Задача Есть выборка $Y_1 \dots Y_n$ из неизвестной плот. $p_0(x)$. Мы предполагаем, что $p_0(x)$ из семейства плотностей $p(x,\mu)$ (но это не обязательно так), $\mu \in \mathbb{R}$, ну и хотим оценить μ, т.е. найти такую плотность $p(x,\mu)$, которая вроде как должна равняться $p_0(x)$.

Идея: $\hat{\mu}(p_0) = \arg\min_\mu d[p_0, p(\cdot;\mu)]$, где d — некое расстояние (или мера близости) между плотностями.

Но эта штука не явл. оценкой, потому что $p_0(x)$ мы не знаем! Знаем только выборку. Но тогда можно взять в кач. оценки $p_0(x)$ эмпирич. плотность $\delta(x-Y)$ ($x=(x_1 \dots x_n)$, $Y=(Y_1 \dots Y_n)$) (сравни с ядерными оценками плотности — там ведь тоже фактически δ-функцией пользовались!)

Сначала разберёмся с расстояниями:

1) Расстояние полной вариации — самое, пожалуй, естественное с вероятн. точки зрения. Пусть ξ, η — сл. в. с плотностями p_ξ и p_η.

Тогда $D(p_\xi, p_\eta) = \sup_{A \in \mathcal{F}} \left| \int_A p_\xi(x)\,dx - \int_A p_\eta(x)\,dx \right|$, $\mathcal{F}$ — σ-алгебра.

Вроде как оч. неудобно вычислять, ведь супремум по всей σ-алгебре берётся. Но есть такая теорема:

Теорема (Шеффе) Если существуют плотности (обобщённые) $p_\xi(x), p_\eta(x)$, $x \in \mathbb{R}^n$ сл. векторов ξ, η, то $D(p_\xi, p_\eta) = \frac{1}{2}\int_{\mathbb{R}^n} |p_\xi(x) - p_\eta(x)|\,dx$

Проблема в том, что непонятно, как вычислять $\hat{\mu}(p_0) = \arg\min_\mu D[p_0, p(\cdot;\mu)]$

Кроме того, $D(\delta(\cdot - Y^n), p(\cdot;\mu)) = \frac{1}{2}\int_{\mathbb{R}^n} |\delta(x-Y^n) - p(x;\mu)|\,dx = 1$ (на 2 инт. надо разбить)

$= \frac{1}{2}\int_{Y^n} |\underbrace{\delta(x-Y) - p(x,\mu)}_{1}|\,dx + \frac{1}{2}\int_{\mathbb{R}^n/Y} \underbrace{|p(x,\mu)|\,dx}_{1}$.

2) Расстояние Кульбака-Лейблера

$$K(p_\xi, p_\eta) = \int_{\mathbb{R}^n} p_\xi(x) \log\frac{p_\xi(x)}{p_\eta(x)}\,dx.$$

Прежде всего, $K(p_\xi, p_\eta)$ расст. не является, т.к. $K(p_\xi, p_\eta) \neq K(p_\eta, p_\xi)$.

Но отсутствие симметрии — это реальная сит. в статистике. Ведь, с одной стороны, есть реал. данные с неизв. плотностью $p_0(x)$ (кладём $p_\xi(x) = p_0(x)$), которой мы управлять не можем, а есть семейство плотностей $p(x,\mu)$ ($= p_\eta(x)$), где мы можем варьировать μ. Но тогда как связано расст. Кульбака-Лейблера с расст. полной вар.?

Теорема (Нер-во Пинскера)

Пусть сл. вектора ξ, η имеют плотности (обобщ.) $p_\xi(x), p_\eta(x)$, $x \in \mathbb{R}^n$, тогда

$$K(p_\xi, p_\eta) \geq 2[D(p_\xi, p_\eta)]^2$$

Крайне удобное св-во расст. К.-Б. таково:

Теорема Пусть $\xi = (\xi_1 \dots \xi_n)$ и $\eta = (\eta_1 \dots \eta_n)$ — выборки. Тогда

$$K(p_\xi, p_\eta) = n K(p_{\xi_1}, p_{\eta_1})$$

Ещё расстояния:

3) Расстояние Хеллингера: $H(p_1, p_2) := \left[\int_{\mathbb{R}^n} \left|\sqrt{p_1(x)} - \sqrt{p_2(x)}\right|^2 dx\right]^{1/2}$

Из нер-ва Коши-Бун. следует, что $H(p_1, p_2)^2 \geq D(p_1, p_2)$.

