ТЕОРЕТИЧНІ ТА ПРИКЛАДНІ ПРОБЛЕМИ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

УДК 519.24

ПРИКЛАДНИЙ СТАТИСТИЧНИЙ АНАЛІЗ МАРКОВСЬКИХ ГАУССОВИХ ПРОЦЕСІВ У ДИСКРЕТНОМУ ЧАСІ

М.В. АНДРЄЄВ

Здійснено перевірки низки статистичних гіпотез щодо приналежності спостережуваної в дискретному часі випадкової вибіркової реалізації будь-якого реального процесу до марковської гауссової послідовності.

У практиці аналізу різних явищ у природі та суспільстві (фізичних, економічних, соціальних та інших) широке розповсюдження мають статистичні методи, які ґрунтуються, головним чином, на фундаментальних результатах теорії ймовірностей і випадкових процесів у неперервному часі. Однак такі процеси в деякому сенсі лише математична абстракція, неадекватна реальним процесам у дискретні моменти впродовж досить значного відрізку часу. Так, для процесів, які описують динаміку ризикових екстремальних ситуацій або ризикових процесів обміну валют на фінансовому ринку [1], страхових ризиків у процесі страхової діяльності [2], характерним є те, що зміна їх станів відбувається у дискретному часі. У зв'язку з цим постає нагальна задача розвинення статистичного апарату дослідження випадкових процесів у дискретному часі або випадкових послідовностей, зокрема марковських гауссових.

У статті досліджується задача статистики марковської гауссової послідовності за відомою її реалізацією на фіксованому інтервалі часу [3]. Розглядається низка статистичних гіпотез, перевірка яких дозволяє робити висновки про можливості опису процесу конкретної природи за допомогою моделі стаціонарного марковського гауссового процесу у дискретному часі. Процеси такого роду широко застосовуються у різних сферах людської діяльності завдяки порівняно простому конструктивному визначенню.

Як відомо, дійсний випадковий процес $X(t)$ є стаціонарним, якщо густина його скінченномірного розподілу не змінюється за будь-якого зсуву впродовж осі часу, тобто для любих n і τ

$$f_{t_1,\ldots,t_n}(x_1,\ldots,x_n) = f_{t_1+\tau_1,\ldots,t_n+\tau}(x_1,\ldots,x_n)\,, \tag{1}$$

де x_i, $i=1,\ldots,n$ — стан процесу відповідно у t_i та ($t_i+\tau$) моменти часу.

З цього визначення з легкістю можна отримати основні властивості характеристик стаціонарних процесів: математичне сподівання $MX(t)$ та дисперсія $DX(t)$ не залежать від часу, тобто постійні, а кореляційна функція $R_X(t_1,t_2)$ залежить від різниці моментів часу t_2-t_1, в яких розглядаються значення X_{t_1} та X_{t_2} випадкового процесу $X(t)$.

СТАТИСТИКА ГАУССОВИХ ПОСЛІДОВНОСТЕЙ

Випадковий процес $X(t)$ називають гауссовим, якщо всі його скінченномірні ймовірносні розподіли є гауссовими або нормальними, тобто для будь-якого n

$$f_{t_1,\dots,t_n}(x_1,\dots,x_n)=\frac{1}{\sqrt{2\pi R\,\sigma_1\dots\sigma_n}}\,e^{-\frac{1}{2R}\sum_{i=1}^{n}\sum_{k=1}^{n}R_{ik}\frac{(x_i-m_i)(x_k-m_k)}{\sigma_i\,\sigma_k}}\,, \qquad (2)$$

де $$m_k=MX(t_k),\quad \sigma_k^2=DX(t_k)\,,\quad R=\begin{vmatrix} r_{11} & \cdots & r_{1n}\\ \dots & \dots & \dots \\ r_{n1} & \cdots & r_{nn}\end{vmatrix};$$

r_{ik} — коефіцієнт кореляції дійсних випадкових величин $X(t_i),\ X(t_k)$; R_{ik} — алгебраїчне доповнення елемента r_{ik} у визначнику R, побудованого за кореляційною функцією $R_X(t)$, яку в дискретному часі можна отримати за допомогою оберненого перетворення Фур'є спектральної густини дійсного нормального процесу [4]

$$R_X(t)=\int_{-\pi}^{\pi}e^{itf}\,\Gamma_{XX}(f)\,df\,, \qquad (3)$$

де $\Gamma_{XX}(f)$ — спектральна густина потужності процесу $X(t)$ (інтеграл по нескінченному інтервалу замінено інтегралом по відрізку довжини 2π).

