Глава 7

Специальные методы построения статистических решений

7.1 Байесовский подход в статистическом прогнозировании

Сформулируем в этом разделе кратко общую модель байесовского прогнозирования, а для двух частных случаев – трендовой и авторегрессионной моделей временных рядов – построим байесовские прогнозирующие статистики.

7.1.1 Общая модель байесовского прогнозирования

Сформулируем общую вероятностную модель байесовского статистического прогнозирования.

Пусть задано некоторое вероятностное пространство $(\Omega, \mathcal{F}, \mathrm{P})$. На этом вероятностном пространстве определены следующие три случайных элемента (см. рис. 7.1):

• ненаблюдаемый вектор параметров $\theta = (\theta_i) \in \Theta \subseteq \mathbb{R}^m$, истинное значение которого не известно и является случайным вектором с априорной плотностью распределения вероятностей (п. р. в.) $\pi^0(\theta)$;

• стохастически зависящий от θ случайный вектор наблюдений $x = (x_t) \in X \subseteq \mathbb{R}^T$ (описывающий, например, «прошлое» и «настоящее» состояния исследуемого явления) с условной п. р. в. $p^0(x \mid \theta)$ при заданном векторе параметров θ;

• неизвестная, подлежащая прогнозированию величина $y \in Y \subseteq \mathbb{R}$ (к примеру, «будущее» состояние исследуемого явления), стохастически зависящая от векторов θ, x и распределенная согласно условной п. р. в. $g^0(y \mid x, \theta)$ при фиксированных векторах наблюдений x и параметров θ.

Задача заключается в построении прогноза $\hat{y}$ для величины y по наблюдениям x с использованием гипотетической информации, заданной в виде плотностей распределения вероятностей $\pi^0(\cdot)$, $p^0(\cdot)$, $g^0(\cdot)$. Стохастическая связь вектора параметров θ с вектором наблюдений x и величиной y позволяет прогнозировать y по x в условиях неопределенности θ, то есть строить статистический прогноз: $\hat{y} = f(x)$, где $\hat{y} \in Y$ — значение прогноза для $y \in Y$, а $f(\cdot)\colon X \longrightarrow Y$ — некоторая статистика, то есть борелевская функция T переменных.

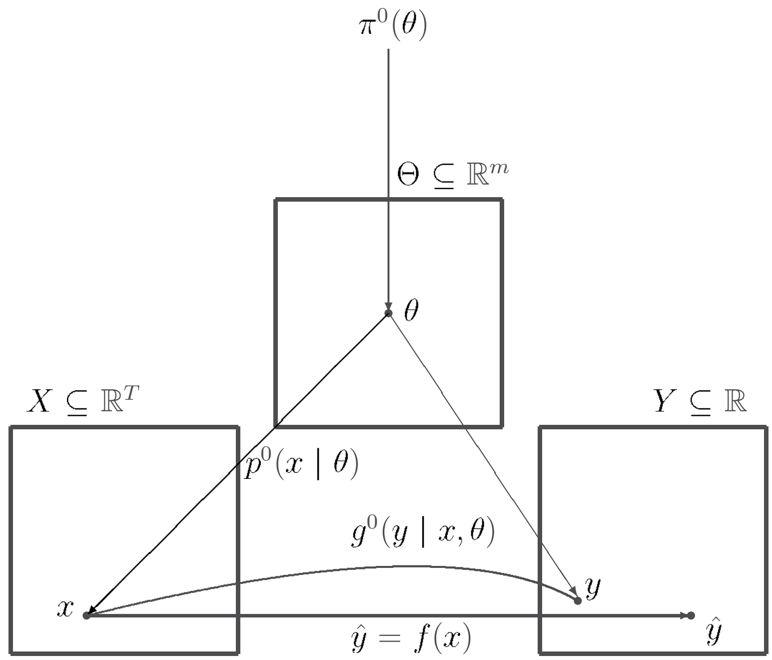

Рис. 7.1: Математическая модель байесовского прогнозирования

В случае, когда пространство Y является дискретным, задача статистического прогнозирования совпадает с задачей классификации. Вследствие допустимости применения в указанной ситуации хорошо разработанных методов теории статистической классификации решение такой задачи существенно упрощается по сравнению с общим случаем. Это привело к достаточной полноте исследований в области эффективности байесовской классификации, а также её устойчивости при наличии искажений.

