

ПРИКЛАДНИЙ СТАТИСТИЧНИЙ АНАЛІЗ ВИСОКОЧАСТОТНИХ ВИПАДКОВИХ ПРОЦЕСІВ У ДИСКРЕТНОМУ ЧАСІ

М.В. АНДРЕЄВ

Досліджується спектральний аналіз високочастотних стаціонарних та нестаціонарних випадкових послідовностей на основі перевірки відповідних статистичних гіпотез у частотній області.

ВСТУП

Високочастотні випадкові процеси можна розглядати як математичні моделі різного роду випадкових явищ, які часто зустрічаються у різних сферах людської діяльності. Такими процесами описуються, наприклад, сигнали, що отримують внаслідок вібраційних випробувань у техніці [1], процеси ціноутворення фінансових активів — акцій та інших цінних паперів на фондовому ринку [2] тощо.

У практиці застосувань зазначених моделей, відповідно процесів, можна спостерігати єдину, для кожного своєї реалізацію, що обумовлює статистичну постановку задачі їх оцінки та оптимізаційну задачу адаптивного керування відповідними випадковими моделями та процесами.

У статті узагальнюються результати практичної діяльності автора стосовно статистичного аналізу різних типів стаціонарних і нестаціонарних високочастотних випадкових послідовностей, що зустрічаються при дослідженні систем автоматичного керування технічними, біологічними, економічними, соціальними та іншими системами, а також задач, пов'язаних з конструктивним заданням та експериментальним дослідженням випадкових процесів у дискретному часі.

I. АДАПТИВНІ АЛГОРИТМИ ОЦІНКИ СПЕКТРАЛЬНОЇ ГУСТИНИ ВИСОКОЧАСТОТНИХ ВИПАДКОВИХ ПРОЦЕСІВ

Розглядається задача адаптивного вибору часових інтервалів усереднення для оцінки спектральної густини стаціонарних та нестаціонарних високочастотних випадкових процесів. Від величини інтервалу усереднення залежить довірчий інтервал для оцінки спектральної густини. Математична формалізація задачі та методи дослідження потужності критеріїв еквівалентності спектрів ґрунтуються на основі математичних моделей оцінки спектральної густини. Вибір інтервалів усереднення ґрунтується на основі лінійних порогових критеріїв перевірки еквівалентності спектрів і відповідно цьому вибір залежить від коригування статистичних даних, а тому є адаптивним.

При постановці статистичної задачі та її розв'язанні виникає необхідність у висвітленні таких основних питань:

- аналіз основних особливостей отримання оцінок спектральної густини стаціонарних і нестаціонарних високочастотних випадкових процесів;
- аналіз деяких типових моделей сигналів, що зустрічаються в техніці та фінансах;
- дослідження критеріїв класифікації процесів за типами нестаціонарності, що є основою для адаптивного вибору інтервалів усереднення спектральної густини;
- дослідження адаптивних алгоритмів оцінювання спектральної густини на основі оптимізації потужності критерія еквівалентності спектрів.

1.1. Постановка задачі оцінки спектральної густини високочастотних випадкових процесів

Спектральна густина є перетворенням Фур'є кореляційної функції випадкового процесу; вона показує, як дисперсія випадкового процесу розподілена за частотою.

Обмежимося відомостями про оцінки спектральної густини на основі дискретного перетворення Фур'є.

Вибірковий спектр сигналу $x(t)$ на відрізку $-T/2 \leq t \leq T/2$ визначається як

$$c_{xx}(f) = \frac{\Delta}{N} \left| \sum_{t=-n}^{n-1} x_t \exp(-i2\pi ft \cdot \Delta) \right|^2, \quad -1/2\Delta \leq f < 1/2\Delta, \quad (1.1)$$

де f — частота; Δ — часовий інтервал дискретизації $x(t)$, $N = 2n$ — число вибірових значень $x(t)$:

$$x_t = x(t), \quad t = -n\Delta, \dots, 0, \Delta, \dots, (n-1)\Delta, \quad T = N\Delta. \quad (1.2)$$

Вибірковий спектр $c_{xx}(f)$ в деякій точці f розглядається як значення випадкової величини $C_{XX}(f)$, що є оцінкою спектральної густини.

Спектральна густина $\Gamma_{XX}(f)$ визначається як

$$\Gamma_{XX}(f) = \lim_{N \rightarrow \infty} M\{C_{XX}(f)\}, \quad -1/2\Delta \leq f < 1/2\Delta, \quad (1.3)$$

що математичне сподівання $MC_{XX}(f)$ визначається згідно з [3]:

$$M\{C_{XX}(f)\} = \int_{-\infty}^{\infty} T \left(\frac{\sin \pi Tg}{\pi Tg} \right)^2 \cdot \Gamma_{XX}(f-g) dg. \quad (1.4)$$

Вираз (1.4) показує, що усереднення оцінки $C_{XX}(f)$ відповідає начебто прогляданню теоретичного спектра через спектральне вікно $w(f)$, що визначається як

$$w(f) = T \left(\frac{\sin \pi Tg}{\pi Tg} \right)^2. \quad (1.5)$$

Оскільки $w(f)$ в (1.5) при великих T веде себе як δ -функція Дірака, то з (1.4) випливає, що

$$\lim_{T \rightarrow \infty} M\{C_{XX}(f)\} = \Gamma_{XX}(f), \quad (1.6)$$

тобто оцінка $C_{XX}(f)$ є асимптотично незміщеною оцінкою $\Gamma_{XX}(f)$. Для скінченного T із (1.4) випливає, що $C_{XX}(f)$ у загальному випадку є зміщеною оцінкою $\Gamma_{XX}(f)$, і це зміщення визначається як

$$B(f) = M\{C_{XX}(f)\} - \Gamma_{XX}(f). \quad (1.7)$$

Розподіл оцінки $C_{XX}(f)$ наближено апроксимується χ^2 -розподілом таким чином, що величина $2C_{XX}(f)/\Gamma_{XX}(f)$ розподілена майже як χ^2 з двома ступенями вільності, причому оцінки для частот, кратних $1/T$, практично некорельовані. Відповідно дисперсія оцінки $C_{XX}(f)$ оцінюється як

$$D\{C_{XX}(f)\} \cong \Gamma_{XX}^2(f). \quad (1.8)$$

Для зменшення дисперсії оцінки спектра використовують згладжування вибіркового спектра $C_{XX}(f)$ за допомогою вибору спектрального вікна $w(f)$ або часового вікна при оцінці спектра, вираженого через оцінку кореляційної функції.

