Empirikus folyamatok approximációi

Ph.D. értekezés tézisei

Szűcs Gábor

Témavezető: Dr. Csörgő Sándor

Matematika- és Számítástudományi Doktori Iskola
Szegedi Tudományegyetem, Bolyai Intézet
Szeged, 2011

1. Bevezetés

A disszertációban független és azonos eloszlású változók segítségével felírt bizonyos empirikus folyamatok aszimptotikus viselkedését vizsgáljuk. A legtöbb esetben az approximációs módszert fogjuk alkalmazni. Ez azt jelenti, a vizsgált empirikus folyamatot és egy alkalmas Gauss folyamat reprezentánsait egy kényelmes választott valószínűségi mezőn konstruáljuk meg, méghozzá oly módon, hogy az empirikus folyamat és a Gauss folyamatok távolsága nullához konvergáljon, amint a mintaméret megy a végtelenbe. Ezáltal az alkalmazott Gauss folyamat segítségével tanulmányozhatjuk a vizsgált empirikus folyamat aszimptotikus viselkedését.

A disszertációban empirikus folyamatoknak két típusát vizsgáljuk. A 3. fejezetben a paraméteres eloszláscsaládokon definiált becsült paraméteres empirikus folyamat paraméteres bootstrap és nemparaméteres bootstrap változatát tanulmányozzuk. A fejezet fő célja gyenge approximációt bizonyítani a folyamatokra, és azáltal belátni, hogy azok konvergálnak eloszlásban. Ezután ismertetünk egy algoritmust, melynek segítségével összetett illeszkedési hipotéziseket tesztelhetünk. A módszer gyakorlati alkalmazását egy szimulációs tanulmányon mutatjuk be.

A 4. fejezetben a nemnegatív értékű valószínűségi változók valószínűségi generátorfüggvényével definiált empirikus folyamatokat tanulmányozunk. Célunk egy olyan elméleti háttér kidolgozása, melynek segítségével hatékonyan vizsgálhatjuk ezen folyamatokat. Ennek segítségével bebizonyítunk egy erős approximációs tételt és egy iterált logaritmustételt az empirikus generátor folyamatra és deriváltjaira. Továbbá, definiálni fogjuk a valószínűségi generátor folyamat bootstrap és becsült paraméteres változatait, melyek alkalmazásával konfidencia sávot szerkeszthetünk az ismeretlen valószínűségi generátorfüggvényhez, illetve összetett illeszkedési hipotéziseket tesztelhetünk.

A 2. fejezetben néhány szükséges technikai eszközt vezetünk be. Ismertetjük a KMT approximáció legfontosabb eredményeit, bebizonyítunk egy elméleti háttértételt a bootstrap módszerre, valamint kiterjesztjük a véges intervallumon tekintett sztochasztikus integrál fogalmát a valós egyenesen értelmezett sztochasztikus integrálra.

A szerző három cikket jegyez a disszertáció témájában. Szűcs (2008) tartalmazza a becsült paraméteres empirikus folyamat paraméteres bootstrap változatának konvergenciáját. A kapcsolatos tétel a nemparaméteres bootstrap folyamatra és a bemutatott szimulációs tanulmány a témája egy elfogadott dolgozatnak, hivatkozásért lásd Szűcs (20??). Végül, Szűcs (2005) tartalmazza a valószínűségi generátor folyamatokra vonatkozó eredményeket arra az esetre, amikor a háttérváltozó nemnegatív egész értékű. Az eredmények általánosítása tetszőleges nemnegatív értékű változóra még nem publikált.

2. Néhány alapvető fogalom

A fejezetben három elméleti fogalmat mutatunk be. Az első a független és a [0,1] intervallumon egyenletes eloszlású $U_1, \dots, U_n$ változók alapján felírt $\beta_n(u)$, $0 \le u \le 1$, egyenletes empirikus folyamatra vonatkozó KMT approximáció. Eszerint egy alkalmasan választott valószínűségi mezőn a változók definiálhatóak olyan módon, hogy Brown hidaknak egy alkalmas $B_1, B_2, \ldots$ sorozatára

$$\sup_{0\le u\le 1} \big|\beta_n(u) - B_n(u)\big| = \mathcal{O}\big(n^{-1/2}\log n\big)\,, \qquad n\to\infty\,, \qquad \text{m.b.} \tag{1}$$

A konstrukció Komlós Jánostól, Major Pétertől és Tusnády Gábortól származik, és alapvető lesz a munkánk során.

