Лекция 12.Байесовский подход

Буре В.М., Грауэр Л.В.

ШАД

Санкт-Петербург, 2013

Содержание

Общая схема байесовского подхода к статистическому оцениванию

Пусть в описании закона распределения анализируемой случайной величины, функции регрессии, временного ряда и т.п. участвует s-мерный параметр $\theta = (\theta_1, \dots, \theta_s)^T$.
Задача состоит в построении наилучшей, в определеннном смысле, статистической оценки $\hat{\theta}$ параметра θ по имеющимся наблюдениям $X_{[n]} = (X_1, \dots, X_n)$.
Байесовский подход основан на двух положениях

- Степень нашей уверенности в справедливости некоторого утверждения численно выражается в вероятности.
- При принятии решения в качестве исходной информации используется одновременно информация двух типов: априорная и содержащаяся в исходных статистических данных.

Априорная информация представлена в виде некоторого априорного распределения вероятностей анализируемого неизвестного параметра, которое описывает степень его уверенности в том, что этот параметр примет то или иное значение, еще до начала сбора исходных статистических данных.
По мере поступления исходных статистических данных это распределение уточняется, переходя от априорного распределения к апостериорному, по формуле Байеса:

$$P\{A_i|B\} = \frac{P\{A_i\}P\{B|A_i\}}{\sum_{i=1}^{K} P\{A_i\}P\{B|A_i\}}, \tag{1}$$

$A_1, \ldots, A_K$ образуют полную группу событий, $P\{B\} > 0$.

Общая логическая схема байесовского метода оценивания

Априорные сведения о параметре θ основаны на предыстории функционирования анализируемого процесса и на профессиональных теоретических соображениях о его сущности, специфике, особенностях. Априорные сведения представлены в виде функции $p(\theta)$, задающей априорное распределение параметра

- вероятность принять значение θ в дискретном случае,
- плотность распределения в непрерывном случае.

При анализе многомерных параметров $\theta = (\theta_1, \ldots, \theta_s)^T$ при построении априорного распределения обычно предполагают стастистическую независимость компонент $\theta_1, \ldots, \theta_s$

$$p(\theta) = p(\theta_1) \cdot \ldots \cdot p(\theta_s).$$

Исходные статистические данные

Выборка $X_1, \dots, X_n$ получена из генеральной совокупности с функцией распределения $F(x|\theta)$.

Пусть $f(x|\theta)$

- плотность распределения наблюдаемой случайной величины ξ, если ξ — непрерывна, или
- вероятность $P\{\xi = X|\theta\}$, если ξ дискретна,

при условии, что значение неизвестного параметра равно θ.

Функция правдоподобия $L(X_1, \dots, X_n|\theta)$ имеющихся данных определяется соотношением

$$L(X_1, \dots, X_n|\theta) = f(X_1|\theta)f(X_2|\theta) \cdot \ldots \cdot f(X_n|\theta). \tag{2}$$

Вычисление апостериорного распределения $\tilde{p}(\theta|X_1,\ldots,X_n)$ осуществялется с помощью формулы Байеса (1), где
A_i — событие, заключающееся в том, что значение оцениваемого параметра равно θ,
B — событие, заключающееся в том, что значения n наблюдений, зафиксированы на уровнях $X_1,\ldots,X_n$.

$$\tilde{p}(\theta|X_1,\ldots,X_n) = \frac{p(\theta)L(X_1,\ldots,X_n|\theta)}{\int L(X_1,\ldots,X_n|\theta)p(\theta)d\theta} \qquad (3)$$

Знаменатель (3) $\int L(X_1,\ldots,X_n|\theta)p(\theta)d\theta$ играет роль нормирующего коэффициента и не зависит от неизвестного параметра θ.

$$\tilde{p}(\theta|X_1,\ldots,X_n) \propto p(\theta)L(X_1,\ldots,X_n|\theta). \qquad (4)$$

