Alkalmazott Matematikai Lapok **34** (2018), 1–8.

SZIMMETRIA ÉS KONFIDENCIA

CSÁJI BALÁZS CSANÁD

A cikkben áttekintjük az SPS (Sign-Perturbed Sums) becslési módszert, amely a zaj szimmetriáját kihasználva egzakt, nem-aszimptotikus konfidencia tartományokat tud konstruálni regressziós modellek pontbecslései köré.

1. Bevezetés

Regressziós modellek becslése zajos megfigyelési adatokból (pl., idősorokból) fontos statisztikai probléma, amely számos területen előkerül, például a rendszer identifikációban, jelfeldolgozásban, gépi tanulásban és a pénzügyi matematikában.

A standard módszerek – például az előrejelzési hiba-, maximum likelihood- és korrelációs módszer [5] – tipikusan *pontbecsléseket* szolgáltatnak. Egy gyakorlati szempontból is fontos kérdés, hogy mennyire bízhatunk a kapott becslésben, amelyre egy lehetséges válasz, hogy megadott valószínűségű *konfidencia tartományt* építünk a becslés köré. A klasszikus módja ilyen halmazok konstrukciójának, ha a pontbecslés határeloszlását hívjuk segítségül [5]. Az aszimptotikus eredményeken alapuló megközelítések azonban nem nyújtanak szigorú garanciákat véges mintaszám esetén, hacsak nem teszünk erős statisztikai feltevéseket a rendszerről.

A dolgozat célja a nemrég kidolgozott SPS (Sign-Perturbed Sums) módszer [3] bemutatása, amely minimális statisztikai feltevésekkel képes regressziós modellek pontbecslései köré egzakt, nem-aszimptotikus konfidencia halmazokat építeni.

Az SPS egy *félparametrikus* becslési módszer, mivel egy parametrikus (akár dinamikus) modellből indul ki, de a folyamatot hajtó zajra nézve nem tételez fel paraméterezést. Először az egyszerű lineáris regresszió esetére mutatjuk be a módszert, majd kiterjesztjük dinamikus rendszerekre is a konstrukciót.

2. Konfidencia halmazok lineáris regressziós problémákhoz

Adott egy bemenet-kiment párokból álló minta, $\mathcal{D}_n \triangleq \{(\varphi_1, Y_1), \dots, (\varphi_n, Y_n)\}$,

$$Y_t \triangleq \varphi_t^{\mathrm{T}} \theta^* + N_t, \tag{1}$$

$t \in \{1, \dots, n\}$, ahol Y_t a kimenet (magyarázott változó), φ_t a bemenet (magyarázó változó, regresszor) és N_t pedig a (nem-megfigyelhető) zaj, a t-edik megfigyelés esetén. A cél az ismeretlen „igazi” θ^* paraméter becslése. Feltesszük, hogy a bemenetek, $\{\varphi_t\} \subset \mathbb{R}^d$, determinisztikusak, $\theta^* \in \mathbb{R}^d$ konstans, és a zaj, $\{N_t\}$, független, szimmetrikus (a nulla körül) eloszlású[1] (valós értékű) valószínűségi változókból áll; azaz N_t és $-N_t$ eloszlása megegyezik, minden t-re. Az egyszerűség kedvéért szintén feltesszük, hogy $n > d$ és a $\Phi \triangleq [\,\varphi_1, \dots, \varphi_n\,]$ mátrix teljes (sor) rangú.

