ГЛАВА 8

БАЙЕСОВСКИЕ МЕТОДЫ ЭКОНОМЕТРИЧЕСКОГО АНАЛИЗА

8.1. БАЙЕСОВСКИЙ ПОДХОД К ОЦЕНИВАНИЮ ПАРАМЕТРОВ И ПРОГНОЗИРОВАНИЮ

Для статистического оценивания параметров моделей и прогнозирования будущих значений наблюдаемых показателей в современной эконометрике часто применяется байесовский подход [1] (см. также Зельнер, Bauwens, Lubrano & Richard). Это обусловлено двумя важными особенностями его использования [Ллойд & Ледерман, Berger].

Во-первых, применение байесовского подхода позволяет учесть не только данные статистических наблюдений, но и доступную априорную информацию (например, данные о пределах изменения каких-либо анализируемых показателей, знания экспертов и т.п.). В результате повышается фактическая точность статистических выводов (достигается оптимальность по критерию минимума среднего квадрата ошибки) и расширяется область их возможного применения (например, обеспечивается меньший объем выборки, чем требуется для использования классических методов, которые исключают априорную информацию из рассмотрения). В ряде случаев использование байесовского подхода наделяет получаемые статистические решения важным для практики свойством – робастностью (устойчивостью к искажениям модели) [Bayesian Robustness].

Во-вторых, современный уровень развития компьютерной техники и информационных технологий позволяет с приемлемыми затратами проводить необходимые сложные вычисления, присущие байесовским методам. Это позволяет реализовать их программно и использовать при решении задач на практике, получать количественные результаты.

Для простоты вначале рассмотрим использование байесовского подхода при оценивании параметров моделей (см. рис. 8.1). Итак, априорная информация о случайном параметре θ (он может быть как скалярным, так и векторным: $\boldsymbol{\theta} = (\theta_i) \in \Theta \subseteq \Re^m$, $m \in N$) должна быть представлена в виде некоторой функции π (θ). Если область Θ определения параметра θ дискретна, то π(θ) интерпретируется как вероятность того, что параметр примет значение, равное θ.

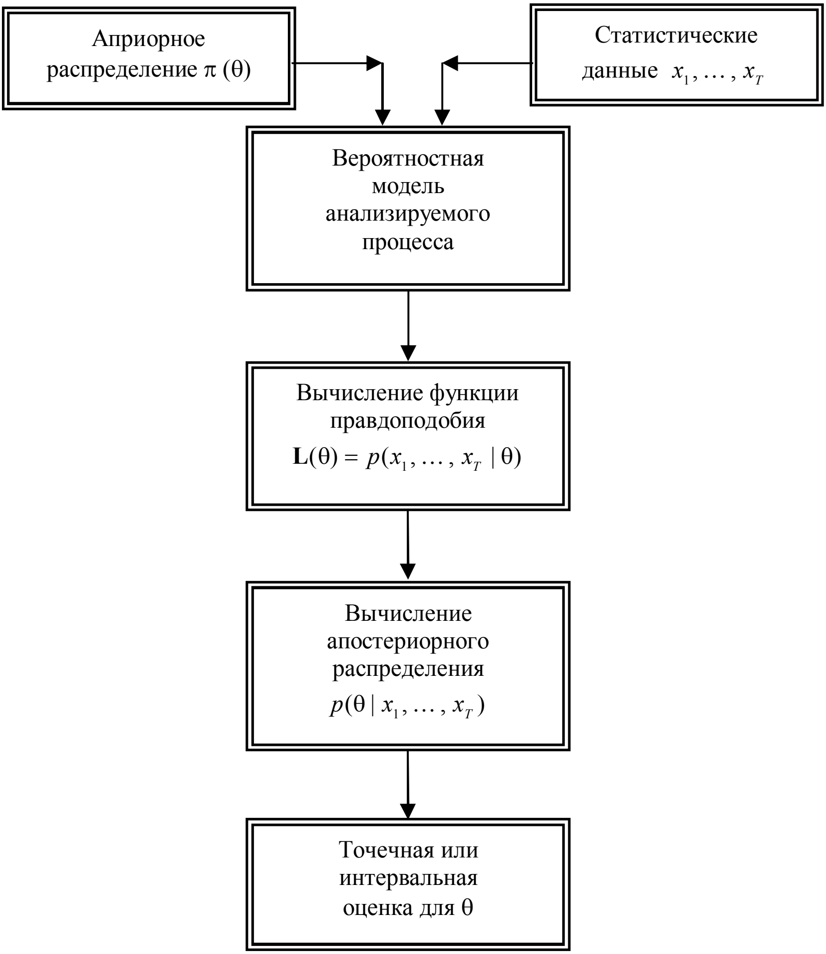

Рис. 8.1. Схема байесовского подхода при оценивании параметров модели

Если параметр по смыслу непрерывен, то $\pi(\theta)$ – плотность распределения вероятностей (п.р.в.) в точке θ.

Следует отметить, что на практике задание функции $\pi(\theta)$ является весьма непростой задачей. В эконометрических приложениях часто априорная информация о распределении параметра θ невелика. Например, в результате предварительной работы с экспертами могут быть известны лишь $\theta_{i_-}, \theta_{i_+}$ – соответственно наименьшее и наибольшее возможные значения i-го параметра ($i = \overline{1, m}$). В таких случаях принято полагать априорное распределение θ_i равномерным: $\mathbf{L}\{\theta_i\} = \mathbf{R}[\theta_{i_-}, \theta_{i_+}]$. Тем не менее, использование даже такой априорной информации избавляет от необходимости рассматривать в качестве оценок заведомо невозможные значения.

Под статистическими данными здесь мы понимаем выборку $x_1, \ldots, x_T$ наблюдений за каким-либо показателем. Используя их, мы можем уточнить вероятностную модель анализируемого процесса, добавив к уже имеющейся априорной информации о параметре полученную информацию о наблюдаемом показателе x, который стохастически связан с θ. Эта связь проявляется в задании условного распределения вероятностей $P\{x \mid \theta\}$.

Напомним, что в абсолютно непрерывном случае (относительно x) в силу независимости и одинаковой распределенности (н.о.р.) выборочных значений $x_1, \ldots, x_T$ функция правдоподобия [6] может быть вычислена по формуле

$$\mathbf{L}(\theta) = p(x_1, \ldots, x_T \mid \theta) = p(x_1 \mid \theta) \ldots p(x_T \mid \theta), \; \theta \in \Theta. \qquad (8.1)$$

В случае, когда показатель x дискретен, все условные п.р.в. в (8.1) следует заменить на соответствующие условные вероятности.

Чтобы вычислить апостериорное распределение параметра θ (в непрерывном случае – п.р.в. $p(\theta \mid x_1, \ldots, x_T)$), воспользуемся известными [**Ллойд & Ледерман**] классическими результатами.

Теорема 8.1. (**Байес**). Если события $H_1, \ldots, H_m$ образуют полную группу событий, $m \geq 2$, и при этом $P\{A\} > 0$ для некоторого события A, то $\forall i \in \{1, \ldots, m\}$

$$P\{H_i \mid A\} = \frac{P\{A \mid H_i\} \cdot P\{H_i\}}{\sum_{j=1}^{m} P\{A \mid H_j\} \cdot P\{H_j\}}. \qquad (8.2)$$

События $H_1, \ldots, H_m$ называют гипотезами, а соотношение (8.2) – *формулой Байеса для событий*.

Теорема 8.2 (**Байес**). Если существует п.р.в. $p(z) = \int_Y p(z \mid y) p(y) dy$ в точке z, $p(z) > 0$, где Y – множество значений СВ y, а также существуют п.р.в. $p(y),\ y \in Y$, и п.р.в. $p(z \mid y),\ y \in Y$ в точке z, то справедливо соотношение:

$$p(y \mid z) = \frac{p(z \mid y) p(y)}{p(z)}. \qquad (8.3)$$

Отметим еще, что равенство (8.3) называется *формулой Байеса для плотностей.*

Итак, в ситуации, когда случайный параметр θ может изменяться непрерывно, пользуясь теоремой 8.2, получим

$$p(\theta \mid x_1, \ldots, x_T) = \frac{L(\theta)\pi(\theta)}{\int_\Theta L(\theta)\pi(\theta) d\theta} \qquad (8.4)$$

Для дискретного случая должна быть проведена замена п.р.в. на соответствующие вероятности.

