

П.И. Бидюк, В.В. Павлов, А.С. Борисевич, Л.Т. Гасанова

ОЦЕНИВАНИЕ РЕГРЕССИОННЫХ МОДЕЛЕЙ С ПОМОЩЬЮ МЕТОДА МОНТЕ-КАРЛО ДЛЯ МАРКОВСКИХ ЦЕПЕЙ

Проанализированы основные тенденции исследований в области байесовского оценивания математических и статистических моделей процессов произвольной природы с использованием численных методов Монте-Карло. Описывается преимущество байесовского подхода по сравнению с известными, заключающееся в использовании априорной информации относительно параметров модели. Предложено два правила выбора априорного распределения, которые охватывают наиболее распространенные случаи. Описанный метод дает возможность получать достаточно большие объемы исходной информации, а также точнее описывать структуру и основные характеристики исследуемой модели.

Введение. В настоящий момент актуальной задачей оценивания математических и статистических моделей является применение байесовской методологии к процессам различной природы. Байесовский подход, как альтернатива классическому статистическому подходу, позволяет точнее и полнее оценивать модели, дает возможность получать хорошие результаты в тех случаях, когда использование классических статистических методов очень ограничено (например, случаи с короткой выборкой статистических данных). Байесовский подход открывает новые, довольно широкие возможности применения методов математического моделирования, а разработанные алгоритмы оценивания на основе генерирования случайных чисел способствуют решению поставленных задач с помощью современных вычислительных процедур.

Классический подход направлен на получение эффективных алгоритмов оценивания и изучение их асимптотических свойств, которые служат основой для формирования статистического вывода на основе данных относительно большого объема. В случае коротких выборок использование результатов асимптотической теории представляется недостаточно обоснованным. Байесовский подход к формированию статистического вывода основывается на других теоретических предпосылках. Байесовские методы отличаются от классических иным подходом к интерпретации истинных параметров модели. Классический подход исходит из того, что истинные параметры — это не случайные величины, а аппроксимирующие их оценки случайные, поскольку являются функциями наблюдений, которые содержат случайные процессы [1]. Байесовский подход относится к числу тех подходов, которые более широко трактуют истинные параметры модели, т.е. случайность рассматривается как имманентное свойство реального физического мира с учетом того, что сам физический объект непрерывно испытывает случайные изменения. Поэтому ищут неслучайные оценки, которые довольно близко аппроксимируют какую-нибудь статистику случайного параметра, например его среднее значение или моду. При практическом применении оцененной

модели разница практически несущественна — исследователь работает с моделью, которая имеет детерминированные коэффициенты. Вероятностные свойства модели используются для определения погрешности прогноза и анализа чувствительности модели, вычисления функции потерь и т.д. Очевидно, что подобные вычисления можно выполнять при использовании обоих подходов.

Байесовская методология исследовалась во многих работах и использовалась в разных областях науки и техники. В частности, А. Зельнер исследовал использование таких методов в эконометрике [1]; В.П. Савчук анализировал надежность технических объектов [2]; известно также много других направлений применения этих методов. Использование байесовской методологии в таких случаях сводилось к большим аналитическим исследованиям, которые иногда требовали основательных знаний математики и статистики [3]. Благодаря развитию компьютерной техники появилась возможность исследовать альтернативные вычислительные алгоритмы, базирующиеся на известных принципах генерирования случайных чисел [4–14] и их использовании для формирования оценок параметров математических и статистических моделей случайных процессов.

Постановка задачи. Цель статьи — проанализировать и раскрыть существующие основные тенденции исследований в направлении байесовского оценивания математических и статистических моделей процессов произвольной природы с использованием численных методов Монте-Карло. Мы предполагаем, что случайные процессы содержат детерминированную составляющую, которая может быть описана некоторой выбранной детерминированной функцией, например авторегрессией, регрессией, полиномом и т.д. Наличие случайной составляющей в исследуемом процессе требует введения в модель случайных переменных с соответствующими распределениями и применения методов анализа и описание случайных процессов.

Суть байесовского подхода. Байесовские методы разработаны в результате систематических попыток сформулировать и решить проблемы статистического анализа поведения процессов и систем разной природы на основе теоремы Байеса. Предпосылкой к использованию этой теоремы являются некоторые соотношения между вероятностями событий разного характера и спецификации любого события на необходимом уровне [3].

Большинство статистических задач, независимо от методов их решения, имеют некоторые общие свойства. Для описания конкретной выборки данных рассматривается несколько вероятностных моделей, как потенциально приемлемые для исследуемой ситуации. После получения данных возникают выраженные в некотором числовом виде знания относительно приемлемости этих моделей.

Отличие байесовской парадигмы от других статистических подходов состоит в том, что еще до получения данных исследователь рассматривает степень своего доверия к возможным моделям и представляет ее в виде вероятностей. Как только данные получены, теорема Байеса позволяет исследовать

дователю определить еще одно множество вероятностей, которые представляют собой пересмотренные степени доверия к возможным моделям-кандидатам с учетом новой информации.

Одним из ключевых преимуществ байесовского подхода является использование любой начальной (априорной) информации относительно параметров модели. Такая информация выражается в виде априорной вероятности или функции плотности вероятности. Затем начальные вероятности «пересматриваются», с помощью выборочных данных, которые находят свое отображение в виде апостериорного распределения оценок параметров или переменных модели.

Рассмотрим случайную переменную X , которая имеет распределение вероятностей, определенное в терминах неизвестного параметра θ , принадлежащего определенному множеству возможных значений параметра Θ . Для заданного значения $X = x$ функция правдоподобия каждого отдельного значения θ задается как $P(x|\theta)$. В непрерывном случае априорные вероятности для множества возможных моделей отвечают, в общем случае, распределению вероятностей на множестве Θ возможных значений параметра. Таким образом, априорные характеристики (суждение) определяют в виде априорной плотности вероятности: $P(\theta)$, $\theta \in \Theta$, такой, что

$$\int_{\Theta} P(\theta) d\theta = 1. \quad (1)$$

Априорное распределение пересматривается на основе выборочных данных $X = x$ для получения апостериорной плотности вероятности $P(\theta|x)$, $\theta \in \Theta$. Соответственно по теореме Байеса, которую называют принципом обратной вероятности, устанавливается взаимосвязь между $P(x|\theta)$, $P(\theta|x)$ и $P(\theta)$:

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)}, \quad \theta \in \Theta, \quad (2)$$

где

$$P(x) = \int_{\Theta} P(x|\theta)P(\theta)d\theta. \quad (3)$$

Учитывая, что в (2) знаменатель не зависит от θ , довольно часто зависимость (2) представляют в виде

$$P(\theta|x) \propto P(x|\theta)P(\theta), \quad (4)$$

где \propto означает пропорциональность.