Отже, для визначення густини функції скінченномірного розподілу будь-якого порядку для гауссового випадкового процесу достатньо знати його математичне сподівання та кореляційну функцію, оцінки яких у часовій області можна отримати таким чином.

Експериментальні дані, які використовують для визначення ймовірносних характеристик оцінок процесу, отримують за допомогою лабораторного експрес-аналізу в моменти часу t_s, $s=1,2,\dots,n$. Оскільки це пов'язано із значними затратами часу, то ймовірносні характеристики гауссового процесу можуть бути отримані за одною лише його вибірковою реалізацією. У цьому випадку природно припустити, що процес ергодичний, тобто для визначення оцінок числових характеристик його скінченномірного розподілу користуються усередненням в часі їх вибіркових значень.

Для дійсного гауссового стаціонарного процесу умова

$$\lim_{|\tau|\to\infty}R_X(\tau)=0 \qquad (4)$$

є достатньою для того, щоб визначити секвенціальні та асимптотично незміщені оцінки математичного сподівання та кореляційної функції за одною його вибірковою реалізацією.

Якщо інтервал часу $(0,T)$, на якому спостерігається вибіркова реалізація $x(t)$ процесу $X(t)$, розбити на N інтервалів довжини Δ, $T = n\Delta$ та ввести позначення

$$s = s\Delta, \quad s = 1,2,..., \quad \tau = \nu = 1,2,..., \tag{5}$$

то оцінки математичного сподівання m_X та кореляційної функції $R_X(t)$ процесу $X(t)$ можна визначити за формулами

$$m_x = \frac{1}{n}\sum_{s=1}^{n} x_s , \tag{6}$$

$$R_x(\nu) = \frac{1}{n-\nu-1}\sum_{s=1}^{n-\nu} \overset{\circ}{x}_s \overset{\circ}{x}_{s+\nu} , \tag{7}$$

де $x_s = x(s\Delta)$; $\overset{0}{x}_s = x_s - m_s$; $\overset{0}{x}_{s+\nu} = x[(s+\nu)\Delta] - m_x$.

При визначенні m_x і $R_x(\nu)$ за формулами відповідно (6) та (7) виникає питання про те, який інтервал дискретності може бути прийнятий без істотного зменшення точності отриманих оцінок. Зазвичай середній інтервал дискретності гауссового стаціонарного процесу $X(t)$ визначають за формулою

$$\Delta = \frac{\pi}{\lambda} e^{\frac{z}{2}[1-2\Phi(z)]} , \tag{8}$$

де $\Phi(z)$ — функція Гаусса–Лапласа;

$$z = \frac{x_{\max} - x_{\min}}{\sigma_x} ;$$

$x_{\max}$, $x_{\min}$ — границі смуги, якій належить вибіркова реалізація $x(t)$ процесу $X(t)$; σ_x — його середньоквадратичне відхилення;

$$\lambda = R''_x(t)\big|_{t=0} . \tag{9}$$

Після визначення достатньої кількості точок, які належать вибірковій кореляційній функції $R_x(t)$ та побудови її графіка, необхідно апроксимувати цю функцію простим аналітичним виразом, зручним для подальшого дослідження. Невідомі коефіцієнти в апроксимаційному виразі визначаються за декількома найбільш характерними точками кривої $R_x(t)$. Оскільки точність оцінки $R_x(t)$, як правило, падає з ростом t, то в якості таких характерних точок не беруться точки, розміщені порівняно далеко від початку координат. Відтак перевіряється умова (4) спадання вибіркової кореляційної функції $R_x(t)$ до нуля, що дозволяє на емпіричному рівні переконатися в ергодичності процесу $X(t)$.

СТАТИСТИКА МАРКОВСЬКИХ ГАУССОВИХ ПОСЛІДОВНОСТЕЙ

Перейдемо до перевірки властивості марковості процесу $X(t)$. За визначенням марковський процес $X(t)$ задовольняє таку умову: якщо в момент часу t_1 відоме значення процесу $x(t_1)$, то будь-який ймовірносний прогноз щодо процесу в майбутньому не залежить від того, як він протікав до моменту t_1, тобто майбутнє не залежить від минулого, коли нам відоме теперішнє. Для марковських процесів у дискретному часі його скінченномірні ймовірносні розподіли будь-якого порядку визначаються умовними перехідними ймовірностями за один крок, так що густина n-мірного його розподілу $f(x_1,...,x_n)$ у символьному вигляді визначається як

$$f(x_1,...,x_n) = p(x_1)p(x_2|x_1),...,p(x_n|x_{n-1}), \tag{10}$$

де $p(x_i|x_{i-1})$ — задана умовна ймовірність переходу x_{i-1} у x_i, $i=1,...,n$ за один крок, а $p(x_1)$ — задана ймовірність x_1-го стану.