Знание апостериорного распределения (8.4) позволяет построить множество точечных оценок (в зависимости от требуемых свойств). Укажем здесь лишь три наиболее важные и часто используемые оценки (для непрерывного случая).

Первая точечная оценка:

$$\widehat{\theta}_E = \mathbf{E}\{\theta \mid x_1, \ldots, x_T\} = \int_\Theta \theta\, p(\theta \mid x_1, \ldots, x_T) d\theta \ - \qquad (8.5)$$

апостериорное среднее значение оцениваемого параметра (условное математическое ожидание). Оптимальность оценки (8.5) состоит в том, что $\widehat{\theta}_E$ минимизирует средний квадрат ошибки оценивания среди всевозможных оценок $\widetilde{\theta}$ параметра θ:

$$\widehat{\theta}_E \arg \min_{\widetilde{\theta}(x_1,\ldots,x_T)} \mathbf{E}\{|\widetilde{\theta}(x_1,\ldots,x_T)-\theta|^2\}.$$

Вторая точечная оценка – мода апостериорного распределения вероятностей:

$$\widehat{\theta}_{Mod} = \arg\max_{\theta\in\Theta} p(\theta\,|\,x_1,\ldots,x_T)\,. \tag{8.6}$$

В случае унимодального симметричного распределения вероятностей (например, гауссового) оценки (8.6) и (8.5) совпадают. В ситуациях, когда решение (8.6) не единственно, оценки $\widehat{\theta}_{Mod}$ могут иметь бóльшую ценность, чем апостериорное среднее (например, два набора параметров, соответствующих различным режимам развития экономики: $\widehat{\theta}_{Mod,1}$, $\widehat{\theta}_{Mod,2}$, весьма полезны для аналитиков; в то же время $\widehat{\theta}_E$ может оказаться ненужной).

Третья точечная оценка $\widehat{\theta}_{Me}$ (в скалярном случае, когда m = 1) является решением уравнения:

$$F(\theta\,|\,x_1,\ldots,x_T) = \frac{1}{2} \; - \tag{8.7}$$

медиана апостериорного распределения, где $F(\cdot)$ – апостериорная функция распределения. Оценка (8.7) сложна для вычисления и для аналитического исследования ее свойств. Тем не менее она интенсивно используется в робастной статистике, чтобы уменьшить влияние искажений модели на точность статистических выводов.

Знание апостериорного распределения (8.4) позволяет строить интервальные оценки для параметра θ [6]. При построении центрального доверительного интервала целесообразно в качестве его центра использовать точечную оценку (8.5).

ЗАДАНИЯ

8.1. Используя теорему 8.1, построить оценку параметра θ в дискретном случае, руководствуясь принципом максимума апостериорной вероятности.

8.2. Пусть некоторая экономическая система может находиться в одном из $k \in N$ состояний, которые характеризуется соответственно своим значением ненаблюдаемого случайного вектора параметров $\boldsymbol{\theta} \in \boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k\} \subset \Re^m$, $m \in N$. Предварительно были получены сведения, что

$$P\{\boldsymbol{\theta} = \boldsymbol{\theta}_i\} = \pi_i, i = 1, \ldots, k.$$

Кроме того, у исследователя появилась возможность сделать $T \in N$ независимых наблюдений $x_1, \ldots, x_T$ за показателем x, который стохастически связан с $\boldsymbol{\theta}$:

$$p(x \mid \boldsymbol{\theta}_i) = n_m(x \mid \boldsymbol{\theta}_i, \boldsymbol{\Sigma}) =$$

$$= \frac{1}{(2\pi)^{m/2}} \mid \boldsymbol{\Sigma}_i \mid^{-\frac{1}{2}} e^{-\frac{1}{2}(x - \boldsymbol{\theta}_i)^{\mathrm{T}} \boldsymbol{\Sigma}_i^{-1} (x - \boldsymbol{\theta}_i)}, i = 1, \ldots, k, \tag{8.8}$$

где $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\mathrm{T}}$, – заданная положительно определенная ковариационная матрица.

Составить компьютерную программу, вычисляющую апостериорные вероятности $P\{\boldsymbol{\theta} = \boldsymbol{\theta}_i \mid x_1, \ldots, x_T\}$, по заданным априорным вероятностям $\pi_i, i = 1, \ldots, k$ и по T сгенерированным наблюдениям согласно (8.8) [20]. Построить оценку для $\boldsymbol{\theta}$ по принципу максимума апостериорной вероятности (см. задание 8.1). Исследовать влияние априорных вероятностей и числа наблюдений T на точность оценивания.

Первоначальный перечень экспериментов: m = 1; k = 3; $\theta_1 = -1$, $\theta_2 = 0$, $\theta_3 = 1$; $\pi_1 = \pi_3 = 0.15$, $\pi_2 = 0.7$; $\Sigma = 1$, $\theta = \theta_2$; T = 2, 3, 5, 10, 50, 100, 1000.

8.3. Пусть некоторая компания, получающая доход от внешних обращений к своим Интернет-ресурсам, интересуется, каким будет среднесуточное число обращений θ. Два независимых эксперта A и B высказывают следующие суждения о распределении СВ θ: $\pi_A(\theta) = n_1(\theta \mid 500, 400)$; $\pi_B(\theta) = n_1(\theta \mid 400, 1600)$. Пусть по прошествии суток было зафиксировано $x_1 = 450$ обращений, причем оба эксперта на основании многочисленных наблюдений за подобными компаниями утверждают, что x_1 – реализация СВ x, при которой $p(x \mid \theta) = n_1(x \mid \theta, 1600)$.

С помощью системы компьютерной алгебры типа MATHEMATICA на одной и той же координатной плоскости построить графики априорных п.р.в. $\pi_A(\theta), \pi_B(\theta)$; график функции правдоподобия $L_1(\theta) = p(x_1 \mid \theta)$; апостериорные п.р.в. $p_A(\theta \mid x_1), p_B(\theta \mid x_1)$. Сгенерировать T–1=99 наблюдений $x_2, \dots, x_{100}$ из распределения вероятностей $\mathbf{N}_1(\theta, 1600)$. Построить графики $\mathbf{L}_{100}(\theta) = p(x_1, \dots, x_T \mid \theta)$; $p_A(\theta \mid x_1, \dots, x_T)$; $p_B(\theta \mid x_1, \dots, x_T)$. Рассмотреть еще случай с экспертным C $\pi_c(\theta) = 0.01 \cdot 1_{[450,550]}(\theta)$.

Первоначальный перечень экспериментов: $\theta = 480, 490, 410$. Изменяя T, изучить его влияние на получаемые функцию правдоподобия и апостериорную п.р.в.

8.2. БАЙЕСОВСКОЕ ОЦЕНИВАНИЕ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ

Проиллюстрируем применение байесовского подхода, изложенного в разделе 8.1, вначале для простейшего случая. Пусть наблюдается выборка $x_1, \dots, x_T$, полученная из нормального распределения:

$$\mathbf{L}\{x_i\} = \mathbf{N}_1(\mu, \sigma^2), i = 1, \dots, T, \tag{8.9}$$

причем дисперсия σ^2 предполагается известной. Тогда совместная п.р.в. выборки, согласно (8.9), задается выражением

$$\mathbf{L}(\mu) = p(\mathbf{x} \mid \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{T/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{T}(x_i - \mu)^2}, \tag{8.10}$$

где $\mathbf{x} = (x_1, \dots, x_T)^{\mathrm{T}}$. Пусть параметр μ – МО распределения (8.9) – случайная величина (СВ), про которую априорно известно, что

$$\mathbf{L}\{\mu\} = \mathbf{N}_1(m, \sigma^2 / v), \tag{8.11}$$

причем параметры m, v предполагаются заданными. Наша цель – построить байесовскую оценку для неизвестного МО μ с использованием выборки **x**, а также информации, представленной (8.10), (8.11). Несмотря на кажущуюся простоту, поставленная задача представляется важной, поскольку является составной частью многих реальных эконометрических задач.