Один з підходів до вибору спектрального вікна передбачає мінімізацію середньоквадратичної похибки $\varepsilon^2(f)$, що визначається формулою

$$\varepsilon^2(f) = D\{\bar{C}_{XX}(f)\} + B^2(f), \quad (1.9)$$

де $\bar{C}_{XX}(f)$ — згладжена оцінка спектра, дисперсія якої має вигляд

$$D\{\bar{C}_{XX}(f)\} \cong \Gamma_{XX}^2(f) \frac{I}{T}, \quad I = \int_{-\infty}^{\infty} w^2(g) dg. \quad (1.10)$$

Із формули (1.10), з урахуванням (1.8), випливає, що відношення I/T дорівнює відношенню зменшення дисперсії, обумовленому згладжуванням.

Розподіл величини $\bar{C}_{XX}(f)$ за фіксованого f можна апроксимувати за допомогою $a\chi_\nu^2$ -розподілу, де константи a і ν обчислюються відповідно як

$$\nu = \frac{2T}{I}, \quad (1.11)$$

$$a = \frac{\Gamma_{XX}(f)}{\nu}, \quad (1.12)$$

яке еквівалентно тому, що випадкова величина $\nu C_{XX}(f)/\Gamma_{XX}(f)$ має χ^2 -розподіл з ν ступенями вільності.

Ширина спектрального вікна визначається як

$$b = 1/I = 1 / \int_{-\infty}^{\infty} w(g) dg, \quad (1.13)$$

причому виконується рівність

$$D\{\bar{C}_{XX}(f)\}b = \Gamma_{XX}^2(f) = \text{const}. \quad (1.14)$$

Процедура згладжування Бартлетта. Весь інтервал T розбивається на k інтервалів T_n ($T_n \times k = T$). Для кожного T_n обчислюється вибіркового спектра, і як оцінку спектра на частоті f приймають середнє значення k вибіркового спектра на частоті f :

$$\bar{C}_{XX}(f) = \frac{1}{k} \sum_{i=1}^k C_{XX}^{(i)}(f), \quad (1.15)$$

де $C_{XX}^{(i)}(f)$ - оцінка спектра на i -му інтервалі, $i = \overline{1, k}$.

Спектральне вікно Бартлетта визначається як

$$w^B(f) = T_n \left(\frac{\sin \pi f T_n}{\pi f T_n} \right)^2, \quad (1.16)$$

що відповідає зменшенню дисперсії згладженої оцінки спектра

$$\frac{I^B}{T} = 0,667 \frac{T_n}{T}, \quad (1.17)$$

ν^B — число ступенів вільності визначається як

$$\nu^B = 3 \frac{T}{T_n}. \quad (1.18)$$

Таким чином, з (1.17) випливає, що при $T \rightarrow \infty$ і $T_n = \text{const}$ дисперсія згладженої оцінки (1.15) необмежено зменшується за відповідного збільшення кількості інтервалів.

На практиці є труднощі з вибором інтервалу розбиття T_n , оскільки виникає необхідність, щоб процес $X(t)$ на інтервалі $T = k \times T_n$ був стаціонарним (а також ергодичним).

1.2. Спектральна структура нестационарного процесу

Спектральна густина тут визначається як подвійне перетворення Фур'є нестационарної кореляційної функції $R_{XX}(t_1, t_2)$, тобто

$$S_X(f_1, f_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{XX}(t_1, t_2) e^{2\pi i(f_1 t_1 - f_2 t_2)} dt_1 dt_2, \quad (1.19)$$

де $S_{XX}(f_1, f_2)$ — подвійна за частотою, а кореляційна функція має вигляд

$$R_{XX}(t_1, t_2) = M\{X(t_1), X(t_2)\}. \quad (1.20)$$

Існує другий підхід до вивчення спектральної структури нестационарних процесів, що заключається у виконанні перетворення Фур'є за аргументом τ нестационарної автокореляційної функції $R_X(\tau, t)$, заданої у вигляді

$$R_X(\tau, t) = M \left\{ X \left(t - \frac{\tau}{2} \right) X \left(t + \frac{\tau}{2} \right) \right\}. \quad (1.21)$$

Такий підхід призводить до визначення частотно-часової спектральної густини

$$\varphi_X(f, t) = \int_{-\infty}^{\infty} R_X(\tau, t) e^{-2\pi i f \tau} d\tau. \quad (1.22)$$

Функція $\varphi_X(f, t)$ описує розподіл середнього значення квадрата $M\{X^2(t)\}$ на площині (f, t) , що забезпечує цілком прийнятну фізичну інтерпретацію цієї функції, тобто інтегрування $\varphi_X(f, t)$ за всіма f дає середнє значення квадрата процесу в момент t .

Як $S_x(f_1, f_2)$, так і $\varphi_X(f, t)$ не можуть бути отримані усередненням за часом окремих реалізацій $x(t)$ як у випадку випадкових стаціонарних процесів. Однак існує виняток для локально стаціонарних процесів. Зокрема, якщо процес має сепарабельну нестационарну кореляційну функцію, що визначається як

$$R(\tau, T) = R_1(t) \cdot R_2(\tau), \quad (1.23)$$

тоді $\varphi_X(f, t)$ можна подати у вигляді

$$\varphi_X(f, t) = M\{X^2(t)\} S_X^H(f), \quad (1.24)$$

а при центруванні процесу $X(t)$, тобто при $M\{X(t)\} = 0$, подаємо частотно-часову спектральну густину випадкового процесу, нестационарного за дисперсією

$$\varphi_X(f, t) = D\{X(t)\} S_X^H(f), \quad (1.25)$$

де $S_X^H(f)$ — нормована спектральна густина, тобто

$$\int_{-\infty}^{\infty} S_X^H(f) df = 1. \quad (1.26)$$

Якщо затухання $R_2(\tau)$ в (1.23) проходить досить швидко порівняно зі зміною $R_1(t)$, то в цьому випадку спектральну густину можна оцінити за окремою реалізацією, оцінюючи спочатку функцію $M\{X^2(t)\}$ або $D\{X(t)\}$ за реалізацією на короткому інтервалі часу, що відповідає часу кореляції $R_2(\tau)$, а відтак оцінити функцію $S_X^H(f)$ усередненням реалізації $x(t)$, як і у випадку стаціонарного процесу.