2.1. Legkisebb négyzetek becslés és konfidencia ellipszoidjai

Az egyik standard pontbecslés a jól-ismert *legkisebb négyzetek* (LS) becslés,

$$\hat{\theta}_n \;\triangleq\; \operatorname*{arg\,min}_{\theta\in\mathbb{R}^d} \mathcal{V}(\theta \,|\, \mathcal{D}_n) \;=\; \operatorname*{arg\,min}_{\theta\in\mathbb{R}^d} \frac{1}{2}\|Y - \Phi^{\mathrm{T}}\theta\|_2^2, \tag{2}$$

ahol $Y \triangleq [Y_1, \dots, Y_n]^{\mathrm{T}}$. A *normál egyenlet* megoldásával megkapható $\hat{\theta}_n$, azaz

$$\nabla_\theta \mathcal{V}(\hat{\theta}_n \,|\, \mathcal{D}_n) \;=\; \Phi\Phi^{\mathrm{T}}\hat{\theta}_n - \Phi Y \;=\; 0, \tag{3}$$

aminek a fenti feltételek mellett az analitikus megoldása $\hat{\theta}_n = (\Phi\Phi^{\mathrm{T}})^{-1}(\Phi Y)$.

Egy fontos kérdés, hogy mennyire megbízható a kapott pontbecslés? Erre egy válasz, ha tudunk például egy megadott $p \in (0,1)$ valószínűségű $\widehat{\Theta}_{\mathcal{D}_n,p}$ konfidencia tartományt készíteni $\hat{\theta}_n$ köré, azaz amelyre $\hat{\theta}_n \in \widehat{\Theta}_{\mathcal{D}_n,p}$ és $\mathbb{P}\big(\theta^* \in \widehat{\Theta}_{\mathcal{D}_n,p}\big) \geq p$.

Ilyen halmazok konstrukciójára egy standard módszer [5], ha kihasználjuk, hogy az LS becslés (átskálázott) hibája aszimptotikusan Gauss eloszlású, azaz

$$\sqrt{n}\,(\hat{\theta}_n - \theta^*) \xrightarrow{\;d\;} \mathcal{N}(0, \sigma^2 R^{-1}), \qquad \text{ahogy} \qquad n \to \infty, \tag{4}$$

amely fennáll pl., ha korlátosak a regresszorok és létezik egy olyan pozitív definit R mátrix, amely az $R_n \triangleq \frac{1}{n}\Phi_n\Phi_n^{\mathrm{T}}$ határértéke[2], valamint $\{N_t\}$ független, azonos eloszlású (f.a.e.) változókból áll, amelyekre $\mathbb{E}[N_t] = 0$ és $\mathbb{E}[N_t^2] = \sigma^2$, $0 < \sigma^2 < \infty$.

A pontbecslés határeloszlásának felhasználásával egy adott p valószínűséghez egy $\hat{\theta}_n$ középpontú (közelítő) konfidencia ellipszoid konstruálható,

$$\widetilde{\Theta}_{n,p} \;\triangleq\; \left\{\, \theta \in \mathbb{R}^d \,:\, (\theta - \hat{\theta}_n)^{\mathrm{T}} R_n\, (\theta - \hat{\theta}_n) \;\leq\; \frac{q\,\hat{\sigma}_n^2}{n} \,\right\}, \tag{5}$$

ahol $p = F_{\chi^2(d)}(q)$, itt $F_{\chi^2(d)}$ a d szabadságfokú χ^2 eloszlás eloszlásfüggvénye, és $\hat{\sigma}_n^2$ a zaj varianciájának becslése az LS megoldás reziduálisainak segítségével,

$$\hat{\sigma}_n^2 \;\triangleq\; \frac{1}{n-d}\sum_{t=1}^{n}(Y_t - \varphi_t^{\mathrm{T}}\hat{\theta}_n)^2. \tag{6}$$

Ekkor nyilván $\hat{\theta}_n \in \widetilde{\Theta}_{n,p}$, valamint $\mathbb{P}(\,\theta^* \in \widetilde{\Theta}_{n,p}\,) \;\approx\; p$. Az így kapott konfidencia halmazok azonban véges minták esetén nem garantáltak és csak heurisztikus megközelítésnek tekinthetőek; kis mintaszám esetén tipikusan pontatlanok [5].

[1] A függetlenség gyengíthető, a szimmetria a kritikus az SPS-hez. Számos nevezetes eloszlás lehet ilyen, például, Gauss, Laplace, Cauchy, Bernoulli, Student t, egyenletes.