Параметр m априорного распределения (8.11) может интерпретироваться как статистическая оценка μ, полученная до наблюдения выборки **x**. Согласно (8.11), средний квадрат ошибки для этой оценки равен σ^2/v. Таким образом, чем больше величина v, тем точнее априорная информация о параметре μ.

Для примера рассмотрим, как может сформироваться априорная информация. Пусть до наблюдения выборки **x** наблюдались N выборочных значений $y_1, \ldots, y_N$, полученные из распределения (8.9). Тогда в нашем случае естественно предположить $m = \frac{1}{N}\sum_{i=1}^{N} y_i, \quad v = N.$ При увеличении объема выборки N априорная информация уточняется (уменьшается величина σ^2/v).

Теорема 8.3. [43] Если априорное распределение неизвестного случайного параметра μ задано (8.11), а функция правдоподобия – соотношением (8.10), то апостериорная п.р.в. параметра μ равна

$$p(\mu\,|\,x) = \frac{1}{(2\pi\sigma^2/(v+T))^{1/2}} e^{-\frac{(\mu-\hat{\mu})^2}{2\sigma^2}(v+T)}, \qquad (8.12)$$

где

$$\hat{\mu} = \frac{v}{v+T} m + \frac{T}{v+T}\bar{x}, \quad \bar{x} = \frac{1}{T}\sum_{i=1}^{T} x_i. \qquad (8.13)$$

Итак, теорема 8.3 утверждает нормальность апостериорного распределения параметра μ с МО $\hat{\mu}$ и дисперсией $\frac{\sigma^2}{v+T}$. Из курса математической статистики [6] известно, что наилучшей оценкой параметра μ (по критерию минимума среднего квадрата ошибки оценивания) является условное (апостериорное) МО (8.13). Эта величина – взвешенное среднее двух величин: 1) выборочного среднего $\overline{x}$ (с весом, пропорциональным T), 2) априорного среднего m (с весом, пропорциональным v).

Знание (8.12) позволяет судить о неопределенности точечной оценки (8.13). Например, $P\left\{\mu\in\left[\hat{\mu}-\frac{2\sigma}{\sqrt{v+T}}\right],\left[\hat{\mu}+\frac{2\sigma}{\sqrt{v+T}},\right]\right\}\geq 0.95$ согласно обобщению следствия из неравенства Чебышева относительно дисперсии [6]. Кроме того, имеется возможность строить интервальные оценки для параметра μ.

Пусть требуется построить центральный доверительный интервал (ЦДИ) $[\mu_-,\mu_+]$ для параметра μ, такой что $P\{\mu\in[\mu_-,\mu_+]\}=1-\alpha$, где $1-\alpha$ – уровень доверия, $\alpha\in(0,1)$. Обозначим через $\Phi^{-1}(1-\alpha/2)$ – квантиль уровня $1-\alpha/2$ стандартного нормального закона $\mathbf{N}_1(0,1)$. Тогда искомые случайные границы ЦДИ вычисляются как $\mu_{\mp}=\hat{\mu}_{\mp}\Phi^{-1}(1-\frac{\alpha}{2})\cdot\frac{\sigma}{\sqrt{v+T}}$. Ценность интервальных оценок в эконометрических задачах велика. Они позволяют судить о том, в каких пределах будет изменяться интересующий показатель с заданной вероятностью.

ЗАДАНИЯ

8.4. С помощью имитационного моделирования получить выборку объема 100 наблюдений из распределения $\mathbf{N}_1(217,100)$. Результат сохранить в текстовом файле.

8.5. Некоторое предприятие интересуется средней ценой продаж продукции своего конкурента. Эксперт утверждает, что средняя цена СВ с распределением вероятностей $\mathbf{N}_1(220,400)$. Аналитики предприятия получили дополнительную информацию в виде ста независимых наблюдений цены продаж (данные из задания 8.4). Кроме того, им известно, что цена продажи распределена по нормальному закону с дисперсией 100. Вычислить для средней величины продаж байесовскую оценку и построить 95 % ЦДИ.

8.6. С помощью серий вычислительных экспериментов исследовать точность байесовской оценки из задания 8.5 в зависимости от точности задания априорной информации и от числа имеющихся наблюдений (объема выборки).

8.7. При малых объемах выборки провести вычислительные эксперименты по сравнительному анализу точности оценивания среднего значения по МНК и с помощью байесовского подхода.

8.8. Для экспериментальных данных (файл data 8_8) проверить гипотезу о соответствии наблюдений нормальному распределению. Исходя из смыслового описания (файл desc 8_8) данных, задать априорное распределение неизвестного МО распределения вероятностей наблюдений. Построить байесовскую оценку для МО и ЦДИ. Проанализировать точность оценивания и факторы, от которых она зависит.

8.9. Аналитически построить точечную и интервальную байесовскую оценки вектора МО в случае многомерных наблюдений, когда $\mathbf{x}_1, \ldots, \mathbf{x}_T \in \Re^n$, $\mathbf{L}\{x_i\} = \mathbf{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad i = 1, \ldots, T; \quad \mathbf{L}\{\boldsymbol{\mu}\} = \mathbf{N}_n(\mathbf{m}, \mathbf{M})$, $\mathbf{m} \in \Re^n$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\mathrm{T}}$, $\mathbf{M} = \mathbf{M}^{\mathrm{T}} \succ 0$ полагаются известными, $|\boldsymbol{\Sigma}| \neq 0$.

8.10. С помощью вычислительных экспериментов исследовать точность оценок, полученных в задании 8.9, по аналогии с заданиями 8.6, 8.7.

8.3. ОЦЕНИВАНИЕ ПАРАМЕТРОВ МОДЕЛИ ЛИНЕЙНОЙ РЕГРЕССИИ

Линейная регрессионная модель весьма часто используется в эконометрике для восстановления стохастических зависимостей между экономическими показателями. Рассмотрим модель одномерной линейной регрессии

$$y_t = \mathbf{x}_t^{\mathrm{T}} \boldsymbol{\beta} + u_t, \tag{8.14}$$

где $\mathbf{L}\{u_t\} = \mathbf{N}_1(0, \sigma^2), \quad t = 1, \ldots, T, \quad \mathbf{x}_t$ – k-вектор экзогенных переменных, $\boldsymbol{\beta}$ – k-вектор коэффициентов. Пусть $\mathbf{y} = (y_1, \ldots, y_T)^{\mathrm{T}}$ –

вектор, составленный из T наблюдений. Обозначим через $\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_T^{\mathrm{T}} \end{pmatrix}$ матрицу, составленную из k-векторов экзогенных переменных.

8.3.1. СЛУЧАЙ ИЗВЕСТНОЙ ДИСПЕРСИИ СЛУЧАЙНЫХ ОШИБОК

Будем считать, что вектор параметров $\boldsymbol{\beta}$ случаен. Пусть дисперсия σ^2 последовательности $\{u_t\}$ известна. Тогда при фиксированных $\boldsymbol{\beta}$, $\mathbf{X}$ условная п.р.в. вектора $\mathbf{y}$ есть

$$p(\mathbf{y} \mid \boldsymbol{\beta}, \mathbf{X}) = \prod_{t=1}^{T} \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-\frac{1}{2\sigma^2}(y_t - \mathbf{x}_t^{\mathrm{T}}\boldsymbol{\beta})^2} =$$

$$= \frac{1}{(2\pi\sigma^2)^{T/2}} e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}. \qquad (8.15)$$

Пусть известно, что априорное распределение случайного вектора параметров $\boldsymbol{\beta}$ есть $\mathbf{N}_k(\mathbf{m}, \sigma^2\mathbf{M})$, где $\mathbf{m} \in \Re^k$ – k-вектор МО, $\sigma^2\mathbf{M} = \sigma^2\mathbf{M}^{\mathrm{T}} \succ 0 - (k \times k)$-ковариационная матрица. Математически это означает, что

$$p(\boldsymbol{\beta}) = \frac{1}{(2\pi\sigma^2)^{k/2}} |\mathbf{M}|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\mathbf{m})^{\mathrm{T}}\mathbf{M}^{-1}(\boldsymbol{\beta}-\mathbf{m})}. \qquad (8.16)$$

Отметим, что, по аналогии с разделом 8.2, вектор $\mathbf{m}$ – априорная оценка вектора $\boldsymbol{\beta}$; чем больше диагональные элементы матрицы $\mathbf{M}$, тем меньше «точность» соответствующих компонент вектора $\mathbf{m}$.