Значительно более сложную задачу представляет собой задача прогнозирования в случае, когда множество Y имеет мощность континуума. В этой ситуации для нахождения и исследования свойств байесовских прогнозирующих статистик эффективно используется аппарат байесовской прогнозной плотности.

Определение 7.1 ***Байесовской прогнозной плотностью*** *называется условная п. р. в.*

$$q^0(y \mid x) = \int_{\Theta} g^0(y \mid x, \theta)\pi^0(\theta \mid x)d\theta, \tag{7.1}$$

где апостериорная п. р. в. $\pi^0(\theta \mid x)$ *вычисляется через известные плотности согласно теореме Байеса:*

$$\pi^0(\theta \mid x) = \frac{p^0(x \mid \theta)\pi^0(\theta)}{p^0(x)}, \ p^0(x) = \int_{\Theta} p^0(x \mid \theta)\pi^0(\theta)d\theta. \tag{7.2}$$

Для упрощения обозначений будем полагать, что $X = \{x : \ p^0(x) > 0\}$.

Отметим, что иногда рассматривается частный случай сформулированной выше вероятностной модели, когда x и y — условно независимые при заданном векторе θ, и определение байесовской прогнозной плотности дается лишь для этого случая:

$$q^0(y \mid x) = \int\limits_{\Theta} g^0(y \mid \theta)\pi^0(\theta \mid x)d\theta. \tag{7.3}$$

Выражение (7.3) очевидно является частным случаем (7.1).

Для удобства применения теории экстремумов будем предполагать, что множества X и Θ являются компактными. Кроме того, как принято в робастной статистике, при вычислении экстремальных значений по пространству $\mathbb{R}^k$, $k \in \mathbb{N}$, условимся пополнять его бесконечно удаленными точками, то есть заменять его множеством $\bar{\mathbb{R}}^k$, где $\bar{\mathbb{R}} = [-\infty, +\infty]$.

Через $\mathbf{E}_0\{\cdot\}$, $\mathbf{D}_0\{\cdot\}$ будем обозначать соответственно математическое ожидание и дисперсию.

Определение 7.2 ***Байесовской прогнозирующей статистикой*** *будем называть статистику, представляющую собой условное математическое ожидание, вычисленное по байесовской прогнозной плотности:*

$$\hat{y} = f_0(x) = \mathbf{E}_0\{y \mid x\} = \int\limits_{Y} yq^0(y \mid x)dy,\ x \in X, \tag{7.4}$$

а ***байесовским прогнозом*** *— значение статистики (7.4) при заданном векторе наблюдений x.*

Точность прогнозирования величины y по вектору наблюдений x будем характеризовать среднеквадратическим риском (среднеквадратической ошибкой) прогнозирования.

Определение 7.3 ***Среднеквадратическим риском*** *прогнозирующей статистики $f(\cdot)\colon X \longrightarrow Y$ будем называть значение функционала*

$$r_0(f(\cdot)) = \mathrm{E}_0(f(x) - y)^2 = \int\limits_{X}\int\limits_{Y} s^0(x, y)\left(f(x) - y\right)^2 dydx \geq 0, \tag{7.5}$$

где совместная п. р. в. $s^0(x, y)$ вычисляется через известные п. р. в. следующим образом:

$$s^0(x, y) = \int\limits_{\Theta} g^0(y \mid x, \theta)p^0(x \mid \theta)\pi^0(\theta)d\theta. \tag{7.6}$$

Свойство 7.1 *Минимум функционала (7.5) достигается для байесовской прогнозирующей статистики (7.4):*

$$r_0 = \inf_{f(\cdot)} r_0(f(\cdot)) = r_0(f_0(\cdot)). \tag{7.7}$$

Определение 7.4 *Наименьшее возможное значение (7.7) функционала среднеквадратического риска (7.5) называется* ***байесовским риском прогнозирования****.*

Отметим, что рассмотренный выше подход к прогнозированию обладает одним очень важным преимуществом: для получения прогноза величины y нет необходимости предварительно решать задачу оценивания m-мерного вектора параметров θ, которая на практике сама по себе оказывается достаточно сложной задачей.