Построение байесовских точечных и интервальных оценок основано на использовании знания апостериорного распределения $\tilde{p}(\theta|X_1,\ldots,X_n)$ (3).
В качестве байесовских точечных оценок $\hat{\theta}^B$ используют среднее или модальное значение распределения $\tilde{p}$:

$$\hat{\theta}^B_{mean} = E(\theta|X_1,\ldots,X_n) = \int \theta\tilde{p}(\theta|X_1,\ldots,X_n)d\theta, \tag{5}$$

$$\hat{\theta}^B_{mod} = \arg\max_{\theta} \tilde{p}(\theta|X_1,\ldots,X_n), \tag{6}$$

Байесовская оценка (5) является наилучшей в смысле доставления минимума апостериорному байесовскому риску:

$$\begin{aligned} R_B(X_1,\ldots,X_n) &= E\{(\hat{\theta}(X_1,\ldots,X_n)-\theta)^2|X_{[n]}\} = \\ &= \int (\hat{\theta}(X_1,\ldots,X_n)-\theta)^2\tilde{p}(\theta|X_{[n]})d\theta \end{aligned} \tag{7}$$

Для построения байесовского доверительного интервала для парметра θ необходимо вычислить по формуле (3) апостериорный закон распределения параметра θ ($\tilde{p}(\theta|X_1, \ldots, X_n)$), а затем по заданной доверительной вероятности γ определить критические значения p_1, p_2, которые дают соотвественно левый и правый концы доверительного интервала.

- Как выбрать параметрическое семейство $p(\theta, D)$ априорного распределения оцениваемого параметра?
- Как подобрать численные значения D_0 параметра D, определяющие конкретный вид априорного распределения?
- Как вычислять апостериорное распределение $\tilde{p}(\theta, X_1, \ldots, X_n)$?

Априорные распределения, сопряженные с наблюдаемой генеральной совокупностью

Определение 1

Семейство априорных распределений $G\{p(\theta, D)\}$ *называется сопряженным по отношению к наблюдаемой генеральной совокупности* $f(X, \theta)$ *(или к функции правдоподобия* $L(X_1, \ldots, X_n|\theta)$*), если и апостериорное рапрсделение* $\tilde{p}(\theta, X_1, \ldots, X_n)$*, вычисленное по формуле* (3)*, принадлежит этому же семейству* G*.*

Теорема 1 (Условие существования сопряженного семейства априорных распределений)

Если функция правдоподобия $L(X_1, \ldots, X_n|\theta)$ *представима в форме*

$$L(X_1, \ldots, X_n|\theta) = v(T_1(X_1, \ldots, X_n), \ldots, T_m(X_1, \ldots, X_n); \theta) \cdot \psi(X_1, \ldots, X_n), \tag{8}$$

где $T_j(X_1, \ldots, X_n)$*,* $j = 1, \ldots, m$*, и* $\psi(X_1, \ldots, X_n)$ *— некоторые функции от наблюдений* $X_1, \ldots, X_n$*, не зависящие от параметров* θ*, то сущетсвует семейство* $G = \{p(\theta; D)\}$ *априорных распределений, соспряженное с* $L(X_1, \ldots, X_n|\theta)$*.*

Теорема 2

Если в байесовском подходе стартовать с априорного распределения, не несущего никакой дополнительной по отношению к имеющимся статистическим данным полезной информации об оцениваемых параметрах, то первый же переход от нее по формуле (3) к апостериорному распределению приведет к семейству распределений, сопряженному с наблюдаемой генеральной совокупностью.

Распределения, отражающие скудость априорных знаний

В случае отсутствия какой-либо полезной априорной информации о значениях оцениваемого параметра рекомендуется следовать следующим рекомендациям:

- если оцениваемый скалярный параметр θ может принимать значения на конечном интервале $[\theta_{min}, \theta_{max}]$ или на бесконечном интервале от $-\infty$ до $+\infty$, то априорную функцию плотности $p(\theta)$ следует считать постоянной на соотвествующем интервале;
- если из смысла оцениваемого параметра вытекает, что он может принимать любые положительные значения, то следует считать постоянной на всей числовой прямой $(-\infty, +\infty)$ функцию плотности распределения логарифма от значения параметра, т.е. $p(\ln\theta) = const$ при $\theta \in (0; +\infty)$.

Такие априорные распределения называют *распределениями, отражающими скудость априорных знаний* или *"САЗ-априорными распределениями"*.