Теорема 8.4. [43] В рамках модели (8.14)–(8.16) апостериорная п.р.в. вектора параметров равна

$$p(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}) = \frac{1}{(2\pi\sigma^2)^{k/2}} \left|\mathbf{M}^{-1} + \mathbf{X}^{\mathrm{T}}\mathbf{X}\right|^{\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\mathbf{m}^*)^{\mathrm{T}}(\mathbf{M}^{-1}+\mathbf{X}^{\mathrm{T}}\mathbf{X})(\boldsymbol{\beta}-\mathbf{m}^*)}, \qquad (8.17)$$

где

$$\mathbf{m}^* = (\mathbf{M}^{-1} + \mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}(\mathbf{M}^{-1}\mathbf{m} + \mathbf{X}^{\mathrm{T}}\mathbf{y}) . \qquad (8.18)$$

Теорема 8.4 фактически утверждает, что апостериорное распределение вектора параметров $\boldsymbol{\beta}$ является нормальным:

$$\mathbf{L}\{\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}\} = \mathbf{N}_k(\mathbf{m}^*, \sigma^2(\mathbf{M}^{-1} + \mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}) .$$

Часто на практике используется априорное предположение о том, что вектор параметров – нулевой: $\mathbf{m} = 0$. При этом степень уверенности в такой информации считается одинаковой для всех коэффициентов (компонент $\boldsymbol{\beta}$): $\mathbf{M}^{-1} = \lambda \cdot \mathbf{I}_k$ для некоторого заданного $\lambda > 0$. В этом случае байесовская оценка (8.18) превращается в

$$\mathbf{m}^* = (\lambda \cdot \mathbf{I}_k + \mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} . \qquad (8.19)$$

Полученная в частном случае оценка (8.19) относится к важному классу так называемых гребневых (*ridge*) оценок. В отличие от оценок, полученных, например, методом наименьших квадратов, такие оценки определены даже при $|\mathbf{X}^{\mathrm{T}}\mathbf{X}| = 0$; кроме того, они обладают свойством устойчивости к вычислительным погрешностям округления [18].

8.3.2. СЛУЧАЙ НЕИЗВЕСТНОЙ ДИСПЕРСИИ СЛУЧАЙНЫХ ОШИБОК

В п. 8.3.1 дисперсия σ^2 последовательности $\{u_t\}$, $t = 1, \ldots, T$ в модели линейной регрессии (8.14) предполагалась известной. На практике такое предположение выполняется редко. В рамках байесовского подхода обычно предполагают σ^2 случайной, и задают ее распределение вероятностей. Для этих целей оказывается удобным использовать гамма-распределение.

Гамма-распределением с параметрами α, λ называется распределение вероятностей абсолютно непрерывной СВ с п.р.в.

$$p(x)=\begin{cases}\dfrac{\lambda}{\Gamma(\alpha)}(\lambda x)^{\alpha-1}e^{-\lambda x}, & x\ge 0,\\ 0 \quad , & x<0;\end{cases}\quad \alpha>0,\quad \lambda>0. \qquad (8.20)$$

В (8.20) $\Gamma(\alpha)=\int_0^{+\infty} x^{\alpha-1}e^{-x}dx$ – гамма-функция, определенная для $\alpha>0$. Она обладает следующими важными свойствами:

$$\Gamma(\alpha+1)=\alpha\Gamma(\alpha),\quad \Gamma(1)=\Gamma(2)=1,\quad \Gamma(n)=(n-1)!,\quad n\in N.$$

Гамма-распределение с параметрами α,λ принято обозначать через $\gamma(\alpha,\lambda)$. Если для некоторой СВ ξ распределение вероятностей $\mathbf{L}\{\xi\}=\gamma(\alpha,\lambda)$, то

$$\mathbf{E}\{X\}=\alpha/\lambda,\quad \mathbf{D}\{X\}=\alpha/\lambda^2. \qquad (8.21)$$

Отметим еще, что при $\lambda=1,\ \alpha=N/2$, где N – натуральное, распределение $\gamma(\alpha,\lambda)$ превращается в χ^2-распределение с N степенями свободы [6].

Удобно задавать априорное распределение вероятностей для величины, обратной σ^2 : $\mathbf{L}\{\sigma^{-2}\}=\Gamma(\frac{N}{2},\frac{\lambda}{2})$, т. е.

$$p(\sigma^{-2})=\frac{(\lambda/2)^{N/2}(\sigma^{-2})^{\frac{N}{2}-1}e^{-\lambda\sigma^2/2}}{\Gamma(N/2)}. \qquad (8.22)$$

Согласно (8.21), априорная оценка для σ^{-2} равна N/λ. Если априорная информация получена в результате предварительного наблюдения некоторой выборки, то параметр N равен ее объему. При заданном отношении N/λ бо́льшие значения N соответствуют бо́льшей уверенности в априорных сведениях.

Зададим условное распределение вероятностей $\boldsymbol{\beta}$ при фиксированном σ^2 в том же виде, что и (8.16):

$$p(\boldsymbol{\beta} \mid \sigma^2) = \frac{1}{(2\pi\sigma^2)^{k/2}} \left|\mathbf{M}\right|^{-\frac{1}{2}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\mathbf{m})^{\mathrm{T}}\mathbf{M}^{-1}(\boldsymbol{\beta}-\mathbf{m})} . \qquad (8.23)$$

Теорема 8.5. [43] Пусть имеет место модель (8.14), причем справедливы предположения (8.15), (8.22), (8.23). Тогда апостериорное распределение вероятностей для вектора $\boldsymbol{\beta}$ – k-мерное t-распределение Стьюдента с N^* степенями свободы, вектором МО $\mathbf{m}^*$ и матрицей масштаба $(\lambda^* / N^*)\mathbf{M}^*$:

$$p(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}) = \frac{\Gamma((k+N^*)/2)}{(\pi N^*)^{k/2}\Gamma(N^*/2)} \left| (\lambda^* / N^*)\mathbf{M}^* \right|^{-\frac{1}{2}} \times$$

$$\times (1 + (1/N^*)(\boldsymbol{\beta}-\mathbf{m}^*)^{\mathrm{T}}((\lambda^* \mid N^*)\mathbf{M}^*)^{-1}(\boldsymbol{\beta}-\mathbf{m}^*))^{-(k+N^*)/2}; \qquad (8.24)$$

$$N^* = N + T;$$

$$\lambda^* = \lambda + (\mathbf{y}-\mathbf{X}\mathbf{b})^{\mathrm{T}}(\mathbf{y}-\mathbf{X}\mathbf{b}) + (\mathbf{b}-\mathbf{m})^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{M}^*\mathbf{X}^{\mathrm{T}}\mathbf{X}(\mathbf{b}-\mathbf{m});$$

$$\mathbf{M}^* = (\mathbf{M}^{-1} + \mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}; \mathbf{m}^* = \mathbf{M}^*(\mathbf{M}^{-1}\mathbf{m} + \mathbf{X}^{\mathrm{T}}\mathbf{y}); \mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} .$$

Теорема 8.5 позволяет вычислить байесовскую оценку вектора коэффициентов $\boldsymbol{\beta}$:

$$\mathbf{E}\{\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}\} = \mathbf{m}^* = \mathbf{M}^*(\mathbf{M}^{-1}\mathbf{m} + \mathbf{X}^{\mathrm{T}}\mathbf{y}) . \qquad (8.25)$$

Таким образом, в условиях теоремы 8.5 оценка (8.25) совпадает с оценкой, полученной при известной дисперсии σ^2.

ЗАДАНИЯ

8.11. Для модели линейной регрессии (8.14)–(8.16) в случае, когда дисперсия σ^2 известна, построить оценку вектора параметров $\boldsymbol{\beta}$ по методу наименьших квадратов. Сравнить полученную оценку с байесовской оценкой (8.18) в предельном случае $\mathbf{M}^{-1} \to \mathbf{O}_{k \times k}$, т. е. когда априорные сведения малоинформативны.

8.12. Исследовать поведение байесовской оценки (8.18) в случае $\mathbf{X}^{\mathrm{T}}\mathbf{X} \to \mathbf{O}_{k\times k}$, когда выборка «содержит мало информации» о векторе β.