Розглянемо статистичну задачу про приналежність спостереженої вибіркової реалізації гауссового стаціонарного процесу марковському процесу. Нехай на інтервалі часу $(0,T)$ спостерігається реалізація $x(t)$ деякого стаціонарного гауссового процесу $X(t)$, $t=0,1,2,...$ із фазовим простором X. Необхідно винести рішення про прийняття однієї з гіпотез.

- H_0^1 : випадковий процес $X(t)$ є марковським процесом.
- H_1^1: $X(t)$ таким не являється.

За гіпотезою H_0^1 процес має таку властивість: коли відомо, що $X(t)=i$ ($i \in X$, $t=1,2,...,n$), то випадкові величини $X(t-1)$ та $X(t+1)$ є незалежні (позначимо їх для простоти відповідно X_i^1 та X_i^2). Прості гіпотези H_i^1, $i=0,1$ є неповними щодо перевірки властивості марковості процесу $X(t)$. Повна перевірка цієї властивості полягає у встановленні незалежності у даному випадку двох сімейств випадкових величин $X(0),X(1),...,X(t-1)$ та $X(t+1),X(t+2),...,X(T)$.

Обмежимось найпростішим випадком, з якого, одначе, стає ясною методика збору статистичного матеріалу щодо перевірки складних гіпотез про незалежність сімейств випадкових величин. Зведемо цю задачу до більш простих задач, застосовуючи метод декомпозиції.

Для перевірки незалежності випадкових величин X_i^j, $j=1,2$ введемо дискретний фазовий простір I шляхом дискретизації неперервного фазового простору X. При цьому правило про прийняття гіпотез H_0^2, H_1^2 щодо марковості або немарковості $X(t)$ можна сформулювати таким чином.

- H_0^2 : X_i^1 та X_i^2 незалежні для всіх $i \in I$.
- H_1^2 : X_i^1 та X_i^2 залежні хоча б при одному $i \in I$.

В силу стаціонарності процесу $X(t)$ емпіричні розподіли величин X_i^1 та X_i^2, $i \in I$ будуємо за частотою появи у вибірковій реалізації $x(t)$ процеса $X(t)$ значень, які спостерігаються та безпосередньо передують стану i, а також слідують за ним.

Емпіричний розподіл X_i^j має вигляд

$$\begin{array}{ll} \text{варіанти} & x_i^j:\ x_1^j,\, x_2^j, \ldots, x_s^j, \\ \text{частоти} & n_i^j:\ n_1^j,\, n_2^j, \ldots, n_s^j. \end{array} \tag{11}$$

Переходимо до перевірки гіпотез щодо гауссовості або нормальності розподілів X_i^j.

- H_0^3 : X_i розподілена за нормальним законом.
- H_1^3 : X_i розподілена за ненормальним законом.

Для перевірки гіпотези H_0^3 за заданим рівнем значущості скористаємось критерієм погодження К.Пірсона, в якому поряд з емпіричними мають бути обчислені і теоретичні частоти n'. За припущенням, що справедлива гіпотеза H_0^3, обчислимо теоретичні частоти у такий спосіб.

1. Весь інтервал спостережень X поділимо на s часткових інтервалів (x_i, x_{i+1}) однакової довжини. Знайдемо середини часткових інтервалів $x_i^* = \dfrac{x_i + x_{i+1}}{2}$ і в якості частоти n_i варіанти x_i^* приймемо число варіант, які попали в i-й інтервал. В результаті отримаємо послідовність рівновіддалених варіант і відповідних їм частот.

$$\begin{array}{l} x_1^*\ x_2^* \ldots x_s^*, \\ n_1\ n_2\ \ldots n_s, \\ \sum\limits_{l=1}^{s} n_l = 1. \end{array} \tag{12}$$

2. Обчислимо, наприклад, методом добутків, вибіркову середню $\bar{x}^*$ та вибіркове середньоквадратичне відхилення σ^*.

3. Пронормуємо випадкову величину і обчислимо кінці інтервалів (z_i, z_{i+1}).

$$z_i = \frac{x_i - \bar{x}^*}{\sigma^*}, \quad z_{i+1} = \frac{x_{i+1} - \bar{x}^*}{\sigma^*},$$

причому найменше значення z, тобто z_1, покладемо рівним $-\infty$, а найбільше, тобто z_s, — рівним ∞.

