Аналіз часових рядів
(Time Series Analysis)

Аналіз часових рядів – це один з розділів аналізу даних, який на основі впорядкованої у часі доступної послідовності спостережень над випадковим процесом, займається його дослідженням та побудовою математичної моделі для нього.

Як правило, отримані виміри не є незалежними і однаково розподіленими і тому потребують розробки відповідних методів дослідження. А отримана модель може бути використана у подальшому для розв'язання задач прогнозування.

Розглянемо деякий випадковий процесом з дискретним часом

$$\eta_t, t \in \mathbb{N}. \tag{1}$$

Нехай доступні спостереження над цим процесом у відповідні моменти часу

$$y_t, t \in \mathbb{N}. \tag{2}$$

Можливий розгляд також випадку, коли $t \in \mathbb{Z}$.

Зауваження. В більш загальному випадку може розглядатися випадковий процесом з неперервним часом

$$\eta_t, t \ge t_*.$$

Тоді спостереження над

$$\eta_{t_1}, \eta_{t_2}, \eta_{t_3}, \ \dots \qquad (t_* \le t_1 < t_2 < t_3 < \ \dots),$$

отримані у ці впорядковані у часі моменти, можна відповідним чином перенумерувати як

$$y_1, y_2, y_3, \ \dots \ .$$

Частіше всього такі виміри отримуються через однакові інтервали часу, тобто з деяким постійним кроком Δ у часі $(\Delta > 0)$, а саме:

$$t_i = t_* + (i-1)\Delta, \ i \in \mathbb{N}.$$

Означення. *Часовим рядом* називається впорядкована у часі послідовність спостережень над випадковим процесом.

Задачі аналізу часових рядів виникають в різних галузях, а саме: у фінансах, економіці, бізнесі, медицині, біології, промисловості тощо.

Наведемо деякі приклади часових рядів.

Приклад 1. Щогодинні спостереження на валютній біржі над значеннями обмінного курсу для деякої валютної пари. Наприклад: USD/UAH, EUR/UAH, EUR/USD, CNY/USD, JPY/USD, GBP/USD.

Приклад 2. Спостереження на товарній біржі над значеннями щоденних цін закриття на золото, срібло, платина, паладій, природній газ, нафта Brent, нафта WTI (Western Texas Intermidiate).

Приклад 3. На фондовій біржі щоденні котирування закриття акцій компаній, наприклад: Alphabet Inc. (яка володіє Google), Microsoft, IBM, NVidia, Apple, Tesla, Ford.

Приклад 4. Щогодинні значення температури повітря, тиску, вологості, швидкості вітру на якійсь метеостанції у м. Києві.

Приклад 5. Щоденні значення температури тіла, кров'яного тиску у пацієнта лікарні зранку на 8:00 та ввечері на 20:00.

Приклад 6. Кількість населення у певній країні на 1 січня кожного року.

В якості математичної моделі часового ряду розглянемо:

$$\eta_t = f(t) + \xi(t), t \in \mathbb{N}, \tag{3}$$

де $f(t)$ - детермінована (систематична) складова часового ряду, яку ще називають *трендом*,

$\xi(t)$ - стохастична складова часового ряду.

До одних з основних задач аналізу часових рядів можна віднести:

- статистичне оцінювання функції тренду,
- аналіз стохастичної складової часового ряду.

Перейдемо до розгляду першого класу задач.

Часові ряди з поліноміальними трендами

Розглянемо в якості математичної моделі для часового ряду

$$\eta_t = f(t) + \xi(t), t = \overline{1,T}, \tag{4}$$

причому вважаємо, що справедливі такі припущення

1. $M\xi_t = 0, D\xi_t = \sigma^2, \sigma^2 > 0, t = \overline{1,T},$ (5)
2. $M\xi_t\xi_s = 0,\ \forall t \neq s,$ (6)
3. функцію тренду можна апроксимувати поліномом степені p, тобто $f(t) = \sum_{i=0}^{p} \alpha_i t^i,\ p \ll T.$ (7)

Потрібно по доступним спостереженням y_t над цим процесом η_t знайти оцінку функції тренду $f(t),\ t = \overline{1,T}$.