4) Расстояние хи-квадрат: $\chi^2(p_1, p_2) := \int_{\mathbb{R}^n} \frac{|p_1(x) - p_2(x)|^2}{p_2(x)} dx$.

Это первый нетривиал. член разлож. в ряд Тейлора $K(p_1, p_2)$ (логарифм правдоподобия)

Связь расст. Кульбака–Лейблера и ММП

Пусть мы решили использовать для получения оценки модели расст. Кульбака-Лейб. Если бы мы знали $p_0(z)$ – плотность $Y = (Y_1 \dots Y_n)$ (выборка), то искали бы параметр неизвест. распределения как точку глоб. max:

$$\hat{\mu}(p_0) = \arg\min_{\mu} \int_{\mathbb{R}^n} p_0(z) \log \frac{p_0(z)}{p(z,\mu)} dz = \arg\max_{\mu} \int_{\mathbb{R}^n} p_0(z) \log p(z,\mu) dz.$$

Но, как и ранее, p_0 мы не знаем, и поэтому в качестве оценки p_0 берём $\delta(z - Y)$, получаем:

$$\hat{\mu}(Y) = \arg\max_{\mu} \int_{\mathbb{R}^n} \delta(z - Y) \log p(z,\mu) dz = \arg\max_{\mu} \log p(Y, \mu) \quad - \text{так это } \textbf{не}$$

ищется максимум логарифмической функции правдоподобия! пусть это макс. есть

Но тогда $\log p(Y,\mu) = \sum_{i=1}^{n} \log p(Y_i, \mu) \Rightarrow \hat{\mu}(Y) \xrightarrow{P} \arg\max_{\mu} E \log p(Y_1, \mu) \in$

можно сказать, что так будет

$= \arg\min_{\mu} E \log \frac{p_0(Y_1)}{p(Y_1,\mu)} = \arg\min_{\mu} K(p_0, p(\cdot, \mu))$.

Т.е. при больших n и метод макс. пр-я выбирает плотность из зад. параметрич. сем-ва, наход. на миним. расст. Кульбака–Лейблера от истинной, но неизвестной плотности наблюдений.

Основная книга – Вальд, „Последовательный анализ".

В процедуре проверки гипотез мы, как правило, действуем по след. алгоритму: по всем наблюдениям, которые у нас есть, строим статистику критерия, исходя из значений которой, принимаем или отвергаем основную гипотезу.

Пусть у нас наблюдения поступают в процессе проверки, тогда в последовательном анализе мы фактически делаем проверку для каждого k, начиная с 1, что позволяет достигнуть уровня значимости α и мощности β, используя в среднем меньшее количество наблюдений.

Итак, пусть $f_m(\theta) = \prod_{i=1}^{m} p(X_i, \theta)$ – ф-ция правдоподобия выборки $(X_1 \dots X_m)$

Пусть проверяем простую гипотезу $H_0: \theta = \theta_0$ против альтернативы $H_1: \theta = \theta_1$

Тогда на каждой стадии эксперимента (т.е. $\forall m$) действуем по след. алгоритму:

1) Если $B < \frac{f_m(\theta_1)}{f_m(\theta_0)} < A$, то эксперим. продолжается и производится доп. наблюд.

2) Если $\frac{f_m(\theta_1)}{f_m(\theta_0)} \geq A$, то отклоняем H_0 и заканч. процесс проверки

3) Если $\frac{f_m(\theta_1)}{f_m(\theta_0)} \leq B$, то принимаем H_0 и заканч. процесс проверки.

Эти вычисления проще проводить с помощью логарифма отнош. пр-й:

$$\ln \frac{f_m(\theta_1)}{f_m(\theta_0)} = \sum_{i=1}^{m} \ln \frac{p(X_i, \theta_1)}{p(X_i, \theta_0)} =: \sum_{i=1}^{m} Z_i,$$ тогда сравнив. $\sum_{i=1}^{m} Z_i$ с $\ln A$ и $\ln B$.