Выбор и анализ априорного распределения. Метод определения априорного распределения зависит от конкретной поставленной задачи. Различают случаи, когда распределение априорной информации известно, если даже распределение параметра «неинформативно». В первом случае, если известен вид функции априорного распределения, то, используя байесовский подход (равенство (2) или (4)), можно найти вид апостериорного распределения.

Существуют случаи, когда априорное и апостериорное распределения относятся к одному и тому же классу распределений. Такие распределения называют сопряженными. Использование их для численных методов в байесовском подходе означает, что существует замкнутая форма решения для условных апостериорных распределений [14]. Например X_1, \dots, X_n рассматриваются как случайная выборка из нормального распределения с неизвестным средним значением μ и известной дисперсией σ^2 . Также допускается, что априорное распределение μ является нормальным со средним значением μ_0 и дисперсией σ_0^2 . Тогда апостериорное распределение μ при выборочных значениях X_1, \dots, X_n и заданном априорном распределении также будет нормальным со средним значением μ_* и дисперсией σ_*^2 , которые определяют следующим образом:

$$\mu_* = \frac{\sigma^2 \mu_0 + n \sigma_0^2 \bar{X}}{\sigma^2 + n \sigma_0^2} \quad \text{и} \quad \sigma_*^2 = \frac{\sigma^2 \sigma_0^2}{\sigma^2 + n \sigma_0^2}, \quad (5)$$

где $\bar{X} = \sum_{i=1}^n X_i / n$ — выборочное среднее.

В байесовском анализе довольно часто удобно использовать параметр точности $\eta = 1/\sigma^2$ (т.е. обратный к дисперсии σ^2). Если для априорного распределения $\eta_0 = 1/\sigma_0^2$, то для апостериорного $\eta_* = 1/\sigma_*^2$. Теперь равенство (5) можно представить в следующем виде:

$$\eta_* = \eta_0 + n\eta \quad \text{и} \quad \mu_* = \frac{\eta_0}{\eta_*} \mu_0 + \frac{n\eta}{\eta_*} \bar{X}. \quad (6)$$

Таким образом, для полученной нормальной случайной выборки информация о μ содержится в выборочном среднем \bar{X} , которое является достаточной статистикой для μ . Точность описания распределения определяется отношением $X: \eta/\sigma^2 = n\eta$, т.е. суммой двух компонент: точностью представления априорного распределения и выборочных данных, а апостериорное среднее является взвешенным средним априорного среднего, и выборочного среднего с весовым коэффициентом, пропорциональным точности. Приведенный результат свидетельствует о том, что вклад априорного распределения уменьшается с возрастанием размера выборки n . Подробное описание получения этого результата можно найти в [1].

Довольно часто рассматривают так называемую байесовскую асимптотику, т.е. если $n \rightarrow \infty$ для апостериорного распределения имеет значение только правдоподобие. Это достаточно просто можно показать с помощью равенства (4), если его представить следующим образом [2]:

$$P(\theta | x) \propto P(x | \theta) P(\theta) = P(\theta) e^{\ln P(x|\theta)}. \quad (7)$$

Если предположить, что априорное распределение $P(\theta)$ и функция правдоподобия $P(x|\theta)$ невырожденные и имеют непрерывную производную, и $P(x|\theta)$ имеет единственный максимум $\hat{\theta}_{\max}$, являющийся оценкой максимального правдоподобия, то $\ln[P(x|\theta)]$ имеет порядок n , а $P(\theta)$ не зависит от объема выборки. Таким образом, интуитивно понятно, что множитель правдоподобия при больших значениях объема выборки будет доминировать в априорном распределении.

Во многих случаях начальная информация о значениях оцениваемых параметров совсем неизвестна, т.е. неизвестный вид априорного распределения $P(\theta)$. Другими словами, параметр θ «неинформативный». Для такого случая предложено два правила выбора априорного распределения, которые охватывают наиболее распространенные случаи [12]. Если существует параметр на конечном интервале или на интервале от $-\infty$ к $+\infty$, то его априорная вероятность считается равномерно распределенной. Если же можно обосновать, что параметр принимает значение на интервале от 0 до ∞ , то вероятность его логарифма следует считать равномерно распределенной.

Первое правило Джеффриса для представления неопределенности значения формулируется следующим образом:

$$P(\theta)d\theta \sim d\theta, \quad -\infty < \theta < +\infty, \quad (8)$$

т.е. $P(\theta) \sim \text{const}$. Это прямоугольное распределение (или же функция плотности вероятности) является несобственным, поскольку $\int_{-\infty}^{\infty} P(\theta)d\theta = \infty$. Известно, что если $-\infty < \theta < +\infty$ — достоверное событие, то вместо 1 для записи вероятности такого события используется ∞ .

Второе правило Джеффриса применимо к параметрам, природа которых позволяет предположить, что они принимают значение от 0 до ∞ . По аналогии распределение логарифма параметра будет равномерным, т.е. если $\theta = \log \sigma$, то априорная функция плотности вероятности для θ будет иметь вид (8). Поскольку $d\theta = d\sigma/\sigma$, то (8) предполагает использование равенства

$$P(\sigma)d\sigma \sim \frac{d\sigma}{\sigma}, \quad 0 < \sigma < +\infty \quad (9)$$

в качестве несобственной функции плотности вероятности, которая будет представлять неопределенность значения параметра σ . Таким образом, для неинформативного параметра, например для дисперсии, априорное распределение задают в виде $P(\sigma) \propto 1/\sigma$.

