Lineárna regresia

0.1 Vektor ako časová závislosť

V tejto kapitole budeme často pojem vektor zamieňať pojmom proces. Dôvodom je, že teoretické výsledky o vektorových priestoroch, ktoré sme zopakovali v predchádzjúcej kapitole budeme používať pri vyšetrovaní vlastností (analyzovaní) procesov závislých od času. Takéto procesy sa niekedy nazývajú časové rady.

Aby sme si lepšie uvedomili, že objektom nášho skúmania je proces, ktorého hodnoty sú merané postupne, v po sebe nasledujúcich časoch, označíme proces nasledovne:

$$\mathbf{f} = (f(t_0), f(t_1), \ldots, f(t_{N-1}))$$

To znamená, že proces $\mathbf{f}$ chápeme na jednej strane ako vektor, na druhej strane ako funkciu času v okamihoch $t_0, t_1, \ldots, t_{N-1}$. Časové okamihy môžeme tiež považovať za vektor

$$\mathbf{t} = (t_0, t_1, \ldots, t_{N-1})$$

Časy môžeme označiť ich poradovým číslom (aj keď intervaly medzi nimi nemusia byť rovnaké, t.j. ekvidištantné).
Potom vektor $\mathbf{f}$ môžeme zapísať ako

$$\mathbf{f} = (f_0, f_1, \ldots, f_{N-1})$$

Základnou úlohou v tejto kapitole bude nájsť v procese nejakú závislosť od času a modelovať proces pomocou tejto časovej funkcie.
Najjednoduchším modelom bude konštantná funkcia.

Príklad 0.1.1 *Nech* $\mathbf{f}_0, \mathbf{f}_1, \mathbf{f}_2$ *sú vektory tvaru:*

$$\begin{aligned}\mathbf{f}_0 &= (c_0, c_0, \ldots, c_0)\\ \mathbf{f}_1 &= (c_1, c_1, \ldots, c_1)\\ \mathbf{f}_2 &= (c_2, c_2, \ldots, c_2)\end{aligned}$$

Zistite koľkorozmerný je priestor, ktorý určujú tieto vektory.

Riešenie:
Ľubovoľný vektor $\mathbf{f}_n$ spomezi vektorov $\mathbf{f}_0, \mathbf{f}_1, \mathbf{f}_2$ môžeme vyjadriť ako:

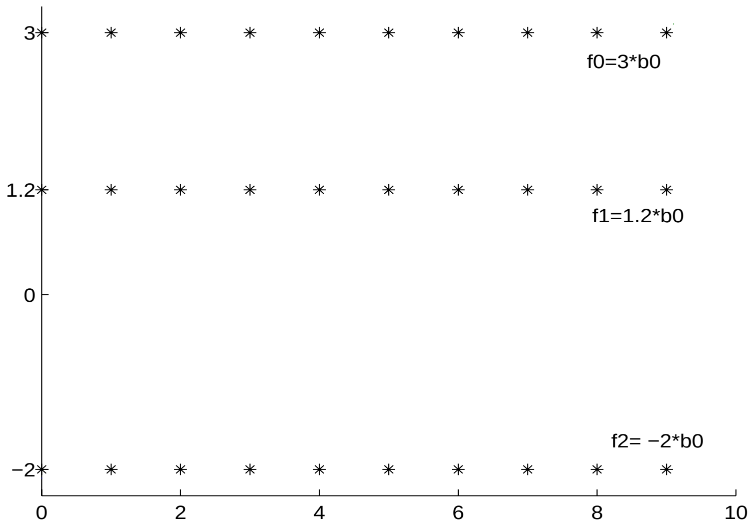
$$\begin{aligned}\mathbf{f}_{n\mathcal{E}} &= c_n \cdot \mathbf{b}_0 = c_n \cdot (1, 1, \ldots, 1) = (c_n, \ldots, c_n)\\ \mathbf{f}_{n\mathcal{B}} &= (c_n)\end{aligned}$$

Vidíme, že každý z vektorov $\mathbf{f}_n$ sa dá napísať ako násobok vektora $\mathbf{b}_0$ a teda sa jedná o jednorozmerný podpriestor generovaný bázou

$$\mathcal{B} = (\mathbf{b}_0) = ((1, 1, \ldots, 1))$$

$\square$

V príklade 0.1.1 boli zadané vektory, ktoré sa dali vyjadriť v tvare $c \cdot (1, 1, \ldots, 1)$. To znamená, že všetky tieto vektory ležia v podpriestore generovanom vektorom $(1, 1, \ldots, 1)$. Keď si tento podpriestor predstavíme ako priebeh procesu v čase, dostávame konštantný priebeh.
Funkcia času zodpovedajúca tomuto modelu by bola $\varphi(t_k) = c$. Na obrázku 1 sú tri vektory, ktoré ležia v podpriestore $\mathbb{V}$ generovanom vektorom $(1, 1, \ldots, 1)$, s hodnotami $c_0 = -2, c_1 = 1.2, c_2 = 3$.



Obrázok 1: Násobky vektora (1,1,...,1)

Konštanty $c_0 = -2, c_1 = 1.2, c_2 = 3$ a vektor $(1, 1, \ldots, 1)$ dobre popisujú procesy $\mathbf{f}_0, \mathbf{f}_1, \mathbf{f}_2$). Teda na popis týchto vektorov stačí omnoho menší počet hodnôt.

Nie každý vektor sa dá vyjadriť ako násobok vektora $(1, 1, \ldots, 1)$. Keď sa budeme snažiť popísať menším počtom hodnôt iné vektory, dopustíme sa chyby. Nevyjadríme presne pôvodný vektor, ale len jeho aproximáciu. V nasledujúcom odseku ukážeme, ako zabezpečiť, aby sme sa dopustili čo najmenšej chyby.

0.2 Aproximácia vektora v podpriestore

V tomto odseku riešime úlohu ako v podpriestore $\mathbb{U}$ nájsť vektor $\mathbf{y}$, ktorý najlepšie aproximuje zadaný vektor $\mathbf{v}$ a leží v $\mathbb{U}$.
Hovoríme, že vektor $\mathbf{y}$ ležiaci v podpriestore $\mathbb{U}$ **sa najviac podobá na** proces $\mathbf{v}$

(alebo vektor $\mathbf{y}$ je najbližšie k vektoru $\mathbf{v}$, vektor $\mathbf{y}$ aproximuje vektor $\mathbf{v}$), práve vtedy, ak vzdialenosť medzi týmito vektormi je minimálna:

$$d(\mathbf{f}, \mathbf{y}) \to \min_{\mathbf{f} \in \mathbb{U}}$$

Jednou z metód ako úlohu riešiť je použitie diferenciálneho počtu, ako ukazuje nasledujúci príklad.

Príklad 0.2.1 *Nájdite vektor* $\mathbf{y}$ *ležiaci na osi x, taký že je najlepšou aproximáciou vektora* $\mathbf{v} = (4, 2)$.

Riešenie:
Vektory na osi x sú násobky vektora $\mathbf{b} = (1, 0)$, teda majú tvar $(c, 0)$, kde c je ľubovoľné reálne číslo.
Najbližšie k vektoru $\mathbf{v} = (4, 2)$ bude ten z vektorov $(c, 0)$, ktorý minimalizuje vzdialenosť medzi nimi

$$d(\mathbf{v}, \mathbf{y}) \to \min$$

$$d((4, 2), (c, 0)) \to \min$$

V Euklidovom vektorovom priestore budeme minimalizovať vzdialenosť

$$\sqrt{(4 - c)^2 + (2 - 0)^2}$$

Hľadáme bod c, v ktorom funkcia $\sqrt{(4 - c)^2 + (2 - 0)^2}$ dosiahne minimum.
Stačí ak nájdeme minimum funkcie $g(c) = (4 - c)^2 + (2 - 0)^2$, nakoľko druhá odmocnina je na príslušnej oblasti rastúca funkcia.

$$g'(c) = 2 \cdot (4 - c)(-1), \quad g''(c) = 2$$

Odtiaľ dostávame, že $c = 4$ je stacionárny bod a kladná druhá derivácia zabezpečí, že je to minimum.
Teda najbližšie k vektoru $\mathbf{v} = (4, 2)$ je na osi x vektor $\mathbf{y} = (4, 0)$.
□
Tento výsledok sa dal dosiahnuť intuitívne. Rovnakým spôsobom teraz dokážeme vyriešiť nasledujúcu úlohu.

Úloha 0.2.1 *Nájdite vektor* $\mathbf{y}$ *ležiaci v podpriestore generovanom vektorom* $\mathbf{b} = (1, 1, 1, 1, 1)$ *taký, že je najlepšou aproximáciou vektora* $\mathbf{v} = (4, 2, 3, 2, 3)$.

Formálnejšie: Let $\mathbb{U}$ *be the subspace of* $\mathbb{E}_5$ *generated by the vector* $\mathbf{b}$*, i.e.* $\mathbb{U} = [\mathbf{b}]$. *Find the best approximation* $\mathbf{b} \in \mathbb{U}$ *of vector* $\mathbf{v} = (4, 2, 3, 2, 3)$.

Metóda, ktorú sme použili v príklade 0.2.1 sa značne skomplikuje, keď bude podpriestor, v ktorom aproximáciu hľadáme, viacrozmerný. Namiesto extrému funkcie jednej premennej bude potrebné nájsť minimum funkcie viac premenných, čo môže byť náročná úloha.

Príklad 0.2.2 *Nech je proces* $\mathbf{f}$ *zadaný 100 hodnotami. Ako tento proces popíšeme len tromi hodnotami* c_0, c_1, c_2*?*

Riešenie:
Proces $\mathbf{f}$ vyjadríme vo vektorovom priestore, ktorého báza je $\mathcal{B} = (\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2)$ a c_0, c_1, c_2 sú koeficienty v báze $\mathcal{B}$.

$$\mathbf{f} = (f_0, f_1, \ldots, f_{100}) = c_0\mathbf{b}_0 + c_1\mathbf{b}_1 + c_2\mathbf{b}_2$$

Vektor $\mathbf{f}$ je vyjadrený v trojrozmernom vektorovom priestore určenom bázou $\mathcal{B}$. Niektoré 100 zložkové vektory naozaj možno napísať ako lineárnu kombináciu vektorov z $\mathcal{B}$. Nie vždy je to však možné, vtedy sa snažíme nájsť v priestore určenom bázou $\mathcal{B}$ taký vektor $\tilde{\mathbf{f}}$, ktorý je najbližší k vektoru $\mathbf{f}$.
□

Existuje však aj iná metóda, vychádzajúca z geometrického pravidla, že kolmá vzdialenosť je najkratšia, alebo, že odvesna v pravouhlom trojuholníku je vždy kratšia ako prepona ako vidno na obrázku 2. V nasledujúcom trvdení je dokázané, že v Euklidovom vektorovom priestore, je k vektoru $\mathbf{f}$, spomedzi všetkých vektorov v podpriestore, najbližšie jeho kolmý priemet do tohoto podpriestoru.

Tvrdenie 0.2.1 *Nech $\mathbb{V}$ je N-rozmerný vektorový priestor s jednotkovou bázou $\mathcal{E}$ a $\mathbb{U}$ je jeho M-rozmerný vektorový podpriestor s bázou $\mathcal{B}$.*
Nech $\mathbf{f} = (f_0, f_1, \ldots, f_{N-1})_{\mathcal{E}} \in \mathbb{V}$. Spomedzi vektorov $\mathbf{y} \in \mathbb{U}$ bude najbližší k vektoru $\mathbf{f}$, jeho kolmý priemet do podpriestoru $\mathbb{U}$, teda vektor $\tilde{\mathbf{f}} \in \mathbb{U}$, taký, že $\mathbf{f} - \tilde{\mathbf{f}}$ je kolmý na všetky vektory $\mathbb{U}$.