Рассмотрим теперь два важных для практики частных случая сформулированной выше общей вероятностной модели байесовского прогнозирования, возникающих при прогнозировании временных рядов (ВР) — трендовую и авторегрессионную модели ВР. Построим байесовские прогнозирующие статистики для этих случаев и исследуем их свойства.

7.1.2 Байесовское прогнозирование временных рядов с трендом

Математическая модель

Обозначим через $S_{i,j}$ $(i, j \in \mathbb{N},\ i < j)$ следующее упорядоченное множество:

$$S_{i,j} = \{1, 2, \dots, i-1, i, j\}.$$

Вероятностная модель ВР с трендом имеет вид

$$x_t = \theta' \psi(t) + \xi_t,\ \ t \in S_{T,T+\tau}, \tag{7.8}$$

где T — длина наблюдаемого временного ряда (число наблюдений); $\tau > 0$ — длина интервала прогнозирования (на сколько отсчетов времени вперед требуется прогноз); $\{\xi_t\ :\ \ t \in S_{T,T+\tau}\}$ — случайные ошибки наблюдения, независимые в совокупности и одинаково распределенные:

$$\{\xi_t\ :\ t \in S_{T,T+\tau}\}\ \ -\ \text{i. i. d.},\ \ \mathcal{L}\{\xi_t\} = \mathcal{N}_1(0, \sigma^2); \tag{7.9}$$

$\psi(t) = (\psi_1(t), \psi_2(t), \dots, \psi_m(t))'$ — заданная неслучайная вектор-функция дискретного времени, $\psi_1(t)$, $\psi_2(t)$, ..., $\psi_m(t)$ — линейно независимые базисные функции тренда; $\theta = (\theta_1, \theta_2, \dots, \theta_m)' \in \Theta = \mathbb{R}^m$ — истинное значение ненаблюдаемого вектора параметров, который полагается случайным с m-мерной гауссовской п. р. в.

$$\pi^0(\theta) = n_m(\theta \mid a, A) = (2\pi)^{-\frac{m}{2}} |A|^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\theta - a)' A^{-1} (\theta - a)\}. \tag{7.10}$$

В (7.10) $a = \mathrm{E}_0\{\theta\}$ — m-вектор математического ожидания,

$$A = \mathrm{Cov}(\theta, \theta) = A' \succ 0\ -$$

$(m \times m)$-ковариационная матрица распределения вероятностей вектора параметров θ.

Из (7.8), (7.10) следует, что $X = \mathbb{R}^T$, $Y = \mathbb{R}$. Задача заключается в построении прогноза величины $y = x_{T+\tau}$ по T наблюдаемым значениям временного ряда (7.8) при выполнении условий (7.9), (7.10).

Отметим, что модель временного ряда с трендом (7.8) — (7.10) в математической литературе имеет и другое название — параметрическая линейная (по параметрам θ) регрессия.

Построение байесовской прогнозирующей статистики

Прежде всего заметим, что для рассматриваемой модели (7.8) — (7.10) случайные элементы x и y являются условно независимыми при заданном векторе θ. Как следствие, байесовская прогнозная плотность допускает более простое по сравнению с общим случаем представление (7.3). Получим вначале следующий вспомогательный результат.

Лемма 7.1 *Пусть $p \in \mathbb{N}$, Ξ — произвольная $(p \times p)$-симметричная положительно определенная матрица:*

$$\Xi = \Xi' \succ 0. \tag{7.11}$$

Тогда для произвольного вектора $\xi \in \mathbb{R}^p$ справедливо неравенство

$$\xi'(\Xi + \xi\xi')^{-1}\xi < 1. \tag{7.12}$$

Введем следующие обозначения:

$$\Psi(t) = \psi(t)\psi'(t),\ t \in S_{T,T+\tau};$$

$$\Gamma = \sum_{t \in S_{T,T+\tau}} \Psi(t);\ \zeta(x) = \sum_{t=1}^{T} x_t \psi(t). \tag{7.13}$$