[2] Itt kivételesen – a határérték miatt – expliciten kiírtuk, hogy Φ függ n-től.

2.2. SPS lineáris regresszió esetén

Most rátérünk az SPS módszer [3] ismertetésére, amellyel minimális statisztikai feltevésekkel véges minták esetén is garantált konfidencia tartományokat konstruálhatunk az LS becslés köré. Első közelítésben tekinthetünk az SPS-re úgy, mint egy hipotézisvizsgálatra, amely egy adott θ paraméter esetén azt a null hipotézist vizsgálja, hogy $\theta = \theta^*$, a $\theta \neq \theta^*$ alternatív hipotézissel szemben.

A *bootstrap-* és *Monte Carlo* tesztek alapgondolatához hasonlóan abból indulunk majd ki, hogy ha $\theta = \theta^*$, akkor egyrészt (i) megkaphatjuk a zaj változók pontos értékét a rendszer „invertálásával”, valamint (ii) a zaj eloszlásának bizonyos regularitását kihasználva (a jelen esetben a szimmetriát) alternatív mintákat generálhatunk, amelyek statisztikailag „hasonlóan viselkednek” majd, mint az eredeti minta. Ha azonban $\theta \neq \theta^*$, akkor a rendszer invertálásával kapott zajbecslések torzítottak lesznek és így az alternatív minták statisztikailag „eltérően viselkednek” majd, mint az eredeti minta. Azt, hogy mennyire „hasonlóan viselkednek” a generált (perturbált) értékek az eredeti mintához viszonyítva, a minták egy skalár értékű kiértékelése után egy rang-teszttel döntjük majd el.

Első lépésként bevezetünk m (véletlenített) kvadratikus $\mathbb{R}^d \to \mathbb{R}$ függvényt,

$$Z_i(\theta) \triangleq \|R_n^{-\frac{1}{2}} \Phi \Lambda_i \big(Y - \Phi^{\mathrm{T}} \theta\big)\|_2^2, \tag{7}$$

$i \in \{0, 1, \ldots, m-1\}$, ahol $\Lambda_0 \triangleq I$ és $\Lambda_i \triangleq \mathrm{diag}(\alpha_{i,1}, \ldots, \alpha_{i,n})$, ha $i \neq 0$; $\{\alpha_{i,t}\}$ független, azonos eloszlású Rademacher valószínűségű változók[3]; a $\mathrm{diag}(\cdot)$ függvény pedig egy diagonális mátrixot képez az argumentumaiból. A Z_0 -t *referencia függvénynek*, míg a többi $\{Z_i\}$ függvényt *előjel-perturbált* függvényeknek nevezzük.

Vegyük észre a kapcsolatot a $\{Z_i\}$ függvények definíciója és a költségfüggvény gradiense, $\nabla_\theta \mathcal{V}$, között. A $\{Z_i\}$ függvényeket interpretálhatjuk úgy, hogy $\mathcal{V}$ gradiensében a reziduálisok előjelét véletlenítjük Rademacher változók segítségével, majd a kapott vektor „nagyságát” kiértékeljük egy súlyozott norma segítségével.

Ha $\theta = \theta^*$, akkor $Y - \Phi^{\mathrm{T}} \theta^* = N$, ahol $N = [N_1, \ldots, N_n]^{\mathrm{T}}$, és a szimmetria feltevésből tudjuk, hogy N és $\Lambda_i N$ eloszlása minden i-re megegyezik. Ekkor

$$Z_0(\theta^*) \,=\, \|R_n^{-\frac{1}{2}} \Phi N\|_2^2 \,\stackrel{d}{=}\, \|R_n^{-\frac{1}{2}} \Phi \Lambda_i N\|_2^2 \,=\, Z_i(\theta^*), \tag{8}$$

de a $\{Z_i(\theta^*)\}$ valószínűségi változók természetesen nem teljesen függetlenek. Be lehet látni azonban, hogy *feltételesen* f.a.e. tulajdonságúak az $\{|N_t|\}$ által generált σ-algebrára nézve. Ekkor viszont *felcserélhetőek* is és így minden lehetséges rendezésük[4], $Z_{i_0}(\theta^*) \prec \cdots \prec Z_{i_{m-1}}(\theta^*)$, ugyanolyan $1/m!$ valószínűséggel áll elő.