Dôkaz:
Nech $\mathbf{y}$ je ľubovoľný vektor $\in \mathbb{U}$, potom vektor $\mathbf{y} - \tilde{\mathbf{f}} \in \mathbb{U}$. Vektor $\mathbf{f} - \tilde{\mathbf{f}}$ je kolmý na každý vektor z $\mathbb{U}$, teda aj na vektor $\mathbf{y} - \tilde{\mathbf{f}}$. Podľa Pytagorovej vety pre navzájom kolmé vektory platí

$$||(\mathbf{f} - \tilde{\mathbf{f}}) - (\mathbf{y} - \tilde{\mathbf{f}})||^2 = ||\mathbf{f} - \tilde{\mathbf{f}}||^2 + ||\mathbf{y} - \tilde{\mathbf{f}}||^2$$

$$||\mathbf{f} - \mathbf{y}||^2 = ||\mathbf{f} - \tilde{\mathbf{f}}||^2 + ||\mathbf{y} - \tilde{\mathbf{f}}||^2$$

$$d(\mathbf{f}, \mathbf{y}) = d(\mathbf{f}, \tilde{\mathbf{f}}) + d(\mathbf{y}, \tilde{\mathbf{f}})$$

Pretože vzdialenosť je nezáporná ($d(\mathbf{y}, \tilde{\mathbf{f}}) \geq 0$), platí

$$d(\mathbf{f}, \mathbf{y}) \geq d(\mathbf{f}, \tilde{\mathbf{f}})$$

Vektor $\mathbf{y}$ bol ľubovoľný, preto platí, že $d(\mathbf{f}, \tilde{\mathbf{f}})$ je minimálna. □

Príklad 0.2.3 *Nájdite v podpriestore $\mathbb{U} = [(1,0,2), (-2,1,1)]$ vektor, ktorý je najbližšie k vektoru $\mathbf{f} = (1,2,3)$.*

Riešenie:
Najbližšie k vektoru $\mathbf{f} = (1,2,3)$ bude vektor $\tilde{\mathbf{f}}$, pre ktorý platí

$$\tilde{\mathbf{f}} = c_0 \cdot (1,0,2) + c_1 \cdot (-2,1,1)$$

$$(\mathbf{f} - \tilde{\mathbf{f}}) \perp (1,0,2)$$

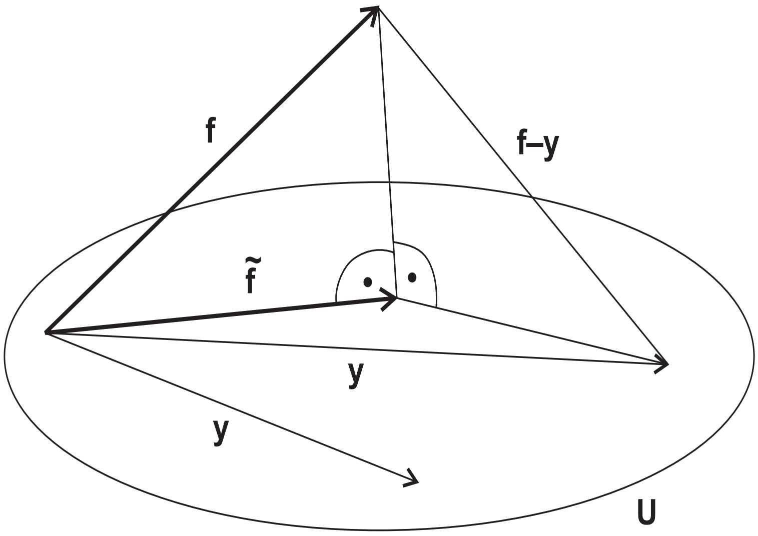
$$(\mathbf{f} - \tilde{\mathbf{f}}) \perp (-2,1,1)$$

Podmienky môžeme upraviť na tvar

$$\tilde{\mathbf{f}} = c_0 \cdot (1,0,2) + c_1 \cdot (-2,1,1)$$

$$\langle \mathbf{f} - \tilde{\mathbf{f}}, (1,0,2) \rangle = 0$$

$$\langle \mathbf{f} - \tilde{\mathbf{f}}, (-2,1,1) \rangle = 0$$



Obrázok 2: Najbližší k vektoru $\mathbf{f}$ spomedzi vektorov $\mathbf{y} \in U$ je vektor $\tilde{\mathbf{f}}$

Odtiaľ po dosadení za $\mathbf{f}$

$$\tilde{\mathbf{f}} = c_0 \cdot (1,0,2) + c_1 \cdot (-2,1,1)$$

$$\langle (1,2,3),(1,0,2)\rangle = c_0 \cdot \langle (1,0,2),(1,0,2)\rangle + c_1 \cdot \langle (-2,1,1),(1,0,2)\rangle,$$

$$\langle (1,2,3),(-2,1,1)\rangle = c_0 \cdot \langle (1,0,2),(-2,1,1)\rangle + c_1 \cdot \langle (-2,1,1),(-2,1,1)\rangle$$

Teda platí

$$\tilde{\mathbf{f}} = c_0 \cdot (1,0,2) + c_1 \cdot (-2,1,1)$$

$$7 = c_0 \cdot 5 + c_1 \cdot 0$$

$$3 = c_0 \cdot 0 + c_1 \cdot 6$$

Nakoniec

$$\tilde{\mathbf{f}} = \frac{7}{5} \cdot (1,0,2) + \frac{3}{6} \cdot (-2,1,1) = (0.4, 0.5, 3.3)$$

□

Úloha 0.2.2 *Nájdite v podpriestore* $\mathbb{U} = [(1,0,2),(-2,1,1)]$ *vektor, ktorý je najbližšie k vektoru* $\mathbf{f} = (-5,2,0)$.

0.3 Metóda najmenších štvorcov

Pomerne známou metódou na výpočet krivky, ktorá čo najlepšie aproximuje dáta, je metóda najmenších štvorcov. Úloha je zvyčajne formulovaná pre priamku a riešená metódami diferenciálneho počtu. Aj v tomto prípade však môžeme úlohu previesť do jazyka vektorových priestorov. Minimalizáciou vzdialenosti medzi vektorom Euklidoveho vektorového priestoru $\mathbf{f}$ a jeho priemetom $\tilde{\mathbf{f}}$

do podpriestoru dostaneme metódu najmenších štvorcov. Na obrázku 3 je zobrazený priemet do podpriestoru, ktorý určuje priamku $y = a + b \cdot t$.
Nech:

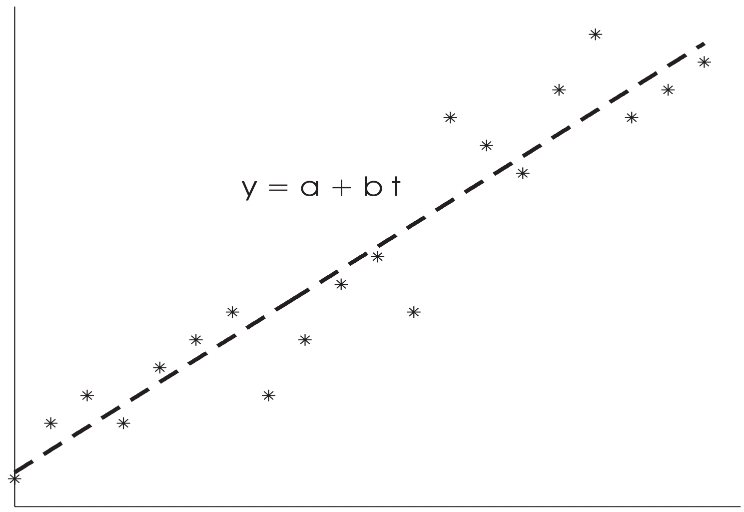
$$\mathbf{f} = (f_0, f_1, \ldots, f_{N-1}), \quad \tilde{\mathbf{f}} = (\tilde{f}_0, \tilde{f}_1, \ldots, \tilde{f}_{N-1})$$

Potom

$$d(\mathbf{f}, \tilde{\mathbf{f}}) = ||\mathbf{f} - \tilde{\mathbf{f}}|| = \sqrt{\langle \mathbf{f} - \tilde{\mathbf{f}}, \mathbf{f} - \tilde{\mathbf{f}} \rangle} =$$

$$= \sqrt{(f_0 - \tilde{f}_0)^2 + \cdots + (f_{N-1} - \tilde{f}_{N-1})^2} = \sqrt{\sum_{k=0}^{N-1} (f_k - \tilde{f}_k)^2} \to \min$$

$$d(\mathbf{f}, \tilde{\mathbf{f}})^2 = \sum_{k=0}^{N-1} (f_k - \tilde{f}_k)^2 \to \min$$



Obrázok 3: Dáta aproximované priamkou

0.4 Chyba aproximácie

Pri aproximácii procesu nejakým iným procesom je dôležité vedieť akej chyby sme sa pri odhade dopustili. Pokiaľ poznáme oba procesy, pôvodný aj jeho odhad, vieme presne povedať aká je chyba odhadu.
Nech $\mathbf{f}_{\mathcal{E}} = (f_0, f_1, \ldots, f_{N-1})$ je proces vyjadrený v jednotkovej báze N-rozmerného vektorového priestoru. Hodnotu $||\mathbf{f}||^2 = d^2(\mathbf{f}, \mathbf{0})$ nazveme **energia procesu f**.

Nech $\tilde{\mathbf{f}}$ je aproximácia procesu $\mathbf{f}$ v podpriestore určenom bázou $\mathcal{B}$. Vektor $\mathbf{e}_{\mathcal{E}} = (e_0, e_1, \ldots, e_{N-1}) = \mathbf{f} - \tilde{\mathbf{f}}$ nazveme **chybový vektor**.
Vzdialenosť

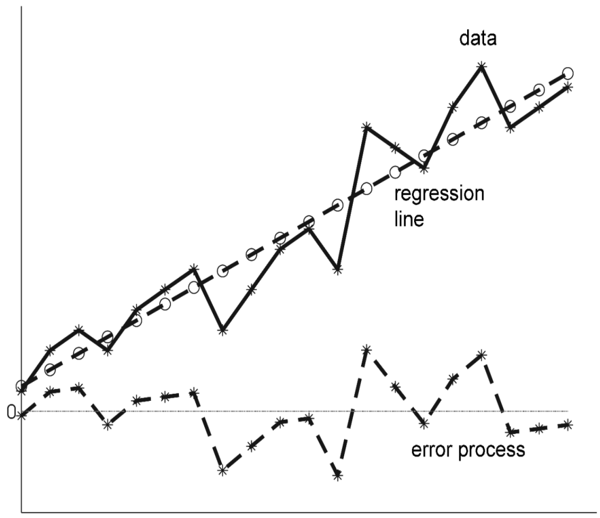
$$d(\mathbf{f}, \tilde{\mathbf{f}}) = \sqrt{\sum_{k=0}^{N-1} (f_k - \tilde{f}_k)^2} = \sqrt{\sum_{k=0}^{N-1} e_k^2} = ||\mathbf{e}||$$

nazveme **veľkosť chybového vektora**.

$$\begin{aligned}\tilde{\mathbf{f}}_{\mathcal{B}} &= (c_0, c_1, \ldots, c_{M-1})\\ \tilde{\mathbf{f}}_{\mathcal{E}} &= (\tilde{f}_0, \tilde{f}_1, \ldots, \tilde{f}_{N-1})\\ \mathbf{e}_{\mathcal{E}} &= (e_0, e_1, \ldots, e_{N-1})\end{aligned}$$

$$d^2(\mathbf{f}, \tilde{\mathbf{f}}) = \sum_{k=0}^{N-1} e_k^2$$

Hodnota $||\mathbf{e}||^2$ je **energia chybového procesu**.
Euklidova norma procesu je úmerná veľkosti jeho energie. Proces je tým väčší, čím väčšiu má energiu.



Obrázok 4: Pôvodný proces, jeho priemet do dvojrozmerného podpriestoru, chybový proces (proces po odstránení lineárneho trendu)

Príklad 0.4.1 *Nech priemet časti procesu* $\mathbf{y} = (1, 3, 4, 3, 5, 6, 7, 4, 6, 8)$ *z obrázku* 4 *je približne proces*

$$\tilde{\mathbf{y}} = (1.23, 2.04, 2.85, 3.67, 4.48, 5.29, 6.11, 6.92, 7.73, 8.54)$$

Aká je veľkosť chyby?