Теорема 7.1 *Если вероятностная модель определяется (7.8) — (7.10), то байесовская прогнозирующая статистика (7.4) имеет следующее представление:*

$$f_0(x) = \frac{\psi'(T+\tau)\left(\sigma^2 A^{-1} + \Gamma\right)^{-1}\left(\sigma^2 A^{-1} a + \zeta(x)\right)}{1 - \psi'(T+\tau)\left(\sigma^2 A^{-1} + \Gamma\right)^{-1}\psi(T+\tau)},\ x \in \mathbb{R}^T. \tag{7.14}$$

Исследование свойств байесовской прогнозирующей статистики

Укажем вначале одно важное свойство байесовской прогнозирующей статистики, имеющее большое значение при ее практическом использовании и программной реализации.

Следствие 7.1 *В условиях теоремы 7.1 байесовская прогнозирующая статистика (7.14) устойчива к вычислительным погрешностям, обусловленным особенностями компьютерной арифметики.*

Отметим, что в теории статистического оценивания параметров существует обширный класс оценок — так называемые «гребневые» (ridge) оценки параметров, которые, как правило, не обладают свойством оптимальности в смысле минимума среднего квадрата ошибки оценивания, но являются устойчивыми к вычислительным погрешностям. Преимущество байесовской прогнозирующей статистики (7.14) состоит в том, что она обладает свойством оптимальности и свойством устойчивости к вычислительным погрешностям.

Исследуем теперь два частных случая асимптотического поведения байесовской прогнозирующей статистики (7.14). Положим

$$m = 1, \ \psi(t) \equiv 1, \tag{7.15}$$

то есть рассмотрим одномерный (по параметру) случай ВР с трендом. (В этом случае тренд определяется константой — параметром $\theta = \theta_1$; матрица A — число.)

Следствие 7.2 *Если выполнены условия теоремы для случая (7.15), то при $A \longrightarrow 0$ ($A^{-1} \longrightarrow \infty$) справедлив следующий асимптотический результат:*

$$\hat{x}_{T+\tau} = f_0(x) \stackrel{\text{п. н.}}{\longrightarrow} a.$$

Следствие 7.3 *Пусть для случая (7.15) выполнены условия теоремы. Тогда в асимптотике $A \longrightarrow \infty$ ($A^{-1} \longrightarrow 0$) имеет место следующее соотношение:*

$$\hat{x}_{T+\tau} \stackrel{\text{п. н.}}{\longrightarrow} \frac{1}{T}\sum_{t=1}^{T} x_t.$$

Следствие 7.2 фактически утверждает, что при увеличении априорной информации о параметре ($\theta \stackrel{\text{п. н.}}{\longrightarrow} a$) байесовская прогнозирующая статистика для ВР с трендом (7.8) — (7.10) совпадает с априорным средним значением параметра. Следствие 7.3 утверждает, что для другого крайнего случая (при отсутствии априорной информации) байесовская прогнозирующая статистика совпадает с выборочным средним значением наблюдений.

7.1.3 Байесовское прогнозирование авторегрессионных временных рядов

Рассмотрим теперь другую часто встречающуюся в практических задачах вероятностную модель — авторегрессионный временной ряд. Отметим только, что в настоящее время байесовские методы анализа авторегрессионных временных рядов интенсивно используются для улучшения разборчивости речи и качества аудиосигналов, передаваемых по каналам связи.

Математическая постановка задачи

Выше исследовалась вероятностная модель, в которой вектор x и величина y были условно статистически независимыми при заданном векторе θ. Авторегрессионная модель ВР является более сложной моделью, в которой указанное свойство не выполняется.