4. Обчислимо теоретичні ймовірності P_i попадання X в інтервали (x_i, x_{i+1}) за рівністю $P_i = \Phi(z_{i+1}) - \Phi(z_i)$, де $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_0^z e^{\frac{u^2}{2}} dz$ — функція Гаусса–Лапласа, та , нарешті, знайдемо шукані теоретичні частоти $n_i' = nP_i$.

Після обчислення теоретичних частот обчислимо спостережуване значення критерію

$$\left(\chi^2_{\text{спост}}\right)^j = \sum_i \frac{n_i^j - (n_i')^j}{(n_i')^j}, \tag{13}$$

а за таблицею критичних точок розподілу χ^2 на даному рівні значущості α та числі ступенів вільності $k = s - 3$ знайдемо критичну точку $\chi^2_{\text{кр}}(\alpha, k)$. Якщо $\left(\chi^2_{\text{спост}}\right)^j < \left(\chi^2_{\text{кр}}\right)^j$, то гіпотезу H_0^3 приймають. Зазначимо, що об'єм виборки має бути досить великим, принаймні, не меншим 50.

Отже, у випадку прийняття гіпотези H_0^3 випадкові величини X_i^j, $j = 1, 2$ розподілені за нормальним законом і, згідно теореми Крамера, для обгрунтування їх незалежності досить переконатись у рівності їх математичних сподівань та дисперсій. Для порівняння числових характеристик розподілів величин X_i^j, $j = 1, 2$ користуються їх емпіричними функціями розподілу.

Для порівняння дисперсій величин X_i^j знайдемо виправлені вибіркові дисперсії $S^2_{x_i^1}$ та $S^2_{x_i^2}$. За виправленими дисперсіями на заданому рівні значущості α перевіримо гіпотези

- $H_0^4 : D\left(X_i^1\right) = D\left(X_i^2\right)$,
- $H_1^4 : D\left(X_i^1\right) \neq D\left(X_i^2\right)$.

В якості критерію перевірки нульової гіпотези H_0^4 приймемо відношення більшої виправленої дисперсії до меншої — критерій Фішера $F = \frac{S_Б^2}{S_м^2}$.

Статистика F за умови справедливості нульової гіпотези має розподіл Фішера-Снедекора із ступенями вільності $k_1 = k_2 = s - 1$. Цей розподіл залежить тільки від числа ступенів вільності.

Побудуємо двосторонню критичну область, виходячи з того, що ймовірність попадання критерію в цю область за припущенням справедливості гіпотези H_0 має бути рівною прийнятому рівню значущості α. Виявляється, що найбільша потужність (ймовірність попадання критерію в критичну область за справедливості альтернативної гіпотези H_1^4) досягається тоді, коли ймовірність попадання критерію в кожний із двох інтервалів критичної області дорівнює $\frac{\alpha}{2}$. Відтак, за таблицею критичних точок розподілу Фішера-Снедекора, на рівні значущості $\frac{\alpha}{2}$ і числі ступенів вільності $k_1 = k_2 = s - 1$ знайдемо критичну точку $F_{\text{кр}}\left(\frac{\alpha}{2}, s-1, s-2\right)$.

Якщо $F^{\text{j}}_{\text{спост}} < F^{\text{j}}_{\text{кр}}$, то немає жодних підстав відхиляти нульову гіпотезу H_0^4.

При порівнянні математичних сподівань величин X_i^j, $j = 1,2$, дисперсії яких однакові і невідомі, користуються критерієм Стьюдента з $k = 2(n-1)$ ступенями вільності.

Для того аби за заданим рівнем значущості α перевірити гіпотезу

- $H_0^5 : M\,X_i^1 \;=\; M\,X_i^2$

за альтернативної гіпотези

- $H_1^5 : M\,X_i^1 \neq\; M\,X_i^2$,

обчислимо спостережуване значення критерію

$$T_{\text{спост}} = \frac{\bar{x}_i^1 - \bar{x}_i^2}{\sqrt{(n-1)\left(S^2_{\bar{x}_i^1} - S^2_{\bar{x}_i^2}\right)}}\sqrt{n(n-1)}\,, \tag{14}$$

де x_i^j, S_{x^j}, $i = 1,2$ — вибіркові математичні сподівання та дисперсії величин x_i^j, $i = 1,2$. За таблицею критичних точок розподілу Стьюдента, на заданому рівні значущості та числі ступенів вільності $\upsilon = 2(n-1)$ знаходимо критичну точку $t_{\text{кр}}(\alpha, k)$. Якщо $|T_{\text{спост}}| < t_{\text{кр}}(\alpha, k)$, то відхиляти нульову гіпотезу немає рації.