Отметим такое важное свойство, как чувствительность априорного распределения. При выборе разных априорных распределений получают разные апостериорные распределения. В таком случае полезно оценить степень влияния отличий. На практике апостериорное распределение используют как априорное, при этом возведя его в некоторую степень α , где $0 < \alpha < 1$.

Вычислительные методы байесовского анализа. Байесовский подход довольно широко использует информацию о вероятностном распределении параметров. До конца 80-х годов прошлого века в качестве вычислительных методов для вывода байесовских оценок параметров и их апостериорных распределений использовали аналитические методы: сопряженные априорные распределения и аппроксимацию [14]. С начала 90-х годов благодаря бурному развитию компьютерных технологий начали распространяться новые методы вычислений, которые базируются на непосредственном генерировании (моделировании выборки) необходимых измерений по апостериорным распределениям.

Генерирование случайных величин с заданным распределением — современный подход, который дает возможность работать в условиях, когда существуют асимптотические аналитические результаты для свойств оценок и их статистических распределений, но неизвестно, какие свойства будут иметь оценки при малых выборках данных. Выделяют два общих подхода к моделированию: историческое моделирование и моделирование по принципу Монте-Карло [14]. Метод Монте-Карло, предложенный Дж. фон Нейманом и С. Уламом в 1940-х гг., относится к моделированию процессов с использованием генератора случайных чисел.

Методы моделирования Монте-Карло делятся на итеративные и неитеративные. К неитеративному относят метод генерирования выборки по важности (importance sampling) и метод отбраковки или принятие выборки (rejection or acceptance sampling) [5].

Неитеративные методы Монте-Карло. Рассмотрим некоторое вероятностное распределение или вероятностную плотность распределения $p(x)$ для дискретного или непрерывного процесса. Распределение $p(x)$ называют еще целевым распределением, или целевой плотностью распределения. Необходимо сгенерировать выборочные значения $\{X_i\}_{i=1}^N$ по распределению $p(x)$ и найти их статистическую оценку, т.е. оценить математическое ожидание функции $\phi(x)$ по распределению:

$$E_p[\phi] = \int \phi(x)p(x) dx. \quad (10)$$

Допустим, что x — вектор из \mathbb{R}^n с компонентами X_i , а $\phi(x)$ — некоторая функция. Для выборочных значений $\{X_i\}_{i=1}^N$ можно записать

$$\hat{\Phi} = \frac{1}{N} \sum_i \phi(X_i). \quad (11)$$

Поскольку вектор $\{X_i\}_{i=1}^N$ сгенерирован по $p(x)$, то математическое ожидание $\hat{\Phi}$ равняется $E_p[\phi]$; при увеличении количества выборочных значений N дисперсия $\hat{\Phi}$ пропорционально уменьшается.

Моделирование выборки по важности — это не метод генерирования выборочных значений $\{X_i\}_{i=1}^N$ по распределению $p(x)$, а только метод оце-

нивания ожидания $\phi(x)$, представленного выражением (10). Основная идея метода состоит в моделировании по другим, более простым аппроксимирующим распределениям, например $q(x)$, которое довольно близко к целевому распределению $p(x)$. Алгоритм моделирования состоит из следующих основных шагов:

1) выбираем значения X_i , сгенерированные по $q(x)$;

2) поскольку моделирование происходило по «ошибочному» распределению, формируем весовой коэффициент

$$\omega_i = \frac{p(X_i)}{q(X_i)}; \quad (12)$$

3) находим оценку искомой величины $E_p[\phi]$ из выражения

$$\hat{\Phi} = \frac{1}{N} \sum_i \omega_i \phi(X_i) \quad \text{или} \quad \hat{\Phi} = \frac{\sum_i \omega_i \phi(X_i)}{\sum_i \omega_i}. \quad (13)$$

Основной недостаток такого метода заключается в том, что относительное возрастание дисперсии, обусловленное непостоянными весовыми коэффициентами, имеет большую зависимость от выбора аппроксимирующего распределения $q(x)$, а также то, что все выборочные значения нивелируются небольшим количеством значений с большими весовыми коэффициентами.

Метод отбраковки, или принятия выборки состоит в моделировании по другим аппроксимирующим распределениям, но перерасчет весового коэффициента выполняется в процессе генерирования, при этом сохраняется только часть смоделированных точек измерений. Основной алгоритм генерирования выборки состоит из следующих шагов:

1) отыскиваем масштабирующую константу M , такую что

$$\forall x \quad p(x) \leq M q(x); \quad (14)$$

2) независимо выбираем значения X_i , сгенерированные по $q(x)$ и U_i , равномерно распределенные на интервале $[0, 1]$;

3) если $M U_i q(X_i) < p(X_i)$, то принимаем сгенерированные значения X_i , в противном случае — отбрасываем (отбраковываем, отсюда и название метода) и возвращаемся к шагу 2;

4) шаги 2), 3) повторяем до тех пор, пока не найдем N значений X_i .

Недостатком этого алгоритма, как и в предшествующем случае, являются проблемы при больших размерах выборки; алгоритм вырабатывает «приемлемые» случайные числа только за $1/M$ времени, если распределения совпадают. Если M слишком большое, то смоделированные значения отбрасываются.

Таким образом, описанные выше методы работают удовлетворительно только в тех случаях, когда избранное для генерирования распределение

$q(x)$ подобно целевому распределению $p(x)$. Однако в большинстве сложных задач очень тяжело создать отдельное распределение, которое удовлетворяло бы всем свойствам.

Итерационные методы Монте-Карло для марковских цепей. Итерационные методы моделирования Монте-Карло базируются на идее построения марковской цепи и, в отличие от предшествующих методов, используют варианты (кандидаты) распределения $q(x)$ (плотностей распределения), которые зависят только от текущего состояния X_t [5, 13]. К этим методам можно отнести генерирование выборки по Гиббсу, алгоритм Метрополиса, Метрополиса–Хастингса и др.