Riešenie:
Chybový proces je proces

$$\mathbf{e} = (-0.23, 0.96, 1.15, -0.67, 0.52, 0.71, 0.89, -2.92, -1.73, -0.54)$$

Veľkosť chybového vektora je $||\mathbf{e}|| = 7.0457$ a jeho energia je $||\mathbf{e}||^2 = 49.6414$. $\square$

Úloha 0.4.1 *Nájdite dva procesy, ktoré majú rovnakú veľkosť chyby, ale rôzne chybové procesy.*

Tvrdenie 0.4.1 *Ak sú dva procesy na seba kolmé, potom energia súčtového procesu sa rovná súčtu energií jednotlivých procesov.*

Dôkaz:

$$||\mathbf{f}+\mathbf{g}||^2 = \langle \mathbf{f}+\mathbf{g}, \mathbf{f}+\mathbf{g}\rangle = \langle \mathbf{f}, \mathbf{f}+\mathbf{g}\rangle + \langle \mathbf{g}, \mathbf{f}+\mathbf{g}\rangle = \langle \mathbf{f}, \mathbf{f}\rangle + \langle \mathbf{f}, \mathbf{g}\rangle + \langle \mathbf{g}, \mathbf{f}\rangle + \langle \mathbf{g}, \mathbf{g}\rangle =$$

$$= \langle \mathbf{f}, \mathbf{f}\rangle + 0 + 0 + \langle \mathbf{g}, \mathbf{g}\rangle = ||\mathbf{f}||^2 + ||\mathbf{g}||^2$$

Toto tvrdenie je vlastne inak formulovaná Pytagorova veta. □

Predchádzajúce tvrdenie v tvare

$$||\mathbf{f}||^2 = ||\tilde{\mathbf{f}} + \mathbf{e}||^2 = ||\tilde{\mathbf{f}}||^2 + ||\mathbf{e}||^2$$

hovorí o tom, že priemet do podpriestoru má menšiu energiu ako pôvodný proces. Tento fakt sa využíva napríklad pri kompresných metódach.

Zopakujme, že úlohou v tejto kapitole je nájsť v procese nejakú závislosť od času a modelovať proces pomocou tejto časovej funkcie. Modely, ktorými budeme priebeh procesu vyjadrovať budú lineárne vzhľadom na neznáme koeficienty. Metóda, ktorá k nameraným dátam nájde spomedzi všetkých lineárnych kombinácií nejakých konkrétnych funkcií, tú najlepšiu v zmysle Euklidovej metriky sa nazýva **lineárna regresia**. Nasledujú jednotlivé úlohy lineárnej regresie rozdelené podľa tvaru regresnej krivky.

0.5 Priamka, rovnobežná s časovou osou $(y = c)$

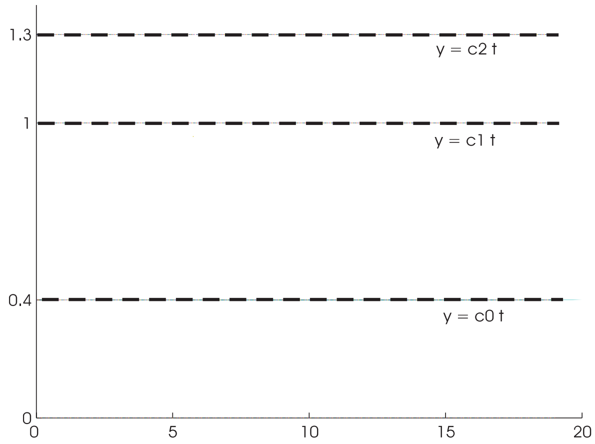
Je daný proces $\mathbf{f} = (f_0, f_1, \ldots, f_{N-1})$ popísaný N hodnotami, ktoré nadobudol v časoch $t_0, t_1, \ldots, t_{N-1}$ (tieto časy sú zadané a nemusia byť ekvidištantné). Riešme úlohu popísať tento proces jednou hodnotou, ktorá určí také umiestnenie priamky rovnobežnej s časovou osou, že priamka namerané dáta popisuje najlepšie.

$$\mathbf{f}_{\mathcal{E}} = (f_0, f_1, \ldots, f_{N-1}) \quad \tilde{\mathbf{f}}_{\mathcal{B}} = (c) \quad \tilde{\mathbf{f}}_{\mathcal{E}} = (c, c, \ldots, c)_{\mathcal{E}}$$

$$\tilde{\mathbf{f}} = c \cdot \mathbf{b}, \quad \text{kde} \quad \mathbf{b} = (1, 1, \ldots, 1)$$

$$c = \frac{\langle \mathbf{f}, \mathbf{b}\rangle}{\langle \mathbf{b}, \mathbf{b}\rangle} = \frac{\langle (f_0, f_1, \ldots, f_{N-1}), (1, 1, \ldots, 1)\rangle}{\langle (1, 1, \ldots, 1), (1, 1, \ldots, 1)\rangle} = \frac{\sum_{k=0}^{N-1} f_k}{N} = \frac{1}{N}\sum_{k=0}^{N-1} f_k$$

Hodnota, ktorú sme dostali, je aritmetický priemer. Uvedeným modelom sú dobre popísané procesy, ktorých hodnoty sú približne rovnaké.



Obrázok 5: Priamky, rovnobežné s časovou osou

Príklad 0.5.1 *Na internetovej stránke http://www.adis-computers.sk sme našli štatistiku návštevnosti stránky za posledných 100 dní. Jedná sa o počet unikátnych ip adries denne. Namerané hodnoty procesu sú*

$$\begin{aligned}\mathbf{f} = (&20, 34, 20, 35, 24, 16, 34, 29, 28, 27, 25, 20, 22, 28, 34, 33, 30, 32, 27, 20,\\ &36, 31, 33, 39, 30, 24, 33, 28, 43, 33, 30, 27, 22, 25, 29, 26, 28, 29, 24, 8,\\ &14, 30, 28, 21, 20, 31, 9, 28, 31, 18, 29, 33, 30, 20, 21, 26, 28, 33, 27, 28,\\ &15, 29, 32, 36, 33, 43, 24, 21, 25, 33, 40, 30, 31, 31, 17, 27, 24, 33, 26, 24,\\ &23, 22, 14, 26, 38, 27, 27, 30, 13, 17, 25, 31, 30, 31, 26, 14, 20, 26, 32, 30)\end{aligned}$$

Aproximujte tento proces procesom s konštantným priebehom.

Riešenie:
Konštantnú zložku procesu vyjadríme ako

$$\tilde{\mathbf{f}} = c \cdot \mathbf{b} = c \cdot (1, 1, \ldots, 1) \quad \text{kde} \quad c = \frac{\langle \mathbf{f}, \mathbf{b} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle}$$

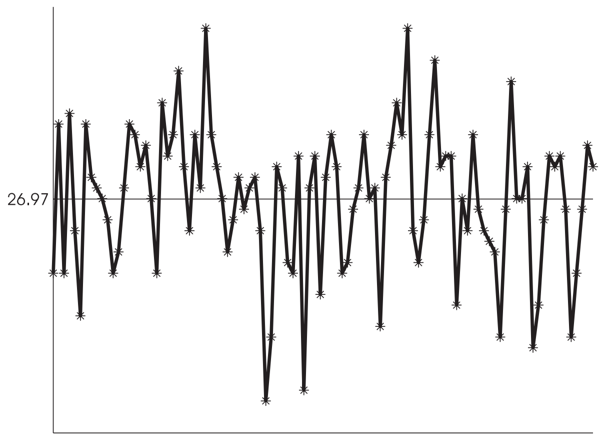
$$c = \frac{1}{100} \sum_{k=0}^{99} f_k = 26.97$$

Priebeh procesu a jeho aproximácia konštantným procesom sú zobrazené na obrázku 6. □

0.6 Priamka prechádzajúca počiatkom: $(y = c \cdot t)$

Ak hodnoty procesu rastú lineárne a priamka lineárnej závislosti prechádza počiatkom, bude mať proces $\tilde{\mathbf{f}}$ hodnoty $\tilde{f}_k = c \cdot t_k$. Teda proces s nameranými hodnotami $\mathbf{f}_\mathcal{E} = (f_0, f_1, \ldots, f_{N-1})$ ktoré nadobudol v časoch $t_0, t_1, \ldots, t_{N-1}$, popíšeme jednou hodnotou c

$$\tilde{\mathbf{f}} = (\tilde{f}_0, \tilde{f}_1, \ldots, \tilde{f}_{N-1})_\mathcal{E} = (c \cdot t_0, c \cdot t_1, \ldots, c \cdot t_{N-1}) = c(t_0, t_1, \ldots, t_{N-1})$$



Obrázok 6: Návštevnosť internetovej stránky a jej aproximácia konštantou

$$\tilde{\mathbf{f}} = c \cdot \mathbf{b}_0 = (c \cdot t_0, c \cdot t_1, \ldots, c \cdot t_{N-1})$$

Ku procesu $\mathbf{f}$ hľadáme taký proces $\tilde{\mathbf{f}}$, pre ktorý platí

$$d(\mathbf{f}, \tilde{\mathbf{f}}) \to \min$$

Nahradíme proces $\tilde{\mathbf{f}} = c \cdot \mathbf{b}$, pretože jeho hodnoty ležia na priamke:

$$d(\mathbf{f}, c \cdot \mathbf{b}) \to \min$$

Na vypočítanie koeficientu c použijeme podmienku $(\mathbf{f} - c \cdot \mathbf{b}) \perp \mathbf{b}$. Podmienka ortogonálnosti vyjadrená pomocou skalárneho súčinu má tvar:

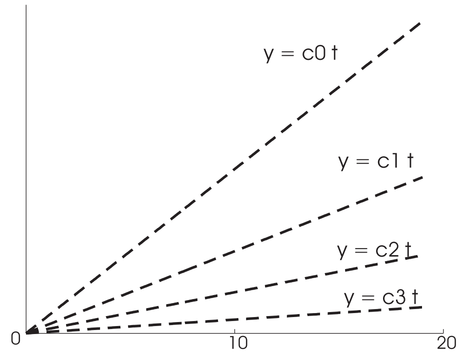
$$\langle \mathbf{f} - c \cdot \mathbf{b}, \mathbf{b} \rangle = 0$$

Využijeme vlastnosti skalárneho súčinu a dostaneme

$$\langle (f_0 - c \cdot t_0, f_1 - c \cdot t_1, \ldots, f_{N-1} - c \cdot t_{N-1}), (t_0, t_1, \ldots, t_{N-1}) \rangle = 0$$

$$c = \frac{\langle \mathbf{f}, \mathbf{b} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle} = \frac{\sum\limits_{k=0}^{N-1} f_k \cdot t_k}{\sum\limits_{k=0}^{N-1} t_k^2}$$

Príklad 0.6.1 *Na internete sme našli históriu vývoja devízových rezerv Národnej banky Slovenska za obdobie od* 29.1.1993 *do* 28.2.2001. *Sledované hodnoty sa zapisovali a sledovali vždy k poslednému pracovnému dňu v danom mesiaci.*



Obrázok 7: Sklon priamky $y = c \cdot t$ závisí od koeficientu c

Namerané hodnoty procesu sú v miliónoch USD:

$$\mathbf{f} = (197, 176, 184, 175, 256, 243, 336, 386, 567, 538, 508, 450, 401, 382, 475, 534, 610, 689, 1125, 1294, 1464, 1565, 1598, 1745, 17411813, 1969, 2022, 2186, 2622, 2630, 2708, 2813, 2873, 3046, 3418, 3307, 3398, 3459, 3407, 3355, 3377, 3504, 3677, 3655, 3602, 3595, 3473, 3434, 3472, 3453, 3347, 2974, 3019, 3010, 3181, 3151, 3411, 3446, 3285, 3161, 3202, 3143, 3349, 3723, 3790, 3770, 3622, 3110, 2987, 2939, 2923, 2860, 2910, 2814, 2732, 2515, 2953, 2864, 2805, 2935, 2952, 2869, 3425, 3444, 3564, 3699, 4198, 4085, 4037, 3988, 4433, 4214, 3994, 4024, 4101, 3973, 3971, 3890, 4074)$$

Aproximujte tento proces procesom s lineárnym nárastom s jedným parametrom.

Riešenie:
Lineárny nárast procesu vyjadríme pomocou jedného koeficientu ako

$$\tilde{\mathbf{f}} = c \cdot \mathbf{b} = c \cdot (0, 1, 2, \ldots, 99) \quad \text{kde} \quad c = \frac{\langle \mathbf{f}, \mathbf{b} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle}$$

$$c = \frac{\sum_{k=0}^{99} k \cdot f_k}{\sum_{k=0}^{99} k^2} = 49.0672$$

Priebeh procesu a jeho aproximácia lineárnym procesom s jedným parametrom sú zobrazené na obrázku 8.

□

Úloha 0.6.1 *Vyhľadajte na internete stav devízových rezerv NBS k dnešnému dátumu a porovnajte túto hodnotu s odhadom, ktorý urobíte pomocou modelu z príkladu* 0.6.1.

