-\varepsilon n} \leq \mathbf{P}(S_n/n \in A) \leq e^{\varepsilon n}$$

при таких A, что A_{int} непусто или

$$0 \leq \mathbf{P}(S_n/n \in A) \leq e^{\varepsilon n}$$

при A с пустой внутренностью. Второе утверждение бессмысленно и не накладывает никаких ограничений на A. Первое лишь говорит, что $\mathbf{P}(S_n/n \in A)$ не может стремиться к нулю экспоненциально быстро — это довольно очевидно следует из того, что S_n/n также имеет распределение Коши и в множество с непустой внутренностью попадает с положительной вероятностью.
2) Для распределения Бернулли

$$R(h) = pe^h + (1-p), \quad \Lambda(\theta) = \sup_h(\theta h - \ln R(h)).$$

Производная $\theta h - \ln R(h)$ есть $\theta - pe^h/(pe^h + 1 - p)$. При $\theta < 0$ функция убывает по h, супремум по h достигается при $h \to \infty$ и есть $+\infty$. Аналогично при $\theta > 1$ функция возрастает по h, супремум достигается при $h \to +\infty$ и равен $+\infty$. При $h \in (0, 1)$ существует единственный максимум, достигающийся при

$$h = \ln\left(\frac{(1-p)\theta}{p(1-\theta)}\right)$$

и равный

$$\Lambda(\theta) = \theta \ln \frac{\theta}{p} + (1-\theta) \ln \frac{1-\theta}{1-p}.$$

Здесь $\Lambda(p) = 0$, Λ не убывает при $\theta > p$ и не возрастает при $\theta < p$. Таким образом, если множество A не пересекается во внутренним точкам с $(0, 1)$, то вероятность попасть в него будет $o(n)$ (в действительности мы видим, что она 0), если содержит среди внутренних точек точку p, то вероятность будет больше $e^{-\varepsilon n}$ при любом ε и достаточно больших n. Вполне естественно, потому что она стремится к 1.

По существу помимо тех случаев, который мы уже рассматривали в теореме Петрова, содержательных результатов мы не получим, а получим только некоторы грубые оценки.