При этом нарушение условий нормировки функции плотности вероятности не доставляет "технических неудобств".

Определим вид априорной плотности $p(\theta)$ для случая $p(\ln\theta) = const$

$$f_\theta(y) = \frac{\delta F_\theta(y)}{\delta y} = \frac{\delta F_{\ln\theta}(\ln y)}{\delta \ln y}\frac{\delta \ln y}{\delta y} = f_{\ln\theta}(\ln y)\frac{1}{y} \propto \frac{1}{y}.$$

Так как $f_{\ln\theta}(\ln y) = p(\ln y) = conts$,

$$p_{saz}(\theta) \propto \frac{1}{\theta}.$$

Для параметров θ с возможными значениями, заполняющимим всю числовую прямую, априорная плоность

$$p_{saz}(\theta) = const$$

Общий подход к выводу семейства априорных распределений, сопряженных с наблюдаемой генеральной совокупностью

Шаг 1. Проверка условия (8) существования семейства априорных распределений, сопряженных с функцией правдоподобия L для наблюдаемой генеральной совокупности.
Шаг 2. Если функция правдоподобия L допускает представление (8), то осуществляется вывод САЗ-апостериорного распределения $\tilde{p}_{saz}(\theta|X_1, \ldots, X_n)$ по формуле

$$\tilde{p}_{saz}(\theta|X_1, \ldots, X_n) \propto p_{saz}(\theta) L(X_1, \ldots, X_n|\theta). \tag{9}$$

Пересчет значений параметров при переходе от априорного сопряженного распределения к апостериорному

Пусть $\{p(\theta, D)\}$, $D = (d_1, \ldots, d_q)^T$, — семейство априорных распределений, сопряженных с функцией правдоподобия $L(x_1, \ldots, x_n|\theta)$ имеющихся наблюдений, и пусть D_0 — известные значения параметров D в анализируемом случае.
Тогда с помощью ряда тождественных преобразований правая часть соотношения

$$\tilde{p}(\theta|X_1, \ldots, X_n) \propto p(\theta; D_0)L(X_1, \ldots, X_n|\theta) \tag{10}$$

приводится, с точностью до множителей, не зависящих от θ, к виду $p(\theta; D(X_1, \ldots, X_n))$, где каждая из функций $d_j(X_1, \ldots, X_n)$, $j = 1, \ldots, q$, вектора $D(X_1, \ldots, X_n))$ является функцией D_0 и $\{X_1, \ldots, X_n\}$.

Пример 1

Пусть $\xi \in N(\theta, \sigma_0^2)$ — нормально распределенная случайная величина с неизвестным математическим ожиданием и известной дисперсией.

$$L(X_1, \dots, X_n|\theta) = \prod_{i=1}^{n} f(X_i|\theta) = e^{-\frac{n}{2\sigma_0^2}(\bar{x}-\theta)^2} \cdot \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\bar{x})^2}.$$

$$v(T_1(X_1, \dots, X_n); \theta) = e^{-\frac{n}{2\sigma_0^2}(\bar{x}-\theta)^2}, \quad T_1(X_1, \dots, X_n) = \bar{x}.$$

$$\psi(X_1, \dots, X_n) = \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (x_i-\bar{x})^2}.$$

Выполняются условия теоремы 1, следовательно, семейство априорных, сопряженных с L, существует.
Определим $p_{saz}(\theta) = const$, тогда

$$\tilde{p}_{saz}(\theta|X_1, \dots, X_n) = p_{saz}(\theta) L(X_1, \dots, X_n|\theta) \propto e^{-\frac{n}{2\sigma_0^2}(\bar{x}-\theta)^2}.$$

Таким образом, семейство

$$p(\theta; D) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(\theta-\theta_0)^2}{2\Delta_0^2}}$$

является сопряженным с $L(x_1, \ldots, x_n|\theta) \propto e^{-\frac{n}{2\sigma_0^2}(\bar{x}-\theta)^2}$.
Обозначим $d_1 = \theta_0$, $d_2 = \Delta_0^2$

$$\tilde{p}(\theta|x_1, \ldots, x_n) \propto e^{-\frac{(\theta-d_1)^2}{2d_2}} \cdot e^{-\frac{n}{2\sigma_0^2}(\bar{x}-\theta)^2} \propto e^{-\frac{(\theta-\tilde{d}_1)^2}{2\tilde{d}_2}} \tag{11}$$