Obrázok 8: Devízové rezervy NBS a ich aproximácia priamkou s jedným parametrom

0.7 Priamka neprechádzajúca počiatkom ($y = c_0 + c_1 \cdot t$)

Ak hodnoty procesu rastú lineárne a priamka lineárnej závislosti neprechádza počiatkom, bude mať proces $\tilde{\mathbf{f}}$ hodnoty $\tilde{f}_k = c_0 + c_1 \cdot t_k$. Predpokladáme, že proces s nameranými hodnotami $\mathbf{f} = (f_0, f_1, \ldots, f_{N-1})$, ktoré nadobudol v časoch
$t_0, t_1, \ldots, t_{N-1}$, ležia približne na priamke, ktorá neprechádza počiatkom.
Budeme ich aproximovať hodnotami

$$\tilde{\mathbf{f}} = (\tilde{f}_0, \tilde{f}_1, \ldots, \tilde{f}_{N-1}) = (c_0 + c_1 \cdot t_0, c_0 + c_1 \cdot t_1, \ldots, c_0 + c_1 \cdot t_{N-1})$$

$$f_k = c_0 + c_1 \cdot t_k$$

$$\begin{aligned} \tilde{\mathbf{f}}_{\mathcal{B}} &= (c_0, c_1) \\ \tilde{\mathbf{f}}_{\mathcal{E}} &= c_0(1, 1, \ldots, 1) + c_1(t_0, t_1, \ldots, t_{N-1}) = \\ &= (c_0 + c_1 \cdot t_0, c_0 + c_1 \cdot t_1, \ldots, c_0 + c_1 \cdot t_k) = c_0 \cdot \mathbf{b}_0 + c_1 \cdot \mathbf{b}_1 \end{aligned}$$

kde báza

$$\mathcal{B} = (\mathbf{b}_0, \mathbf{b}_1), \text{ kde } \mathbf{b}_0 = (1, 1, \ldots, 1), \mathbf{b}_1 = (t_0, t_1, \ldots, t_{N-1})$$

Procesy $\mathbf{b_0}$ a $\mathbf{b_1}$ nie sú na seba kolmé. Hľadáme proces $\tilde{\mathbf{f}}$, ktorý leží v podpriestore $\mathbb{U}$ určenom bázou $\mathcal{B}$ a je najbližší k procesu $\mathbf{f}$. Ukázali sme, že to bude proces $\tilde{\mathbf{f}}$, pre ktorý platí, že $\mathbf{f} - \tilde{\mathbf{f}}$ bude kolmý na $\mathbb{U}$. To znamená, že bude kolmý na každý proces z $\mathbb{U}$, teda aj na každý proces bázy $\mathcal{B} = (\mathbf{b}_0, \mathbf{b}_1)$. Teda proces $\tilde{\mathbf{f}} = c_0\mathbf{b}_0 + c_1\mathbf{b}_1 \quad \text{a} \quad \mathbf{f} - \tilde{\mathbf{f}} \perp \mathbb{U}$.
Vychádzame zo vzťahov $\mathbf{f} - \tilde{\mathbf{f}} \perp \mathbf{b_0}$ a $\mathbf{f} - \tilde{\mathbf{f}} \perp \mathbf{b_1}$.
Upravíme:

$$\langle \mathbf{f} - \tilde{\mathbf{f}}, \mathbf{b}_0 \rangle = 0\,; \quad \langle \mathbf{f} - \tilde{\mathbf{f}}, \mathbf{b}_1 \rangle = 0$$

$$\begin{gathered} \langle \mathbf{f} - c_0 \cdot \mathbf{b}_0 - c_1 \cdot \mathbf{b}_1, \mathbf{b}_0 \rangle = 0 \\ \langle \mathbf{f} - c_0 \cdot \mathbf{b}_0 - c_1 \cdot \mathbf{b}_1, \mathbf{b}_1 \rangle = 0 \end{gathered}$$

$$\begin{aligned}\langle \mathbf{f}, \mathbf{b}_0\rangle - \langle c_0\mathbf{b}_0, \mathbf{b}_0\rangle - \langle c_1\mathbf{b}_1, \mathbf{b}_0\rangle = 0\\ \langle \mathbf{f}, \mathbf{b}_1) - \langle c_0\mathbf{b}_0, \mathbf{b}_1\rangle - \langle c_1\mathbf{b}_1, \mathbf{b}_1\rangle = 0\end{aligned}$$

$$\begin{aligned}\langle \mathbf{f}, \mathbf{b}_0\rangle - c_0\langle \mathbf{b}_0, \mathbf{b}_0\rangle - c_1\langle \mathbf{b}_1, \mathbf{b}_0\rangle = 0\\ \langle \mathbf{f}, \mathbf{b}_1\rangle - c_0\langle \mathbf{b}_0, \mathbf{b}_1\rangle - c_1\langle \mathbf{b}_1, \mathbf{b}_1\rangle = 0\end{aligned}$$

$$\begin{aligned}c_0\langle \mathbf{b}_0, \mathbf{b}_0\rangle + c_1\langle \mathbf{b}_1, \mathbf{b}_0\rangle = \langle \mathbf{f}, \mathbf{b}_0\rangle\\ c_0\langle \mathbf{b}_0, \mathbf{b}_1\rangle + c_1\langle \mathbf{b}_1, \mathbf{b}_1\rangle = \langle \mathbf{f}, \mathbf{b}_1\rangle\end{aligned}$$

Dostaneme sústavu dvoch lineárnych rovníc o dvoch neznámych.

$$\begin{aligned}c_0 \cdot N + c_1 \cdot \sum_{k=0}^{N-1} t_k = \sum_{k=0}^{N-1} f_k\\ c_0 \cdot \sum_{k=0}^{N-1} t_k + c_1 \cdot \sum_{k=0}^{N-1} t_k^2 = \sum_{k=0}^{N-1} f_k \cdot t_k\end{aligned}$$

Riešením tejto sústavy budú koeficienty procesu $\tilde{\mathbf{f}}$ v báze $\mathcal{B} = (\mathbf{b_0}, \mathbf{b_1})$.

Príklad 0.7.1 *Aproximujte proces z príkladu* 0.6.1 *procesom s lineárnym nárastom s dvomi parametrami.*

Riešenie:
Lineárny nárast procesu vyjadríme pomocou dvoch koeficientov ako

$$\tilde{\mathbf{f}} = c_0 \cdot \mathbf{b}_0 + c_1 \cdot \mathbf{b}_1 = c_0 \cdot (1, 1, \dots, 1) + c_1 \cdot (0, 1, 2, \dots, 99)$$

Koeficienty c_0, c_1 vypočítame ako riešenie sústavy rovníc

$$\begin{aligned}100 \cdot c_0 + c_1 \cdot \sum_{k=0}^{99} k = \sum_{k=0}^{99} f_k\\ c_0 \cdot \sum_{k=0}^{99} k + c_1 \cdot \sum_{k=0}^{99} k^2 = \sum_{k=0}^{99} f_k \cdot k\end{aligned}$$

$$\begin{aligned}100 \cdot c_0 + 4950 \cdot c_1 = 264773\\ 4950 \cdot c_0 + 328350 \cdot c_1 = 16111220\end{aligned}$$

Riešením tejto sústavy sú koeficienty $c_0 = 862.61, c_1 = 36.06$.

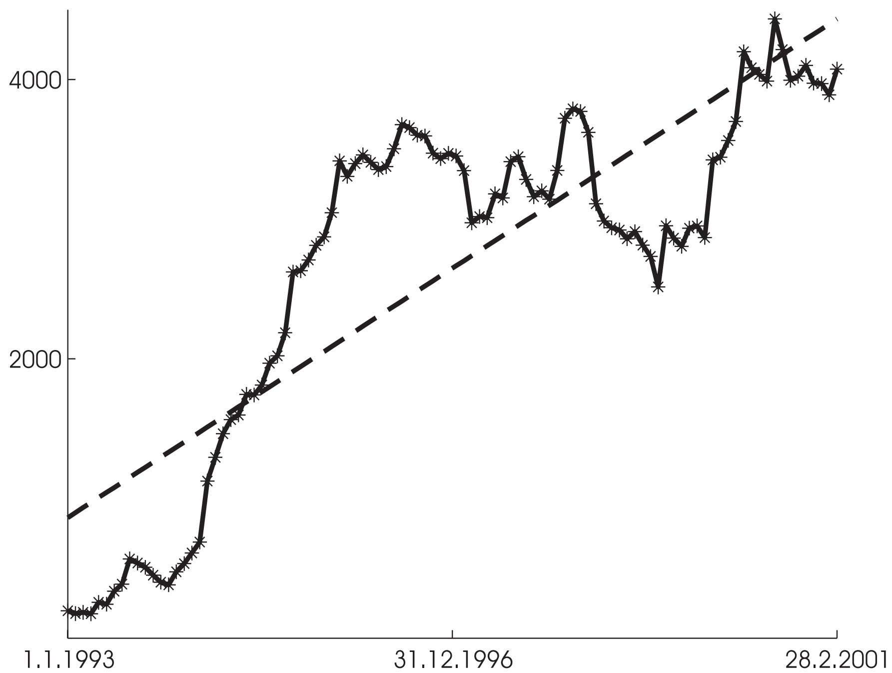
Priebeh procesu a jeho aproximácia lineárnym procesom s dvomi parametrami sú zobrazené na obrázku 9.

□

Úloha 0.7.1 *Stav devízových rezerv NBS k dnešnému dátumu porovnajte s odhadom, ktorý urobíte pomocou modelu z príkladu* 0.7.1.*Chyba tohoto odhadu bude zrejme menšia, ako chyba odhadu z úlohy* 0.6.1.

0.8 Lineárna regresia ako priemet do podpriestoru

Nech $\mathbf{f}_{\mathcal{E}} = (f_0, f_1, \dots, f_{N-1})$ je proces vyjadrený v jednotkovej báze N-rozmerného vektorového priestoru. Tento proces aproximujeme procesmi menejrozmerného vektorového podpriestoru $\mathbb{U}$. Nasledujú príklady priemetu procesu $\mathbf{f}$ v podpriestoroch generovaných bázami:



Obrázok 9: Devízové rezervy NBS a ich aproximácia priamkou s dvomi parametrami

- $\mathcal{B}_1 = (\mathbf{b}) = ((1, 1, \dots, 1))$ (jednorozmerný podpriestor):

$$\tilde{\mathbf{f}}_{\mathcal{E}} = (c, c, \dots, c)$$
$$\tilde{\mathbf{f}}_{\mathcal{B}_1} = (c)$$

- $\mathcal{B}_2 = (\mathbf{b}_0 = (1, 1, \dots, 1), \mathbf{b}_1 = (t_0, t_1, \dots, t_{N-1}))$ (dvojrozmerný podpriestor):

$$\tilde{\mathbf{f}}_{\mathcal{E}} = (c_0 + c_1 \cdot t_0, c_0 + c_1 \cdot t_1, \dots, c_0 + c_1 \cdot t_{N-1}) \qquad \tilde{\mathbf{f}}_{\mathcal{B}_2} = (c_0, c_1)$$

- $\mathcal{B}_3 = ((1, 1, \dots, 1), \dots, (t_0^{M-1}, t_1^{M-1}, \dots, t_{N-1}^{M-1}))$ (M-rozmerný podpriestor):

$$\begin{aligned}\tilde{\mathbf{f}}_{\mathcal{E}} =&(c_0 + c_1 t_0 + c_2 \cdot t_0^2 + \dots c_{M-1} \cdot t_0^{M-1}, c_0 + c_1 t_1 + c_2 \cdot t_1^2 + \dots + c_{M-1} \cdot t_1^{M-1},\\ &\dots, c_0 + c_1 t_{N-1} + c_2 \cdot t_{N-1}^2 + \dots c_{M-1} \cdot t_{N-1}^{M-1})\end{aligned}$$

$$\tilde{\mathbf{f}}_{\mathcal{B}_3} = (c_0, c_1, \dots, c_{M-1})$$

- $\mathcal{B}_4 = ((\varphi_0(t_0), \dots, \varphi_0(t_{N-1})), \dots, (\varphi_{M-1}(t_0), \dots, \varphi_{M-1}(t_{N-1})))$ (M-rozmerný podpriestor):

$$\begin{aligned}\tilde{\mathbf{f}}_{\mathcal{E}} =&(c_0 \cdot \varphi_0(t_0) + c_1 \cdot \varphi_1(t_0) + c_2 \cdot \varphi_2(t_0) + \dots + c_{M-1} \cdot \varphi_{M-1}(t_0),\\ &c_0 \cdot \varphi_0(t_1) + c_1 \cdot \varphi_1(t_1) + c_2 \cdot \varphi_2(t_1) + \dots + c_{M-1} \cdot \varphi_{M-1}(t_1),\\ &\dots\\ &c_0 \cdot \varphi_0(t_{N-1}) + c_1 \cdot \varphi_1(t_{N-1}) + \dots + c_{M-1}\varphi_{M-1}(t_{N-1}))\\ =&\left(\sum_{n=0}^{M-1} c_n \cdot \varphi_n(t_0), \dots, \sum_{n=0}^{M-1} c_n \cdot \varphi_n(t_k), \dots, \sum_{n=0}^{M-1} c_n \cdot \varphi_n(t_{N-1})\right)\end{aligned}$$

$$\tilde{\mathbf{f}}_{\mathcal{B}_4} = (c_0, c_1, \dots, c_{M-1})$$

Úloha 0.8.1 *Nájdite v literatúre, alebo na internete návod, ako riešiť úlohu lineárnej regresie pomocou diferenciálneho počtu funkcie viac premenných. Vyriešte pomocou tohoto návodu úlohu z príkladu* 0.7.1.