A második eszköz Efron bootstrap módszere, melynek segítségével megbecsülhetjük egy n elemű mintára felírt τ_n statisztika eloszlását. Vegyük észre, hogy ilyen becslés a szokásos statisztikai technikákkal nem adható, ugyanis egyetlen minta birtokában a τ_n változóra csupán egy megfigyelés áll rendekésünkre. A bootstrap alapötlete az, hogy ha a mintaelemek ismeretlen $F(x)$ eloszlásfüggvényét egy $\hat{F}_n(x)$, $x\in\mathbb{R}$, függvénnyel becsüljük, és tekintünk az eredeti mintára nézve feltételesen független $X^*_{1,n},\ldots,X^*_{m_n,n}$ változókat, melyek feltételes eloszlásfüggvénye $\hat{F}_n$, akkor $\tau^*_{m_n,n} = \tau_{m_n}(X^*_{1,n},\ldots,X^*_{m_n,n})$ eloszlása „hasonlít" τ_n eloszlásához. Elméleti számítások vagy Monte Carlo szimuláció révén $\tau^*_{m_n,n}$ eloszlása tetszőleges pontossággal megkapható, és ezáltal egy jobb vagy rosszabb becslést nyerhetünk τ_n eloszlására.

Munkánk során a bootstrap technika két változatát alkalmazzuk, a paraméteres és a nemparaméteres bootstrapet. A nemparaméteres esetben $\hat{F}_n$ az $X_1,\ldots,X_n$ változók empirikus eloszlásfüggvénye, tehát a bootstrap mintát visszatevéses mintavételezéssel kapjuk az eredeti megfigyelésekből. A paraméteres bootstrap csak akkor alkalmazható, ha az X_i megfigyelések eloszlásfüggvénye egy $F(x,\theta)$, $\theta\in\Theta$, paraméteres eloszláscsalád $F(x,\theta_0)$ eleme, $x\in\mathbb{R}$. Ekkor, a θ_0 paraméter egy $\hat{\theta}_n$ becslését tekintve, a paraméteres bootstrap mintát úgy definiáljuk, mint feltételesen független véletlen változókat közös $F(x,\hat{\theta}_n)$ feltételes eloszlásfüggvénnyel.

Végül, ki kell terjesztenünk a véges intervallumon értelmezett lokálisan négyzetesen integrálható martingálokra vett sztochasztikus integrált az egész valós egyenesen vett sztochasztikus integrálra. Bizonyítunk egy állítást az integrál létezésére, és leírjuk az olyan folyamatok eloszlását, melyek bizonyos kétváltozós függvényeknek a standard Wiener folyamatra vett integráljaként állnak elő.

3. Bootstrap paraméterbecsült empirikus folyamatok

Bevezetés és előzmények

Tekintsük eloszlásoknak egy $\mathcal{F} = \{F(x,\theta) : x\in\mathbb{R}, \theta\in\Theta\subseteq\mathbb{R}^d\}$ családját, valamint $X_1, X_2, \ldots$ független változókat közös $F(x,\theta_0)$, $x\in\mathbb{R}$, eloszlásfüggvénnyel, $\theta_0\in\Theta$. Jelölje $F_n(x)$, $x\in\mathbb{R}$, a sorozat első n elemének, mint mintának az empirikus eloszlásfüggvényét, és legyen $\hat{\theta}_n$ a θ_0 paraméter becslése. Ekkor a becsült paraméteres empirikus folyamat

$$\hat{\alpha}_n(x) = n^{1/2}\big[F_n(x) - F(x,\hat{\theta}_n)\big]\,, \qquad x\in\mathbb{R}\,. \tag{2}$$

Mióta Durbin (1973) bebizonyította, hogy $\hat{\alpha}_n(x)$ gyengén konvergál egy $G(x)$, $x\in\mathbb{R}$, Gauss folyamathoz, amint a mintaméret tart a végtelenbe, a becsült paraméteres empirikus folyamat széles körben használt eszköz összetett illeszkedési hipotézisek tesztelésére. Sajnos a folyamatra épülő statisztikai módszerek általában nem eloszlásmentesek, és a kapcsolatos kritikus értékeket nem lehet elméleti úton meghatározni. Szerencsére ezen nehézségek kiküszöbölhetőek bootstrap módszer alkalmazásával.