Пусть наблюдается T значений $x = (x_1, x_2, \dots, x_T)'$ стационарного временного ряда, описываемого разностным уравнением авторегрессии порядка $m \in \mathbb{N}$ (АР(m)) с независимыми остатками $\{\xi_t\}$:

$$x_t + \sum_{i=1}^{m} \theta_i x_{t-i} = \xi_t,\ t \in S_{T,T+1} = \{1, 2,\ \dots, T+1\}; \tag{7.16}$$

$$\{\xi_t\}\ -\ \text{i. i. d.},\ \mathcal{L}\{\xi_t\} = \mathcal{N}_1(0, \sigma^2),\ t \in S_{T,T+1}. \tag{7.17}$$

Ненаблюдаемый истинный вектор параметров $\theta = (\theta_1, \theta_2,\ \dots, \theta_m)'$ полагается случайным с m-мерной гауссовской п. р. в. (7.10):

$$\pi^0(\theta) = n_m(\theta \mid a, A) = (2\pi)^{-\frac{m}{2}} |A|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\theta - a)' A^{-1} (\theta - a)\right\},$$

$$\theta \in \Theta = \mathbb{R}^m,$$

где вектор a и матрица A — параметры нормального распределения вероятностей.

Начальные значения

$$x_{-m+1}, x_{-m+2}, \dots, x_0 \tag{7.18}$$

авторегрессионного процесса (7.16), а также дисперсия σ^2 предполагаются известными. Из (7.16), (7.17) следует, что $X = \mathbb{R}^m$, $Y = \mathbb{R}$. Задача состоит в байесовском прогнозировании будущего значения временного ряда $y = x_{T+1}$ ($\tau = 1$) по вектору x наблюдаемых значений в условиях вероятностной модели (7.16), (7.17), (7.10).

Байесовская прогнозирующая статистика

Примем следующее обозначение:

$$\tilde{x}_{t-1} = (x_{t-1}, x_{t-2}, \dots, x_{t-m})' \in \mathbb{R}^m,\ t \in S_{T,T+1}. \tag{7.19}$$

Теорема 7.2 *Если гипотетическая вероятностная модель определяется (7.16), (7.17), (7.10), то при заданных начальных значениях (7.18) гипотетический байесовский прогноз определяется прогнозирующей статистикой*

$$\hat{y} = f_0(x) =$$

$$= \frac{\tilde{x}'_T \left(\sigma^2 A^{-1} + \sum_{i=0}^{T} \tilde{x}_i \tilde{x}'_i\right)^{-1} \left(\sum_{i=1}^{T} x_i \tilde{x}_{i-1} - \sigma^2 A^{-1} a\right)}{1 - \tilde{x}'_T \left(\sigma^2 A^{-1} + \sum_{i=0}^{T} \tilde{x}_i \tilde{x}'_i\right)^{-1} \tilde{x}_T}. \tag{7.20}$$

Анализ свойств байесовской прогнозирующей статистики

Исследуем свойства байесовской прогнозирующей статистики $f_0(\cdot)$, построенной выше.

Следствие 7.4 *Если в условиях теоремы 7.2* $A \to \mathbf{0}_m$, *то*

$$\hat{y} = f_0(x) \xrightarrow{\text{п. н.}} -\tilde{x}'_T a.$$

Как видно из следствия 7.4, байесовский прогноз тем ближе к оптимальному прогнозу по вектору априорных средних коэффициентов, чем точнее априорная информация о векторе θ.

Следствие 7.5 *Если в условиях теоремы 7.2* $\left|\sum_{t=0}^{T-1} \tilde{x}_t \tilde{x}'_t\right| \neq 0$, *то при* $A^{-1} \to \mathbf{0}_m$

$$\hat{y} = f_0(x) \xrightarrow{\text{п. н.}} f_{\text{ОМП}}(x) = \tilde{x}'_T \left(\sum_{t=0}^{T-1} \tilde{x}_t \tilde{x}'_t\right)^{-1} \sum_{t=1}^{T} x_t \tilde{x}_{t-1}. \tag{7.21}$$

Следствие 7.5 утверждает, что при отсутствии априорной информации о случайном векторе коэффициентов авторегрессии гипотетический байесовский прогноз авторегрессионного временного ряда совпадает с подстановочным прогнозом, основанным на оценке максимального правдоподобия (ОМП). Для данной модели ОМП вектора коэффициентов совпадает с оценкой по методу наименьших квадратов (МНК-оценкой).

Следствие 7.6 *В условиях теоремы 7.2 байесовская прогнозирующая статистика (7.20) устойчива к вычислительным погрешностям, обусловленным особенностями компьютерной арифметики.*