Úloha 0.8.2 *Zistite, aký algoritmus je použitý na výpočet regresnej krivky v rôznych softvéroch. Porovnajte riešenie rovnakej úlohy v aspoň troch z nich. Zistite, ktorý urobí najmenšiu a ktorý najväčšiu chybu odhadu.*

0.9 Všeobecný zápis lineárnej regresie

Všeobecne sa snažíme proces $\mathbf{f}$ nahradiť procesom $\tilde{\mathbf{f}}$ vyjadreným v báze $\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_{M-1}$

$$\tilde{\mathbf{f}} = \sum_{n=0}^{M-1} c_n \mathbf{b_n}$$

pričom

$$\mathbf{b_0} = (\varphi_0(t_0), \varphi_0(t_1), \dots, \varphi_0(t_k), \dots, \varphi_0(t_{N-1}))$$

$$\mathbf{b}_1 = (\varphi_1(t_0), \varphi_1(t_1), \ldots, \varphi_1(t_k), \ldots, \varphi_1(t_{N-1}))$$

$$\mathbf{b}_{M-1} = (\varphi_{M-1}(t_0), \varphi_{M-1}(t_1), \ldots, \varphi_{M-1}(t_k), \ldots, \varphi_{M-1}(t_{N-1}))$$

Teda proces $\tilde{\mathbf{f}}$ je lineárnou kombináciou procesov $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_{M-1}$ a proces $\mathbf{f} - \tilde{\mathbf{f}}$ je kolmý na všetky bázické procesy.

$$(\mathbf{f} - \tilde{\mathbf{f}}) \perp \mathbf{b}_m \quad \text{pre} \quad m = 0, 1, \ldots M - 1$$

$$\langle \mathbf{f} - \tilde{\mathbf{f}}, \mathbf{b}_m \rangle = 0$$

$$\langle \mathbf{f}, \mathbf{b}_m \rangle = \langle \tilde{\mathbf{f}}, \mathbf{b}_m \rangle$$

$$\langle \mathbf{f}, \mathbf{b}_m \rangle = \left\langle \sum_{n=0}^{M-1} c_n \cdot \mathbf{b}_n, \mathbf{b}_m \right\rangle$$

$$\langle \mathbf{f}, \mathbf{b}_m \rangle = \sum_{n=0}^{M-1} c_n \cdot \langle \mathbf{b}_n, \mathbf{b}_m \rangle \quad \text{pre} \quad m = 0, 1, \ldots M - 1$$

$$\sum_{k=0}^{N-1} f_k \varphi_m(t_k) = \sum_{n=0}^{M-1} c_n \cdot \sum_{k=0}^{N-1} \varphi_n(t_k) \varphi_m(t_k) \quad \text{pre} \quad m = 0, 1, \ldots M - 1$$

Príklad 0.9.1 *V časoch* $0, 1, 3, 4$ *nadobudol proces* $\mathbf{f}$ *hodnoty* $7, 3, 5, 7$. *Zistite, ktorá z nasledujúcich funkcií (modelov) najlepšie vyjadruje priebeh procesu a aká bude približná hodnota procesu v časoch* 2 *a* 5.

a) $\varphi(t_k) = c$

b) $\varphi(t_k) = c \cdot t_k$

c) $\varphi(t_k) = c \cdot t_k^2$

d) $\varphi(t_k) = c_0 + c_1 \cdot t_k$

e) $\varphi(t_k) = c_0 \cdot t_k + c_1 \cdot t_k^2$

f) $\varphi(t_k) = c_0 + c_1 \cdot \mathrm{e}^{t_k}$

Riešenie:

a) Pre model $\varphi(t_k) = c$, teda pre konštatný priebeh zvolíme bázu

$$\mathcal{B} = ((1, 1, 1, 1))$$

Proces $\mathbf{f}$ odhadneme procesom $\tilde{\mathbf{f}} = (c)_{\mathcal{B}} = (c, c, c, c)_{\mathcal{E}}$.
Predpokladáme teda, že odhad má tvar $\tilde{\mathbf{f}} = c \cdot \mathbf{b}$ a hodnotu c vypočítame úpravou vzťahu

$$\langle \mathbf{f} - \tilde{\mathbf{f}}, \mathbf{b} \rangle = 0$$

Využitím vlastností skalárneho súčinu dostaneme

$$\langle \mathbf{f}, \mathbf{b} \rangle = \langle \tilde{\mathbf{f}}, \mathbf{b} \rangle$$

$$\langle \mathbf{f}, \mathbf{b} \rangle = c \cdot \langle \mathbf{b}, \mathbf{b} \rangle$$

Preto platí

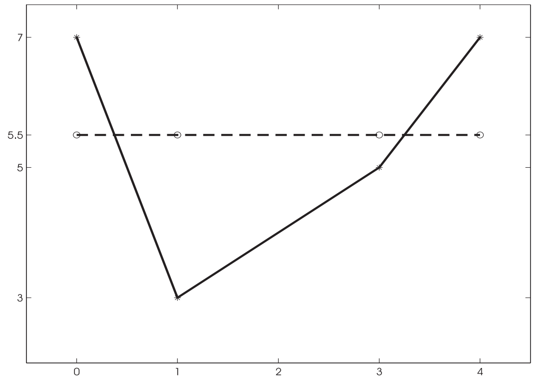
$$c = \frac{\langle \mathbf{f}, \mathbf{b} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle} = \frac{22}{4} = 5.5$$

Proces $\mathbf{f} = (7, 3, 5, 7)$ odhadneme (priblížime, aproximujeme) procesom $\tilde{\mathbf{f}} = (5.5, 5.5, 5.5, 5.5)$. Chyba tohoto odhadu (chybový vektor) je

$$\mathbf{e} = \mathbf{f} - \tilde{\mathbf{f}} = (1.5, -2.5, -0.5, 1.5)$$

Veľkosť tejto chyby je

$$\|\mathbf{e}\| = \|\mathbf{f} - \tilde{\mathbf{f}}\| = \sqrt{(1.5)^2 + (-2.5)^2 + (-0.5)^2 + (1.5)^2} = 3.3166$$



Obrázok 10: Aproximácia procesu priamkou $\varphi(t_k) = 5.5$

V čase 2 bude odhadnutá hodnota procesu rovná 5.5 (interpolácia).
V čase 5 bude odhadnutá hodnota procesu rovná 5.5 (extrapolácia).

$$\varphi(t_k) = c, \quad \varphi(2) = 5.5, \quad \varphi(5) = 5.5$$

b) Pre model $\varphi(t_k) = c \cdot t_k$, teda pre lineárny nárast zvolíme bázu, ktorá v časoch t_0, t_1, t_2, t_3 postupne nadobudne hodnoty $c \cdot t_0, c \cdot t_1, c \cdot t_2, c \cdot t_3$. Pre jednoduchosť zvolíme $c = 1$ a dostaneme bázu

$$\mathcal{B}_1 = ((0, 1, 3, 4))$$

Proces $\mathbf{f}$ odhadneme procesom $\tilde{\mathbf{f}}_{\mathcal{B}_1} = (c)$, $\tilde{\mathbf{f}}_{\mathcal{E}} = (0, c, 3 \cdot c, 4 \cdot c)$.
Hodnotu c vypočítame ako v predchádzajúcej časti zo vzťahu

$$\langle \mathbf{f}, \mathbf{b} \rangle = c \cdot \langle \mathbf{b}, \mathbf{b} \rangle$$

Preto platí

$$c = \frac{\langle \mathbf{f}, \mathbf{b} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle} = \frac{46}{26} = 1.77$$

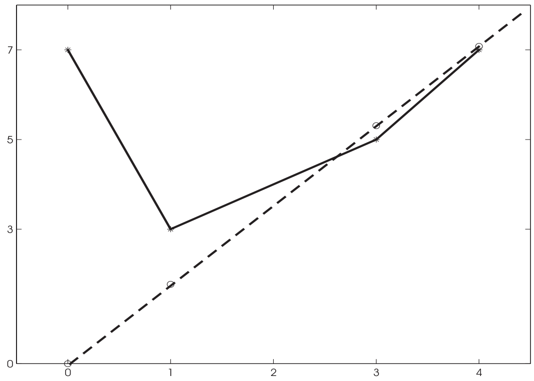
Proces $\mathbf{f} = (7, 3, 5, 7)$ odhadneme (priblížime, aproximujeme) procesom $\tilde{\mathbf{f}}_\mathbf{1} = (0, 1.77, 5.31, 7.08)$. Chyba tohoto odhadu (chybový vektor) je

$$\mathbf{e_1} = \mathbf{f} - \tilde{\mathbf{f}}_\mathbf{1} = (7, 1.23, -0.31, -0.08)$$

Veľkosť tejto chyby je

$$\|\mathbf{e}_1\| = \|\mathbf{f} - \tilde{\mathbf{f}}_1\| = \sqrt{(7)^2 + (1.23)^2 + (-0.31)^2 + (-0.08)^2} = 7.11$$

Veľkosť chyby ukazuje, že táto aproximácia je horšia ako v predchádzajúcom modeli. Dôvodom, že chyba je veľká je, že v tomto modeli musí regresná priamka $c \cdot t_k$ prechádzať bodom $(0,0)$, ktorý predstavuje v čase 0 hodnotu procesu 0. Tento bod je ďaleko od skutočnej hodnoty (=7) procesu v bode 0. Príslušné regresné priamky sú na obrázkoch 10 a 11. V



Obrázok 11: Aproximácia procesu priamkou $\varphi(t_k) = 1.77t_k$

časoch 2 a 5 budú odhadnuté hodnoty procesu rovné 3.54 a 8.85.

$$\varphi(t_k) = c \cdot t_k, \quad \varphi(2) = 1.77 \cdot 2 = 3.54, \quad \varphi(5) = 1.77 \cdot 5 = 8.85$$

c) Pre model $\varphi(t_k) = c \cdot t_k^2$, teda pre kvadratický priebeh zvolíme bázu, ktorá v časoch t_0, t_1, t_2, t_3 postupne nadobudne hodnoty $c \cdot t_0^2, c \cdot t_1^2, c \cdot t_2^2, c \cdot t_3^2$. Pre jednoduchosť zvolíme $c = 1$ a dostaneme bázu

$$\mathcal{B}_2 = ((0, 1, 9, 16))$$

Proces $\mathbf{f}$ odhadneme procesom $\tilde{\mathbf{f}}_{\mathcal{B}_2} = (c)$, $\tilde{\mathbf{f}}_{\mathcal{E}} = (0, c, 9 \cdot c, 16 \cdot c)$. Hodnotu c vypočítame ako

$$c = \frac{\langle \mathbf{f}, \mathbf{b} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle} = 0.47$$

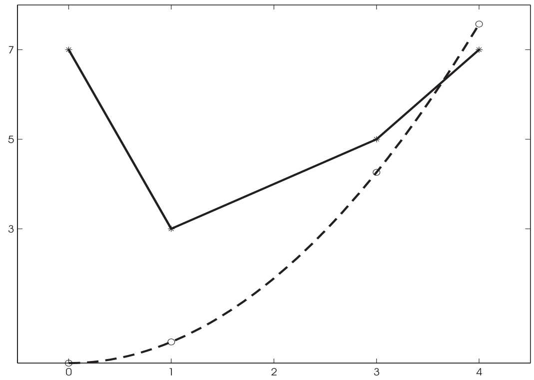
Proces $\mathbf{f} = (7, 3, 5, 7)$ odhadneme procesom $\tilde{\mathbf{f}}_\mathbf{2} = (0, 0.47, 4.26, 7.57)$. Chyba tohoto odhadu je

$$\mathbf{e_2} = \mathbf{f} - \tilde{\mathbf{f}}_\mathbf{2} = (7, 2.53, 0.74, -0.57)$$

Veľkosť tejto chyby je

$$\|\mathbf{e}_2\| = \|\mathbf{f} - \tilde{\mathbf{f}}_2\| = \sqrt{(7)^2 + (2.53)^2 + (0.74)^2 + (-0.57)^2} = 7.5007$$

Veľkosť chyby ukazuje, že táto aproximácia je lepšia ako v predchádzajúcom modeli. Chyba je ale stále veľká, pretože v tomto modeli musí regresná priamka $c \cdot t_k$ prechádzať bodom $(0, 0)$, ktorý predstavuje v čase 0 hodnotu procesu 0. Tento bod je ďaleko od skutočnej hodnoty (=7) procesu v bode 0. Regresná krivka pre tento model (parabola) je na obrázku 12.