Tekintsünk $X^*_{1,n},\ldots,X^*_{m_n,n}$ bootstrap mintaelemeket az $X_1,\ldots,X_n$ megfigyelésekre nézve. Legyen $F^*_{m_n,n}(x)$, $x\in\mathbb{R}$, a bootstrap változók empirikus eloszlásfüggvénye, és legyen θ^*_n paraméterbecslés a bootstrap mintaelemek segítségével. A bootstrap becsült paraméteres empirikus folyamat a bootstrap mintaelemekre felírt becsült paraméteres empirikus folyamat, tehát

$$\bar{\alpha}^*_{m_n,n}(x) = n^{1/2}\big[F^*_{m_n,n}(x) - F(x,\theta^*_n)\big]\,, \qquad x\in\mathbb{R}\,.$$

A folyamatra a $\hat{\alpha}^*_{m_n,n}$ jelölést használjuk a paraméteres és a $\tilde{\alpha}^*_{m_n,n}$ jelölést a nemparaméteres bootstrap esetben. A bootstrap alkalmazásának motivációja az az ötlet, hogy ha $\hat{\alpha}^*_{m_n,n}$ és/vagy $\hat{\alpha}^*_{m_n,n}$ eloszlásban konvergál ugyanazon határfolyamathoz, mint $\hat{\alpha}_n$, akkor egy $\psi(\hat{\alpha}_n)$ statisztika kritikus értékei becsülhetőek, mint a kapcsolatos $\psi(\hat{\alpha}^*_{m_n,n})$ és/vagy $\psi(\tilde{\alpha}^*_{m_n,n})$ funkcionál empirikus kvantilisei. Stute et al. (1993), és ettől az eredménytől függetlenül, Babu és Rao (2004) bizonyította $\hat{\alpha}^*_{m_n,n}$ gyenge konvergenciáját az $m_n=n$ esetben folytonos eloszláscsaládokra, tehát ekkor a paraméteres bootstrap alkalmazható. Ezzel szemben Babu és Rao rámutatott, hogy az $\tilde{\alpha}^*_{m_n,n}$ folyamat nem konvergál eloszlásban, ugyanis bias korrekcióra van szükség

Feltevések és eredmények

Tekintsük az $\mathcal{F}=\{F(x,\theta): x\in\mathbb{R}, \theta\in\Theta\subseteq\mathbb{R}^d\}$ eloszláscsaládot, és legyen θ_0 egy rögzített és $\theta=(\theta^{(1)},\ldots,\theta^{(d)})$ egy tetszőleges vektor a Θ halmazban. Legyen $X_1,X_2,\ldots$ és $Y_1,Y_2,\ldots$ független és azonos eloszlású változóknak sorozatai rendre $F(x,\theta_0)$ és $F(x,\theta)$ eloszlásfüggvénnyel, $x\in\mathbb{R}$. Tekintsünk egy

$$\theta_n = \theta_n(Y_1,\ldots,Y_n)$$

statisztikát, mint az általános θ paraméter becslését, és legyen $\hat{\theta}_n=\theta_n(X_1,\ldots,X_n)$ a θ_0 paraméter becslése a $X_1,\ldots,X_n$ minta alapján. Legyen m_n a bootstrap mintaméret, és legyen $\hat{\theta}^*_n$ és $\tilde{\theta}^*_n$ paraméterbecslés a paraméteres illetve a nemparaméteres bootstrap minta alapján. A vektorokat sorvektorként értelmezzük, továbbá rendre V^T és $V^{(k)}$ jelöli a V vektor transzponáltját és k. komponensét.