8.13. Построить $T = 100$ реализаций СВ $y_t : y_1, \dots, y_{100}$ в соответствии с моделью (8.14), где положить $\sigma^2 = 1$, $k = 2$; в (8.16) приравнять $\mathbf{m} = (3, 7)^{\mathrm{T}}$, $\mathbf{M} = \mathbf{I}_2$. Матрицу $\mathbf{X}$ задать произвольным образом, но так чтобы $|\mathbf{X}^{\mathrm{T}}\mathbf{X}| \neq 0$. Предусмотреть при выполнении задания возможность изменения параметров модели.

8.14. По вектору наблюдений $\mathbf{y} = (y_1, \dots, y_T)^{\mathrm{T}}$ построить байесовскую оценку параметров в модели (8.14), согласно (8.18), полагая дисперсию σ^2 и п.р.в. (8.16) известными. Воспользоваться данными, полученными в задании 8.13. Исследовать влияние точности, с которой задана априорная п.р.в., и длительности наблюдения T на точность оценки (8.18). Сравнить по точности оценки (8.18) и (8.19). Провести сравнительное исследование оценок (8.18), (8.19) и МНК-оценки параметров регрессии в случае $|\mathbf{X}^{\mathrm{T}}\mathbf{X}| \to 0$.

8.15. Построить $T = 100$ реализаций СВ $y_t : y_1, \dots, y_{100}$ в соответствии с моделью (8.14), где $k = 2$, σ^2 – реализация СВ, распределенной согласно (8.22), где $N = \lambda = 4$; в (8.23) выбрать $\mathbf{m} = (3, 7)^{\mathrm{T}}$, $\mathbf{M} = \mathbf{I}_2$. Матрицу $\mathbf{X}$ выбрать удовлетворяющей условию $|\mathbf{X}^{\mathrm{T}}\mathbf{X}| \neq 0$. Предусмотреть возможность изменения параметров модели .

8.16. По вектору $\mathbf{y}$, построенному в задании 8.15, вычислить байесовскую оценку (8.25) параметров регрессионной модели (8.14) в предположениях (8.22), (8.23). Исследовать изменения точности оценки (8.25) при изменениях в задании априорной информации и изменениях числа используемых наблюдений.

8.17. Используя экспериментальные экономические данные (файл data 8_17), построить регрессионную модель (8.14) и вычислить байесовскую оценку (8.25). Особое внимание уделить заданию параметров априорных распределений вероятностей.

8.18.* Для заданий 8.14 и 8.16 построить 95 % доверительные области для вектора параметров $\boldsymbol{\beta}$ модели (8.14). Отобразить полученные множества графически.

8.4. ОБЩАЯ МОДЕЛЬ БАЙЕСОВСКОГО ПРОГНОЗИРОВАНИЯ

Рассмотренные выше методы позволяют строить «подстановочные» прогнозы экономических показателей, основанные на статистических оценках параметров, от которых эти показатели зависят. Отметим, однако, что даже при использовании оптимальных статистических оценок параметров «подстановочный» прогноз может быть не оптимальным.

Сформулируем общую вероятностную модель байесовского статистического прогнозирования. Ее применение, во-первых, позволяет учесть в прогнозе имеющуюся априорную информацию об анализируемом показателе. Во-вторых, в рамках этой модели построенные прогнозы обладают свойством оптимальности. В-третьих, здесь не требуется находить решения задач оценивания параметров модели; на практике эти задачи могут оказаться сложнее, чем построение самого прогноза. В четвертых, такая модель позволяет строить не только точечные, но и интервальные прогнозы. Кроме того, полученные прогнозы обладают свойством робастности (устойчивости) к некоторым типам искажений модельных предположений.

Пусть задано некоторое вероятностное пространство ($\boldsymbol{\Omega}, \Im, \mathbf{P}$). На нем определены следующие три случайных элемента (рис. 8.2).

Ненаблюдаемый вектор параметров $\boldsymbol{\theta} = (\theta_i) \in \Theta \subseteq \Re^m$, истинное значение которого не известно и является случайным вектором с априорной п.р.в. $\pi(\boldsymbol{\theta})$;

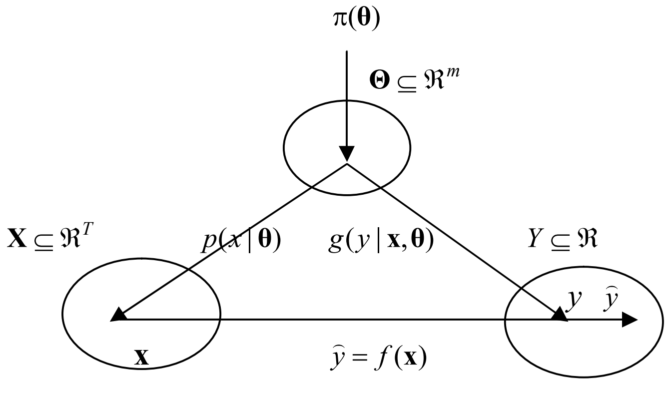

Рис. 8.2. Математическая модель байесовского прогнозирования

2.Стохастически зависящий от $\boldsymbol{\theta}$ вектор наблюдений $\mathbf{x}=(x_t)\in X\subseteq\Re^T$ (описывающий, например, «прошлые» и «настоящее» значения анализируемого эконометрического показателя) с условной п.р.в. $p(\mathbf{x}\,|\,\boldsymbol{\theta})$ при заданном векторе параметров $\boldsymbol{\theta}$;

3.Неизвестная, подлежащая прогнозированию величина $y\in Y\subseteq\Re$ (к примеру, «будущее» значение анализируемого показателя), стохастически зависящая от векторов $\boldsymbol{\theta},\mathbf{x}$ и распределенная согласно условной п.р.в. $g(y\,|\,\mathbf{x},\boldsymbol{\theta})$ при фиксированных векторах наблюдений **x** и параметров $\boldsymbol{\theta}$.

Задача заключается в построении прогноза $\widehat{y}$ для величины y по наблюдениям **x** с использованием информации, заданной в виде п.р.в. $\pi(\cdot),p(\cdot),g(\cdot)$. Стохастическая связь вектора параметров $\boldsymbol{\theta}$ с вектором наблюдений **x**, величиной y позволяет прогнозировать y по **x** в условиях неопределенности $\boldsymbol{\theta}$, т. е. строить статистический прогноз: $\widehat{y}=f(\mathbf{x})$, где $\widehat{y}\in Y$ – значение прогноза для $y\in Y$, а $f(\cdot):X\to Y$ – некоторая статистика, т. е. борелевская функция T переменных, задающая прогноз и называемая прогнозирующей статистикой.

Байесовской прогнозной плотностью называется условная п.р.в.

$$g(y\,|\,\mathbf{x})=\int_{\Theta}g(y\,|\,\mathbf{x},\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}\,|\,\mathbf{x})d\boldsymbol{\theta},\qquad(8.26)$$

где апостериорная п.р.в. $\pi(\boldsymbol{\theta}\,|\,\mathbf{x})$ вычисляется через известные плотности согласно теореме Байеса:

$$\pi(\boldsymbol{\theta}\,|\,\mathbf{x}) = \frac{p(\mathbf{x}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{p(\mathbf{x})},\ p(\mathbf{x}) = \int_{\Theta} p(\mathbf{x}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{8.27}$$

Для упрощения дальнейших обозначений в текущей главе будем полагать, что $X = \{\mathbf{x} : p(\mathbf{x}) > 0\}$.

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

$$\hat{y} = f_0(\mathbf{x}) = \mathbf{E}\{y\,|\,\mathbf{x}\} = \int_{Y} yq^0(y\,|\,\mathbf{x})dy,\ \mathbf{x} \in X, \tag{8.28}$$

а *байесовским прогнозом* – значение статистики (8.28) при заданном векторе наблюдений **x**.