где

$$\tilde{d}_1(x_1, \ldots, x_n) = \frac{\frac{1}{\sigma_0^2/n}\bar{x} + \frac{1}{\Delta_0^2}\theta_0}{\frac{1}{\sigma_0^2/n} + \frac{1}{\Delta_0^2}}, \tag{12}$$

$$\tilde{d}_2(x_1, \ldots, x_n) = \left(\frac{1}{\sigma_0^2/n} + \frac{1}{\Delta_0^2}\right)^{-1}, \tag{13}$$

Пример 2

Пусть $\xi \in B(M, \theta)$ биномиально распределенная случайная величина

$$f(x|\theta) = P\{\xi = x|\theta\} = C_M^x \theta^x (1-\theta)^{M-x}, \quad x = 0, 1, \ldots, M.$$

$$L(X_1, \ldots, X_n|\theta) = \prod_{i=1}^{n} C_M^{x_i} \theta^{x_i} (1-\theta)^{M-x_i} = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{nM - \sum_{i=1}^n x_i} \prod_{i=1}^{n} C_M^{x_i}.$$

В данном случае $T(X_1, \ldots, X_n) = \sum\limits_{i=1}^{n} x_i$ и семейство априорных сопряженных распределений существует.
Определим $p_{saz}(\theta) = 1$ для $\theta \in (0;\ 1)$, тогда

$$\tilde{p}_{saz}(\theta|X_1, \ldots, X_n) \propto \theta^{\sum_{i=1}^n x_i} (1-\theta)^{nM - \sum_{i=1}^n x_i} \tag{14}$$

С точностью до нормирующего множителя, не зависящего от θ, правая часть (14) представляет собой плотность бета-распределения.

Таким образом, семейство

$$p(\theta; D) \propto \theta^{a-1}(1-\theta)^{b-1}$$

является сопряженным с $L(x_1, \ldots, x_n|\theta) \propto \theta^{\sum_{i=1}^n x_i}(1-\theta)^{nM-\sum_{i=1}^n x_i}$
Формула (3) дает

$$\tilde{p}(\theta|x_1, \ldots, x_n) \propto \theta^{a-1}(1-\theta)^{b-1} \cdot \theta^{\sum_{i=1}^n x_i}(1-\theta)^{nM-\sum_{i=1}^n x_i} = \\ = \theta^{a+\sum_{i=1}^n x_i - 1}(1-\theta)^{b+nM-\sum_{i=1}^n x_i - 1}. \quad (15)$$

Правая часть (15) определяет с точностью до нормирующего множителя бета-распределение с параметрами