Рассмотрим стохастический процесс $\{X_t\}$, где $X_t \in \Theta$. Если при заданном значении X_t значение X_h (при $h > t$) не зависит от значений X_s при $s < t$, то он марковский. Иначе говоря, $\{X_t\}$ — марковский процесс, если его условная функция распределения удовлетворяет равенству

$$P(X_h / X_s, s \leq t) = P(X_h / X_t), \quad h > t. \quad (15)$$

Если $\{X_t\}$ — дискретный стохастический процесс, то его основная характеристика имеет вид

$$P(X_h / X_t, X_{t-1}, \dots) = P(X_h / X_t), \quad h > t. \quad (16)$$

Если A — подмножество Θ , то функция

$$P_t(\theta, h, A) = P(X_h \in A / X_t = \theta), \quad h > t, \quad (17)$$

выступает как функция переходной вероятности марковского процесса. Если переходная вероятность зависит от $h - t$, а не от t , то процесс имеет стационарную переходную вероятность.

Основой моделирования с помощью марковской цепи служит построение марковского процесса, для которого стационарное распределение переходов определяется функцией $P(\theta / X)$. Процесс моделирования довольно длительный; он продолжается до тех пор, пока распределение текущих значений процесса не приблизится к стационарному распределению переходов. Таким образом, для заданного распределения $P(\theta / X)$ может быть сконструировано большое количество марковских цепей с заданными параметрами. Методы, которые используют моделирование случайных величин марковской цепью для получения распределения $P(\theta / X)$, относят к методам Монте-Карло для марковских цепей (МКМЦ).

Метод генерирования выборки Гиббса представляет собою утонченный способ формирования выборки из общих распределений многомерных переменных путем применения многомерных выборок из определенных одномерных условий. Например, для двумерной общей плотности распределения $f(x, y)$ при расчете используют условные плотности распределений $f(x / y)$ и $f(y / x)$. Основной алгоритм генерирования можно описать следующим образом:

- 1) выбор начальных значений X_0 и Y_0 в соответствии с принятым распределением;
- 2) генерирование Y_{t+1} по условному распределению $f(y/X_t)$;
- 3) генерирование X_{t+1} по условному распределению $f(x/Y_{t+1})$;
- 4) многократное повторение шагов 2, 3 для того, чтобы цепь достигла сходимости к своему стационарному распределению.

В результате получаем набор значений (X_t, Y_t) , которые при $t = 1, \dots, N$, $N \rightarrow \infty$, будут совпадать с общим распределением $f(x, y)$. На практике используют достаточно большое количество измерений N , при котором отвергают первые M ($M < N$) случайных значений итераций Гиббса, которые считают образцами для испытаний на отказ (обычно значение M выбирают равным 10–20 % от N). Испытание на отказ используется для обеспечения близости выборочных значений к общему распределению $f(x, y)$. В общем случае сходимость метода Гиббса можно записать так:

$$\frac{1}{N - M} \sum_{t=M+1}^N g(X_t, Y_t) \rightarrow \int g(x, y) f(x, y) dx dy, \quad N \rightarrow \infty. \quad (18)$$

Заметим также, что рассмотренный метод представляет собой прохождение по длинной цепи с сохранением всех случайных измерений после испытаний на отказ для получения выборочных значений Гиббса. Кроме того, можно запускать несколько относительно коротких цепей, используя различные начальные значения при относительно маленьких N . Случайное значение последней итерации Гиббса в каждой цепи используется для формирования текущего значения выборки Гиббса.

Итак, выборка Гиббса имеет преимущество при декомпозиции многомерной задачи оценивания задачами меньшей размерности путем применения полных условных распределений параметров. Эта особенность делает выборку Гиббса простой и широко применимой. Однако довольно часто недостаточно сводить все значения выборки Гиббса к одномерной задаче. Если параметры сильно коррелированы, приходится измерять их совместно. Для того чтобы убедиться в сходимости метода Гиббса, часто процедуру повторяют несколько раз с разными начальными значениями.

Другой метод — алгоритм Метрополиса — применяется, например, в тех случаях, когда вероятностное распределение известно, за исключением нормирующей константы. Допустим, что необходимо получить случайную выборку из распределения с целевой плотностью $f(x)$, которая содержит сложную нормирующую константу; при этом прямое получение выборки или трудоемкое или невыполнимое. В соответствии с процедурой выбирают аппроксимирующее распределение, для которого легко генерировать случайные значения, т.е. определяют кандидата на функцию плотности, $q(v, x)$. Алгоритм Метрополиса генерирует последовательность случайных измерений из аппроксимирующего распределения $q(v, x)$, сходящегося к $f(x)$. Алгоритм можно представить следующими шагами:

- 1) выбрать плотность аппроксимирующего распределения $q(v, x)$;

2) задать текущее состояние цепи: выбирая значение X_t , сгенерировать значения V по распределению $q(v, X_t)$ (которое называют еще «скачкообразным»), такое распределение должно быть симметричным, т.е. $q(v, x) = q(x, v)$ для всех v и x .

3) вычислить показатель пропускной способности (или значение скачка):

$$r = \frac{f(V)}{f(X_t)}; \quad (19)$$

4) если $r \geq 1$, то $X_{t+1} = V$; если же $r < 1$, то

$$X_{t+1} = \begin{cases} V & \text{с вероятностью } r, \\ X_t & \text{с вероятностью } 1 - r. \end{cases} \quad (20)$$

Повторение алгоритма несколько раз при некоторых регулярных нежестких условиях обеспечивает сходимость последовательности $\{X_t\}$ к распределению $f(x)$.

Правило принятия и отклонения сгенерированного значения для этого алгоритма может базироваться на таких условиях:

(i) если скачок от X_t к V увеличивает плотность распределения, то $V = X_{t+1}$;

(ii) если скачок уменьшает плотность распределения, то $X_{t+1} = V$ с вероятностью, равной показателю r , и устанавливают $X_{t+1} = X_t$ в противном случае с вероятностью $(1 - r)$.

Примерами симметричной плотности $q(v, x)$ могут служить нормальное распределение и распределение t -Стюдента для средних значений параметров модели.