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

Среднеквадратическим риском прогнозирующей статистики $f(\cdot) : X \to Y$ будем называть значение функционала

$$r(f(\cdot)) = \mathbf{E}\{(f(\mathbf{x}) - y)^2\} = \iint_{XY} s(\mathbf{x}, y)(f(\mathbf{x}) - y)^2\, dyd\mathbf{x} \geq 0, \tag{8.29}$$

где совместная п.р.в. $s(\mathbf{x}, y)$ вычисляется через известные п.р.в. следующим образом:

$$s(\mathbf{x}, y) = \int_{\Theta} g(y\,|\,\mathbf{x}, \boldsymbol{\theta})p(\mathbf{x}\,|\,\boldsymbol{\theta})d\boldsymbol{\theta},\ \mathbf{x} \in X,\ y \in Y\ . \tag{8.30}$$

Лемма 8.1. Минимум функционала (8.29), (8.30) достигается для байесовской прогнозирующей статистики (8.28):

$$r_0 = \inf_{f(\cdot)} r(f(\cdot)) = r(f_0(\cdot)). \tag{8.31}$$

Наименьшее возможное значение (8.31) функционала (8.29), (8.30) называется *байесовским риском прогнозирования* r_0 .

ЗАДАНИЯ

8.19. Пусть $\mathbf{x} = (x_1, \ldots, x_T)^{\mathrm{T}}$ – выборка объема T из распределения вероятностей $\mathbf{N}_1(\theta, 1)$; θ – СВ, $\mathbf{L}(\theta) = \mathbf{N}_1(5, 1)$. Смоделировать $\mathbf{x}$ при T = 100. Пусть $y = x_{T+1}$ – неизвестное значение СВ, имеющей распределение $\mathbf{N}_1(\theta, 4)$. С помощью системы компьютерной алгебры типа MATHEMATICA построить график байесовской прогнозной плотности (8.26). Меняя T и $\mathbf{D}\{\theta\}$, исследовать влияние этих параметров на плотность (8.26).

8.20. В условиях задания 8.19 вычислить байесовский прогноз (8.28) и байесовский риск (8.31). Для этого воспользоваться системой компьютерной алгебры. Проверить правильность получаемых результатов при $\mathbf{D}\{x_i \mid \theta\} \to 0, i \in \{1, \ldots, T\}$.

8.21. Выполнить задание 8.19 при условии, что $\mathbf{x}$ получена при фиксированном θ из гамма-распределения с параметрами θ, 2.

8.5. БАЙЕСОВСКОЕ ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ С ТРЕНДОМ

Обозначим через $S_{i,j}$ следующее упорядоченное множество $(i, j \in N, i < j)$:
$S_{i,j} = \{1, 2, \ldots, i-1, i, j\} \subset N$.

Рассмотрим модель временного ряда (ВР) с трендом:

$$x_t = \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{\psi}(t) + \xi_t, t \in S_{T,T+\tau}, \tag{8.32}$$

где T – длина наблюдаемого ВР (число наблюдений); τ – 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:

$$\{\xi_t : t \in S_{T,T+\tau}\} \quad \text{– н.о.р. СВ, } \mathbf{L}\{\xi_t\} = \mathbf{N}_1(0, \sigma^2); \tag{8.33}$$

$\boldsymbol{\psi}(t) = (\psi_1(t), \ldots, \psi_m(t))^{\mathrm{T}}$ – заданная неслучайная вектор-функция дискретного времени, $\psi_1(t), \ldots, \psi_m(t)$ – линейно независимые базисные функции тренда; $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^{\mathrm{T}} \in \Theta = \Re^m$ – истинное значение ненаблюдаемого вектора параметров, который полагается случайным (раздел 8.4) с m-мерной гауссовой п.р.в.

$$\pi(\boldsymbol{\theta}) = n_m(\boldsymbol{\theta} \mid \boldsymbol{\alpha}, \mathbf{A}) = (2\pi)^{-\frac{m}{2}} \mid \mathbf{A} \mid^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\alpha})^{\mathrm{T}} \mathbf{A}^{-1} (\boldsymbol{\theta}-\boldsymbol{\alpha})}; \qquad (8.34)$$

в (8.34) $\boldsymbol{\alpha} = \mathbf{E}\{\boldsymbol{\theta}\}$ – m-вектор математического ожидания, $\mathbf{A} = \mathbf{Cov}(\boldsymbol{\theta}, \boldsymbol{\theta}) = \mathbf{A}^{\mathrm{T}} \succ 0$ – $(m \times m)$-ковариационная матрица распределения вероятностей вектора параметров $\boldsymbol{\theta}$.

Из (8.32), (8.34) следует, что $X = \Re^T, Y = \Re$. Задача заключается в построении прогноза величины $y = x_{T+\tau}$ по T наблюдаемым значениям ВР (8.32) при выполнении условий (8.33), (8.34).

Отметим, что модель ВР с трендом (8.32)–(8.34) в математической литературе имеет и другое название – параметрическая линейная (по параметрам $\boldsymbol{\theta}$) регрессия (см. раздел 8.3). Для рассматриваемой модели случайные элементы $\mathbf{x}$ и y условно независимы при заданном векторе $\boldsymbol{\theta}$. Как следствие, байесовская прогнозная плотность (8.26) допускает более простое, по сравнению с общим случаем, представление

$$q(y \mid \mathbf{x}) = \int_{\Theta} g(y \mid \boldsymbol{\theta})\, \pi(\boldsymbol{\theta} \mid \mathbf{x}) d\boldsymbol{\theta}. \qquad (8.35)$$

Применим следующие обозначения:

$$\boldsymbol{\Psi}(t) = \boldsymbol{\psi}(t)\boldsymbol{\psi}^{\mathrm{T}}(t), t \in S_{T,T+\tau};\ \boldsymbol{\Gamma} = \sum_{t \in S_{T,T+\tau}} \boldsymbol{\Psi}(t);\ \boldsymbol{\zeta}(x) = \sum_{t=1}^{T} x_t \boldsymbol{\psi}(t). \qquad (8.36)$$

Теорема 8.6. [19] В обозначениях (8.36) для модели ВР с трендом (8.32) при выполнении (8.33), (8.34) байесовская прогнозирующая статистика (8.28) имеет следующее представление:

$$f_0(\mathbf{x}) = \frac{\boldsymbol{\Psi}^{\mathrm{T}}(T+\tau)(\sigma^2 \mathbf{A}^{-1} + \boldsymbol{\Gamma})^{-1}(\sigma^2 \mathbf{A}^{-1}\boldsymbol{\alpha} + \boldsymbol{\zeta}(x))}{1 - \boldsymbol{\Psi}^{\mathrm{T}}(T+\tau)(\sigma^2 \mathbf{A}^{-1} + \boldsymbol{\Gamma})^{-1}\boldsymbol{\Psi}(T+\tau)}, \quad \mathbf{x} \in \Re^T. \qquad (8.37)$$

Отметим, что статистика (8.37) устойчива к вычислительным погрешностям, обусловленным особенностями компьютерной арифметики [18].

Исследуем теперь два частных случая асимптотического поведения байесовской прогнозирующей статистики (8.37). Пусть

$$m = 1, \quad \boldsymbol{\psi}(t) = 1, \tag{8.38}$$

т. е. будем рассматривать одномерный (по параметру) ВР с трендом. (В этом случае тренд определяется константой – параметром $\boldsymbol{\theta} = \theta_1$; матрица $\mathbf{A}$ – число).

Следствие 1. Если выполнены условия теоремы 8.6 для случая (8.38), то при $\mathbf{A} \to 0\,(\mathbf{A}^{-1} \to \infty)$ справедлив следующий асимптотический результат:

$$m = 1, \quad \boldsymbol{\psi}(t) = 1, \tag{8.39}$$

Следствие 2. Пусть для случая (8.38) выполнены условия теоремы 8.6. Тогда в асимптотике $\mathbf{A} \to \infty\,(\mathbf{A}^{-1} \to 0)$ имеет место следующее соотношение:

$$\widehat{x}_{T+\tau} - \frac{1}{T}\sum_{t=1}^{T} x_t \xrightarrow{п.н.} 0. \tag{8.40}$$

Следствие 1 фактически утверждает, что при уточнении априорной информации о параметре $(\boldsymbol{\theta} \xrightarrow{п.н.} \alpha)$ байесовская прогнозирующая статистика для ВР с трендом (8.32)–(8.34) стремится к априорному среднему значению параметра. Следствие 8.2 утверждает, что для другого крайнего случая (при отсутствии априорной информации) байесовская прогнозирующая статистика совпадает с выборочным средним значением наблюдений.