$$\tilde{a} = a + \sum_{i=1}^{n} x_i, \quad (16)$$

$$\tilde{b} = b + nM - \sum_{i=1}^{n} x_i. \quad (17)$$

№ пп	З.р.в. наблюдаемой генеральной совокупности	Сопряженный априорный з.р.в. $p(\theta)$, выражения для $\mathbf{E}\theta$ и $\mathbf{D}\theta$	Апостериорный з.р.в. $p(\theta \mid x_1, x_2, \ldots, x_n)$, выражения для его параметров
1	$(\theta;\sigma^2)$-нормальный, $f(x\mid\theta)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\theta)^2}{2\sigma^2}}$ (значение дисперсии σ^2 известно)	$(\theta_0;\sigma_0^2)$-нормальный; $\mathbf{E}\theta=\theta_0;\ \mathbf{D}\theta=\sigma_0^2$ (θ_0 и σ_0^2 — заданы)	$(\theta_0';\sigma_0'^2)$-нормальный, где $\theta_0'=\frac{\bar{x}+\gamma\theta_0}{1+\gamma}$ и $\sigma_0'^2=\sigma^2/n(1+\gamma)$, а $\gamma=\sigma^2/n\sigma_0^2$
2	Экспоненциальный $f(x\mid\theta)=\begin{cases}\theta e^{-\theta x} & \text{при } x\geq 0\\ 0 & \text{при } x<0\end{cases}$	$p(\theta)=\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}\quad(\theta>0)$ — гамма-распределение; $\mathbf{E}\theta=\alpha/\beta;\ \mathbf{D}\theta=\alpha/\beta^2$ (α и β — заданы)	Гамма-распределение с параметрами $\alpha'=\alpha+n$, $\beta'=\beta+\sum_{i=1}^{n}x_i$
3	$[0;\theta]$-равномерный: $f(x\mid\theta)=\begin{cases}\frac{1}{\theta} & \text{для } 0\leq x\leq\theta\\ 0 & \text{для } x\notin[0;\theta]\end{cases}$	$p(\theta)=\begin{cases}\frac{\alpha\theta_0^\alpha}{\theta^{\alpha+1}} & \text{при } x\geq 0\\ 0 & \text{при } x<0\end{cases};(\theta>0)$ — распределение Парето; $\mathbf{E}\theta=\frac{\alpha\theta_0}{\alpha-1};\ \mathbf{D}\theta=\frac{\alpha\theta_0^2}{(\alpha-1)^2(\alpha-2)}$ (α и θ_0 — заданы)	Распределение Парето с параметрами $\alpha'=\alpha+n$, $\theta_0'=\max\{\theta_0;x_1,x_2,\ldots,x_n\}$
4	Распределение Пуассона: $P\{\xi=x\}=\frac{\theta^x}{x!}e^{-\theta}$ $x=0,1,2,\ldots$	$p(\theta)=\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}\quad(\theta>0)$ — гамма-распределение; $\mathbf{E}\theta=\frac{\alpha}{\beta};\ \mathbf{D}\theta=\frac{\alpha}{\beta^2}$ (α и β — заданы)	Гамма-распределение с параметрами $\alpha'=\alpha+\sum_{i=1}^{n}x_i$, $\beta'=\beta+n$

№ пп	З.р.в. наблюдаемой генеральной совокупности	Сопряженный априорный з.р.в. $p(\theta)$, выражения для $E\theta$ и $D\theta$	Апостериорный з.р.в. $p(\theta \mid x_1, x_2, \dots, x_n)$, выражения для его параметров
5	Биномиальное распределение: $P\{\xi = x\} = C_N^x \theta^x (1-\theta)^{N-x}$ (значение параметра N известно)	$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}$ $(0 \le \theta \le 1)$ — бета-распределение; $E\theta = \frac{a}{a+b}$; $D\theta = \frac{ab}{(a+b)^2(a+b+1)}$ (a и b — заданы)	Бета-распределение с параметрами $a' = a + \sum_{i=1}^{n} x_i$, $b' = b + nN - \sum_{i=1}^{n} x_i$
6	Отрицательное биномиальное распределение $P\{\xi = x\} = C_{x-1}^{k-1} \theta^k (1-\theta)^{x-k}$ (значение параметра k известно) $x = k, k+1, \dots$	$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}$ $(0 \le \theta \le 1)$ — бета-распределение; $E\theta = \frac{a}{a+b}$; $D\theta = \frac{ab}{(a+b)^2(a+b+1)}$ (a и b — заданы)	Бета-распределение с параметрами $a' = a + kn$, $b' = b + \sum_{i=1}^{n} x_i - kn$
7	Распределение Парето $f(x \mid \theta) = \begin{cases} \frac{\theta x_0^{\theta}}{x^{\theta+1}} & \text{при } x \ge x_0 \\ 0 & \text{при } x < x_0 \end{cases}$ (значение параметра x_0 — известно)	$p(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$ $(\theta > 0)$ — гамма-распределение; $E\theta = \frac{\alpha}{\beta}$; $D\theta = \frac{\alpha}{\beta^2}$ (α и β — заданы)	Гамма-распределение с параметрами $\alpha' = \alpha + n$, $\beta' = \beta + n \ln\left(\frac{g_n}{x_0}\right)$, где $g_n = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$

Байесовский прогноз зависимой переменной, основанный на нормальной линейной модели множественной регрессии