В 1970 г. Хастингс обобщил алгоритм Метрополиса: во-первых, скачкообразное распределение, т.е. аппроксимирующую плотность распределения, не обязательно выбирать симметричной; во-вторых, правило прыжка изменяется таким образом:

$$r = \frac{f(V)/q(V, X_t)}{f(X_t)/q(X_t, V)} = \frac{f(V)q(X_t, V)}{f(X_t)q(V, X_t)}. \quad (21)$$

Такую модификацию называют алгоритмом Метрополиса–Хастингса. Аналогично с помощью настоящего алгоритма можно генерировать случайные измерения с любой функцией плотности, при этом симметричность распределения не требуется.

Как и в случае выборки Гиббса, необходимо использовать достаточное количество значений N и исключать из рассмотрения первые M значений. Алгоритм Метрополиса–Хастингса можно представить так:

$$\frac{1}{N - M} \sum_{t=M+1}^N g(X_t) \rightarrow \int g(x)f(x)dx, \quad N \rightarrow \infty. \quad (22)$$

Для алгоритма Метрополиса–Хастингса выбор аппроксимирующей плотности распределения очень важен. Плотность не должна иметь слишком большую или маленькую дисперсию, должна быть достаточно близкой к целевому распределению. Если показатель пропускной способности слишком большой, то цепь будет вырабатывать много маленьких шагов в окрестности локальных выбросов, увеличивая корреляцию и время сходимости. Таким образом, независимые выборочные значения получают лишь с большими интервалами. Если же этот показатель слишком низкий, то цепь будет «застевать» в отдельных местах. Оптимальной пропускной способностью считается 20–50 %.

Заметим, что индексы $i=1\dots N$ обозначают независимые выборочные значения, которые генерируются по соответствующему распределению. Обозначения $i=1\dots N$ отвечают последовательностям состояний марковской цепи. Алгоритм Метрополиса–Хастингса, как обобщенный вариант алгоритма Метрополиса и выборки Гиббса, в отличие от метода отбраковки выборки, при моделировании N итераций не вырабатывает N независимых выборочных значений по целевым распределениям, а сами выборочные значения коррелируют. Это объясняется тем, что методы МКМЦ включают в себя марковский процесс, в котором сгенерирована последовательность состояний $\{X_t\}$, каждый элемент X_t которой имеет распределение вероятности, зависящее от предшествующего значения X_{t-1} . В течение значительного промежутка времени марковская цепь должна существенно переместиться, чтобы эффективно сгенерировать независимые выборочные значения целевого распределения.

Выбор начальной функции плотности позволяет избежать корреляционной зависимости между выборочными. В частности, если в качестве предложения выбрать плотности $q(v, x) = q(v)$, то получим независимую выборку. Если положить $q(v, x) = f(x - v)$, то получим алгоритм Метрополиса–Хастингса со случайным блужданием. Довольно полезным бывает применение композиций и комбинаций, т.е. использование нескольких видов предложений распределения, которые выбирают случайным образом. Например, формируют независимую выборку по априорным распределениям для дальнейшего анализа апостериорного распределения на больших шагах.

Для определения нижней границы количества итераций, необходимых для того, чтобы выборочные значения были независимыми, для метода Метрополиса–Хастингса довольно часто применяют эмпирическое правило. Если наибольшая длина в пространстве вероятных состояний равняется L , то для получения независимых выборочных значений по алгоритму Метрополиса–Хастингса, где аппроксимирующее распределение моделируется случайным блужданием с размером шага ε , он должен выполняться за наименьшее $T \cong (L/\varepsilon)^2$ количество итераций.

Назовем преимущества описанных выше методов МКМЦ:

— чтобы решить поставленную задачу, необходимо «прогнать» этот метод несколько раз;

— используя алгоритм Метрополиса–Хастингса, можно избежать расчета предельных распределений и нормирующих констант в плотностях распределений;

— можно получить функции параметров $\phi(\theta)$, исходя из имитированных (смоделированных) распределений $\phi(\theta_i)$;

— воспроизведение многих естественных форм и других нестандартных особенностей процессов и явлений, которые моделируются.

К недостаткам методов МКМЦ относят:

— нередуцированность, апериодичность и апостериорность распределений;

— диагностика сходимости методов все еще остается эвристической;

— выборочные значения, полученные методами МКМЦ, могут быть сильно автокоррелированными, поэтому необходимо иметь большой размер выборки, порядка (10^3-10^6) сгенерированных значений.

Преодоление описанных недостатков считается творческим процессом. Диагностика сходимости может включать в себя исследование графиков и тестирование равенства распределений между разными частями цепи. Автокорреляции могут быть частично откорректированы новой параметризацией и группированием выборочных значений. Цепь можно «перезапустить», т.е. повторить выполнение алгоритма, используя разные начальные значения и/или разные генераторы случайных чисел. Для проверки сходимости также можно воспользоваться показателем уменьшения масштаба, который вычисляется по следующему выражению:

$$\hat{R} = \sqrt{\frac{\text{общая дисперсия}}{\text{дисперсия внутри цепи}}}. \quad (23)$$

Если значение \hat{R} приближается к 1, то это обеспечивает сходимость. Однако вопрос сходимости методов МКМЦ все еще остается открытым.

Проверка гипотез и выбор модели. Во многих случаях возникает задача сравнения альтернативных гипотез и моделей. При наличии точно сформулированных гипотез или моделей, которые сравниваются, практический способ их сравнения зависит от цели анализа, состояния априорной информации и от наличия функции потерь. Рассмотрим критерии выбора альтернативных моделей в общем случае, например множество параметрических моделей M_1, M_2, \dots, M_m , которые описывают X_1, \dots, X_n условными распределениями $f(x|\theta, M_i)$, $i = 1, \dots, m$. Предельное распределение данных предполагает, что

$$p(x|M_i) = \int_{\Theta} f(x|\theta, M_i) P(\theta|M_i) d\theta. \quad (24)$$

Главным механизмом формирования статистического вывода является байесовский фактор (коэффициент), который представляет собой отношение апостериорных вероятностей к априорным:

$$BF(i, j) = \frac{P(M_i | x) / P(M_j | x)}{P(M_i) / P(\theta | M_j)} = \frac{P(x | M_i)}{P(x | M_j)}, \quad i, j = 1, \dots, m. \quad (25)$$

Преимущество одной модели над другой определяется в соответствии со значением байесовского фактора. Если он существенно превышает 1, то принимается соответствующее решение. Для выбора модели используют критерий наибольшей предельной плотности распределения $p(x | M_i)$, который отвечает условию $BF(i, j) > 1$.