ЗАДАНИЯ

8.22. В соответствии с (8.34) получить реализацию случайного вектора $\boldsymbol{\theta}$ при $m = 3, \boldsymbol{\alpha} = (2, -3, 4)^{\mathrm{T}}, \mathbf{A} = \mathbf{I}_3$. Используя полученное значение, построить T = 100 значений ВР $x = (x_1, \ldots, x_T)$ и еще 5 значений: $x_{T+1}, \dots, x_{T+5}$. Предположить дисперсию $\sigma^2 = 1$, а базисные функции тренда равными $\psi_1(t) = 1, \psi_2(t) = t, \psi_3(t) = t^2$. Предусмотреть возможность задания всех характеристик модели при работе компьютерной программы.

8.23. По T = 100 первым значением ВР $x_1, \dots, x_T$, полученным в задании 8.22, построить байесовский прогноз (8.37) величин $x_{T+1}, \dots, x_{T+5}$; сравнить прогнозы с соответствующими истинными значениями, полученными в предыдущем задании. Провести тот же набор экспериментов с другими базисными функциями тренда.

8.24. С помощью вычислительных экспериментов проверить выполнение (8.39), (8.40) в соответсвующей асимптотике.

8.25. По статисическим данным, представляющим объемы продаж за последовательные 17 месяцев на одном из предприятий (файл data 8_25), построить байесовский прогноз (8.37) объема продаж на 6 месяцев вперед.

8.26.* В задании 8.25 построить при $\tau = 1, \dots, 5$, 95 % доверительные интервалы для прогнозируемого показателя. Представить результаты прогнозирования графически.

8.27.* Сравнить байесовский прогноз (8.37) с подстановочным байесовским прогнозом, когда $\hat{y} = f_1(\mathbf{x}) = \hat{\boldsymbol{\theta}}^{\mathrm{T}} \boldsymbol{\psi}(T+\tau)$, а оценка $\hat{\boldsymbol{\theta}}$ определяется (8.18).

8.6. БАЙЕСОВСКОЕ ПРОГНОЗИРОВАНИЕ АВТОРЕГРЕССИОННЫХ ВРЕМЕННЫХ РЯДОВ

Рассмотрим теперь другую, часто встречающуюся в эконометрических задачах вероятностную модель – авторегрессионный временной ряд [**Зельнер**]. В этой модели предполагается зависимость наблюдений $\{x_t\}$, в отличие от трендовой модели. В результате свойство (8.35) не выполнено, что усложняет анализ.

Пусть наблюдаются T значений $\mathbf{x} = (x_1, \dots, x_T)^{\mathrm{T}}$ стационарного ВР, описываемого разностным уравнением авторегрессии порядка

$m \in N$ АР(m) с независимыми остатками $\{\xi_t\}, t \in S_{T,T+\tau}$ $= \{1, 2, \ldots, T+1\}$:

$$x_t + \sum_{i=1}^{m} \theta_i x_{t-i} = \xi_t, \ \ t \in S_{T,T+\tau}; \tag{8.41}$$

$$\{\xi_t\} - \text{н.о.р. СВ}, \mathbf{L}\{\xi_t\} = \mathbf{N}_1(0, \sigma^2), t \in S_{T,T+\tau}. \tag{8.42}$$

В (8.41) ненаблюдаемый истинный вектор параметров $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)^{\mathrm{T}}$ полагается случайным с m-мерной гауссовой п.р.в. (8.34):

$$\pi(\boldsymbol{\theta}) = n_m(\boldsymbol{\theta} \,|\, \boldsymbol{\alpha}, \mathbf{A}) = (2\pi)^{-\frac{m}{2}} \,|\, \mathbf{A} \,|^{-\frac{1}{2}} \, e^{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\alpha})^{\mathrm{T}} \mathbf{A}^{-1} (\boldsymbol{\theta}-\boldsymbol{\alpha})},$$

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

где, как и в разделе 8.5, вектор $\boldsymbol{\alpha}$ и матрица $\mathbf{A}$ – параметры нормального распределения вероятностей. Начальные значения

$$x_{-m+1}, \ldots, x_0 \tag{8.43}$$

авторегрессионного процесса (8.41), а также дисперсия σ^2 предполагаются известными. Из (8.41), (8.42) следует, что $X = \mathfrak{R}^T$, $Y = \mathfrak{R}$. Задача состоит в байесовском прогнозировании будущего значения временного ряда $y = x_{T+1} (\tau = 1)$ по вектору $\mathbf{x}$ наблюдаемых значений в условиях вероятностной модели (8.41), (8.42), (8.43).

Обозначим:

$$\tilde{\mathbf{x}}_{t-1} = (x_{t-1}, \ldots, x_{t-m})^{\mathrm{T}} \in \mathfrak{R}^m, \quad t \in S_{T,T+1}. \tag{8.44}$$

Теорема 8.7. [**Харин А.Ю.**] Если имеет место модель ВР АР(m), удовлетворяющая (8.41), (8.42), (8.34), то при заданных начальных значениях (8.43) байесовский прогноз определяется статистикой

$$\widehat{y} = f_0(\mathbf{x}) = \frac{x_T^{\mathrm{T}}(\sigma^2\mathbf{A}^{-1} + \sum_{i=0}^{T-1}\widetilde{\mathbf{x}}_i\widetilde{\mathbf{x}}_i^{\mathrm{T}})^{-1}(\sum_{i=1}^{T} x_i\widetilde{\mathbf{x}}_{i-1} - \sigma^2\mathbf{A}^{-1}\boldsymbol{\alpha})}{1 - \widetilde{x}_T^{\mathrm{T}}(\sigma^2\mathbf{A}^{-1} + \sum_{i=0}^{T-1}\widetilde{\mathbf{x}}_i\widetilde{\mathbf{x}}_i^{\mathrm{T}})^{-1}\widetilde{\mathbf{x}}_T}. \quad (8.45)$$

Следствие 1. Если в условиях теоремы 8.7 $\mathbf{A} \to \mathbf{0}_{m\times m}$, то $\widehat{y} = f_0(\mathbf{x})$ удовлетворяет асимптотике: $\widehat{y} = \widetilde{\mathbf{x}}_T^{\mathrm{T}}\boldsymbol{\alpha} \xrightarrow{п.н.} 0.$

Таким образом, байесовский прогноз тем ближе к подстановочному прогнозу по вектору априорных средних коэффициентов, чем точнее априорная информация о векторе $\boldsymbol{\theta}$.

Следствие 2. Если в условиях теоремы 8.7 $|\sum_{i=0}^{T-1}\widetilde{x}_i\widetilde{x}_i^{\mathrm{T}}| \neq 0$, то при $\mathbf{A}^{-1} \to \mathbf{0}_{m\times m}$ для байесовского прогноза $\widehat{y} = f_0(\mathbf{x})$ и «подстановочного» прогноза $f_{ОМП}(\mathbf{x}) = \widetilde{x}_T^{\mathrm{T}}(\sum_{t=0}^{T-1}\widetilde{\mathbf{x}}_t\widetilde{x}_t^{\mathrm{T}})^{-1}\sum_{t=1}^{T} x_t\widetilde{\mathbf{x}}_{t-1}$ справедливо асимптотическое соотношение

$$f_0(\mathbf{x}) - f_{ОМП}(\mathbf{x}) \xrightarrow{п.н.} 0. \quad (8.46)$$

Фактически следствие 2 утверждает, что при отсутствии априорной информации о случайном векторе коэффициентов авторегрессии байесовский прогноз авторегрессионного ВР 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) [2].

Отметим еще, что как и для модели ВР с трендом, байесовский прогноз (8.45) устойчив к вычислительным погрешностям округления.