1.3. Про моделі сигналів, отриманих при плануванні експерименту

Для аналізу і розрахунку параметрів алгоритмів оцінки спектральної густини (СГ) розглядаються наближені моделі сигналів, отримані при плануванні експерименту. Зручно скористатися при цьому співвідношенням між СГ вхідного сигналу $S_x(f)$ та СГ вихідного сигналу $S_y(f)$ лінійної системи з передаточною функцією $H(f)$:

$$S_y(f) = |H(f)| \circ S_x(f). \quad (1.27)$$

Моделі нестационарного вихідного сигналу $y(t)$, що поступає на статистичну обробку, можна визначити заданням змінної в часі $H(f)$ або $S_x(f)$, а відтак розбити на ділянки, де має місце стаціонарність.

Друга модель характеризується нестационарністю за дисперсією.

Визначення 1. Допустимий інтервал T_g випадкового процесу $X(t)$ визначається мінімальною тривалістю реалізації $x(t)$, за якої зміщення оцінок вибіркового спектра не перевищує за абсолютною величиною заданої величини ε_g . У загальному випадку ε_g є функцією як f , так і $\Gamma_{XX}(f)$.

Завдяки використанню (1.4) та (1.6) це визначення формалізується:

$$T_g = \min \{T\} \quad (1.28)$$

за умови

$$\left| \int_{-\infty}^{\infty} T \left(\frac{\sin \pi T_g}{\pi T_g} \right)^2 \Gamma_{XX}(f-g) dg - \Gamma_{XX}(f) \right| \leq \varepsilon_g.$$

Визначення 2. Процес називається слабко нестационарним, якщо він є нестационарним процесом, для якого існує допустимий інтервал.

Визначення 3. Нестационарність дисперсії називаються слабкою, якщо відхилення дисперсії від її середнього значення на допустимому інтервалі не перевищує заданої величини ε_g , тобто

$$\left| \int_{t_0}^{t_0+T_g} D_x(\tau) d\tau - D_x(t) \right| \leq \varepsilon_g, \quad t_0 \leq t \leq t_0 + T_g. \quad (1.29)$$

Для сигналів, що охоплюються цими моделями, досить просто отримати оцінки спектральної густини за допомогою процедури усереднення оцінок нормованих спектрів, отриманих на допустимих інтервалах.

Перевірка гіпотези про те, що конкретна реалізація випадкового процесу належить до моделі стаціонарного процесу або нестационарного за дисперсією зводиться до розв'язання двох задач:

1. Перевірка гіпотези про рівність дисперсій, отриманих на допустимих інтервалах.
2. Перевірка гіпотези про еквівалентність нормованих спектрів, отриманих на допустимих інтервалах.

Із задачі 2 впливає стаціонарність або нестаціонарність за дисперсією, якщо до того ж приймається гіпотеза про рівність дисперсій, то звідси впливає стаціонарність на всьому інтервалі.

1.4. Загальна характеристика адаптивного алгоритму вибору інтервалів усереднення

Тут використовується ідея згладжування (усереднення) за допустимими інтервалами. Як відомо, відносно зменшення середньоквадратичного оцінки за рахунок усереднення некорельованих оцінок характеризується величиною

$$\varepsilon = \frac{1}{\sqrt{m}}, \quad (1.30)$$

де m — кількість оцінок, що усереднюються.

При використанні статистичних критеріїв для вибору кількості оцінок m величина m , а отже і ε будуть випадковими величинами. Величина m визначається числом послідовних ділянок, які за критерієм еквівалентні. Рівень значущості критерію α визначається як ймовірність помилки першого роду, а тому для кількості еквівалентних ділянок можна знайти функцію розподілу випадкової величини m

$$F(m) = P\{m < \mu\} = 1 - (1 - \alpha)^m, \quad (1.31)$$

оскільки з урахуванням ергодичної теорії з надійністю $(1 - \alpha)$ можна говорити про знаходження випадкового процесу в довірчому інтервалі, побудованому за його реалізацією. Очевидно, що за такого підходу середнє значення довірчого інтервалу має співпадати з довжиною допустимого інтервалу.

Із формули (1.31), беручи до уваги (1.30), функція розподілу випадкової величини ε має вигляд

$$F(k) = P\{\varepsilon < k\} = 1 - (1 - \alpha)^{1/k^2}. \quad (1.32)$$

2. ВИБІР ТА АНАЛІЗ КРИТЕРІЇВ ПЕРЕВІРКИ ЕКВІВАЛЕНТНОСТІ СПЕКТРІВ

При виборі та аналізі критеріїв перевірки статистичних гіпотез часто використовують припущення щодо гауссовості (нормальності) розподілу статистичних даних. Однак виникає питання, наскільки ймовірно, що це припущення призводить до помилок. Статистична процедура, нечутлива до відхилень від припущень, що лежать в її основі, називається стійкою або робастною. Виявляється, що критерії стосовно генеральних середніх (тобто t -критерій Стьюдента для середнього значення нормального розподілу та різниці середніх двох нормальних розподілів з однаковими дисперсіями) достатньо нечутливі до відхилень від нормальності, а критерії стосовно дисперсій (тобто ν^2 -критерій для дисперсії нормального розподілу), ν^2 -критерій для відношення дисперсій двох нормальних розподілів) дуже чутливі до відхилень, тобто вони нестійкі або неробастні.

На практиці часто дуже важко обґрунтувати рішення про те, чи будуть стандартні процедури з деяким наближенням справедливими, чи призведуть до невірному результату. У тих випадках, коли порушуються або не виконуються припущення нормальності, розглядаються два інших підходи до прийняття статистичних рішень.

Перший підхід полягає в тому, щоб знайти перетворення, за якого розподіл спостережень став би більш близьким до нормального з тим, щоб до перетворення спостережень можна було застосувати нормальну теорію.

Другий підхід є більш радикальним. Замість того, щоб дотримуватися стандартних методів нормальної теорії (через те, що вони є стійкими і з деяким наближенням їх можна застосовувати при відхиленні від нормальності, або через те, що їх можна застосувати з деяким наближенням після перетворення спостережень), стараються знайти статистичні процедури, які б могли бути застосовані в широкому класі розподілів. Такі процедури називаються вільними від розподілів.