Ha azonban $\theta \neq \theta^*$, akkor már ez a felcserélhetőségi tulajdonság nem áll fenn és $Z_0(\theta)$ egyre nagyobb valószínűséggel fogja dominálni a többi $\{Z_i(\theta)\}_{i \neq 0}$ változót, ahogy egyre távolabb kerülünk az igazi paramétertől, azaz ahogy $\|\theta^* - \theta\|_2 \to \infty$.

---

[3] Azaz 1 és -1 értékeket vehetnek, mindegyiket $1/2$ valószínűséggel.

[4] A „$\prec$” egy szigorú teljes rendezés, amit a standard „$<$”-ből úgy kapunk, hogy egyenlő értékek esetén véletlenszerűen döntjük el, hogy melyiket tekintjük kisebbnek; formális def. lásd [3].

Ahhoz, hogy össze tudjuk mérni a referencia függvény értékét az előjel-perturbált függvényekével, szükségünk lesz a referencia függvény *normalizált rangjára*,

$$\mathcal{R}(\theta) \;\triangleq\; \frac{1}{m}\left(1+\sum_{i=1}^{m-1}\mathbb{I}\left(Z_0(\theta)\prec Z_i(\theta)\right)\right), \tag{9}$$

ahol $\mathbb{I}(\cdot)$ egy indikátor függvény; tehát 1, ha az argumentuma (egy formula) igaz és 0 máskülönben. Tegyük fel, hogy $p = 1 - q/m$ alakban írható, ahol $0 < q < m$ és q, m tetszőleges egész számok. Ekkor az SPS teszt elfogadja a null hipotézist, ha $\mathcal{R}(\theta) \le p$ és elutasítja, ha $\mathcal{R}(\theta) > p$. Mivel mind m, a referencia és előjel-perturbált függvények száma, mind q szabad paraméterek (a felhasználó által meghatározott), így tetszőleges racionális[5] $p \in (0, 1)$ valószínűséghez készíthető egy SPS teszt.

A fentiek szellemében az SPS konfidencia tartományt így definiálhatjuk,

$$\widehat{\Theta}_{n,p} \;\triangleq\; \left\{\, \theta\in\mathbb{R}^d : \mathcal{R}(\theta)\le p \,\right\}. \tag{10}$$

Bebizonyítható [3], hogy az így kapott (véletlenített) konfidencia halmaz *egzakt*,

$$\mathbb{P}\big(\,\theta^*\in\widehat{\Theta}_{n,p}\,\big) \;=\; p. \tag{11}$$

A konfidencia halmaz egzaktsága annak ellenére fennáll, hogy nem használtuk ki a zaj konkrét eloszlását, sőt, még azt is megengedtük, hogy minden megfigyelésre különböző eloszlású zaj hasson, amelyeknek akár végtelen varianciája is lehet.

Az SPS konfidencia halmazai *csillag konvexek*, ahol az LS becslés egy csillag központ [3]. Tehát minden $\theta \in \widehat{\Theta}_{n,p}$ és $\beta \in [0, 1]$ -re, $\beta\,\theta + (1-\beta)\,\hat{\theta}_n \in \widehat{\Theta}_{n,p}$.