Feltevések. A fő eredményekben a következő feltevéseket fogjuk alkalmazni.

(a1) A parciális deriváltak

$$\nabla_\theta F(x,\theta) = \left(\frac{\partial}{\partial\theta^{(1)}}F(x,\theta),\ldots,\frac{\partial}{\partial\theta^{(d)}}F(x,\theta)\right)$$

vektora létezik minden $(x,\theta)\in\mathbb{R}\times\Lambda$ pontban, ahol $\Lambda\subseteq\Theta$ a θ_0 pont egy megfelelő környezete.

(a2) $\nabla_\theta F(x,\theta)$ egyenletesen konvergál a $\nabla_\theta F(x,\theta_0)$, $x\in\mathbb{R}$, függvényhez, amint $\theta\to\theta_0$.

(a3) $\nabla_\theta F(x,\theta_0)$, $x\in\mathbb{R}$, korlátos.

(a4) Létezik $l(\theta_0):\mathbb{R}\to\mathbb{R}^d$ és $\varepsilon_n(\theta_0):\mathbb{R}^n\to\mathbb{R}^d$ Borel mérhető függvény, melyre

$$\hat{\theta}_n-\theta_0=\frac{1}{n}\sum_{i=1}^{n} l(X_i,\theta_0)+n^{-1/2}\varepsilon_n(\theta_0) \qquad \text{m.b.},$$

ahol $\varepsilon_n(\theta_0)=\varepsilon_n(X_1,\dots,X_n,\theta_0)$.

(a5) Létezik $l:\mathbb{R}\times\Lambda\to\mathbb{R}^d$ és $\varepsilon_n:\mathbb{R}^n\times\Lambda\to\mathbb{R}^d$ Borel mérhető függvény, hogy tetszőleges $\theta\in\Lambda$ esetén

$$\theta_n-\theta=\frac{1}{n}\sum_{i=1}^{n} l(Y_i,\theta)+n^{-1/2}\varepsilon_n(\theta) \qquad \text{m.b.},$$

ahol $\varepsilon_n(\theta)=\varepsilon_n(Y_1,\dots,Y_n,\theta)$.

(a6) Létezik Borel mérhető függvény $\varepsilon_{m,n}(\theta_0):\mathbb{R}^{n+m}\to\mathbb{R}^d$, melyre

$$\tilde{\theta}_n^*-\hat{\theta}_n=\frac{1}{m}\sum_{i=1}^{m} l(X_{i,n}^*,\theta_0)-\frac{1}{n}\sum_{i=1}^{n} l(X_i,\theta_0)+m^{-1/2}\varepsilon_{m,n}(\theta_0) \qquad \text{a.s.}$$

ahol l az (a4) pontban definiált függvény, $X_{1,n}^*,\dots,X_{m,n}^*$ a nemparaméteres bootstrap minta, és $\varepsilon_{m,n}(\theta_0)=\varepsilon_{m,n}(X_1,\dots,X_n,X_{1,n}^*,\dots,X_{m,n}^*,\theta_0)$.

(a7) $E\,l(X_i,\theta_0)=0$.

(a8) $E\,l(Y_i,\theta)=0$ tetszőleges $\theta\in\Lambda$ esetén.

(a9) $M(\theta_0)=E\,l(X_i,\theta_0)^T l(X_i,\theta_0)$ véges pozitív szemidefinit mátrix.

(a10) $M(\theta)=E\,l(Y_i,\theta)^T l(Y_i,\theta)$ véges pozitív szemidefinit mátrix minden $\theta\in\Lambda$ esetén.

(a11) Az $M(\theta)$, $\theta\in\Lambda$, függvény folytonos a θ_0 pontban.

(a12) Az $l(x,\theta_0)$, $x\in\mathbb{R}$, vektorfüggvény komponensei korlátos változásúak minden véges intervallumon.

(a13) $l(x,\theta)$ egyenletesen konvergál az $l(x,\theta_0)$, $x\in\mathbb{R}$, függvényhez minden véges intervallumon, amint $\theta\to\theta_0$.