Фактично задача знаходження оцінки функції тренду звелася до знаходження оцінок параметрів $\{\alpha_i\}_{i=0}^{p}$ апроксимуючого полінома, див. (7). Скористаємося результатами регресійного аналізу. У цій ситуації буде корисним перейти від регресорів $\{t^i\}_{i=0}^{p}$ до регресорів у вигляді набору ортогональних поліномів $\{P_i(t)\}_{i=0}^{p}$, де

$$\begin{cases} P_0(t) = 1, \\ P_k(t) = t^k + C_k(k-1)t^{k-1} + C_k(k-2)t^{k-2} + \ldots \\ \qquad \ldots + C_k(1)t + C_k(0), k = \overline{1,T-1}, \end{cases}$$

для яких справедлива умова ортогональності

$$\sum_{t=1}^{T} P_i(t) P_j(t) = 0,\ \forall i \neq j,\ i,j = \overline{0,T-1}. \tag{8}$$

Причому усі набори коефіцієнтів $\{C_k(j)\}_{j=0}^{k-1},\quad k = \overline{0,T-1}$ однозначно визначаються з умови ортогональності (8).

Представимо тепер функцію тренду через ортогональні поліноми

$$f(t) = \sum_{i=0}^{p} \beta_i P_i(t),\ p \ll T. \tag{9}$$

Знайдемо спочатку оцінку МНК $\hat{\beta} = \left(\hat{\beta}_0, \quad \hat{\beta}_1, \quad \ldots, \quad \hat{\beta}_p\right)^T$ вектора параметрів $\beta = \left(\beta_0, \quad \beta_1, \quad \ldots, \quad \beta_p\right)^T$, як розв'язок такої задачі

$$\hat{\beta} = \arg\min_{\beta} \sum_{t=1}^{T} \left(y_t - \sum_{i=0}^{p} \beta_i P_i(t) \right)^2 .$$

В результаті отримаємо

$$\hat{\beta}_j = \frac{\sum_{t=1}^{T} y_t P_j(t)}{\sum_{t=1}^{T} P_j^2(t)}, \;\; j = \overline{0, p}.$$

Між іншим, так як $P_0(t) = 1$, то з останнього випливає, що

$$\hat{\beta}_0 = \frac{\sum_{t=1}^{T} y_t}{T} = \overline{y}.$$

А оцінки МНК $\{\hat{\alpha}_i\}_{i=0}^{p}$ для параметрів $\{\alpha_i\}_{i=0}^{p}$ можна визначити через знайдені оцінки МНК $\left\{\hat{\beta}_i\right\}_{i=0}^{p}$ для параметрів $\{\beta_i\}_{i=0}^{p}$. Для цього достатньо прирівняти записи тренду у представленнях (7) і (9) та порівняти коефіцієнти при однакових степенях $t^k, k = \overline{0, p}$. В результаті отримуємо:

$$\hat{\alpha}_k = \hat{\beta}_k + \sum_{i=k+1}^{p} \hat{\beta}_i C_i(k), \;\; k = \overline{0, p}.$$

У підсумку оцінку $\hat{f}(t)$ для функції тренду $f(t)$ можна підрахувати згідно:

$$\hat{f}(t) = \sum_{i=0}^{p} \hat{\alpha}_i t^i = \sum_{i=0}^{p} \hat{\beta}_i P_i(t). \qquad (10)$$

Додатково припустимо, що $\xi(t)$ нормально розподілена, $t = \overline{1, T}$. Тоді незміщена оцінка $\hat{\sigma}^2$ методу максимальної правдоподібності для σ^2 буде мати вигляд

$$\hat{\sigma}^2 = \frac{1}{T-p-1}\sum_{t=1}^{T}\left(y_t - \hat{f}(t)\right)^2.$$

З'ясуємо, чи не можна понизити степінь апроксимуючого полінома, тобто уточнимо його степінь. Нехай апроксимуючий поліном має степінь p, тоді перевіримо гіпотезу про те, чи не можна в якості апроксимуючого полінома, поліном меншої степені.

Для цього перевіримо гіпотезу

$$H_0 : \beta_p = 0,$$

з деяким малим рівнем значущості $\gamma > 0$.