Для оцінки еквівалентності спектрів скористаємось саме цими підходами.

2.1. Вибір класу критеріїв

Задачу перевірки еквівалентності спектрів можна сформулювати таким чином.

Задано дві послідовності оцінок спектрів $\{C_1(f)\}$ і $\{C_2(f)\}$ в точках $f = \frac{k}{T}$, $k = 1, 2, \dots, N/2$ з числом ступенів вільності кожної із оцінок ν_1 і ν_2 відповідно. Необхідно перевірити гіпотезу H_0 на рівні значущості α про те, що істинні значення спектрів $\Gamma_1(f)$ і $\Gamma_2(f)$ $\left(|f| \leq \frac{1}{2\Delta}\right)$ цих двох послідовностей співпадають.

Виникає питання, яку інформацію про розподіл оцінок маємо? Як зазначалося, величина $\nu \bar{C}_{XX}(f) | \Gamma_{XX}(f)$ має χ^2 -розподіл з ν ступенями вільності (1.11) і (1.12). Оскільки на практиці істинне значення спектра $\Gamma_{XX}(f)$ невідоме, то для порівняння спектрів зазвичай використовують статистику

$$\nu_{\nu_1, \nu_2}^2 = \frac{C_1(f)}{C_2(f)}, \quad (2.1)$$

яка має розподіл відношення дисперсій (ν^2 — розподіл Фішера) за умови еквівалентності спектрів

$$\varphi_{\nu^2}(\nu_1, \nu_2)(Y) = \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} Y^{(\nu_1-2)/2} \left(1 + \frac{\nu_1}{\nu_2} Y\right)^{-(\nu_1+2)/2}, \quad Y > 0, \quad (2.2)$$

$$\varphi_{\nu_2}(\nu_1, \nu_2)(Y) = 0, \quad Y < 0, \quad (2.3)$$

де $\Gamma(x)$ — гамма-функція.

Числові характеристики цього розподілу мають вигляд

$$M\{Y\} = \frac{\nu_2}{\nu_2 - 2} \quad (\nu_2 > 2), \quad D\{Y\} = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}, \quad (\nu_2 > 4). \quad (2.4)$$

Розподіл (2.2) значно відрізняється від нормального розподілу. Більш близьким до нормального є z -розподіл:

$$z = \ln \nu = \frac{1}{2} \ln \nu^2 = \frac{1}{2} \ln \left(\frac{C_1(f)}{C_2(f)} \right). \quad (2.5)$$

При $\nu_1 \rightarrow \infty$ та $\nu_2 \rightarrow \infty$ випадкова величина z розподілена асимптотично нормально з математичним сподіванням $\frac{\nu_1 - \nu_2}{2\nu_1\nu_2}$ та дисперсією $\frac{\nu_1 + \nu_2}{2\nu_1 - \nu_2}$. Ця апроксимація має місце для великих виборок при $\nu_1 > 30$, $\nu_2 > 30$.

Отже, розподіл відношення спектрів близький до нормального, тому логічно перейти до дослідження статистик і критерію перевірки еквівалентності спектрів, що враховують цей розподіл.

Важливим моментом у виборі критерію є простота обчислень. Широко застосовуються статистики у вигляді лінійних функцій від елементарних статистик

$$y = \sum_{i=1}^{N_E} S_{E_i}, \quad (2.6)$$

де S_{E_i} — елементарні статистики; N_E — кількість елементарних статистик.

Статистика S_{E_i} , як правило, є функцією від невеликої кількості вихідних даних, стосовно яких перевіряється деяка гіпотеза, при цьому вважається відомим розподіл S_{E_i} або його перші два моменти. Значно спрощується аналіз і розрахунок критерію, якщо елементарні статистики можна вважати незалежними, оскільки легко оцінити перші два моменти розподілу сумарної статистики. За першими двома моментами з урахуванням розподілу елементарних статистик можна апроксимувати розподіл статистики y і на основі цього розподілу проводити аналіз і розрахунок критерію. Зазвичай стараються, щоб елементарні статистики задовольняли умовам центральної граничної теореми для того, щоб мати можливість апроксимувати розподіл статистики (2.6) нормальним розподілом.

Якщо елементарна статистика розподілена на півпрямій, то широко використовується апроксимація за допомогою χ^2 -розподілу, яка необхідна

для визначення допустимого інтервалу $\{y_{p_1} \leq y \leq y_{p_2}\}$ статистики y , для якого

$$P\{y_{p_1} \leq y \leq y_{p_2}\} = p_2 - p_1 = 1 - \alpha, \quad (2.7)$$

де y_{p_1} і y_{p_2} — квантили вибіркового розподілу; α — рівень значущості, що задається.

Оцінка квантилів для (2.7) за апроксимацією нормальним законом розподілу має вигляд

$$y_p = N_E M_E + u_p \sqrt{D_E N_E}, \quad (2.8)$$

де u_p — квантиль стандартного нормального розподілу; M_E і D_E — відповідно математичне сподівання і дисперсія статистики S_E за умови справедливості гіпотези, що перевіряється.

За апроксимації χ^2 -розподілом можна скористатися наближеною формулою для випадку, коли ймовірнісні квантили близькі до 0 або 1, а саме

$$\chi_p^2(m) \cong m \left(1 - \frac{2}{9m} + u_p \sqrt{\frac{2}{9m}} \right)^3, \quad (2.9)$$

де m — число ступенів вільності.

Використовуючи (2.9), неважко отримати для квантилів такий вираз:

$$y_p = N_E M_E \left(1 - \frac{D_E}{9M_E^2 N_E} + u_p \sqrt{\frac{D_E}{9M_E^2 N_E}} \right)^3. \quad (2.10)$$

Як бачимо з (2.8) і (2.10), визначення допустимого інтервалу для статистики (2.6) не пов'язане з великими труднощами, особливо за апроксимацією нормальним законом розподілу.