Рассмотрим множественную линейную регрессионную модель

$$Y = X\beta + \varepsilon, \tag{18}$$

где $Y = (y_1, \dots, y_n)^T$, $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$,

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}$$

— матрица порядка $n \times (k+1)$.
Случайный вектор $\varepsilon^T \sim N(0, h^{-1}E_n)$

Введем прогнозные (на q тактов времени вперед) значения $\tilde{X}$ и $\tilde{Y}$

$$\tilde{X} = \begin{pmatrix} 1 & \tilde{x}_{(n+1)1} & \tilde{x}_{(n+1)2} & \cdots & \tilde{x}_{(n+1)k} \\ 1 & \tilde{x}_{(n+2)1} & \tilde{x}_{(n+2)2} & \cdots & \tilde{x}_{(n+2)k} \\ \cdots & \cdots & \cdots & \cdots & \\ 1 & \tilde{x}_{(n+q)1} & \tilde{x}_{(n+q)2} & \cdots & \tilde{x}_{(n+q)k} \end{pmatrix}$$

$\tilde{Y} = (\tilde{y}_{n+1}, \ldots, \tilde{y}_{n+q})^T$, а также остатки $\tilde{\varepsilon} = (\varepsilon_{n+1}, \ldots, \varepsilon_{n+q})^T$. Тогда с учетом исходной модели (18)

$$\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}, \quad \tilde{\varepsilon}_q \sim N(0, h^{-1}E_q) \tag{19}$$

Для построения точечных и интервальных оценок для $\tilde{Y}$ по заданным значениям X, $\tilde{X}$, Y необходимо располагать *прогнозной функцией плотности вероятности* $p(\tilde{Y}|X,\tilde{X},Y)$:

$$p(\tilde{Y}|X,\tilde{X},Y) = \int_\beta \int_h p(\tilde{Y},\beta,h|X,\tilde{X},Y)d\beta dh = \\ = \int_\beta \int_h p(\tilde{Y}|\beta,h,X,\tilde{X},Y)p(\beta,h|X,\tilde{X},Y)d\beta dh \quad (20)$$

С учетом того, что

$$p(\tilde{Y}|\beta,h,X,\tilde{X},Y) = p(\tilde{Y}|\beta,h,\tilde{X}) \propto h^{\frac{q}{2}} e^{-\frac{h}{2}(\tilde{Y}-\tilde{X}\beta)^T(\tilde{Y}-\tilde{X}\beta)} \quad (21)$$

$p(\beta,h|X,\tilde{X},Y) = p(\beta,h|X,Y)$ —гамма-нормальное распределение с параметрами $\tilde{\beta}_0$, $\tilde{\Lambda}_0$, $\tilde{a}$ и $\tilde{b}$, определяемыми по параметрам β_0, Λ_0, a и b априорного гамма-нормального распределения $p(\beta,h)$ по формулам

$$\tilde{\theta}_0 = (\Lambda_0 + X^TX)^{-1}(\Lambda_0\theta_0 + X^TY); \qquad \tilde{\Lambda}_0 = \Lambda_0 + X^TX$$

$$\tilde{a} = a + n/2; \qquad \tilde{b} = b + 0.5[(Y - X\tilde{\theta}_0)^TY + (\theta_0 - \tilde{\theta}_0)^T\Lambda_0\theta_0]$$

интегригруя (20), получаем

$$p(\tilde{Y}|X,\tilde{X},Y) \propto \left[1+\frac{1}{v}(\tilde{Y}-\tilde{X}\tilde{\beta}_0)^T B(\tilde{Y}-\tilde{X}\tilde{\beta}_0)\right]^{-\frac{v+q}{2}}, \qquad (22)$$

где $v = n - k - 1$ и $B = \frac{\tilde{a}}{\tilde{b}}\left[E_q - \tilde{X}(\Lambda_0 + X^T X + \tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\right]$.

Таким образом условное распределение $\tilde{Y}$ при заданных значениях X, $\tilde{X}$, Y описывается обобщенным многомерным t-распределением с $n-k-1$ степенями свободы, параметром сдвига $\tilde{X}\tilde{\beta}_0$ и матрицей точности B.