(a14) $\varepsilon_n(\theta_0)\xrightarrow{P}0$.

(a15) $\varepsilon_{m_n}(\hat{\theta}_n)\xrightarrow{P}0$.

(a16) $\varepsilon_{m_n,n}(\theta_0)\xrightarrow{P}0$.

Jegyezzük meg, hogy Burke, Csörgő, Csörgő és Révész (1979) bizonyított egy gyenge approximációs tételt az $\hat{\alpha}_n$ folyamatra. Megmutatták, hogy ha az (a1)–(a4), (a7), (a9), (a12) és (a14) feltételek teljesülnek, akkor létezik az $X_1, X_2, \ldots$ változóknak olyan reprezentációja, valamint a G folyamatnak olyan $G_1, G_2, \ldots$ másolatai, hogy

$$\sup_{x\in\mathbb{R}} \big|\hat{\alpha}_n(x) - G_n(x)\big| \xrightarrow{P} 0\,, \qquad n\to\infty\,. \tag{3}$$

Ebből az eredményből azonnal következik az $\hat{\alpha}_n$ folyamat gyenge konvergenciája a G folyamathoz. A határfolyamat Gauss és

$$G(x) = B\big(F(x,\theta_0)\big) - \left[\int_{\mathbb{R}} l(x,\theta_0)\, dB\big(F(x,\theta_0)\big)\right] \nabla_\theta F(x,\theta_0)^T\,, \qquad x\in\mathbb{R}\,, \tag{4}$$

alakban írható fel, ahol $B(u)$, $0\le u\le 1$, egy Brown híd. A bootstrappelt empirikus folyamatok területén elért fontosabb eredményinket az alábbi tételek a tartalmazzák. A 3.2. Tétel foglalkozik a paraméteres bootstrap esettel, és a 3.3. Tétel szól a nemparaméteres változatról.

3.2. Tétel. *Tegyük fel, hogy a bootstrap mintaméretre $m_n\to\infty$, továbbá tegyük fel, hogy az $\mathcal{F}$ eloszláscsalád, a θ_0 rögzített paraméter és az alkalmazott paraméterbecslő eljárás kielégíti az (a1)–(a3), (a5), (a8), (a10)–(a15) feltevéseket. Ekkor, egy alkalmas valószínűségi mezőn, konstruálhatóak X_i és $X_{i,\theta}$, $\theta\in\Theta$, $i=1{,}2,\ldots$, változók rendre $F(x,\theta_0)$ illetve $F(x,\theta)$ eloszlásfüggvénnyel, valamint definiálható Brown hidaknak egy $B_1, B_2, \ldots$ sorozata úgy, hogy az $X_1, X_{1,\theta}, X_2, X_{2,\theta}, \ldots$ változók függetlenek minden θ esetén, és az*

$$\big(X^*_{1,n},\ldots,X^*_{m_n,n}\big) = \big(X_{1,\hat\theta_n},\ldots,X_{m_n,\hat\theta_n}\big)\,, \qquad n=1{,}2,\ldots\,,$$

paraméteres bootstrap minta alapján felírt $\hat{\alpha}^*_{m_n,n}$ paraméteres bootstrap becsült paraméteres empirikus folyamat teljesíti a

$$\sup_{x\in\mathbb{R}} \big|\hat{\alpha}^*_{m_n,n}(x) - G_{m_n}(x)\big| \xrightarrow{P} 0\,, \qquad n\to\infty\,,$$

approximációt, ahol a $G_1, G_2, \ldots$ sorozatot a (4) formulával definiáljuk a $B_1, B_2, \ldots$ hidak alkalmazásával.