Обмежимося лінійними статистиками вигляду (2.6) з таких міркувань. Розглянемо статистику

$$d^2 = \left(\frac{2}{n_1} + \frac{2}{n_2} \right)^{-1} \sum_{i=1}^{N_f} \left[\ln \frac{C_1(f_i)}{C_2(f_i)} \right]^2, \quad (2.11)$$

яка має χ^2 -розподіл з N_f ступенями вільності; n_1 і n_2 — числа ступенів вільності відповідних розподілів оцінок спектрів $C_1(f_i)$ і $C_2(f_i)$. У (2.11) елементарною статистикою є

$$S_E = \left[\ln \frac{C_1(f_i)}{C_2(f_i)} \right]^2. \quad (2.12)$$

Функція логарифма в (2.12) асимптотично є статистикою квадрата відношень нормально розподілених випадкових величин. Звідси впливає аси-

мнотична оптимальність властивості цієї статистики. Однак розподіл (2.5) близький до нормального уже за числа ступенів вільності розподілів оцінок спектрів більшого 30.

2.2. Вибір елементарних статистик

Регулярних методів, що дозволяють однозначно синтезувати елементарну статистику, оптимальну в деякому сенсі з точки зору простоти обчислень потужності отриманого критерію, наразі не існує. У зв'язку з цим розглянемо одну із методик вибору елементарних статистик, а саме:

1. На основі якісного аналізу задачі інтуїтивно вибирається деякий набір елементарних статистик.

2. Проводиться оцінка потужності вибраних елементарних статистик.

3. На основі оцінки потужності критеріїв з урахуванням простоти обчислень проводиться вибір елементарних статистик.

Для оцінки потужності критеріїв необхідно визначити альтернативні гіпотези щодо розподілів спектральних густин $\Gamma_1(f)$ і $\Gamma_2(f)$. Це можна здійснити за допомогою введення функції відношення спектрів $\theta(f)$:

$$\theta(f) = \frac{\Gamma_2(f)}{\Gamma_1(f)}. \quad (2.13)$$

Ставиться задача перевірки нульової гіпотези H_0 щодо функції $\theta(f)$:

$$\theta(f) \Big|_{H_0} \equiv \text{const} = 1. \quad (2.14)$$

Відповідно альтернативні гіпотези H_i , $i \neq 0$, можна визначити за допомогою функцій $\theta_i(f)$, відмінних від тотожної одиниці, тобто

$$\theta_i(f) = \theta(f) \Big|_{H_i} \neq 1. \quad (2.15)$$

Використовуючи визначення функції $\theta(f)$, умова незміщення для порогового критерію зі статистикою (2.5) означає, що ймовірність виходу статистики за довірчий інтервал має не зменшуватись за будь-якої альтернативної гіпотези, тобто за будь-якої функції $\theta(f)$, відмінної від тотожної одиниці.

Відтак виникає питання про вибір вигляду довірчого інтервала, тобто, чи буде використовуватися двосторонній або односторонній критерій. Двосторонній критерій доцільно використовувати тільки тоді, коли статистика критерію за альтернативних гіпотез має тенденцію як до зменшення, так і до збільшення. Враховуючи лінійність статистики (2.5), ця умова буде виконуватися лише тоді, коли елементарні статистики матимуть тенденцію як до збільшення, так і до зменшення за альтернативних гіпотез.

Для статистики $\theta(f)$ це означає, наприклад, що при збільшенні $\theta(f)$ і елементарна статистика має тенденцію до збільшення, тобто її розподіл зміщується в напрямку більших значень, а при зменшенні $\theta(f)$ — до зменшення, тобто її розподіл зміщується в напрямку менших значень. Однак при цьому буде спостерігатися компенсація відхилень елементарних ста-

тистик за альтернативних гіпотез, що веде до зменшення потужності критерію і робить його зміщеним. Цей факт особливо суттєвий при перевірці еквівалентності нормованих спектрів, оскільки мають місце умови нормування:

$$\sum_f C_1(f) = 1, \quad \sum_f C_2(f) = 1. \quad (2.16)$$

Ефект компенсації пропаде, якщо використати елементарні статистики з односторонньою тенденцією зміни за відхилень $\theta(f)$ від одиниці. На основі цих статистик будуються односторонні критерії, при цьому ці статистики називаються також односторонніми.

Для них має виконуватися умова

$$M\{S_E | \theta(f)=1\} = \min_{(\max)} [M\{S_E\}]. \quad (2.17)$$

З метою задоволення умов центральної граничної теореми умову (2.17) слід доповнити умовою

$$D\{S_E\} < \infty. \quad (2.18)$$

Перейдемо до вибору елементарних статистик.

Якщо відома альтернативна гіпотеза у вигляді функції $\theta(f)$, то можна переконатися, що статистика

$$v^2 = \theta(f) \frac{C_1(f)}{C_2(f)} \quad (2.19)$$

наближено має v^2 -розподіл. Цим же розподілом описується і статистика

$$\xi(f) = \frac{C_1(f)}{C_2(f)} = \frac{v^2}{\theta(f)}. \quad (2.20)$$

Із формули (2.20) випливає, що безпосередньо $\xi(f)$ не є односторонньою статистикою для відхилення $\theta(f)$ від одиниці. Оскільки v^2 -розподіл зосереджений близько одиниці, то слід очікувати, що односторонніми будуть такі елементарні статистики:

$$\eta_M = |\xi - 1|, \quad 0 \leq \eta_M \leq \infty, \quad (2.21)$$

$$\eta_X = \max \left\{ \xi, \frac{1}{\xi} \right\}, \quad 0 \leq \eta_X \leq \infty, \quad (2.22)$$

$$\eta_Y = \min \left\{ \xi, \frac{1}{\xi} \right\}, \quad 0 \leq \eta_Y \leq 1. \quad (2.23)$$

Всі ці оцінки зміщені, тобто для них не виконується умова (2.17).

Враховуючи, що число ступенів вільності незгладжених спектральних оцінок дорівнює 2, а дисперсія ν^2 -розподілу скінченна, якщо $\nu_2 > 4$, то можна показати, що статистика ξ в (2.21) і (2.22) має будуватися на основі згладжених спектральних оцінок. Для (2.23) можна використати і незгладжені оцінки.

Щодо потужності критеріїв, то найменшою потужністю характеризується статистика (2.23), а найбільшою — (2.22), оскільки для останньої спостерігається найбільша відносна чутливість за зміни $\theta(f)$.

Наразі виникає задача оцінки параметрів цих елементарних статистик, а саме їх математичного сподівання та дисперсії в залежності від параметра θ .