Точный байесовский прогноз для компонент вектора $\tilde{Y}$

$$\hat{y}_{n+m} = (\hat{\beta}^B)^T X_{n+m}, \quad m = 1, \ldots, q. \tag{23}$$

Интервальный байесовский прогноз для компонент вектора $\tilde{Y}$ с доверительной вероятностью γ, $m = 1, \ldots, q$,

$$y_{n+m} \in \left(\hat{y}_{n+m} - t_{\frac{\gamma}{2}}(n-k-1)\frac{1}{\sqrt{c_m}}; \;\; \hat{y}_{n+m} + t_{\frac{\gamma}{2}}(n-k-1)\frac{1}{\sqrt{c_m}} \right). \tag{24}$$

Проверка статистических гипотез

Пусть $(x_1, \dots, x_n)$ — выборка из генеральной совокупности ξ с законом распределения $f(x, \theta)$ известным с точностью до неизвестного параметра θ.

$$(x_1, \dots, x_n)|\theta \sim L(x_1, \dots, x_n|\theta)$$

Проверим нулевую гипотезу H_0 о принадлежности неизвестного параметра θ некоторому множеству Θ_0 против альтернативной гипотезы H_1 о принадлежности параметра θ множеству Θ_1, где

$$\Theta_0 \cap \Theta_1 = \varnothing \quad \Theta_0 \cup \Theta_1 = \Theta.$$

Предположим, что имеется априорная информация о распределении вероятности параметра θ

$$\pi_0 = Pr\{\theta \in \Theta_0\}, \quad \pi_1 = Pr\{\theta \in \Theta_1\} \tag{25}$$

Пусть $Pr(H_0)$, $Pr(H_1)$ — априорные вероятности справедливости гипотез H_0 и H_1,соотвественно.

Пусть

$$p_0 = Pr\{\theta \in \Theta_0 | X_{[n]}\}, \quad p_1 = Pr\{\theta \in \Theta_1 | X_{[n]}\} \tag{26}$$

— апостериорные вероятности по данным наблюдений $(x_1, \ldots, x_n)$ того, что параметр θ принадлежит множествам, соотвествующим нулевой гипотезе: p_0, и альтернативной: p_1.
Априорные шансы H_0 против H_1 — π_0/π_1, апостериорные — p_0/p_1.

Байесовским фактором B_{01} гипотезы H_0 против гипотезы H_1 называется отношение апостериорных шансов к априорным шансам

$$B_{01} = \frac{p_0/p_1}{\pi_0/\pi_1} = \frac{p_0\pi_1}{p_1\pi_0}. \tag{27}$$

Так как $\pi_1 = 1 - \pi_0$ и $p_1 = 1 - p_0$, имеем

$$B_{01} = \frac{p_0(1-\pi_0)}{(1-p_0)\pi_0}. \tag{28}$$

$$B_{10} = \frac{1}{B_{01}}. \tag{29}$$

В случае **двух простых гипотез**

$$\Theta_0 = \theta_0, \quad \Theta_1 = \theta_1$$

апостериорные вероятности

$$p_i \propto \pi_i p(x_1, \ldots, x_n | \theta_i), \quad i = 0, 1. \tag{30}$$

Тогда

$$\frac{p_0}{p_1} = \frac{\pi_0 p(x_1, \ldots, x_n | \theta_0)}{\pi_1 p(x_1, \ldots, x_n | \theta_1)} \tag{31}$$

и Байесовский фактор принимает вид

$$B_{01} = \frac{p(x_1, \ldots, x_n | \theta_0)}{p(x_1, \ldots, x_n | \theta_1)}, \tag{32}$$

что есть просто отношение правдоподобия.