3.3. Tétel. *Tegyük fel, hogy a bootstrap mintaméretre $m_n\to\infty$, és tegyük fel, hogy az $\mathcal{F}$ eloszláscsalád, a θ_0 rögzített paraméter és az alkalmazott paraméterbecslő eljárás kielégíti az (a1)–(a4), (a6), (a7), (a9), (a12), (a14) és (a16) feltevéseket. Egy alkalmas valószínűségi mezőn definiálhatóak független $X_1, X_2, \ldots$ változók $F(x,\theta_0)$ eloszlásfüggvénnyel, továbbá ezekhez kapcsolódó $X^*_{1,n},\ldots,X^*_{m_n,n}$, $n=1{,}2,\ldots$, nemparaméteres bootstrap mintaelemek, valamint Brown hidaknak egy $B_1, B_2, \ldots$ sorozata olyan módon, hogy az $\tilde{\alpha}^*_{m_n,n}$ nemparaméteres bootstrap becsült paraméteres empirikus folyamatra*

$$\sup_{x\in\mathbb{R}} \left|\tilde{\alpha}^*_{m_n,n}(x) - \Big(\frac{m_n}{n}\Big)^{1/2}\hat{\alpha}_n(x) - G_{m_n}(x)\right| \xrightarrow{P} 0\,, \qquad n\to\infty\,,$$

ahol $\hat{\alpha}_n$ a becsült paraméteres empirikus folyamat, és a $G_1, G_2, \ldots$ sorozatot a (4) formulával definiáljuk a $B_1, B_2, \ldots$ hidak alkalmazásával.

Mivel a G_{m_n} folyamat eloszlása minden n esetén megegyezik G eloszlásával, azonnal adódik a 3.2. és a 3.3. Tétel alábbi következménye.

3.4. Következmény. *A 3.2. és a 3.3. Tétel feltételei mellett*

$$\hat{\alpha}^*_{m_n,n}(x) \qquad \text{és} \qquad \tilde{\alpha}^*_{m_n,n}(x) - \Big(\frac{m_n}{n}\Big)^{1/2} \hat{\alpha}_n(x)\,, \qquad x \in \mathbb{R}\,,$$

gyengén konvergál a $G(x)$, $x \in \mathbb{R}$, folyamathoz a $D[-\infty,\infty]$ térben.

A tételek bizonyításához két dolgozatból merítettünk ötleteket. A véletlen elemek konstrukciója Csörgő és Mason (1989) módszerére épül, tőlük származik az ötlet, hogy az eredeti és a bootstrap mintaelemeket két KMT mező szorzatán reprezentáljuk. Szerencsére ez a megoldás nem igényel semmilyen regularitási feltételt. Emellett alkalmaztuk a Burke, Csörgő, Csörgő és Révész (1979) által a nem bootstrappelt $\hat{\alpha}_n$ folyamatra kidolgozott approximációs technikát, ahonnan megörököltük a (3) eredmény feltételeit is a módszerrel együtt. Nevezetesen, meg kell követelnünk a $\nabla_\theta F(x,\theta)$ függvény létezését és simaságát a θ_0 pont egy környezetében, valamint feltevéseket kell tennünk az alkalmazott paraméterbecslő eljárás regularitására a θ_0 pontban. A nemparaméteres bootstrap esetben nincs is szükség sokkal több feltevésre, csupán a $\tilde{\theta}^*_n$ bootstrap becslésre kell egy olyan összeg reprezentációt adnunk, mint ami már a rendelkezésünkre áll a nem bootstrappelt esetben. Ezzel szemben a paraméteres bootstrap esetben a bootstrappelt minta az $F(x,\hat{\theta}_n)$, $x \in \mathbb{R}$, eloszlásból jön, és emiatt a paraméterbecslő eljárásra tett feltevéseket ki kell terjesztenünk a θ_0 pont egy környezetére. Végigtekintve a témával foglalkozó korábbi cikkeket azt látjuk, hogy Stute et al. (1993) és Babu és Rao (2004) hasonló feltevésekkel dolgozott.