Obrázok 12: Aproximácia procesu regresnou krivkou $\varphi(t_k) = 0.4734 \cdot t_k^2$

V časoch 2 a 5 budú odhadnuté hodnoty procesu rovné 1.89 a 11.83.

$$\varphi(t_k) = c \cdot t_k^2, \quad \varphi(2) = 0.47 \cdot 2^2 = 1.89, \quad \varphi(5) = 0.47 \cdot 5^2 = 11.83$$

d) Pre model $\varphi(t_k) = c_0 + c_1 \cdot t_k$, teda pre lineárny nárast s konštantnou zložkou zvolíme bázu, ktorá v časoch t_0, t_1, t_2, t_3 postupne nadobudne hodnoty $c_0 + c_1 \cdot t_0, c_0 + c_1 \cdot t_1, c_0 + c_1 \cdot t_2, c_0 + c_1 \cdot t_3$. Pre jednoduchosť zvolíme najprv $c_0 = 1, \quad c_1 = 0$, potom $c_0 = 0, \quad c_1 = 1$ a dostaneme bázu

$$\mathcal{B}_3 = ((1, 1, 1, 1), (0, 1, 3, 4))$$

Proces $\mathbf{f}$ odhadneme procesom

$$\tilde{\mathbf{f}}_{\mathcal{B}_3} = (c_0, c_1)$$
$$\tilde{\mathbf{f}}_{\mathcal{E}} = (c_0, c_0, c_0, c_0) + (0, c_1, 3 \cdot c_1, 4 \cdot c_1)$$

Teda $\tilde{\mathbf{f}} = c_0 \cdot \mathbf{b_0} + c_1 \cdot \mathbf{b_1}$. Hodnoty c_0 a c_1 vypočítame úpravou vzťahov

$$\langle \mathbf{f} - \tilde{\mathbf{f}}, \mathbf{b_0} \rangle = 0$$
$$\langle \mathbf{f} - \tilde{\mathbf{f}}, \mathbf{b_1} \rangle = 0$$

$$\langle \mathbf{f}, \mathbf{b_0} \rangle = \langle c_0 \cdot \mathbf{b_0} + c_1 \cdot \mathbf{b_1}, \mathbf{b_0} \rangle$$
$$\langle \mathbf{f}, \mathbf{b_1} \rangle = \langle c_0 \cdot \mathbf{b_0} + c_1 \cdot \mathbf{b_1}, \mathbf{b_1} \rangle$$

Preto sú c_0 a c_1 riešeniami sústavy

$$\langle \mathbf{f}, \mathbf{b_0} \rangle = c_0 \cdot \langle \mathbf{b_0}, \mathbf{b_0} \rangle + c_1 \cdot \langle \mathbf{b_1}, \mathbf{b_0} \rangle$$

$$\langle \mathbf{f}, \mathbf{b_1} \rangle = c_0 \cdot \langle \mathbf{b_0}, \mathbf{b_1} \rangle + c_1 \cdot \langle \mathbf{b_1}, \mathbf{b_1} \rangle$$

$$22 = 4c_0 + 8c_1, \quad 46 = 8c_0 + 26c_1$$

$$c_0 = 5.1, \quad c_1 = 0.2$$

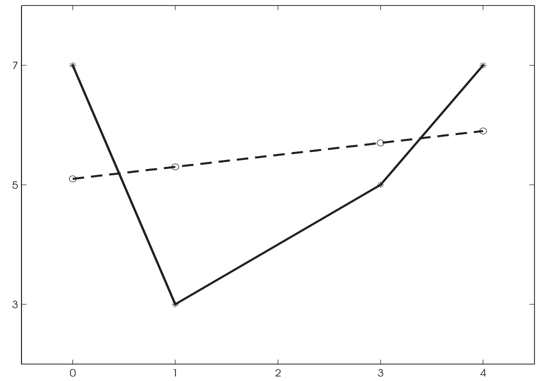
Proces $\mathbf{f} = (7, 3, 5, 7)$ odhadneme (priblížime, aproximujeme) procesom $\tilde{\mathbf{f}}_3 = (5.1, 5.3, 5.7, 5.9)$. Chyba tohoto odhadu (chybový vektor) je

$$\mathbf{e}_3 = \mathbf{f} - \tilde{\mathbf{f}}_3 = (1.9, -2.3, -0.7, 1.1)$$

Veľkosť tejto chyby je

$$\|\mathbf{e}_3\| = \|\mathbf{f} - \tilde{\mathbf{f}}_3\| = 3.2558$$

Toto je zatiaľ najlepšia aproximácia. Príslušná regresná priamka je na obrázku 13.



Obrázok 13: Aproximácia procesu priamkou $\varphi(t_k) = 5.1 + 0.2 \cdot t_k$

V časoch 2 a 5 budú odhadnuté hodnoty procesu rovné 5.5 a 6.1.

$$\varphi(t_k) = c_0 + c_1 \cdot t_k, \quad \varphi(2) = 5.5, \quad \varphi(5) = 6.1$$

e) Pre model $\varphi(t_k) = c_0 \cdot t_k + c_1 \cdot t_k^2$ zvolíme bázu, ktorá v časoch t_0, t_1, t_2, t_3 postupne nadobudne hodnoty $c_0 \cdot t_0 + c_1 \cdot t_0^2, c_0 \cdot t_1 + c_1 \cdot t_1^2, c_0 \cdot t_2 + c_1 \cdot t_2^2, c_0 \cdot t_3 + c_1 \cdot t_3^2$. Pre jednoduchosť zvolíme najprv $c_0 = 1, \quad c_1 = 0$, potom $c_0 = 0, \quad c_1 = 1$ a dostaneme bázu

$$\mathcal{B}_4 = ((0, 1, 3, 4), (0, 1, 9, 16))$$

Proces $\mathbf{f}$ odhadneme procesom

$$\tilde{\mathbf{f}}_{\mathcal{B}_4} = (c_0, c_1)$$
$$\tilde{\mathbf{f}}_{\mathcal{E}} = (0, c_0, 3 \cdot c_0, 4 \cdot c_0) + (0, c_1, 9 \cdot c_1, 16 \cdot c_1)$$

Teda $\tilde{\mathbf{f}}_4 = c_0 \cdot \mathbf{b_0} + c_1 \cdot \mathbf{b_1}$. Hodnoty c_0 a c_1 vypočítame riešením sústavy

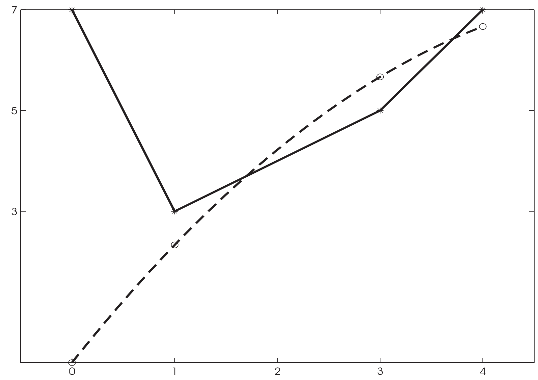
$$46 = 26c_0 + 92c_1, \quad 160 = 92c_0 + 338c_1$$

$$c_0 = 2.56, \quad c_1 = -0.22$$

Proces $\mathbf{f} = (7, 3, 5, 7)$ aproximujeme procesom $\tilde{\mathbf{f}}_\mathbf{4} = (0, 2.33, 5.67, 6.67)$.
Chyba tohoto odhadu (chybový vektor) je

$$\mathbf{e}_4 = \mathbf{f} - \tilde{\mathbf{f}}_4 = (7, 0.67, -0.67, 0.33)$$

Veľkosť tejto chyby je $\|\mathbf{e}_4\| = \|\mathbf{f} - \tilde{\mathbf{f}}_4\| = 7.0711$.
Príslušná regresná priamka je na obrázku 14.



Obrázok 14: Aproximácia procesu krivkou $\varphi(t_k) = 2.56 \cdot t_k - 0.22 \cdot t_k^2$

V časoch 2 a 5 budú odhadnuté hodnoty procesu rovné 4.22 a 7.22.

$$\varphi(t_k) = c_0 \cdot t_k + c_1 \cdot t_k^2, \quad \varphi(2) = 2.56 \cdot 2 - 0.22 \cdot 2^2 = 4.22,$$

$$\varphi(5) = 2.56 \cdot 5 - 0.22 \cdot 5^2 = 7.22$$

f) Pre model $\varphi(t_k) = c_0 + c_1 \, \mathrm{e}^{t_k}$ zvolíme bázu, ktorá v časoch t_0, t_1, t_2, t_3 postupne nadobudne hodnoty $c_0 + c_1 \cdot \mathrm{e}^{t_0}, c_0 + c_1 \cdot \mathrm{e}^{t_1}, c_0 + c_1 \cdot \mathrm{e}^{t_2}, c_0 + c_1 \cdot \mathrm{e}^{t_3}$. Pre jednoduchosť zvolíme najprv $c_0 = 1, \quad c_1 = 0$, potom $c_0 = 0, \quad c_1 = 1$ a dostaneme bázu

$$\mathcal{B}_5 = ((1, 1, 1, 1), (1, 2.72, 20.09, 54.6))$$

Proces $\mathbf{f}$ odhadneme procesom

$$\tilde{\mathbf{f}}_{\mathcal{B}_5} = (c_0, c_1)$$
$$\tilde{\mathbf{f}}_{\mathcal{E}} = (c_0, c_0, c_0, c_0) + (\mathrm{e}^0 \cdot c_1, \mathrm{e}^1 \cdot c_1, \mathrm{e}^3 \cdot c_1, \mathrm{e}^4 \cdot c_1)$$

Teda $\tilde{\mathbf{f}}_5 = c_0 \cdot \mathbf{b_0} + c_1 \cdot \mathbf{b_1}$. Hodnoty c_0 a c_1 vypočítame riešením sústavy

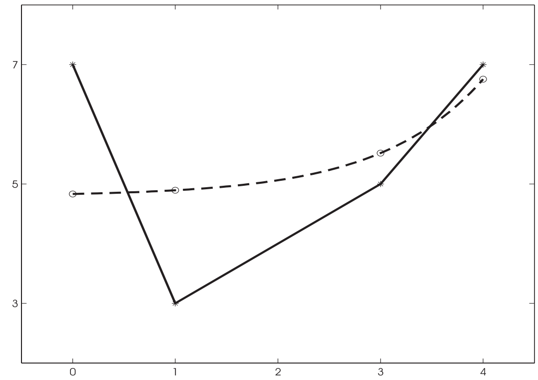
$$22 = 4c_0 + 78.4c_1, \quad 497.77 = 78.4c_0 + 3392.8c_1$$

$$c_0 = 4.8, \quad c_1 = 0.04$$

Proces $\mathbf{f} = (7, 3, 5, 7)$ aproximujeme procesom $\tilde{\mathbf{f}}_5 = (2.17, -1.9, -0.52, 0.25)$. Chyba tohoto odhadu (chybový vektor) je

$$\mathbf{e}_5 = \mathbf{f} - \tilde{\mathbf{f}}_5 = (7, 0.67, -0.67, 0.33)$$

Veľkosť tejto chyby je $\|\mathbf{e}_5\| = \|\mathbf{f} - \tilde{\mathbf{f}}_5\| = 2.94$.
Príslušná regresná priamka je na obrázku 15.