Часто в расчетах используют аппроксимацию (приближение) байесовского фактора, а именно информационный критерий Байеса–Шварца (Schwarz Bayesian information criterion), который определяется так:

$$-2 \ln BF \approx \Delta BIC = -2 \ln \frac{\sup_{M_i} f(x | \theta, M_i)}{\sup_{M_j} f(x | \theta, M_j)} - (p_j - p_i) \ln n, \quad (26)$$

где n — объем выборки; $\dim M_i = p_j$, $i, j = 1, \dots, m$, — размерность модели.

Байесовские методы сравнения и выбора альтернативных моделей и гипотез составляют унифицированное множество принципов, которые считают функциональными и применимыми к широкому классу задач. Такие методы позволяют учитывать при сравнении и проверке гипотез и моделей априорную информацию. Как показано в [1], байесовский подход к проверке гипотез и моделей — единственный подход, который предусматривает выбор действия на основе максимизации ожидаемой полезности.

Прогнозное распределение. Важнейшим аспектом байесовского подхода, как полноценного подхода к формированию статистического вывода, является прогнозирование. Байесовская парадигма дает достаточный объем информации для прогнозирования. В частности, главным инструментом выступает вероятностное распределение, которое используется в качестве прогнозного.

Если x^* — новое измерение, то можно записать совместное распределение x^* и параметров θ при условии заданной выборки данных $\{X_t\}$:

$$P(x^*, \theta | x) = P(x^* | \theta, x) P(\theta | x). \quad (27)$$

Прогнозное распределение $P(x^* | x)$ получаем, проинтегрировав равенство (27) по θ . Используя (2) и (3), запишем

$$P(x^* | x) = \int_{\Theta} P(x^*, \theta | x) d\theta = \int_{\Theta} P(x^* | \theta) P(\theta | x) d\theta. \quad (28)$$

Прогнозное распределение $P(x^* | x)$ не зависит от θ . Равенство (28) указывает на то, что прогнозное распределение (или прогнозирующая функция распределения вероятности) рассматривается как среднее условных прогнозных распределений $P(x^* | \theta)$, причем весовой функцией служит апостериорное распределение для θ , т.е. $P(\theta | x)$.

Примеры применения МСМС методов к моделям стохастической волатильности. Модель стохастической волатильности Тейлора относится к классу моделей, в которых учитываются изменения дисперсии и ковариаций. Для дискретного времени модель волатильности в достаточно простом варианте имеет такой вид:

$$y_t = \exp\left(\frac{h_t}{2}\right) \varepsilon_t,$$

$$\varepsilon_t \sim \text{Norm}(0, 1), \quad h_{t+1} = \mu + \varphi(h_t - \mu) + \eta_t, \quad (29)$$

$$\eta_t \sim \text{Norm}(0, \sigma_\eta^2),$$

где y_t — средняя откорректированная доходность актива на момент времени t ; h_t — логарифм волатильности на момент времени t ; $\text{Norm}(\cdot, \cdot)$ — нормальное распределение. Параметр μ , как среднее значение h_t , представляет собой масштабирующий коэффициент; φ отображает устойчивость волатильности; σ_η — волатильность логарифмированной переменной. Предполагается, что ε_t и η_t не коррелируют между собой, а процесс h_t стационарный ($|\varphi| < 1$) и имеет начальное условие $h_0 \text{Norm}[\mu, \sigma_\eta^2 / (1 - \varphi^2)]$ [12].

Первое уравнение модели (29) в пространстве состояний определяет условные распределения наблюдений с заданными неизвестными состояниями h_t , т.е. это уравнения измерений:

$$y_t | h_t = \exp\left(\frac{h_t}{2}\right) \varepsilon_t, \quad \varepsilon_t \sim \text{Norm}(0, 1). \quad (30)$$

Неизвестные состояния определяются марковскими переходами во времени и задаются соответствующим уравнением состояния:

$$h_t | h_{t-1}, \mu, \varphi, \sigma_\eta^2 = \mu + \varphi(h_{t-1} - \mu) + \eta_t, \quad \eta_t \sim \text{Norm}(0, \sigma_\eta^2) \quad (31)$$

с начальным условием h_0 .

Байесовская модель состоит из общего априорного распределения из всех ненаблюдаемых переменных, трех параметров $\mu, \varphi, \sigma_\eta^2$, неизвестных состояний h_0, h_1, \dots, h_n и общего распределения наблюдений относительно средней доходности y_1, \dots, y_n . Байесовский вывод основывается на апосте-

риорном распределении ненаблюдаемых параметров при заданных данных, т.е. на определении общей вероятности распределения

$$p(\mu, \phi, \sigma_{\eta}^2, h_0, h_1, \dots, h_n) = p(\mu, \phi, \sigma_{\eta}^2) p(h_0 | \mu, \phi, \sigma_{\eta}^2) \prod_{t=1}^n p(h_t | h_{t-1}, \mu, \phi, \sigma_{\eta}^2), \quad (32)$$

при этом допускает независимость априорных распределений параметров μ , ϕ и σ_{η}^2 .

В качестве априорного распределения выбираем неинформативное распределение для μ : $\mu \sim \text{Norm}(0, 10)$; для ϕ положим $\phi = 2\phi^* - 1$, где $\phi^* \sim \text{Beta}(\alpha, \beta)$, т.е. это бета-распределение с параметрами $\alpha = 20$ и $\beta = 1,5$, которые дают возможность получать значение ϕ в границах $|\phi| < 1$. Для σ_{η}^2 выбираем сопряженное обратное гамма-распределение с такими параметрами: $\sigma_{\eta}^2 \sim \text{Inv Gamma}(5; 0,01)$.

