Спецкурс "Большие уклонения"

Шкляев А.В.

11 ноября 2016 г.

Рассмотрим несколько применений рассмотренной нам теории:

Пример 9.1. Рассмотрим критерий Неймана-Пирсона проверки простой гипотезы H_0 (X_i имеют плотность $f_0(x)$) с простой альтернативой H_1 (X_i имеют плотность $f_1(x)$). Допустим у нас есть выборка размера n, тогда критерий Неймана-Пирсона предлагает действовать следующим образом:
1) отвергать гипотезу H_0, если выборка попала в множество D

$$D = \left\{ \frac{f_1(x_1)...f_1(x_n)}{f_0(x_1)....f_0(x_n)} > \gamma_n \right\},$$

2) принимать ее в противном случае,
где γ_n — некоторое заданное число. Предположим, что размер выборки растет, а $\gamma_n = \gamma^n$, γ — фиксированно.
Рассмотрим ошибку первого рода $\alpha_n = P_0((X_1, ..., X_n) \in D)$, т.е. вероятность того, что гипотеза была верна, а мы ее отвергли. Как ведет себя α_n при $n \to \infty$?
При этом

$$\alpha_n = P_0((X_1, ..., X_n) \in D) = P_0\left(\sum_{i=1}^{n} \ln \frac{f_1(X_i)}{f_0(X_i)} \geq n \ln \gamma\right).$$

Для $Y_1 = \ln \frac{f_1(X_1)}{f_0(X_1)}$

$$R(h) = \int_{\mathbb{R}} f_1^h(x) f_0^{1-h}(x) dx.$$

В частности, $R(1) = 1$.
Следовательно, $E_0 Y_1 = \mu < \infty$. При $\ln\gamma < \mu$ α_n сходится к 1. При $\ln\gamma > \mu$

$$\lim_{n\to\infty} \frac{1}{n} \ln \alpha_n = -\inf_{\theta \geq \ln\gamma} \Lambda(\theta) = -\Lambda(-\ln\gamma).$$

Если же $\mu > -\infty$, а $\mu < \ln\gamma < m^+$, то

$$\alpha_n \sim C(\ln\gamma) n^{-1/2} \exp(-\Lambda(\ln\gamma) n).$$

Аналогично для ошибки второго рода

$$\beta_n = P_1((X_1, ..., X_n) \notin D) = P_1\left(-\sum_{i=1}^{n} \ln \frac{f_1(X_i)}{f_0(X_i)} > -n \ln \gamma\right).$$

Тогда

$$\tilde{R}(h) = E_1 e^{-hY_1} = \int_{\mathbb{R}} f_1^{1-h}(x) f_0^h(x) dx = R(1-h),$$

$\tilde{R}(1) = 1$, $\tilde{\mu} = -E_1 Y_1 < \infty$. Тогда аналогично

$$\beta_n \sim \tilde{C}(-\ln\gamma) n^{-1/2} \exp(-\tilde{\Lambda}(-\ln\gamma) n)$$

при $\ln\gamma > -\tilde{\mu}$. При этом $\tilde{C}(-\ln\gamma) = C(\ln\gamma)$.

$$\tilde{\Lambda}(-\theta) = \inf_h(-\theta h - \ln R(1-h)) = \Lambda(\theta) - \theta$$

Пример 9.2. Давайте зададимся таким статистическим вопросом для задачи из предыдущего примера. Пусть α_n, β_n — вероятности ошибок I, II родов, a, b — положительные константы, $a + b = 1$. Тогда какой минимальной может быть байесовская ошибка $\Delta_n = a\alpha_n + b\beta_n$? Мы знаем, что наилучшим критерием будет критерий Неймана-Пирсона и наименьшую Δ_n будет давать он.
Рассмотрим критерий Неймана-Пирсона с $\gamma = 1$, его вероятность ошибки первого рода обозначим через $\alpha_n^{(0)}$. Рассмотрим любой другой критерий T_n Неймана-Пирсона с $\gamma = \gamma_n^{(T)}$ и $\alpha = \alpha_n^{(T)}$. При $\gamma_n^{(T)} \leq 1$ $\alpha_n^{(0)} \leq \alpha_n^{(T)}$ (поскольку критическое множество у второго критерия больше), откуда

$$\ln \min_{T_n} \Delta_n \geq \ln(a\alpha_n^{(0)}) = \ln a + \ln \alpha_n^{(0)},$$