A bootstrap algoritmus

Tekintsünk $X_1,\ldots,X_n$ független és azonos eloszlású véletlen változókat ismeretlen $F(x)$, $x \in \mathbb{R}$, eloszlásfüggvénnyel, és tekintsük az

$$\mathcal{F} = \big\{F(x,\theta) : x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^d\big\}$$

eloszláscsaládot egy $\theta_n : \mathbb{R}^n \to \Theta$ paraméterbecslő statisztikával. Ebben a környezetben a minta illeszkedése az $\mathcal{F}$ családhoz, tehát a $\mathcal{H}_0 : F \in \mathcal{F}$ null-hipotézis tesztelhető egy $\psi_n = \psi(\hat{\alpha}_n)$ teszt statisztika alkalmazásával. Itt $\hat{\alpha}_n$ a becsült paraméteres empirikus folyamat, és ψ a $D[-\infty,\infty]$ függvénytéren értelmezett valós értékű funkcionál, mely kielégít bizonyos további feltételeket. Mivel $\hat{\alpha}_n$ eloszlásban konvergál a G folyamathoz, a $\varphi = \psi(G)$ változó elméleti kvantilisei alkalmazhatóak, mint aszimptotikusan helyes kritikus értékek a ψ_n statisztikára.

Ahogyan azt már láttuk a fejezet bevezetésében, a legfőbb nehézség az, hogy a φ limeszváltozó kvantiliseit nem tudjuk elméleti úton meghatározni. Szerencsére, ha az $\mathcal{F}$ eloszláscsalád és a θ_n becslő módszer kielégíti a 3.2. és/vagy a 3.3. Tétel feltevéseit, továbbá a ψ funkcionál teljesít bizonyos további feltételeket, amiket nem részletezünk, akkor a

$$\psi^{*p}_{m_n,n} = \psi\big(\hat{\alpha}^*_{m_n,n}\big) \qquad \text{és} \qquad \psi^{*np}_{m_n,n} = \psi\Big(\tilde{\alpha}^*_{m_n,n} - (m_n/n)^{1/2}\hat{\alpha}_n\Big)$$

paraméteres és nemparaméteres funkcionál eloszlásban konvergál a φ változóhoz, és a limeszváltozó $(1-\alpha)$ kvantilise becsülhető a

$$c_n^{*p}(\alpha) = \inf\Big\{x \in \mathbb{R} : P\big(\psi_{m_n,n}^{*p} \leq x \mid X_1, \ldots, X_n\big) \geq 1-\alpha\Big\}$$

kvantilissel a paraméteres bootstrap, és a

$$c_n^{*np}(\alpha) = \inf\Big\{x \in \mathbb{R} : P\big(\psi_{m_n,n}^{*np} \leq x \mid X_1, \ldots, X_n\big) \geq 1-\alpha\Big\}$$

kvantilissel a nemparaméteres bootstrap esetben. A gondolatmenet eredményeként az alábbi bootstrap tesztelő algoritmus kapjuk.

1. Határozzuk meg a $\hat{\theta}_n$ becslést az $X_1, \ldots, X_n$ mintaelemek alapján.
2. Határozzuk meg a ψ_n teszt statisztika értékét.
3. Generáljunk $X_{1,n}^*, \ldots, X_{m_n,n}^*$ független paraméteres vagy nemparaméteres bootstrap megfigyeléseket rendre $F(x, \hat{\theta}_n)$ vagy $F_n(x)$ eloszlásfüggvénnyel.
4. Határozzuk meg a $\hat{\theta}_n^*$ vagy a $\tilde{\theta}_n^*$ becslést a bootstrap minta alapján.
5. Határozzuk meg a $\psi_{m_n,n}^{*p}$ vagy a $\psi_{m_n,n}^{*np}$ bootstrap statisztika értékét.
6. Ismételjük meg a 3.–5. lépést R alkalommal, és legyen $\psi_{n,1}^* \leq \cdots \leq \psi_{n,R}^*$ a $\psi_{m_n,n}^{*p}$ vagy a $\psi_{m_n,n}^{*np}$ változó megfigyelt értékeiből képzett rendezett minta.
7. Legyen $c_{n,\alpha}^*$ a $\psi_{m_n,n}^{*p}$ vagy a $\psi_{m_n,n}^{*np}$ változó $(1-\alpha)$ rendű empirikus kvantilise, tehát, az $\lceil R(1-\alpha)\rceil$-dik legnagyobb elem a rendezett mintában.
8. Vessük el a $\mathcal{H}_0$ null-hipotézist, ha ψ_n nagyobb, mint $c_{n,\alpha}^*$.