Алгоритм покроково:

1. підрахуємо p-ту частинну F-статистику

$$F_p = \frac{\hat{\beta}_p^2}{\hat{\sigma}^2 d_p}, \text{ де } d_p = \left\{\left(X^T X\right)^{-1}\right\}_{pp} = \left(\sum_{t=1}^{T} P_p^2(t)\right)^{-1};$$

2. перевіряємо умову

$$F_p < F_\gamma\left(1, T-p-1\right)$$

- : Stop;

+ : p:=p-1;
if $p \geq 1$ then goto п. 1 else Stop;

Нехай після уточнення, степінь апроксимуючого полінома дорівнює p_*. Тепер перевіримо на значимість параметри β_i при усіх більш нижчих степенях, а саме будемо перевіряти гіпотези виду

$$H_0^{(i)} : \beta_i = 0$$

з деяким малим рівнем значущості $\gamma > 0$, $i = \overline{0, p_* - 1}$.

Для перевірки такої гіпотези можна скористатися або F_i, i-тою частинною F-статистикою, або статистикою t_i. Причому відомо, що $F_i = t_i^2$.

Алгоритм покроково:

1. $i := p_* - 1;$
2. підрахуємо статистику t_i, де

$$t_i = \frac{\hat{\beta}_i}{\hat{\sigma}\sqrt{d_i}}, \text{ де } d_i = \left\{ \left(X^T X \right)^{-1} \right\}_{ii} = \left(\sum_{t=1}^{T} P_i^2(t) \right)^{-1};$$

3. if $|t_i| < t_{\frac{\gamma}{2}}(T - p_* - 1)$ then (i-тий член можна del);
4. $i := i - 1;$
5. if $i \geq 0$ then goto п. 2 else Stop;

Часові ряди з періодичними трендами

Розглянемо в якості математичної моделі для часового ряду

$$\eta_t = f(t) + \xi(t), t = \overline{1,T}, \tag{11}$$

причому вважаємо, що справедливі такі припущення

1. $M\xi_t = 0, D\xi_t = \sigma^2, \sigma^2 > 0, t = \overline{1,T},$ (12)
2. $M\xi_t\xi_s = 0, \ \forall t \neq s,$ (13)
3. тренд $f(t)$ є періодичною функцією з періодами $\frac{T}{k_j}, j = \overline{1,q}$, причому k_j націло ділять T. (14)

Потрібно по доступним спостереженням y_t над цим процесом η_t знайти оцінку функції тренду $f(t)$, $t = \overline{1,T}$.

Згадаємо. У функції $Cos(\lambda t-\varphi)$ період дорівнює $\frac{2\pi}{|\lambda|}$, а частота $\frac{|\lambda|}{2\pi}$.

Тоді логічно функцію тренд $f(t)$ наближати лінійною комбінацію періодичних ортогональних функцій $Cos(\bullet)$ та $Sin(\bullet)$ з періодами виду (14), а саме:

$$f(t)=\alpha(0)+\sum_{j=1}^{q}\left[\alpha(k_j)Cos\left(\frac{2\pi k_j}{T}t\right)+\beta(k_j)Sin\left(\frac{2\pi k_j}{T}t\right)\right]+\alpha\left(\frac{T}{2}\right)(-1)^t\delta,$$

де

$$\delta=\begin{cases}0, \textit{якщо } T-\textit{непарне},\\ 1, \textit{якщо } T-\textit{парне}.\end{cases}$$

Зауваження. Коли T є парним числом, допустимо для $k_j=\frac{T}{2}$. В результаті з'являється доданок з $Cos\left(\frac{2\pi\frac{T}{2}}{T}t\right)=Cos(\pi t)=(-1)^t$.

А $Sin(\pi t)=0$.

Для знаходження оцінки функції тренду $f(t)$ знайдемо оцінку МНК $\hat{\alpha}$ такого вектора невідомих параметрів

$$\alpha=\begin{cases}\left(\alpha(0),\alpha(k_1),\beta(k_1),\dots,\alpha(k_q),\beta(k_q)\right)^T, \textit{якщо } T-\textit{непарне},\\ \left(\alpha(0),\alpha(k_1),\beta(k_1),\dots,\alpha(k_q),\beta(k_q),\alpha\left(\frac{T}{2}\right)\right)^T, \textit{якщо } T-\textit{парне}.\end{cases}$$

Потрібна оцінка МНК $\hat{\alpha}$ є розв'язком такої задачі

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{t=1}^{T} \{ y_t -$$

$$- \alpha(0) - \sum_{j=1}^{q} \left[\alpha(k_j) Cos\left(\frac{2\pi k_j}{T} t\right) + \beta(k_j) Sin\left(\frac{2\pi k_j}{T} t\right) \right] - \alpha\left(\frac{T}{2}\right)(-1)^t \delta \right\}^2 .$$