при $\gamma_n^{(T)} > 0$, напротив,

$$\ln \min_{T_n} \Delta_n \geq \ln(b\beta_n^{(0)}) = \ln b + \ln \beta_n^{(0)},$$

Следовательно,

$$\liminf \frac{1}{n} \ln \min_{T_n} \Delta_n \geq -\Lambda(0).$$

При этом на критерии Н-П $\lim \frac{1}{n} \ln \Delta_n = -\Lambda(0)$. Отсюда имеем утверждение — наилучшая возможная байесовская ошибка критерия будет Δ_n с

$$\lim \frac{1}{n} \ln \Delta_n = -\Lambda(0) = \inf_h \ln R(h)$$

и достигаться она будет на критерии, предлагающем выбрать тот параметр, при котором правдоподобие будет больше.

Пример 9.3. Рассмотрим еще одну статистическую задачу. Пусть $X_1, ..., X_n$ — выборка из некоторого распределения F, мы рассматриваем гипотезу $H_0 : F = F_0$, где F_0 — некоторая непрерывная функция распределения, с альтернативой $H_1 : F \geq F_0$, то есть $F(x) \leq F_0(x)$ при всех x. Для этого используется критерий Колмогорова, основанный на рассмотрении

$$P_{F_0}(D_n^+ \geq x) = \alpha,$$

где x — некоторый заданный нами порог, $D_n^+ = \sup(\hat{F}_n(t) - F_0(t))$, где $\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^n I_{X_i \leq t}$. Если гипотеза верна, то D_n^+ имеет фиксированное распределение при любых непрерывных F_0. Как при фиксированном x будет вести себя $P(D_n^+ \geq x)$ с ростом x? Докажем, что

$$\frac{\ln P(D_n^+ \geq x)}{n} \to -\inf_t \Lambda_t(x + F_0(t)),$$

где

$$\Lambda_t(x) = x \ln\left(\frac{x}{F_0(t)}\right) + (1-x)\ln\left(\frac{1-x}{1-F_0(t)}\right)$$

функция уклонений, построенная по бернуллиевским величинам $Y_i = I_{X_i \leq t}$ с $p = F_0(t)$.

Заметим, что в силу независимости рассматриваемой величины от F_0, можно считать, что $t \in [0,1]$, $F_0(t) = t$.

1) Оценка снизу:

$$P(D_n^+ \geq x) \geq P(\hat{F}_n(t) - t \geq x) = P\left(\sum_{i=1}^n Y_i \geq n(x+t)\right),$$

откуда

$$\liminf_{n\to\infty} \frac{\ln P(D_n^+ \geq x)}{n} \geq -\Lambda_t(x+t)$$

при любом $t \in [0,1]$. Выбирая то t, при котором правая часть наибольшая, получаем требуемую оценку.

2) Оценка сверху:

$$P(D_n^+ \geq x) \leq \sum_{k=0}^{l-1} P\left(\sup_{t\in[k/l,(k+1)/l]}(\hat{F}_n(t) - t) \geq x\right) = \sum_{k=0}^{l-1} P(\hat{F}_n((k+1)/l) \geq k/l + x).$$

Тогда

$$\liminf_{n\to\infty} \frac{\ln P(D_n^+ \geq x)}{n} \geq -\max_{k<l} \Lambda_{(k+1)/l}(x + k/l)$$

при произвольном l. Устремляя l к бесконечности, имеем

$$\max_{k<l} \Lambda_{(k+1)/l}(x + k/l) \to \sup_t \Lambda_t(x),$$

что и требуется.

При этом при альтернативе D_n^+ будет сходиться по вероятности к $\max(F(x) - F_0(x))$. Тем самым, критерий будет име

Аналогичный результат можно получить для $D_n^- = \sup(F_0(t) - \hat{F}_n(t))$, Λ при этом окажется той же. Отсюда тот же результат будет для $D_n = D_n^+ + D_n^-$.