Общий случай

Функция правдоподобия при условии справедливости гипотезы H_i, $i = 0, 1$:

$$L(x_1, \ldots, x_n | H_i) = \int_{\Theta_i} f(x_1, \ldots, x_n | \theta) \pi_i(\theta) d\theta, \quad i = 0, 1$$

Байесовский фактор гипотезы H_0 против гипотезы H_1

$$B_{01} = \frac{L(x_1, \ldots, x_n | H_0)}{L(x_1, \ldots, x_n | H_1)}. \tag{33}$$

Апостериорное распределение гипотез

$$Pr(H_0 | x_1, \ldots, x_n) = \frac{Pr(H_0) L(x_1, \ldots, x_n | H_0)}{Pr(H_0) L(x_1, \ldots, x_n | H_0) + Pr(H_1) L(x_1, \ldots, x_n | H_1)}, \tag{34}$$

$$Pr(H_1 | x_1, \ldots, x_n) = 1 - Pr(H_0 | x_1, \ldots, x_n).$$

Формулу для байесовского фактора можно переписать в виде

$$\frac{Pr(H_0|x_1,\ldots,x_n)}{Pr(H_1|x_1,\ldots,x_n)} = \frac{Pr(H_0)}{Pr(H_1)} \cdot B_{01},$$

откуда получаем соотношение

$$Pr(H_0|x_1,\ldots,x_n) = \left[1 + \frac{Pr(H_1)}{Pr(H_0)}\frac{1}{B_{01}}\right]^{-1}. \qquad (35)$$

Выводы из апостериорных вероятностей

Нулевая гипотеза H_0 принимается, если

$$Pr(H_0|x_1,\ldots,x_n) > Pr(H_1|x_1,\ldots,x_n).$$

Выводы из байесовского фактора

Джефрис предложил следующую шкалу

B_{01}	Сила доказательств
$[1, 3]$	не стоит отмечать
$(3, 10]$	существенная
$(10, 30]$	сильная
$(30, 100]$	очень сильная
> 100	решающая

Решающее правило

Выбираем между a_0: "принимаем H_0"и a_1: "принимаем H_1"
Рассмотрим 0-1 функцию потерь

$$L(\theta, a_i) = \begin{cases} 0, & \text{если} \;\; \theta \in \Theta_i \\ 1, & \text{если} \;\; \theta \in \Theta_j, \;\; j \neq i \end{cases} \tag{36}$$

Оптимальное правило минимизирует ожидаемые апостериорные потери

$$E_{\pi(\theta|X_{[n]})}(L(\theta, a_1)) = \int L(\theta, a_1)\pi(\theta|X_{[n]})d\theta = Pr(H_0|x_1, \ldots, x_n), \tag{37}$$

$$E_{\pi(\theta|X_{[n]})}(L(\theta, a_0)) = \int L(\theta, a_0)\pi(\theta|X_{[n]})d\theta = Pr(H_1|x_1, \ldots, x_n). \tag{38}$$

Тогда предпочитаем $a_0 \succ a_1$ тогда и только тогда, когда

$$E_{\pi(\theta|X_{[n]})}(L(\theta, a_0)) < E_{\pi(\theta|X_{[n]})}(L(\theta, a_1)),$$

что равносильно

$$Pr(H_1|x_1, \dots, x_n) < Pr(H_0|x_1, \dots, x_n),$$

т.е. выбираем наиболее вероятную гипотезу.

Рассмотрим $0 - K_i$ функцию потерь

$$L(\theta, a_i) = \begin{cases} 0, & \text{если} \;\; \theta \in \Theta_i \\ K_i, & \text{если} \;\; \theta \in \Theta_j, \;\; j \neq i \end{cases} \tag{39}$$

Оптимальное решение есть a_1 (отклоняем H_0) тогда и только тогда, когда

$$\frac{Pr(H_0|x_1, \dots, x_n)}{Pr(H_1|x_1, \dots, x_n)} < \frac{K_0}{K_1}$$

Литература

- Chibara L., Hesterberg T.Mathematical statistics with resampling and R. Wiley
- Айвазян С.А., Мхитарян В.С.Прикладная статистика. Основы эконометрики. Т.1, 2001
- Айвазян С.А.Байесовский подход в эконометрическом анализе // Прикладная эконометрика, 2008, № 1(9), стр. 93–108
- Боровков А.А.Математичсекая статистика. Оценка параметров. Проверка гипотез. М.:Наука, 1984
- Jean-Michel Marin, Christian P. RobertBayesian Core: A Practical Approach to Computational Bayesian Statistics. Springer, 2007