Obrázok 15: Aproximácia procesu krivkou $\varphi(t_k) = 4.8 - 0.04\,\mathrm{e}^{t_k}$

V časoch 2 a 5 budú odhadnuté hodnoty procesu rovné 5.06 a 10.11.
$\varphi(2) = 4.8 + 0.04 \cdot \mathrm{e}^2 = 5.06, \quad \varphi(5) = 4.8 + 0.04 \cdot \mathrm{e}^5 = 10.11$ □

Najlepší odhad spomedzi uvedených modelov dáva funkcia $\varphi(t_k) = c_0 + c_1\,\mathrm{e}^{t_k}$. Veľkosť chyby v tomto prípade je 2.94.

Úloha 0.9.1 *Navrhnite funkciu s dvomi parametrami c_0 a c_1, ktorá odhadne proces z príkladu 0.9.1 lepšie než modely, ktoré sú v príklade navrhnuté.*

0.10 Koeficienty v neortogonálnej báze

Koeficienty priemetu procesu do bázy sú rôzne v závislosti od bázy. Pre jeden vektor dostávame rôzne jeho vyjadrenia v rôznych bázach.

$$\tilde{\mathbf{f}} = c_0\mathbf{b}_0 + c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n = c_0'\mathbf{b}_0' + c_1'\mathbf{b}_1' + \cdots + c_n'\mathbf{b}_n'$$

Ak bázové vektory predstavujú časové vyjadrenie funkcie, ktorou je proces modelovaný, je možné vypočítané koeficienty použiť na vyjadrenie funkcie, ktorou proces aproximujeme:

$$\varphi(t_k) = c_0\varphi_0(t_k) + c_1\varphi_1(t_k) + \cdots + c_n\varphi_n(t_k) \qquad (0.10.1)$$

Výpočet koeficientov c_k vyžaduje riešiť sústavu rovníc. V prípade, že bázu najprv ortogonalizujeme, stačí na výpočet každého z koeficientov jeden vzorec.

$$c_k' = \frac{\langle \mathbf{f}, \mathbf{b}_k' \rangle}{\langle \mathbf{b}_k', \mathbf{b}_k' \rangle}$$

Preto sa niekedy neortogonálna báza (vyjadrujúca časový priebeh) najskôr ortogonalizuje. Výpočet koeficientov ortogonálnej bázy bude jednoduchší, ale koeficienty c'_k sa nedajú použiť ako koeficienty funkcie $\varphi(t_k)$. V nasledujúcom príklade ukážeme, ako sa zo známych koeficientov ortogonálnej bázy dajú vypočítať koeficienty pôvodnej neortogonálnej bázy.

Príklad 0.10.1 *Aproximujte proces z príkladu* 0.6.1*, resp. príkladu* 0.7.1*, procesom s lineárnym nárastom s dvomi parametrami. Bázu najprv ortogonalizujte, potom vypočítajte koeficienty v ortogonálnej báze a nakoniec z týchto koeficientov odvoďte koeficienty funkcie*

$$\varphi(t_k) = c_0 + c_1 \cdot t_k$$

Riešenie:
Lineárny nárast procesu vyjadríme pomocou dvoch koeficientov ako

$$\tilde{\mathbf{f}} = c_0 \cdot \mathbf{b}_0 + c_1 \cdot \mathbf{b}_1 = c_0 \cdot (1, 1, \ldots, 1) + c_1 \cdot (0, 1, 2, \ldots, 99)$$

Zodpovedajúca neortogonálna báza je

$$\mathcal{B} = (\mathbf{b}_0, \mathbf{b}_1) = ((1, 1, \ldots, 1), (0, 1, 2, \ldots, 99))$$

Ortogonálnu bázu $\mathcal{B}'$ dostaneme Gram-Schmidtovou ortogonalizáciou bázy $\mathcal{B}$.

$$\mathbf{b}'_0 = \mathbf{b}_0$$

$$\mathbf{b}'_1 = \mathbf{b}_1 - \tilde{\mathbf{b}}_1 = \mathbf{b}_1 - k \cdot \mathbf{b}_0 = \mathbf{b}_1 - 49.5 \cdot \mathbf{b}_0$$

Koeficienty c'_0, c'_1 v báze $\mathcal{B}'$ vypočítame zo vzorcov

$$c'_0 = \frac{\langle \mathbf{f}, \mathbf{b}'_0 \rangle}{\langle \mathbf{b}'_0, \mathbf{b}'_0 \rangle} = 2647.7$$

$$c'_1 = \frac{\langle \mathbf{f}, \mathbf{b}'_1 \rangle}{\langle \mathbf{b}'_1 \cdot \mathbf{b}'_1 \rangle} = 36.06$$

Koeficienty c_0 a c_1 dostaneme zo vzťahu

$$f = c_0 \cdot \mathbf{b}_0 + c_1 \cdot \mathbf{b}_1 = c'_0 \cdot \mathbf{b}'_0 + c'_1 \cdot \mathbf{b}'_1 = 2647.7 \cdot \mathbf{b}_0 + 36.06 \cdot \langle \mathbf{b}_1 - 49.5 \cdot \mathbf{b}_0 \rangle =$$

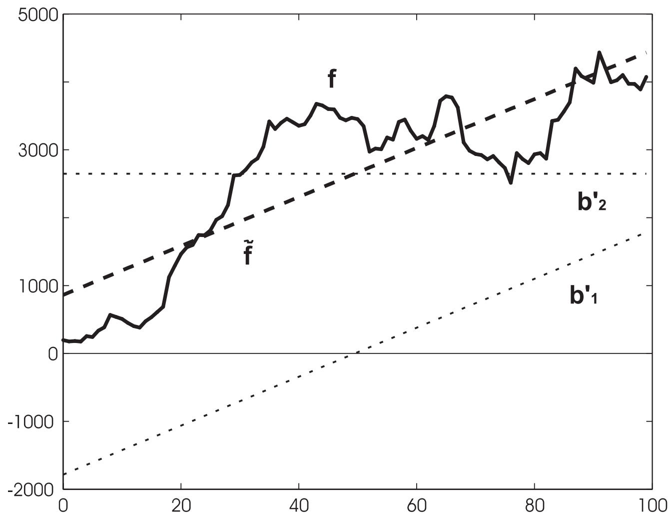
$$= (2647.7 - 1.7851) \cdot \mathbf{b}_0 + 36.06 \cdot \mathbf{b}_1 = 862.61 \cdot \mathbf{b}_0 + 36.06 \cdot \mathbf{b}_1$$

Odtiaľ dostávame $c_0 = 862.61$ a $c_1 = 36.06$ a teda vyjadrenie procesu ako funkcie času bude

$$\varphi(t_k) = 862.61 + 36.06 \cdot t_k$$

Výslednú aproximáciu vidíme na obrázku 16. Rovnaký výsledok sme dostali aj v príklade 0.7.1, keď sme koeficienty c_0 a c_1 hľadali riešením sústavy rovníc.

□



Obrázok 16: Aproximácia dát priamkou $\varphi(t_k) = 862.61 + 36.06 \cdot t_k$, vektor $\mathbf{b}_0'$ reprezentuje konštantný trend a vektor $\mathbf{b}_1'$ reprezentuje lineárny nárast

Príklad 0.10.2 *Nech je proces* $\mathbf{f}$ *aproximovaný pomocou časovej funkcie*

$$\varphi(t_k) = c_0\varphi_0(t_k) + c_1\varphi_1(t_k) + c_2\varphi_2(t_k)$$

Jednotlivé funkcie $\varphi_0(t_k), \varphi_1(t_k), \varphi_2(t_k)$ *sú vyjadrené postupne procesmi pôvodnej bázy*

$$\mathcal{B} = (\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2)$$

Vyjadrenie v pôvodnej báze je

$$\tilde{\mathbf{f}} = c_0\mathbf{b}_0 + c_1\mathbf{b}_1 + c_2\mathbf{b}_2$$

Z bázy $\mathcal{B}$ *je Gram-Schmitovou ortogonalizáciou vytvorená ortogonálna báza:*

$$\mathcal{B}' = (\mathbf{b}'_0, \mathbf{b}'_1, \mathbf{b}'_2)$$

Vyjadrite zo známych koeficientov c_0', c_1', c_2' *procesu v ortogonálnej báze, koeficienty v pôvodnej, neortogonálnej báze!*

Riešenie:
Priemet procesu $\mathbf{f}$ do podpriestoru s ortogonálnou bázou vyjadruje vzťah:

$$\tilde{\mathbf{f}} = c_0'\mathbf{b}_0' + c_1'\mathbf{b}_1' + c_2'\mathbf{b}_2'$$

Vyjadríme vektory ortogonálnej bázy pomocou vektorov pôvodnej bázy.

Potom pôvodné bázické vektory nahradíme jednotlivými funkciami, čím dostaneme funkciu, ktorá aproximuje proces.

Postup ako vyjadriť vektory ortogonálnej bázy pomocou vektorov pôvodnej bázy je možné vyvodiť z postupu ortogonalizácie.

Je zrejmé, že $\mathbf{b}_0 = \mathbf{b}'_0$.
V postupe ortogonalizácie sme použili

$$\mathbf{b}'_1 = \mathbf{b}_1 - k \cdot \mathbf{b}'_0 = \mathbf{b}_1 - k \cdot \mathbf{b}_0\,, \quad \text{kde} \quad k = \frac{\langle \mathbf{b}_1, \mathbf{b}_0 \rangle}{\langle \mathbf{b}_0, \mathbf{b}_0 \rangle}$$

$$\mathbf{b}'_2 = \mathbf{b}_2 - k_0 \cdot \mathbf{b}_0 - k_1 \mathbf{b}'_1\,, \,\text{kde}\, k_0 = \frac{\langle \mathbf{b}_2, \mathbf{b}_0 \rangle}{\langle \mathbf{b}_0, \mathbf{b}_0 \rangle} \,\text{a}\, k_1 = \frac{\langle \mathbf{b}_2, \mathbf{b}'_1 \rangle}{\langle \mathbf{b}'_1, \mathbf{b}'_1 \rangle}$$

Vo vzťahu pre $\mathbf{b}'_2$ je síce použité $\mathbf{b}'_1$, ktoré je ale vyjadrené vyššie.
Čiže

$$\mathbf{b}'_2 = \mathbf{b}_2 - k_0 \cdot \mathbf{b}_0 - k_1(\mathbf{b}_1 - k\mathbf{b}_0) = \mathbf{b}_2 - k_1\mathbf{b}_1 - (k_0 - k_1 \cdot k)\mathbf{b}_0$$

Napríklad funkcia

$$y = a \cdot t_k + b \cdot t_k^2 + c \cdot t_k^3$$

by pomocou koeficientov c'_0, c'_1, c'_2 vypočítaných v ortogonálnej báze vyzerala nasledovne

$$f(t_k) = c'_0 \cdot t_k + c'_1 \cdot (t_k^2 - kt_k) + c'_2 \cdot (t_k^3 - k_1 \cdot t_k^2 - (k_0 - k_1 k) \cdot t_k)$$

$\square$

Príklad 0.10.3 *Je daný proces* $\mathbf{f} = (2, 4, 5, 13)$. *Vytvorte pre funkciu* $\varphi(t_k) = c_0 + c_1 \cdot t_k + c_2 \ln t_k$ *bázu* $\mathcal{B}$ *zodpovedajúcu časom* $\mathbf{t} = (1, 2, 3, 4)$. *Ortogonalizujte pomocou Gram-Schmitovej metódy bázu* $\mathcal{B}$. *Vyjadrite koeficienty* c'_0, c'_1, c'_2 *procesu* $\mathbf{f} = (2, 4, 5, 13)$ *v ortogonálnej báze* $\mathcal{B}'$. *Pomocou koeficientov* c'_0, c'_1, c'_2 *vyjadrite koeficienty* c_0, c_1, c_2*, ktoré je možné použiť na vyjadrenie funkcie* $\varphi(t_k) = c_0 + c_1 \cdot t_k + c_2 \ln t_k$.