2.3. Результати обчислень числових характеристик розподілів елементарних статистик

Для статистик (2.21) та (2.22) при $\nu_1 = 6$ нескладно отримати точні вирази для математичного сподівання та дисперсії. Наведемо ці вирази, допускаючи, що

$$\nu_2 = 6\bar{M}, \quad \bar{M} = 1, 2, \dots, \quad (2.24)$$

тобто $C_1(f)$ у (2.19) відповідає згладженій оцінці з 6 ступенями вільності, отриманій на одному допустимому інтервалі, а $C_2(f)$ відповідає оцінці спектральної густини, отриманій внаслідок адаптивного згладжування за попередніми припустимими інтервалами. З урахуванням заданих ν_1 і ν_2 розподіл (2.2) буде мати вигляд

$$\varphi_y(Y) = \begin{cases} 0, & Y < 0, \\ \frac{3(3\bar{M}+2)(3\bar{M}+1)}{2\bar{M}^2} Y^2 \left(1 + \frac{Y}{\bar{M}}\right)^{-(3\bar{M}+3)}, & Y > 0. \end{cases} \quad (2.25)$$

Початковий момент порядку r має вигляд

$$MY^r = \frac{r(3+r)r(3\bar{M}-r)}{2(3\bar{M}-r)!} \bar{M}^r. \quad (2.26)$$

При проведенні обчислень використовується значення інтеграла

$$\int \frac{y^m dy}{\left(1 + \frac{y}{\bar{M}}\right)^{3\bar{M}+3}} = \bar{M}^{m+1} \sum_{i=0}^m \frac{(-1)^i m! (1 + Y/\bar{M})^{(-3\bar{M}+2+m-i)}}{(m-i)! (-3\bar{M}-2+m-i)} \quad (2.27)$$

для m — цілого, $0 \leq m \leq 3\bar{M} + 1$.

Для зручності вводять змінну

$$A = 1/\theta. \quad (2.28)$$

Наведемо результати обчислень числових характеристик розподілу статистики $\eta_M = |AY - 1|$. Математичне сподівання $M\{\eta_M\}$ має вигляд

$$M\{\eta_M\} = \frac{(6A^2 + 12A + 9)M^2 + (8A + 9)M + 2}{(3M - 1)(AM + 1)} \left(1 + \frac{1}{AM}\right)^{-3M} - \frac{1 - 3M(1 - A)}{3M - 1}. \quad (2.29)$$

В асимптотичному випадку при $M \rightarrow \infty$ маємо

$$M\{\eta_M\} = \frac{2A + 4 + 3/A}{e^{3/A}} + 1 - A. \quad (2.30)$$

Початковий момент 2-го порядку для η_M має вигляд

$$M\{\eta_M^2\} = \frac{M_2(12A_2 - 18A + 9) + M(12A - 9) + 2}{(3M - 1)(3M - 2)}. \quad (2.31)$$

В асимптотичному випадку при $M \rightarrow \infty$ маємо

$$M\{\eta_M^2\} = \frac{4A^2 - 6A + 3}{3}. \quad (2.32)$$

Наведемо результати обчислень параметрів розподілу статистики $\eta_X = \max\{\xi, 1/\xi\}$. Для зручності її апроксимації χ^2 -розподілом розглянемо статистику

$$\eta = \eta_X - 1. \quad (2.33)$$

Математичні сподівання статистик η та η_X відрізняються на одиницю, а дисперсії — рівні.

Математичне сподівання $M\{\eta\}$ має вигляд

$$M\{\eta\} = \frac{3M \left[\frac{M(A^2 + 3A + 3) + A + 2}{(3M - 1)(AM + 1)} \right] \left(1 + \frac{1}{AM}\right)^{-3M} + \frac{3 - 2A}{2A}}{1}. \quad (2.34)$$

В асимптотичному випадку при $M \rightarrow \infty$ маємо

$$M\{\eta\} = \frac{A + 3 + 3/A}{e^{3/A}} + \frac{3 - 2A}{2A}. \quad (2.35)$$

Початковий момент 2-го порядку для η має вигляд

$$M\{\eta^2\} = \frac{6(3 + 2A)(AM + 1)M}{(3M - 2)(3M - 1)} \left(1 + \frac{1}{AM}\right)^{-3M} + \frac{M(2A^2 - 6A + 9) + 3}{2A^2M}. \quad (2.36)$$

В асимптотичному випадку при $M \rightarrow \infty$ маємо

$$M\{\eta^2\} = \frac{2A(A+3)}{3e^{3/A}} + \frac{2A^2 - 6A + 9}{2}. \quad (2.37)$$

Проаналізуємо отримані асимптотичні залежності числових характеристик розподілу статистики η_M , показаних на рис. 1.

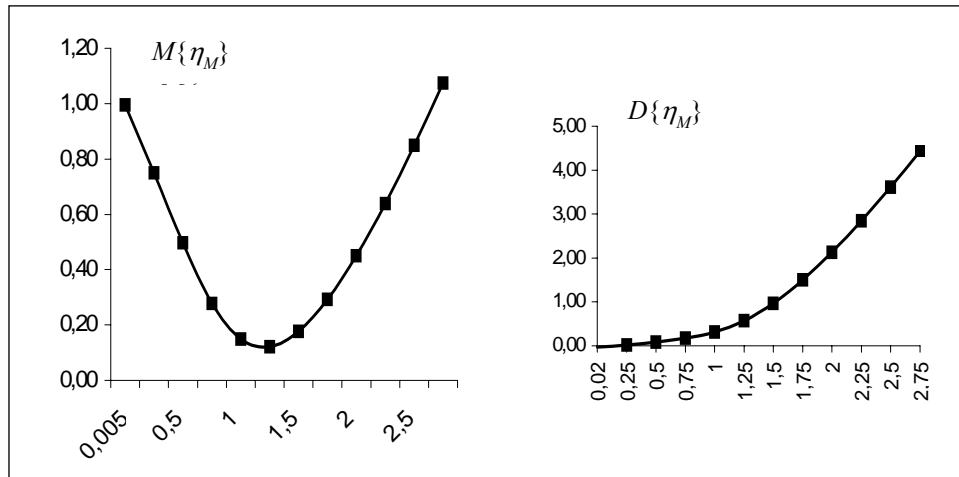


Рис. 1

Із рис.1 видно, що математичне сподівання та дисперсія не досягають мінімальних значень при $A=1$, тобто критерій, що використовує як елементарну статистику η_M , буде зміщеним. Скоригуємо статистику:

$$\eta_{MC} = |A \cdot A_{MC} \cdot Y - 1|, \quad (2.38)$$

причому $A_{MC} = A'$, де A' — значення A , за якого досягається мінімальне значення математичного сподівання статистики η_M . Однак цього недостатньо, щоб критерій був незміщеним, оскільки із збільшенням математичного сподівання при відхиленні A від одиниці дисперсія статистики може зменшитись настільки, що ймовірність виходу за довірчий інтервал буде меншою рівня значущості, що характеризує зміщення критерію. Аналіз з урахуванням математичного сподівання та дисперсії є раціональним, оскільки число елементарних статистик γ_E в статистиці (2.6) наближено описується нормальним законом розподілу.