Для общего распределения (32) величину $p(h_t | h_{t-1}, \mu, \phi, \sigma_{\eta}^2)$ можно определить по уравнениям состояния (31). Правдоподобие $p(y_1, \dots, y_n | \mu, \phi, \sigma_{\eta}^2, h_0, h_1, \dots, h_n)$ определяется уравнением измерений (30) и предположениями условий независимости

$$p(y_1, \dots, y_n | \mu, \phi, \sigma_{\eta}^2, h_0, h_1, \dots, h_n) = \prod_{t=1}^n p(y_t | h_t). \quad (33)$$

Далее по теореме Байеса совместное апостериорное распределение ненаблюдаемых параметров при заданных данных пропорционально априорному распределению и правдоподобию, т.е.

$$p(\mu, \phi, \sigma_{\eta}^2, h_0, \dots, h_n | y_1, \dots, y_n) \propto p(\mu) p(\phi) p(\sigma_{\eta}^2) p(h_0 | \mu, \phi, \sigma_{\eta}^2) \prod_{t=1}^n p(h_t | h_{t-1}, \mu, \phi, \sigma_{\eta}^2) \prod_{t=1}^n p(y_t | h_t). \quad (34)$$

Распределение для y_t можно рассматривать как распределение t -Стьюдента с нулевым средним, переменной дисперсией σ_t^2 и неопределенными степенями свободы ϑ для погрешностей наблюдений: $y_t \sim t\text{-Student}(0, \sigma_t^2, \vartheta)$. Априорное распределение для ϑ можно избрать из распределения Хи-квадрат с помощью преобразования $\vartheta = \vartheta^* + 2$, где $\vartheta^* \sim \chi^2$ (8). Распределение для y_t часто рассматривают как нормальное с нулевым средним и переменной дисперсией σ_t^2 , т.е. $y_t \sim \text{Norm}(0, \sigma_t^2)$.

С помощью данной модели исследуем устойчивость ресурсной базы банка на примере анализа текущих счетов клиентов. Необходимо проанализировать волатильность клиентских счетов на возможность оттока денежных средств и дальнейшего определения их устойчивости. Начальные данные, представляющие собой ежедневные остатки денежных средств клиентов $\{x_t\}$ за полтора года (приблизительно 370 значений), необходимо пре-

вратить во временной ряд, который будет указывать на среднюю откорректированную доходность удерживаемого актива, т.е. на среднюю волатильность остатков, с помощью преобразования

$$y_t = \log x_t - \log x_{t-1} - \frac{1}{n} \sum_{i=1}^n (\log x_i - \log x_{i-1}), \quad t = 1, \dots, n. \quad (35)$$

На рис. 1 изображена динамика временного ряда y_t .

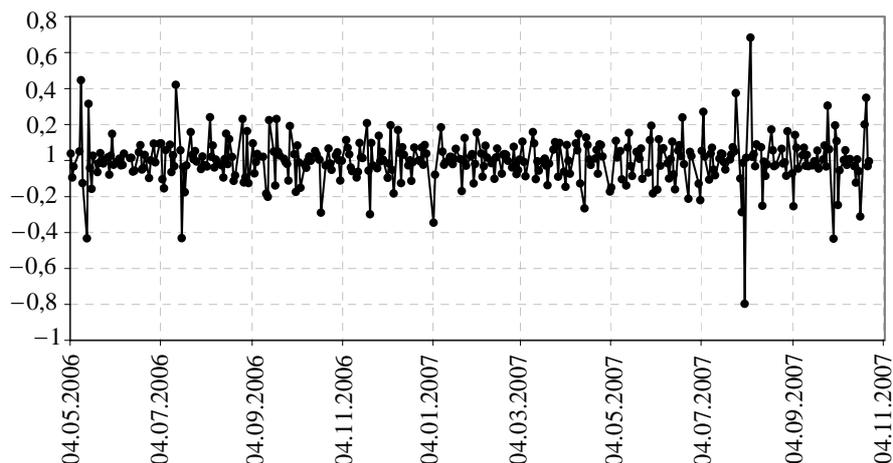


Рис. 1

Для реализации модели использована программная реализация алгоритма Гиббса в системе WinBUGS (версия 1.4.3), которая является свободно доступным программным продуктом. Результаты работы программы и алгоритма генерирования выборки Гиббса представлены в таблице. Значения, которые отвечают 2,5 % и 97,5 %, образуют доверительный интервал для оценок, а начало уточнения оценок означает, что первые 10 тыс. случайных значений отбрасываются, т.е. фактические оценки формируются, начиная с 10001-го сгенерированного значения.

На рис. 2. представлены выборки значений параметров модели и их апостериорные распределения (вид функции плотности распределения), полученные в результате применения метода МКМЦ.

Таблица. Байесовское оценивание параметров модели, вычисленных с помощью WinBUGS

Параметр	Среднее значение	Стандартное отклонение	Погрешность	2,5 %	Медиана	97,5 %	Начало уточнения	Объем данных
$1/\sigma_{\eta}^2$	622,1	230,2	21,24	232,6	595,6	1144,0	10001	10000
μ	-4,486	0,1896	0,00804	-4,739	-4,511	-4,116	10001	10000
ϑ^*	4,252	1,472	0,08974	2,146	3,97	7,966	10001	10000
φ^*	0,9881	0,006572	3,377E-4	0,973	0,989	0,9981	10001	10000