Az SPS konfidencia halmazok *erősen konzisztensek* [2], azaz aszimptotikusan egy valószínűséggel semmilyen hamis paraméter értéket nem tartalmaznak; a halmaz θ^* körüli tetszőleges kicsi gömb belsejébe kerül, ahogy a mintaszám növekszik

$$\mathbb{P}\left(\,\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\left\{\,\widehat{\Theta}_n\subseteq B_\varepsilon(\theta^*)\,\right\}\right) \;=\; 1, \tag{12}$$

minden $\varepsilon > 0$, ahol $B_\varepsilon(\theta^*) \triangleq \{\,\theta\in\mathbb{R}^d : \|\theta-\theta^*\|_2 \le \varepsilon\,\}$. A konzisztencia gyenge plusz feltevések mellett fennáll, pl. mind a zaj szórásnégyzete, mind a regresszorok normája tarthat a végtelenhez, ha a növekedésük bizonyos ráta alatt marad [2].

Hatékony *külső approximációs ellipszoid* is konstruálható az SPS halmazokhoz:

$$\widehat{\Theta}_{n,p} \;\subseteq\; \left\{\,\theta\in\mathbb{R}^d \,:\, (\theta-\hat{\theta}_n)^{\mathrm{T}} R_n (\theta-\hat{\theta}_n) \le r^*\,\right\}, \tag{13}$$

amelyek gyorsan (polinomiális időben) – *szemidefinit programozási* feladatok megoldásával – számolhatóak. Vegyük észre, hogy az így kapott konfidencia ellipszoidoknak ugyanaz lesz a középpontja (ti. az LS becslés) és az alakját meghatározó mátrixa, mint (5) esetén, csak az ellipszoidok sugara fog különbözni. Azonban, míg a határeloszláson alapuló ellipszoidok csak heurisztikák, az SPS halmaz külső approximációján alapuló ellipszoidok garantált konfidenciával rendelkeznek [3].

[5] További véletlenítéssel könnyen kiterjeszthető a teszt irracionális valószínűségekre is, azonban ennek elhanyagolható a gyakorlati jelentősége, ezért ismertetésétől eltekintünk.

3. Konfidencia halmazok dinamikus rendszerekhez

Az alapvető célkitűzés az SPS módszer megalkotásakor az volt, hogy *dinamikus rendszerek* pontbecslései köré tudjunk nem-aszimptotikus garanciákkal rendelkező konfidencia halmazokat készíteni. A lineáris regressziós eset egy természetes általánosítása, ha *általános lineáris (dinamikus) rendszereket* [5] vizsgálunk, azaz

$$Y_t \,\triangleq\, G(z^{-1};\theta^*)\,U_t + H(z^{-1};\theta^*)\,N_t, \tag{14}$$

ahol G és H (stabil) lineáris szűrők, H stabilan invertálható, z^{-1} a késeltetés („lag") operátor, $\theta^* \in \mathbb{R}^d$ az „igazi" paraméter, $G(0;\theta^*) = 0$, $H(0;\theta^*) = 1$, $\{U_t\}$ a (megfigyelt) bemenetek, $\{N_t\}$ a (nem megfigyelt) független és szimmetrikus zaj, a bementek és a zaj függetlenek, valamint a rendszer kezdeti állapota ismert [4].

Ekkor az *előrejelzési hibák* vektora[6] $\widehat{\varepsilon}(\theta) \triangleq [\,\widehat{\varepsilon}_1(\theta),\dots,\widehat{\varepsilon}_n(\theta)\,]^{\mathrm{T}}$, ahol $\widehat{\varepsilon}_t(\theta) \triangleq H^{-1}(z^{-1};\theta)\big(Y_t - G(z^{-1};\theta)\,U_t\big)$ minden t-re; valamint teljesül, hogy $\widehat{\varepsilon}_t(\theta^*) = N_t$.