Крім того, неважко показати, що коли для нормальної випадкової величини встановлено деяку границю, ймовірність перевищення якої дорівнює $\alpha < 0,5$ (відповідає рівню значущості критерію), то за умови, що математичне сподівання та дисперсія не зменшуватимуться, ймовірність $(1-\beta)$ перевищення нормальною величиною заданої границі (відповідає потужності критерію), не буде меншою α (відповідає незміщенню критерію).

Формула залежності потужності критерію від приростів математичного сподівання та дисперсії статистики (2.6) має вигляд

$$1 - \beta = \Phi \left(\frac{\Delta M \cdot N_E - u_{1-\alpha} \sqrt{N_E \cdot D_0}}{\sqrt{N_E (D_0 + \Delta D)}} \right), \quad (2.39)$$

де $\Phi(x)$ — функція нормального розподілу Гаусса; N_E — кількість елементарних статистик в (2.6); ΔM , ΔD — відповідний приріст математичного сподівання та дисперсії елементарної статистики за деякої альтернативної гіпотези; D_0 — дисперсія елементарної статистики за нульової гіпотези (коли спектри еквівалентні); u_p — квантиль нормального розподілу.

У формулі (2.39) необхідно пам'ятати, що ΔM і ΔD для кожної елементарної статистики функціонально пов'язані змінною A . Однак із (2.39) випливає, якщо дисперсія елементарної статистики не змінюється, то збільшення ΔM буде завжди збільшувати потужність критерію, якщо ΔM не змінюється, то збільшення ΔD буде збільшувати потужність, якщо $(1 - \beta) < 0,5$, і зменшувати — якщо $(1 - \beta) > 0,5$.

Формулу (2.39) можна перетворити до вигляду

$$1 - \beta = \Phi \left(\frac{\Delta M \cdot \sqrt{N_E}}{\sqrt{D_0 + \Delta D}} - u_{1-\alpha} \sqrt{\frac{D_0}{D_0 + \Delta D}} \right). \quad (2.40)$$

На основі (2.40) зручно порівнювати критерії за альтернативних гіпотез, заданих у вигляді змін ΔM і ΔD .

Для того, щоб критерій був незміщеним, необхідно, щоб вираз (2.40) досягав мінімального значення за $\Delta M = 0$ та $\Delta D = 0$.

Звідси видно, як проводити коригування або адаптацію статистики. Якщо мінімальні значення математичного сподівання та дисперсії статистики близькі, то доцільно коригувати статистику за математичним сподіванням, оскільки у (2.40) вплив відхилень дисперсій частково компенсується.

З рис. 1 випливає, що статистика η_M характеризується низкою небажаних властивостей.

1. Мінімальні значення $M\{\eta_M\}$ та $D\{\eta_M\}$ досягаються в істотно різних точках, приблизно при $A=1$ та $A=0$ відповідно, а тому важко запропонувати прості методи коригування елементарної статистики η_M .

2. Скінченне відхилення $M\{\eta_M\}$ за зміни A від 0 до 1 свідчить про низьку чутливість статистики η_M у цій області.

На підставі цих властивостей статистика η_M вилучається з подальшого розгляду, відтак розглядається статистика максимуму (2.22).

Для обчислення числових характеристик $M\eta$ та $D\eta$ розподілу η існує програмне забезпечення з комп'ютерною реалізацією і графічним зображенням, звідки видно, що точки A , які мінімізують $M\{\eta\}$ та $D\{\eta\}$, близькі одна до одної. Тому можна очікувати прийнятних результатів щодо незміщення критерію, скориставшись тільки центруванням математичного сподівання статистики η .

Скоригована статистика η матиме вигляд

$$\eta' = \max \{A \cdot A'_C \cdot Y, 1/A \cdot A'_C \cdot Y\} - 1, \quad (2.41)$$

однак статистику η можна скоригувати ще й таким чином:

$$\eta'' = \max \{AY - A''_C, 1/AY - 1/A''_C\}, \quad (2.42)$$

де A''_C знаходиться за допомогою підпрограми пошуку екстремуму $M\{\eta\}$, статистика AY формується за допомогою підпрограми моделювання нормально розподілених випадкових чисел на основі відомого співвідношення:

$$AY_{\nu_1 \nu_2} = \frac{A \nu_2 \chi_{\nu_1}^2}{\nu_1 \chi_{\nu_2}^2}, \quad (2.43)$$

де χ_{ν}^2 — «хі-квадрат» розподіл з ν ступенями вільності.

Для роботи підпрограми задаються такі параметри: число ступенів вільності ν_1 , кількість усереднень M ($\nu_2 = M\nu_1$), границі та крок зміни A і кількість статистичних випробувань N .

Можна переконатися, що для статистики мінімуму (2.23) критерій матиме малу потужність, хоча він характеризується скінченною дисперсією за будь-якого числа ступенів вільності, тобто може безпосередньо використовувати незгладжені оцінки спектрів з $\nu_1 = 2$. Застосування статистики мінімуму (2.23) є мало перспективним за малої потужності критерію, а статистика максимуму (2.22) може використовуватися з числом ступенів вільності $\nu_1 \geq 6$, тобто згладжування невелике, і можливе відхилення розподілу цієї статистики від теоретичного буде незначним.

2.4. Про дослідження потужності критеріїв

Одне з найважливіших питань, що виникає при розробці критеріїв, — вибір альтернативних гіпотез. Як уже зазначалося, для дослідження потужності критеріїв доцільно використати функцію $A(f) = \frac{1}{\theta(f)}$. Як $A(f)$ приймають функцію відношення двох спектрів резонансного характеру.