В результаті отримуємо такі оцінки

$$\begin{cases} \hat{\alpha}(0) = \dfrac{1}{T} \sum\limits_{t=1}^{T} y_t = \overline{y}, \\ \hat{\alpha}(k_j) = \dfrac{2}{T} \sum\limits_{t=1}^{T} y_t Cos\left(\dfrac{2\pi k_j}{T} t\right), \; j = \overline{1,q}, \\ \hat{\beta}(k_j) = \dfrac{2}{T} \sum\limits_{t=1}^{T} y_t Sin\left(\dfrac{2\pi k_j}{T} t\right), \; j = \overline{1,q}, \\ \hat{\alpha}\left(\dfrac{T}{2}\right) = \dfrac{1}{T} \sum\limits_{t=1}^{T} y_t (-1)^t, \; якщо \; T - парне. \end{cases} \quad (15)$$

Це дозволяє виписати остаточну оцінку для функції тренду

$$\hat{f}(t) = \hat{\alpha}(0) +$$

$$+ \sum_{j=1}^{q} \left[\hat{\alpha}(k_j) Cos\left(\frac{2\pi k_j}{T} t\right) + \hat{\beta}(k_j) Sin\left(\frac{2\pi k_j}{T} t\right) \right] + \hat{\alpha}\left(\frac{T}{2}\right)(-1)^t \delta. \quad (16)$$

А оцінки амплітуд та фаз відповідних складових будуть мати такий вигляд:

$$\begin{cases} \hat{\rho}(k_j) = R(k_j) = \sqrt{\hat{\alpha}^2(k_j) + \hat{\beta}^2(k_j)}, \\ \hat{\varphi}(k_j) = arctg\left(\dfrac{\hat{\beta}(k_j)}{\hat{\alpha}(k_j)}\right), \quad j = \overline{1,q}. \end{cases} \quad (17)$$

Означення. Графік залежності $R^2(k)$ від періоду $\frac{T}{k}$ називається *періодограмою*, а графік залежності $R^2(k)$ від частоти $\frac{k}{T}$ називається *спектрограмою.*

Припустимо додатково, що $\xi(t)$ нормально розподілена, $t=\overline{1,T}$. Тоді незміщена оцінка $\hat{\sigma}^2$ методу максимальної правдоподібності для σ^2 буде мати вигляд

$$\hat{\sigma}^2 = \frac{1}{T-2q-1-\delta}\sum_{t=1}^{T}\left(y_t-\hat{f}(t)\right)^2. \qquad (18)$$

Метод рухомого середнього

До цього …

Нехай математична модель часового ряду має вигляд

$$\eta_t = f(t)+\xi(t), t=\overline{1,T}, \qquad (19)$$

причому вважаємо, що справедливі такі припущення

1. $M\xi_t = 0, D\xi_t = \sigma^2, \sigma^2 > 0, t=\overline{1,T},$ (20)
2. $M\xi_t\xi_s = 0,\ \forall t \neq s,$ (21)
3. функцію тренду в околі довільного t на відрізку $[t-m, t+m]$ можна апроксимувати деяким поліномом степені p, а саме

$$f(t+s)=\sum_{i=0}^{p}a_i(t)s^i,\ s=\overline{-m,m}. \qquad (22)$$

Потрібно по доступним спостереженням y_t над цим процесом η_t знайти оцінку функції тренду $f(t)$, $t=\overline{1,T}$.

Фактично задача знаходження оцінки $\hat{f}(t)$ для функції тренду $f(t)$ звелася до знаходження для довільного t набору оцінок $\left\{\hat{a}_i(t)\right\}_{i=0}^{p}$ для параметрів $\left\{a_i(t)\right\}_{i=0}^{p}$ апроксимуючого полінома (22). Скористаємося результатами регресійного аналізу.

Для кожного моменту t знайдемо потрібні оцінки $\left\{\hat{a}_i(t)\right\}_{i=0}^{p}$, $t=\overline{m+1,T-m}$. Шукати будемо за допомогою МНК.