Теперь рассмотрим более общий случай. Пусть имеет место модель (8.41), (8.34), где остатки $\{\xi_t\}, t \in \{1, 2, \dots, T+1\}$ допускают корреляцию между собой. Пусть $\{\xi_t\}$ – стационарная гауссова случайная последовательность (СП), для которой

$$\mathbf{E}\{\xi_t\} = 0, \mathbf{D}\{\xi_t\} = \sigma^2 < \infty, \quad (8.47)$$

$\mathbf{Cov}\{\xi_t, \xi_{t+\tau}\} = \sigma^2 \rho_\xi(\tau), t \in \{1, 2, \dots, T+1\}, \tau \in \{0, 1, \dots, T\}$, где $\rho_\xi(\cdot)$ – заданная корреляционная функция СП $\{\xi_t\}$, $\rho_\xi(0) = 1$. Отметим, что в условиях рассматриваемой модели $X = \Re^T$, $Y = \Re$, $y = x_{T+1}$.

Сохраняя принятые выше обозначения, обозначим ковариационную матрицу СП $\{\xi_t\}, t \in \{1, 2, \dots, T+1\}$:

$$\overline{\mathbf{\Sigma}} = \sigma^2 \begin{pmatrix} 1 & \rho_\xi(1) & \dots & \rho_\xi(T) \\ \rho_\xi(1) & 1 & \dots & \rho_\xi(T-1) \\ \vdots & \vdots & \dots & \vdots \\ \rho_\xi(T) & \rho_\xi(T-1) & \dots & 1 \end{pmatrix} = \sigma^2 \begin{pmatrix} \mathbf{\Sigma} & \boldsymbol{\sigma}_{T+1} \\ \boldsymbol{\sigma}_{T+1} & 1 \end{pmatrix}, \quad (8.48)$$

где

$$\mathbf{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho_\xi(1) & \dots & \rho_\xi(T-1) \\ \rho_\xi(1) & 1 & \dots & \rho_\xi(T-2) \\ \vdots & \vdots & \dots & \vdots \\ \rho_\xi(T-1) & \rho_\xi(T-2) & \dots & 1 \end{pmatrix},$$

$$\boldsymbol{\sigma}_{T+1} = (\rho(T), \rho(T-1), \dots, \rho(1))^{\mathrm{T}}.$$

Кроме того, введем в рассмотрение матрицы:

$$\overline{\mathbf{X}} = \begin{pmatrix} x_0 & x_{-1} & \dots & x_{-m+1} \\ x_1 & x_0 & \dots & x_{-m+2} \\ \vdots & \vdots & \ddots & \vdots \\ x_T & x_{T-1} & \dots & x_{T-m+1} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \dots \\ \tilde{\mathbf{x}}_T^{\mathrm{T}} \end{pmatrix}, \quad (8.49)$$

$$\mathbf{X} = \begin{pmatrix} x_0 & x_{-1} & \dots & x_{-m+1} \\ x_1 & x_0 & \dots & x_{-m+2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T-1} & x_{T-2} & \dots & x_{T-m} \end{pmatrix}.$$

Примем еще следующие обозначения:

$$\gamma = 1 - \boldsymbol{\sigma}_{T+1}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\sigma}_{T+1}, \boldsymbol{\gamma}_{T+1} = -\gamma \boldsymbol{\Sigma}^{-1} \boldsymbol{\sigma}_{T+1};$$

$$\varsigma(\mathbf{x}) = \widetilde{\mathbf{x}}_T - \mathbf{X}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\sigma}_{T+1}, \quad (8.50)$$

$$\mathbf{Z}(\mathbf{x}) = \mathbf{X}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\gamma \mathbf{I}_T + \boldsymbol{\sigma}_{T+1} \boldsymbol{\sigma}_{T+1}^{\mathrm{T}}) - \widetilde{\mathbf{x}}_T \boldsymbol{\sigma}_{T+1}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}.$$

Теорема 8.8.[**Харин А.Ю.**] Для модели, заданной (8.41), (8.47), (8.34), в обозначениях (8.48)–(8.50) байесовская прогнозирующая статистика равна

$$f_0(\mathbf{x}) = \frac{\varsigma^{\mathrm{T}}(\mathbf{x})(\sigma^2 \mathbf{A}^{-1} + \overline{\mathbf{X}}^{\mathrm{T}} \sigma^2 \overline{\boldsymbol{\Sigma}}^{-1} \overline{\mathbf{X}})^{-1} (\mathbf{Z}(\mathbf{x})\mathbf{x} - \sigma^2 \gamma \mathbf{A}^{-1} \boldsymbol{\alpha}) - \boldsymbol{\gamma}_{T+1}^{\mathrm{T}} \mathbf{x}}{\gamma - \varsigma^{\mathrm{T}}(\mathbf{x})(\sigma^2 \mathbf{A}^{-1} + \overline{\mathbf{X}}^{\mathrm{T}} \sigma^2 \overline{\boldsymbol{\Sigma}}^{-1} \overline{\mathbf{X}})^{-1} \varsigma(\mathbf{x})}. \quad (8.51)$$

Прежде всего отметим, что теорема 8.7 следует из теоремы 8.8: достаточно в условиях теоремы 8.8 предположить ковариационную матрицу СП $\{\xi_t\}, t \in \{1, 2, \dots, T+1\}$ диагональной: $\overline{\boldsymbol{\Sigma}} = \sigma^2 \mathbf{I}_{T+1}$.

Отметим также, что широко известная в литературе и часто применяемая в эконометрике модель ВР АР(m) с остатками в виде скользящего среднего порядка $q \in N$ [2] (АРСС(p, q)) н.о.р. СВ $\{ u_t \}$, $t = 1-q, 2-q, \dots,$ $\mathbf{L}(u_t) = \mathbf{N}_1(0, \sigma^2)$

$$\xi_t = \sum_{j=0}^{q} b_j u_{t-j} \; - \quad (8.52)$$

частный случай рассмотренной выше модели. Для нее матрица $\overline{\boldsymbol{\Sigma}}$ будет $(2q+1)$-диагональной матрицей, элементы которой однозначно выражаются через коэффициенты скользящего среднего:

$$\rho_\xi(\tau) = \sum_{j=0}^{q} b_j b_{j+\tau} 1_{[0,q]}(j+\tau), \tau \in \{0, 1, \dots, \tau+1\}. \quad (8.53)$$

Байесовский прогноз, построенный с помощью прогнозирующей статистики (8.51), не является подстановочным байесовским прогнозом (в числителе вычитаемое зависит от всего вектора $\mathbf{x}$, а не только от m-вектора «предыстории» $\widetilde{\mathbf{x}}_T$). Таким образом, в общем случае недопустимо сведение задачи байесовского статистического прогнозирования к задаче байесовского оценивания параметров модели.

ЗАДАНИЯ

8.28. В соответствии с (8.34) получить реализацию случайного вектора $\boldsymbol{\theta}$ при $m = 2, \boldsymbol{\alpha} = (0.7, 0.5)^{\mathrm{T}}, \mathbf{A} = \begin{pmatrix} 0.16 & 0 \\ 0 & 0.04 \end{pmatrix}$. Используя полученное значение и положив $x_{-1} = x_0 = 0$, построить в соответствии с (8.41) ВР АР(2) длины T = 100; смоделировать $x_{T+1}, \dots, x_{T+10}$. Предположить дисперсию $\sigma^2 = 1$. Предусмотреть возможность задания всех характеристик модели.

8.29. По T = 100 наблюдениям ВР АР(2) $x_1, \dots, x_T$, полученным в задании 8.28, построить байесовский прогноз (8.45) величины $y = x_{T+1}$. Используя $x_1, \dots, x_T, \widehat{x}_{T+1}$, построить прогноз $y = x_{T+2}$, и т. д. Сравнить полученные величины с истинными значениями.

8.30. Многократно проводя эксперименты, подобно как в задании 8.29, выяснить влияние T на точность прогноза на один шаг, т. е. величины $y = x_{T+1}$. Исследовать эффект, возникающий при «подстановочном» прогнозировании на большую глубину τ.

8.31. По статистическим данным ежемесячного значения цепного индекса потребительских цен в Республике Беларусь с 01.97 по 12.01 (файл data 8_31), построить байесовский прогноз на 6 месяцев вперед, используя модель АР(6).

8.32. (*) Смоделировать произвольный ВР АРСС(2, 2) длины $T + 1$ = 101 согласно (8.41), (8.52). Используя первые 100 наблюдений и (8.53), построить байесовский прогноз (8.51) для $y = x_{T+1}$. Сравнить полученный результат с байесовским прогнозом (8.45), построенным в предположении (8.42).