Riešenie:

$$\mathcal{B} = ((1, 1, 1, 1), (1, 2, 3, 4), (0, 0.69, 1.1, 1.39))$$

$$\mathcal{B}' = (\mathbf{b}'_0, \mathbf{b}'_1, \mathbf{b}'_2),\ \mathbf{b}'_0 = (1, 1, 1, 1)$$

$$\mathbf{b}'_1 = \mathbf{b}_1 - k \cdot \mathbf{b}_0,\ \text{kde}\, k = \frac{\langle \mathbf{b}_1, \mathbf{b}'_0 \rangle}{\langle \mathbf{b}'_0, \mathbf{b}'_0 \rangle} = 2.5$$

$$\mathbf{b}'_1 = (1, 2, 3, 4) - 2.5 \cdot (1, 1, 1, 1) = (-1.5, -0.5, 0.5, 1.5)$$

$$\mathbf{b}'_2 = \mathbf{b}_2 - k_0 \cdot \mathbf{b}'_0 - k_1 \cdot \mathbf{b}'_1,\ \text{kde}\, k_0 = \frac{\langle \mathbf{b}_2, \mathbf{b}'_0 \rangle}{\langle \mathbf{b}'_0, \mathbf{b}'_0 \rangle} = 0.8 \,\text{a}\, k_1 = \frac{\langle \mathbf{b}_2, \mathbf{b}'_1 \rangle}{\langle \mathbf{b}'_1, \mathbf{b}'_1 \rangle} = 0.46$$

$$\mathbf{b}'_2 = (0, 0.69, 1.1, 1.39) - 0.8 \cdot (1, 1, 1, 1) - 0.46 \cdot (-1.5, -0.5, 0.5, 1.5) =$$

$$= (-0.11, 0.13, 0.08, -0.09)$$

Koeficienty odhadu $\tilde{\mathbf{f}} = (2.3, 2.69, 6.74, 12.28)$ procesu $\mathbf{f} = (2, 4, 5, 13)$ v ortogonálnej báze $\mathcal{B}'$ sú

$$c'_0 = 6, c'_1 = 3.4, c'_2 = -12.7$$

Dosaďme do vyjadrenia procesu pôvodné bázové vektory.

$$\tilde{\mathbf{f}} = c_0' \cdot \mathbf{b}_0' + c_1' \cdot \mathbf{b}_1' + c_2' \cdot \mathbf{b}_2' = c_0' \cdot \mathbf{b}_0 + c_1' \cdot (\mathbf{b}_1 - k \cdot \mathbf{b}_0) + c_2' \cdot (\mathbf{b}_2 - k_0 \cdot \mathbf{b}_0 - k_1 \cdot (\mathbf{b}_1 - k \cdot \mathbf{b}_0) =$$

$$= (c_0' - c_1' \cdot k - c_2' \cdot k_0 + c_2' \cdot k_1 \cdot k)\mathbf{b}_0 + (c_1' - c_2' \cdot k_1)\mathbf{b}_1 + c_2' \cdot \mathbf{b}_2$$

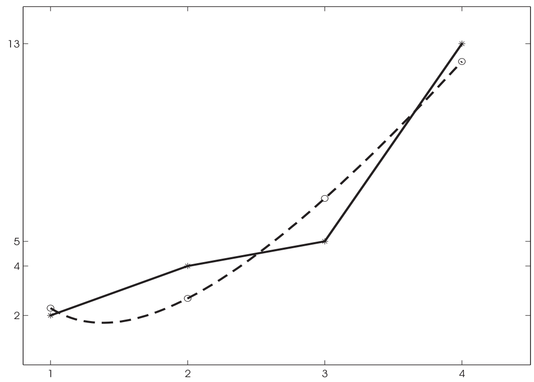
Platí zároveň

$$\tilde{\mathbf{f}} = c_0 \cdot \mathbf{b}_0 + c_1 \cdot \mathbf{b}_1 + c_2 \cdot \mathbf{b}_2$$

A odtiaľ dostávame

$$c_0 = c_0' - c_1' \cdot k - c_2' \cdot k_0 + c_2' \cdot k_1 \cdot k = -6.9, c_1 = c_1' - c_2' \cdot k_1 = 9.2, c_2 = c_2' = -12.7$$

Odhad procesu $\mathbf{f}$ je daný funkciou $\varphi(t_k) = c_0 + c_1 \cdot t_k + c_2 \ln t_k$ s koeficientami $c_0 = -6.9$, $c_1 = 9.2$, $c_2 = -12.7$. Regresná krivka pre odhad procesu je na obrázku 17. □



Obrázok 17: Aproximácia procesu krivkou $\varphi(t_k) = -6.9 + 9.2 \cdot t_k - 12.7 \ln t_k$

0.11 Linearizácia nelineárnych úloh

Metódou lineárnej regresie nevyriešime napríklad

$$\varphi(t_k) = t_k^{c_0} \ln(c_1 \cdot t_k + c_2)$$

lebo tento vzťah sa nedá vyjadriť ako lineárna kombinácia bázických vektorov s koeficientami c_0, c_1, c_2.
Iná je situácia napríklad pre

$$\varphi(t_k) = c_0 \cdot e^{-c_1 \cdot t_k}$$

ktorú vieme linearizovať tak, že výraz logaritmujeme

$$\ln \varphi(t_k) = (\ln c_0) - c_1 \cdot t_k$$

Substitúciou $\ln \varphi(t_k) = \varphi^\star(t_k)$ dostaneme $\varphi^\star(t_k) = -c_1 \cdot t_k$.
Teda ak sa dá $\varphi(t_k)$ prepísať na tvar

$$\varphi^\star(t_k) = \sum_{n=0}^{M-1} c_n \cdot \varphi_n(t_k)$$

môžeme pôvodnú úlohu riešiť ako úlohu lineárnej regresie.
Hovoríme, že nelineárnu úlohu môžeme **linearizovať**.

Poznámka:
Na prevod nelineárnej úlohy na úlohu lineárnej regresie sa môže použiť napríklad logaritmus, arcsin, prevrátená hodnota...

Príklad 0.11.1 *Nech sú v časoch* $\mathbf{t} = (1, 2, 3, 4)$ *namerané hodnoty procesu* $\mathbf{f} = (9, 5, 3, 2)$. *Nelineárnu úlohu aproximovať proces* $\mathbf{f}$ *pomocou funkcie*

$$\varphi(t_k) = \frac{c_0 \cdot t_k}{c_1 + t_k}$$

prerobte na úlohu lineárnej regresie.

Riešenie:
Použijeme prevrátenú hodnotu nelineárnej funkcie $\varphi(t_k)$ a dostaneme lineárnu funkciu:

$$\varphi^\star(t_k) = \frac{1}{\varphi(t_k)} = \frac{c_1 + t_k}{c_0 \cdot t_k} \tag{0.11.1}$$

Rovnakú úpravu urobíme s hodnotami procesu $\mathbf{f} = (9, 5, 3, 2)$ a dostaneme $\mathbf{f}^\star = (0.11, 0.2, 0.33, 0.5)$.
Úpravami vzťahu 0.11.1 dostaneme úlohu lineárnej regresie

$$\varphi^\star(t_k) = \frac{c_1}{c_0 \cdot t_k} + \frac{t_k}{c_0 \cdot t_k} = \frac{c_1}{c_0 \cdot t_k} + \frac{1}{c_0}$$

$$\tilde{\mathbf{f}}^\star = d_0 \cdot \mathbf{b}_0 + d_1 \cdot \mathbf{b}_1$$

kde

$$d_0 = \frac{c_1}{c_0}, \quad d_1 = \frac{1}{c_0}, \quad \mathbf{b}_0 = (\frac{1}{t_0}, \frac{1}{t_1}, \frac{1}{t_2}, \frac{1}{t_3}) = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}), \quad \mathbf{b}_1 = (1, 1, 1, 1)$$

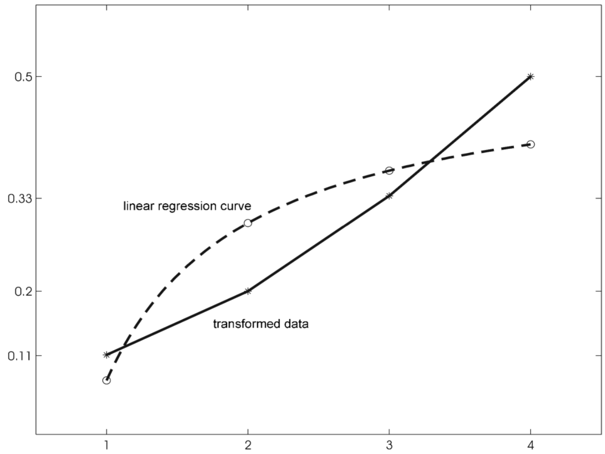
Koeficienty d_0, d_1 sú riešením sústavy rovníc:

$$\begin{aligned} \langle \mathbf{f}^\star, \mathbf{b}_0 \rangle &= d_0 \cdot \langle \mathbf{b}_0, \mathbf{b}_0 \rangle + d_1 \cdot \langle \mathbf{b}_1, \mathbf{b}_0 \rangle \\ \langle \mathbf{f}^\star, \mathbf{b}_1 \rangle &= d_0 \cdot \langle \mathbf{b}_0, \mathbf{b}_1 \rangle + d_1 \cdot \langle \mathbf{b}_1, \mathbf{b}_1 \rangle \end{aligned}$$

$$\begin{aligned} 0.45 &= d_0 \cdot 1.42 + d_1 \cdot 2.08 \\ 1.14 &= d_0 \cdot 2.08 + d_1 \cdot 4 \end{aligned}$$

$$\tilde{\mathbf{f}}^\star = -0.44 \cdot \mathbf{b}_0 + 0.51 \cdot \mathbf{b}_1 = (0.08, 0.3, 0.37, 0.41)$$

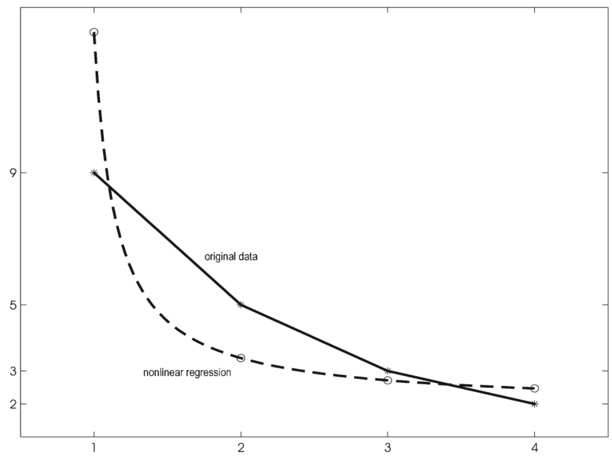
Treba si uvedomiť, že lineárnu regresiu sme robili v pokrivenej, transformovanej oblasti. Odhadovaný vektor $\mathbf{f}^\star$ aj jeho aproximácia $\tilde{\mathbf{f}}^\star$ sa nachádzajú v inom priestore ako pôvodné hodnoty procesu, viď obrázok 18.



Obrázok 18: Aproximácia procesu regresnou krivkou $\varphi^\star(t_k) = \frac{d_0}{t_k} + d_1$

Odhad pôvodného procesu dostaneme po spätnej transformácii procesu $\tilde{\mathbf{f}}^\star$ na proces $\tilde{\mathbf{f}}$, viď obrázok 19.

$$\tilde{\mathbf{f}} = \frac{1}{\tilde{\mathbf{f}}^\star} = (13.26, 3.39, 2.71, 2.47)$$



Obrázok 19: Aproximácia procesu krivkou $\varphi(t_k) = \frac{c_0 \cdot t_k}{c_1 + t_k}$

□

Poznámka:
Je potrebné si uvedomiť, že po linearizácii úlohy už nehľadáme najbližší vektor v metrike pôvodného priestoru. Veľkosť chyby aproximácie bude minimálna iba

v zmenenej metrike nového priestoru (po transformácii) a nebude vyjadrená v hodnotách pôvodného priestoru.

Úloha 0.11.1 *V príklade* 0.11.1 *je minimalizovaná veľkosť chyby aproximácie* $\tilde{\mathbf{f}}^\star$ *procesu* $\mathbf{f}^\star$. *Veľkosť chyby v transformovanom priestore je*

$$\|\mathbf{e}^\star\| = \|\mathbf{f}^\star - \tilde{\mathbf{f}}^\star\| = 0.14$$

Veľkosť chyby nelineárnej aproximácie $\tilde{\mathbf{f}}$ *procesu* $\mathbf{f}$ *v pôvodnom priestore je*

$$\|\mathbf{e}\| = \|\mathbf{f} - \tilde{\mathbf{f}}\| = 4.58$$

Zistite, či veľkosť chyby v pôvodnom priestore je minimálna, alebo či sa dajú nájsť koeficienty c_0 *a* c_1 *také, že chyba aproximácie procesu krivkou*

$$\varphi(t_k) = \frac{c_0 \cdot t_k}{c_1 + t_k}$$

bude menšia ako 4.58.