Пример Пусть $X_1 \dots X_n \sim Bern(p)$, $H_0: p = p_0$, $H_1: p = p_1$, тогда

$$Z_i = \begin{cases} \ln \frac{p_1}{p_0}, & \text{если } X_i = 1 \\ \ln \frac{1-p_1}{1-p_0}, & \text{если } X_i = 0 \end{cases}$$, т.е. получается такое случ. блуждание

Как выбирать константы A и B?

Пусть процесс у нас закончился отклонением гипотезы H_0, т.е. в какой-то момент $\frac{f_m(\theta_1)}{f_m(\theta_0)} \geq A$. Если гипотеза H_0 верна, то вероятность такого исхода (таких исходов, что мы отклоним H_0) $\leq \alpha$, где α – ур. знач., если H_1 верна – то вероятность $1-\beta$ (это нужно отдельно доказывать – нужно док., что процесс окончится), т.е. $\frac{1-\beta}{\alpha} \geq A$ по идее должно быть.

А вероятность получения выборки, на кот. гипотеза H_0 отвергается, при невер. H_0 хотя бы в A раз больше, чем при вер. H_0, поэтому $\frac{1-\beta}{\alpha} \geq A$.

Аналогично, $B \geq \frac{\beta}{1-\alpha}$. И можно док. следующее утверждение:

Утв. Если взять $B = \frac{\beta}{1-\alpha}$, $A = \frac{1-\beta}{\alpha}$, то вер. ошибки I рода последов. критерия будет $\leq \alpha$, а вер. ош. II рода $\leq \beta$.

* Это все, конечно, прекрасно, но проверка сложных гипотез – гораздо более интересная и важная задача. Попробуем найти вероятность того, что последов. анализ закончится принятием гипотезы H_0, когда истинное знач. параметра равно θ, $\theta \neq \theta_0, \neq \theta_1$, эта вероятность называется оперативной характеристикой, обозн. $L(\theta)$

Если пренебречь эффектом превышения границ при оконч. процедуры (об этом позже), то $L(\theta) = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}}$, где $h(\theta)$ находится (и $h(\theta) \neq 0$) как решение след. ур-ния: $E_\theta \left(\frac{p(X_1, \theta_1)}{p(X_1, \theta_0)} \right)^{h(\theta)} = 1$, т.е. (в непр. случае)

$$\int_R \left(\frac{p(x, \theta_1)}{p(x, \theta_0)} \right)^{h(\theta)} p(x, \theta) dx = 1.$$

Замечание. Что делать, если все наблюдения закончились, а вывода мы так и не сделали (т.е. не отвергли H_0, но и не приняли её), то тогда решение следующее: если при n_{max} $\ln B < \sum_{i=1}^{n_{max}} z_i < \ln A$, то если $\sum_{i=1}^{n_{max}} z_i > 0$, то отвергаем H_0, если $\sum_{i=1}^{n_{max}} z_i \leq 0$, то принимаем H_0. Конечно, вер-ти ошибок I и II рода изменятся, но при большом n_{max} это влияние становится незначительным.

Замечание. Так насколько же всё-таки последов. анализ Вальда лучше критерия Неймана-Пирсона? Ну, при фиксированных α и β последовательный анализ требует в среднем в 2 раза меньше наблюдений, чем стандартный критерий Неймана-Пирсона, подробнее см. книгу Вальда.

Как в итоге проверять сложные гипотезы? Дело в том, что операт. х-ка, как правило, является монотон. ф-цией по θ.

Проверим сначала $H_0: \theta = \theta_0$ против $H_1: \theta > \theta_0$.

Можно выбрать такое $\theta_1 > \theta_0$, что принятие основной гипотезы в области $\theta > \theta_1$ приведёт к ошибке, имеющей практическое значение, т.е. область отклонения — $\theta \geq \theta_1$ и область безразличия $\theta_1 > \theta > \theta_0$.

Ну и строим обычный послед. критерий по θ_1 и θ_0 такой, что $L(\theta_0) = 1 - \alpha (\geq)$ и $L(\theta_1) \leq \beta$.

В случае двусторонней альтернативы $H_0: \theta = \theta_0$ и $H_1: \theta \neq \theta_0$ можно выбрать $f_m(\theta) = \frac{1}{2}(f_m(\theta_0 - \delta) + f_m(\theta_0 + \delta))$ вместо $f_m(\theta_1)$ и рассм. послед. критерий $B \leq \frac{f_m(\theta)}{f_m(\theta_0)} \leq A$, причём δ выбирается примерно так же, как и в односторонем случае.