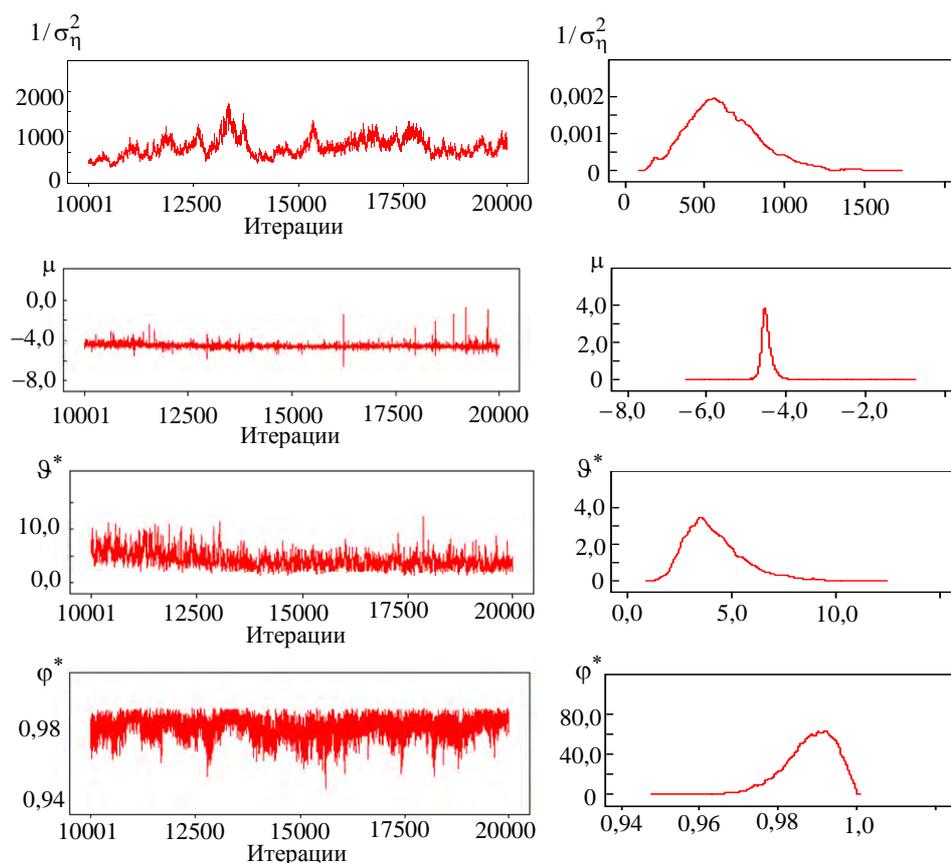


Рис. 2

Выводы. Применение байесовского подхода к формированию статистического вывода дает возможность совсем по иному воспринимать и исследовать оцениваемые модели. Он позволяет оперировать не только полученными оценками, а также соответствующими вероятностными распределениями, применять имеющиеся в разных формах априорные знания исследователя относительно оценок параметров модели. Это дает возможность получать большие объемы исходной информации и точнее описывать структуру и другие характеристики исследуемой модели.

Методика применения вычислительных алгоритмов Монте-Карло, которые базируются на генерировании случайных измерений, тесно связана с байесовской методологией. Этим решается задача генерирования измерений по необходимым вероятностным распределениям. Описанные методы значительно расширяют и улучшают возможности байесовского анализа и сферу его применения в эргатических системах. Одним из преимуществ такого подхода является создание и анализ моделей по данным разной, в частности маленькой, размерности.

Рассмотренный пример показывает, что для оценивания таких сложных нелинейных моделей, как модель стохастической волатильности (подобные модели иногда довольно сложно оценивать с помощью известных классических методов), позволяет получить приемлемые оценки, оперируя априорными и апостериорными распределениями параметров и используя алгоритмы МКМЦ.

В целом методы Монте-Карло для марковских цепей трансформируются в относительно несложные для программирования алгоритмы, направленные на реализацию методов байесовского анализа. При этом следует указать, что методы данного класса требуют углубленного анализа и осмысления. В частности, необходимо исследовать точность оценивания сложных нелинейных моделей, сходимость оценок и их прогнозные характеристики, провести сравнение с другими, более известными методами.

1. *Зельнер А.* Байесовские методы в эконометрии. — М.: Статистика, 1980. — 434 с.
2. *Савчук В.П.* Байесовские методы статистического оценивания: Надежность технических объектов. — М.: Наука, 1989. — 328 с.
3. *Справочник по прикладной статистике /* Под ред. Э. Ллойда, У. Ледермана — М.: Финансы и статистика, 1989. — 525 с.
4. *Bergman N.* Recursive Bayesian estimation: navigation and tracking applications / Linkoping University (Sweden). — 1999. — N 579. — 219 p.
5. *Besag J.* Markov chain Monte Carlo for statistical inference // Working Paper. — 2001. — N 9. — 25 p.
6. *Carlin B.P., Louis T.A.* Bayes and empirical Bayesian methods for data analysis. — London: Chapman and Hall, 1996. — 418 p.
7. *Chib S., Nardari F., Shepard N.* MCMC methods for generalized SVM / TR OX1 1NF. — Oxford, UK, 1998. — 24 p.
8. *Dueker M.* Kalman filtering with truncated normal state variables for Bayesian estimation of macroeconomic models // Working Paper. — 2005. — N 057B. — 55 p.
9. *Geweke J., Tanizaki H.* Bayesian estimation of state-space models using M-H algorithm and Gibbs sampling // Comput. Statistics and Data Analysis. — 2001. — 37, N 2. — P. 151–170.
10. *Бард Й.* Нелинейное оценивание параметров. — М.: Статистика, 1979. — 349 с.
11. *Kim S., Shephard N., Chib S.* Stochastic volatility: Likelihood inference and comparison with ARCH models // Review of Econom. Stud. — 1998. — 65. — P. 361–393.
12. *MacKay D.J.C.* Information theory, inference, and learning algorithms. — Cambridge: Cambridge University Press, 2003. — 640 p.
13. *Nigel Da Costa Lewis* Market risk modeling. Applied statistical methods for practitioners. — London: Risk Waters Group Ltd., 2003. — 238 p.
14. *Tsay R.S.* Analysis of financial time series. — New York: Wiley & Sons, Inc, 2002. — 448 p.

Институт прикладного системного анализа Национального
технического университета Украины «КПИ», Киев,
Международный научно-учебный центр
информационных технологий и систем
НАН Украины и Министерства образования
и науки Украины, Киев

Получено 25.12.2007