Введемо позначення

$$a(t)=\left(a_0(t),a_1(t),\ldots,a_p(t)\right)^T,$$

$$\hat{a}(t)=\left(\hat{a}_0(t),\hat{a}_1(t),\ldots,\hat{a}_p(t)\right)^T.$$

Тоді

$$\hat{a}(t)=\arg\min_{a(t)}\sum_{s=-m}^{m}\left[y_{t+s}-\sum_{i=0}^{p}a_i(t)s^i\right]^2,\quad t=\overline{m+1,T-m}.$$

Отримаємо деякі набори оцінок $\left\{\hat{a}_i(t)\right\}_{i=0}^{p}$, $t=\overline{m+1,T-m}$.

Тоді з співвідношення (22) випливає, що оцінка функції тренду у момент t може бути підрахована згідно

$$\hat{f}(t)=\hat{a}_o(t),\quad t=\overline{m+1,T-m}.$$

Але це ще не все, виявляється, що остання оцінка має спеціальну структуру

$$\hat{a}_o(t)=\sum_{s=-m}^{m}c_s\cancel{(t)}y_{t+s},\quad t=\overline{m+1,T-m},$$

причому справедливо

$$\begin{cases}\sum\limits_{s=-m}^{m}c_s=1,\\ c_s=c_{-s},\ s=\overline{1,m}.\end{cases}$$

У підсумку

$$\hat{f}(t)=\sum_{s=-m}^{m}c_s y_{t+s},\quad t=\overline{m+1,T-m}. \tag{23}$$

Зауважимо, що коефіцієнти $\{c_s\}_{s=-m}^{m}$ залежать тільки від m та p, але не залежать від t та $\{y_t\}_{t=1}^{T}$.

Для найбільш вживаних значень m та p коефіцієнти $\{c_s\}_{s=-m}^{m}$ або складені відповідні таблиці, або для них отримані аналітичні формули.

Аналіз стохастичної складової часового ряду

Припустимо, що по доступним спостереженням y_t над процесом η_t, була знайдена оцінка $\hat{f}(t)$ для функції тренду $f(t)$, $t=\overline{1,T}$.

Тоді для аналізу стохастичної складової часового ряду буде корисний такий часовий ряд

$$e_t = y_t - \hat{f}(t), t=\overline{1,T}.$$

Зрозуміло, що для цього часового ряду e_t умова некорельованості у загальному випадку буде порушена.

Тому в якості моделі для e_t візьмемо модель стаціонарної у широкому розумінні випадкової послідовності з середнім

$$m = Me_t, t=\overline{1,T},$$

автоковаріаційною функцією

$$\sigma(h) = M(e_t - m)(e_{t+h} - m), h \in \mathbb{Z}, t = \overline{1,T-h},$$

та спектральною щільністю

$$g(\lambda)=\frac{1}{\pi}\sum_{h=-\infty}^{\infty}\sigma(h)Cos(\lambda h),\lambda\in[-\pi,\pi].$$

Маючи у розпорядженні значення часового ряду e_t, вибіркові оцінки його характеристик можуть бути підраховані таким чином

$$\hat{m}=\frac{1}{T}\sum_{t=1}^{T}e_t,t=\overline{1,T},\tag{24}$$

$$\hat{\sigma}(h)=\frac{1}{T-h}\sum_{t=1}^{T-h}(e_t-e_0)(e_{t+h}-e_0),h\in\mathbb{Z},t=\overline{1,T-h},\tag{25}$$

де

$$e_0=\begin{cases}m, \text{\textit{якщо} } m-\text{\textit{відоме}},\\ \hat{m}, \text{\textit{якщо} } m-\text{\textit{не відоме}}.\end{cases}\tag{26}$$

Оцінки (24) та (25) є незміщеними, а оцінка (26) є асимптотично незміщеною.

Означення. *Корелограмою* називається вибіркова автоковаріаційна функція $\hat{\sigma}(h)$ нормована на $\hat{\sigma}(0)$, тобто

$$r(h)=\frac{\hat{\sigma}(h)}{\hat{\sigma}(0)}.$$

В свою чергу вибіркова оцінка спектральної щільності має вигляд

$$\hat{g}(\lambda)=\frac{T}{4\pi}R^2(\lambda),\lambda\in[-\pi,\pi],$$

де

$$R^2(\lambda)=A^2(\lambda)+B^2(\lambda),$$

$$A(\lambda)=\frac{2}{T}\sum_{t=1}^{T}(e_t-e_0)Cos(\lambda t),$$

$$B(\lambda)=\frac{2}{T}\sum_{t=1}^{T}(e_t-e_0)Sin(\lambda t).$$