Egy tipikus költségfüggvény az előrejelzési hibák négyzetösszege [5], azaz

$$\mathcal{V}(\theta\,|\,\mathcal{D}_n) \,=\, \frac{1}{2}\widehat{\varepsilon}^{\mathrm{T}}(\theta)\widehat{\varepsilon}(\theta), \qquad \text{ekkor} \qquad \Psi(\hat{\theta}_n)\,\widehat{\varepsilon}(\hat{\theta}_n) \,=\, 0, \tag{15}$$

ahol a $\hat{\theta}_n$ az *előrejelzési hiba becslés*, $\Psi(\theta) \triangleq [\,\psi_1(\theta),\dots,\psi_n(\theta)\,]$, és $\psi_t(\theta)$ pedig a t-edik előrejelzési hiba gradiense. Ismert, hogy ez a gradiens $\psi_t(\theta) = W(z^{-1};\theta)\,Y_t + K(z^{-1};\theta)\,U_t$ alakban is írható, ahol W és K (vektor-értékű) lineáris szűrők [5].

Ha megpróbáljuk ugyanazt a konstrukciót használni mint a 2.2 fejezetben, az sajnos erre az általános esetre nem fog működni, a $Z_0(\theta^*)$ -nak *nem ugyanaz lesz az eloszlása*, mint a többi $\{Z_i(\theta^*)\}_{i\neq 0}$ -nek, így az igazi paraméterhez tartózó függvényértékek felcserélhetőségéhez vezető érvelés sem marad érvényben [4].

A probléma a rendszer dinamikusságából fakad; azaz abból, hogy ha a zajt perturbáljuk, az – a lineáris regressziós esettől eltérően – hatással van az előrejelzési hibák gradiensére is. A megoldást alternatív kimeneti trajektóriák generálása jelenti, amelyeket a perturbált előrejelzési hibákkal hajtunk meg

$$\bar{Y}_{i,t}(\theta) \,\triangleq\, G(z^{-1};\theta)\,U_t + H(z^{-1};\theta)\,(\alpha_{i,t}\,\widehat{\varepsilon}_t(\theta)), \tag{16}$$

$i \in \{1,\dots,m-1\}$, és a jelölés egyszerűsítése miatt bevezetjük, hogy minden t-re $\bar{Y}_{0,t}(\theta) \triangleq Y_t$. Az alternatív trajektóriákkal perturbált gradienseket számolhatunk,

$$\bar{\psi}_{i,t}(\theta) \,\triangleq\, W(z^{-1};\theta)\,\bar{Y}_{i,t}(\theta) + K(z^{-1};\theta)\,U_t, \tag{17}$$

$i \in \{0,\dots,m-1\}$, amelyekkel eljutunk az SPS működőképes általánosításához

$$Z_i(\theta) \,\triangleq\, \|Q_i^{-\frac{1}{2}}(\theta)\Psi_i(\theta)\Lambda_i\widehat{\varepsilon}(\theta)\|_2^2, \tag{18}$$

ahol $\Psi_i(\theta) \triangleq [\,\bar{\psi}_{i,1}(\theta),\dots,\bar{\psi}_{i,n}(\theta)\,]$, és $Q_i(\theta) \triangleq \frac{1}{n}\Psi_i(\theta)\Psi_i^{\mathrm{T}}(\theta)$. Bebizonyítható, hogy az ezekkel a függvényekkel konstruált SPS halmazok már *egzakt* konfidenciával rendelkeznek [4]. A konstrukció kiterjeszthető *visszacsatolt* (pl., szabályozóval rendelkező) rendszerekre [4], valamint *nemlineáris* (pl., GARCH) modellekre is [1].

[6] Tegyük fel, hogy pont annyi adatunk van, hogy n hibatagot kiszámolhassunk.

4. Konklúzió és nyitott problémák

A dolgozatban röviden bemutattuk az SPS módszert, amely minimális statisztikai feltevésekkel képes regressziós modellek adott pontbecslései köré véges minták esetén is garantált – tipikusan egzakt – konfidencia tartományokat készíteni.

Ugyan az SPS módszer viselkedését már elég jól ismerjük lineáris regressziós problémákon, de még sok nyitott kérdés van vele kapcsolatban például dinamikus rendszerek esetén. Ilyen kérdés az SPS teszt másodfajú hibájának jellemzése és hatékony külső approximációk konstruálása különböző dinamikus modellekhez.