Якщо стаціонарний процес $X(t)$ — гауссовий і одночасно марковський, то його визначальною характеристикою являється умовна густина ймовірності переходу за один крок

$$P\left(\overset{\circ}{x}_{n+1}\middle|\,\overset{\circ}{x}_n\right) = \frac{1}{\sigma_\Delta\sqrt{2\pi}}\exp\left\{-\frac{\overset{\circ}{x}_{n+1} - \beta\,\overset{\circ}{x}_n}{2\sigma_\Delta^2}\right\}, \tag{15}$$

де σ_Δ^2 — дисперсія умовного розподілу; $\overset{\circ}{x}_n = x_n - m_x$. Величина β дорівнює значенню нормованої кореляційної функції $r_x(\tau) = \frac{R_x(\tau)}{\sigma_k^2}$ за $\tau = 1$.

ВИСНОВКИ

Означимо можливий напрям застосувань марковського гауссового процесу з незалежними приростами у дискретному часі як математичної моделі ризикового процесу ціноутворення цінних паперів, зокрема, котирування акцій на фондовому ринку. Простота цієї моделі пояснюється порівняно простим застосуванням центральної граничної теореми до реального випадкового процесу котирування акцій як випадкового процесу з незалежними нормальними приростами, для якого стан процесу в будь-який момент часу визначається сумою незалежних приростів процесу до цього моменту часу. Відтак з незначними труднощами обгрунтовується нормальність розподілів відповідних статистик або критеріїв перевірки гіпотез щодо приналежності конкретної вибіркової реалізації цін акцій на фондовому ринку до марковської гауссової послідовності.

Задача адаптивного, або дуального, керування (одночасного оцінювання невідомих станів і параметрів та вибору оптимальних рішень) керованими марковськими гауссовими послідовностями в умовах неповної інформації розглянута в роботі [5]. При цьому марковська гауссова послідовність, яка задається сім'єю перехідних ймовірностей (15), є математичною моделлю одного технологічного процесу масообміну в хімічній промисловості.

Очевидно, що сім'єю перехідних ймовірностей (15) описується процес автокореляції першого порядку, який широко застосовується в технічній, економічній, фінансовій, страховій, соціальній та інших сферах людської діяльності. Спектр цікавих застосувань процесу автокореляції першого порядку продемонстровано в роботі [6], більшість з яких стосується досліджень в області біології, зокрема, в генетиці, де розглянуто проблеми механізму спадковості : успадкування дітьми ознак та властивостей, які притаманні їхнім батькам, тощо.

Марковські гауссові послідовності, які задаються сім'єю перехідних ймовірностей (15), можуть бути використані як математичні моделі процесів суспільного розвитку за умови, що у стані моделі, який характеризує рівень суспільного розвитку, враховується набутий досвід або стратегія, завдяки яким цей рівень досягнуто. Тим самим доходимо висновку, що математичними моделями процесів суспільного розвитку можуть бути керовані марковські гауссові послідовності, стратегія керування якими має набутий соціальний досвід.

ЛІТЕРАТУРА

1. *Андреев Н.В.* Управление риском обмена валют // Вісник АПСВ. Україна: поступ у майбутнє. Спецвипуск до 290-річчя прийняття конституції Пилипа Орлика. — 2000. — С. 88–92.
2. *Андрєєв М.В.* Деякі аспекти оптимізації страхової діяльності // Вісник АПСВ. Україна: шляхами віків. Матеріали конференції, присвяченої 175-річчю з дня народження Георгія Андрузького. Ч. № 2. — 2002. — С. 79–81.
3. *Андреев Н.В., Караченец Д.В., Ткаченко Л.Н.* К статистике гауссовских марковских последовательностей // Адаптивные системы управления. — Киев: ИК АН УССР , 1975. — С. 23–31.

4. *Андрєєв М.В.* Прикладний статистичний аналіз високочастотних випадкових процесів у дискретному часі // Системні дослідження та інформаційні технології. — 2002. — № 4. — С. 61–79.
5. *Караченец Д.В., Андреев Н.В., Массальский Г.Э.* Массообменный процесс как управляемый случайный процесс // Управляемые случайные процессы и системы. — Киев : ИК АН УССР, 1973. — С. 158–175.
6. *Бартлетт М.В.* Введение в теорию случайных процессов. — М.: ИЛ, 1958. — 384 с.

Надійшла 12.11.2002