Поблизу резонансу спектр з точністю до коефіцієнта пропорційності подається у вигляді

$$\Gamma_p(f) = \frac{1}{\varepsilon^2 + d^2}, \quad (2.44)$$

де $\varepsilon = \frac{2\Delta f}{f_p} = \frac{2(f - f_p)}{f_p}$ — розладнання частоти щодо резонансу; f_p — резонансна частота; d — затухання. З урахуванням (2.44) $A(f)$ має вигляд

$$A(f) = \frac{1}{\theta(f)} = \frac{\varepsilon_1^2(f) + d_1^2}{\varepsilon_2^2(f) + d_2^2}, \quad \varepsilon_i = \frac{2(f - f_{p_i})}{f_{p_i}}, \quad i = 1, 2. \quad (2.45)$$

Функція (2.45) наближено визначає відношення двох нормованих спектрів із заданими резонансними частотами f_{p1} і f_{p2} та їх добротностями.

Із графіка видно, що область зміни $A(f)$ досить значна і для оцінки потужності критеріїв необхідно враховувати розподіл $A(f)$.

Це можна врахувати у формулі вигляду (2.40), а саме

$$1 - \beta = \Phi \left(\frac{\sum_{i=1}^{N_E} \Delta M_i(A(f)) - u_{1-\alpha} \sqrt{N_E \cdot D_0}}{\sum_{i=1}^{N_E} (D_0 + \Delta D_i) A(f)} \right). \quad (2.46)$$

У виразі (2.46) враховується конкретний розподіл змін математичних сподівань ΔM_i та дисперсій ΔD_i елементарних статистик в (2.6).

Отже, методика дослідження потужності критеріїв перевірки еквівалентності спектрів на основі статистики (2.6) виглядає таким чином:

1. Для досліджуваної статистики визначається залежність математичного сподівання та дисперсії від відношення спектрів.
2. Із фізичних міркувань вибирається конкретна залежність відношення спектрів від частоти, наприклад, у вигляді (2.45).
3. Оцінюється потужність критерію за формулою (2.46).

Важливим є також питання визначення границі для довірчої області критерію.

З метою дослідження точності визначення границі при апроксимації нормальним та «хі-квадрат»-законами, а також для статистичного дослідження потужності критерію розроблена комп'ютерна реалізація статистичного оцінювання ймовірності перевищення статистикою (2.6) границі Γ_N , визначеною за формулою (2.8), та границі Γ_{χ^2} , визначеною за формулою (2.10). Вихідними даними при цьому є математичне сподівання та дисперсія статистики за умови еквівалентності спектрів; коефіцієнт $A(f)$ є однаковий для всіх елементарних статистик; кількість елементарних статистик N_E ; квантиль $u_{1-\alpha}$ для заданого рівня значущості α критерія та кількість статистичних випробувань N . Точність задання границі визначається порівнянням оцінки потужності $1 - \beta$ при $A(f) = 1$ (тобто рівня значущості) із заданим значенням α .

На рис. 2 показані результати дослідження для статистики максимуму при $M = 1$, $N_E = 20$, $u_{1-\alpha} = 1,28$, тобто $\alpha = 0,1$.

Із рис. 2 видно, що границя Γ_{χ^2} при $A(f) = 1$ дуже близька до значення 0,1, в той час як границя Γ_N дає приблизно у 2 рази занижений рівень значущості. Критерій χ^2 з ймовірністю близькою до 0.5 виявлятиме зміну спектральної густини в середньому у 2 рази частіше за критерій Γ_N .

Підсумовуючи результати досліджень статистичного аналізу високочастотних випадкових процесів у дискретному часі, окреслимо важливий напрямок їх застосувань на прикладі трактування ризикового процесу ціно-

утворення цінних паперів, зокрема процесу котировки акцій на фондовому ринку як високочастотного в залежності від кон'юнктури ринку, стаціонарного або нестаціонарного випадкового процесу.

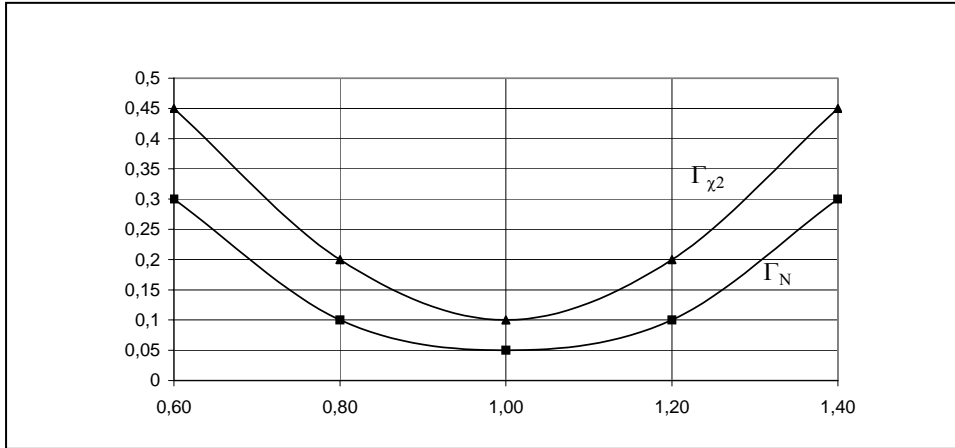


Рис. 2

На основі реальних спостережень конкретної реалізації котировки акцій на певному відрізку часу будується статистика або критерій, що з надійністю $(1-\alpha)$ (α — рівень значущості) попадає в область прийняття нульової гіпотези про рівність спектрів. Питання про розподіл статистик розв'язується з урахуванням результатів ергодичної теорії випадкових процесів, що призводять до нормального або «хі-квадрат» розподілів досліджуваних статистик.

ЛІТЕРАТУРА

1. *Случайные колебания* // Под ред. С. Кренделл. — М.: Мир, 1967. — 356 с.
2. *Andreev A., Andreev M.V. Optimal Investment Strategy on Financial Market Derivatives* // Third International School on Applied Statistics, Financial and Actuarial Mathematics, Sept. 4–13, 2000. — Feodosiya, Crimea, 2000. — P. 1–2.
3. *Бартлетт М.В. Введение в теорию случайных процессов.* — М.: Изд-во иностр. лит., 1958. — 384 с.

Надійшла 12.11.2002