Köszönetnyilvánítás

Csáji B. Cs. munkáját az MTA Bolyai János Kutatási Ösztöndíja, valamint az ED_18-2-2018-0006 és 125698 számú NKFIH pályázatok támogatásával végezte.

Hivatkozások

[1] Csáji, B. Cs.: *Score Permutation Based Finite Sample Inference for Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) Models*, in: *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS), Cadiz, Spain*, 296–304 (2016).

[2] Csáji, B. Cs., Campi, M. C., and Weyer, E.: *Strong Consistency of the Sign-Perturbed Sums Method*, in: *Proceedings of the 53rd IEEE Conference on Decision and Control, Los Angeles, California*, 3352–3357 (2014).

[3] Csáji, B. Cs., Campi, M. C., and Weyer, E.: *Sign-Perturbed Sums: A New System Identification Approach for Constructing Exact Non-Asymptotic Confidence Regions in Linear Regression Models*, IEEE Trans. on Signal Proc., **63(1)**, (2015), 169–181.

[4] Csáji, B. Cs. and Weyer, E.: *Closed-Loop Applicability of the Sign-Perturbed Sums Method*, in: *Proceedings of the 54th IEEE Conference on Decision and Control, Osaka, Japan*, 1441–1446 (2015).

[5] Ljung, L.: *System Identification: Theory for the User*, 2nd edn., Prentice-Hall, Upper Saddle River (1999).

(Beérkezett: 2017. április 10.)

CSÁJI BALÁZS CSANÁD
Magyar Tudományos Akadémia
Számítástechnikai és Automatizálási Kutatóintézet
1111 Budapest, Kende utca 13-17.
csaji.balazs@sztaki.mta.hu

SYMMETRY AND CONFIDENCE

Balázs Csanád Csáji

In this article we briefly overviewed the SPS (Sign-Perturbed Sums) method which was recently developed by Marco Campi, Erik Weyer and Balázs Csanád Csáji. The main draw card of SPS is that it can construct non-asymptotically guaranteed confidence regions for parameters of regression models under minimal statistical assumptions. The fundamental assumption of the SPS method is that the noises affecting the system are distributed symmetrically about zero, but their distributions may change over time.

We started with investigating classical linear regression problems. We recalled the standard least-squares (LS) estimate with its confidence ellipsoids, whose construction is based on the asymptotic Gaussianity of the estimate. We argued that such ellipsoids do not come with rigorous finite sample guarantees and are imprecise for small samples.

Then, we presented the construction of SPS confidence regions for linear regression problems. We highlighted that these regions have (i) exact confidence probabilities; are (ii) star convex with the LS estimate as a star center; are (iii) strongly consistent, that is, for every ball around the true parameter, the confidence regions are asymptotically (as the sample size tends to infinity) almost surely remain inside the ball (thus every false parameter value will be eventually excluded); and (iv) efficient ellipsoidal outer-approximations can be constructed for them by solving semidefinite programming problems. The SPS ellipsoids have the same center (i.e., the LS estimate) and kernel matrix as the classical ellipsoids of the LS theory, only their radii are different. However, while the ellipsoids of the classical theory (based on limiting distributions) are just heuristics for finite samples, the SPS ellipsoids have rigorous non-asymptotic guarantees.

Finally, we generelized the SPS construction for general linear (dynamical) systems. We discussed that a straightforward generalization would not work, as the noises affecting the system and the outputs are not independent for dynamical systems. It was shown that the proper generalization should be built using alternative output trajectories, which are constructed using perturbed prediction error sequences. The resulting SPS confidence regions also have exact confidence probabilities, and the construction can be generalized for closed-loop and certain non-linear systems (such as GARCH models), as well.

There are several open problems concerning the SPS method, especially for its construction for dynamical systems: for example, building efficient outer-approximations for various types of dynamical systems and analyzing the type II errors of SPS tests.

Keywords: confidence regions, resampling methods, linear regression, time series

Mathematics Subject Classification (2000): 62F25, 662G09, 62J05, 62M10