A 3.4. fejezetben megmutatjuk, hogy a bootstrap alkalmazható a vizsgált empirikus folyamatok Kolmogorov–Szmirnov típusú szuprémum funkcionáljaira. A 3.5. fejezetben megvizsgáljuk a tételek regularitási feltételeit, és bebizonyítjuk, hogy ezen feltételek teljesülnek a Poisson és a normális eloszlásra és a maximum likelihood becslésre. Hogy bemutassuk a bootstrap technikát egy gyakorlati alkalmazáson keresztül, az utolsó fejezetben ismertetjük egy szimulációs tanulmány eredményeit. A paraméteres és a nemparaméteres bootstrap alkalmazásával teszteljük negatív binomiális minták illeszkedését a Poisson családhoz, valamint lokáció és skála kontaminált normális változók illeszkedését a normális eloszláshoz.

4. Empirikus valószínűségi generátor folyamatok

Bevezetés és előzmények

Legyen $X, X_1, X_2, \ldots$ nemnegatív értékű független és azonos eloszlású valószínűségi változó $F(x)$, $x \in \mathbb{R}$, eloszlásfüggvénnyel, és legyen

$$g(t) = Et^X = \int_{\mathbb{R}} t^x dF(x) \qquad \text{és} \qquad g_n(t) = \frac{1}{n}\sum_{j=1}^{n} t^{X_j}, \qquad 0 \le t \le 1,$$

az első n elem valószínűségi generátorfüggvénye és a generátorfüggvény empirikus változata. A fejezetben a 0^0 szimbólum 1-nek van definiálva, ugyanis szükségünk lesz a t^x függvénynek az x változóban való folytonosságára. Ekkor az empirikus valószínűségi generátor folyamat

$$\gamma_n(t) = n^{1/2}\big[g_n(t) - g(t)\big], \qquad 0 \le t \le 1.$$

A generátorfüggvények használata statisztikai problémák megoldására nem új ötlet, a karakterisztikus és a momentumgeneráló függvényen alapuló hasonló transzformált folyamatok ismertek. (Például, Csörgő (1981) és Csörgő, Csörgő, Horváth és Mason (1986).) Ezen módszerek elméleti alapja az, hogy bizonyos feltételek teljesülése esetén a transzformált folyamatok eloszlásban konvergálnak valamilyen függvénytéren. A valószínűségi generátor folyamat esetében Rémillard és Theodorescu (2000) mondta ki, hogy γ_n eloszlásban konvergál a $C[0,1]$ téren a

$$Y(t) = \int_{\mathbb{R}} t^x \, dB\big(F(x)\big), \qquad 0 \le t \le 1,$$

folyamathoz tetszőleges nemnegatív egész értékű X változóra. Sajnos a bizonyításukba belecsúszott egy hiba, de az alapötlet jó, és a bizonyítás javítható.

A fejezet célja kidolgozni egy olyan általános és rugalmas eszköztárat, melynek segítségével vizsgálhatjuk az empirikus generátor folyamat és deriváltjai, valamint a bootstrap és/vagy becsült paraméteres változatok aszimptotikus viselkedését. Az eredmények abban az értelemben is általánosak, hogy nem csak az egész értékű esetben alkalmazhatóak, hanem tetszőleges nemnegatív értékű változóra.

Általánosított valószínűségi generátor folyamatok

A 4.2. fejezetben olyan folyamatokat vizsgálunk, melyeket az

$$I_r(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r} \, dK(x)$$

integrál definiál, ahol a $K(x)$ függvény előáll egy lokálisan négyzetesen integrálható $M(x)$ martingál és egy $A(x)$ korlátos változású folyamat összegeként, $x \in \mathbb{R}$. Emellett feltesszük, hogy M és A eltűnik a $(-\infty,0)$ negatív félegyenesen, és a folyamatoknak càdlàg trajektóriái vannak. A 4.2., 4.3. és 4.8. Állításban megmutatjuk, hogy bizonyos feltételek mellett az I_r folyamat definiált a $(-1{,}1]$ intervallum valamely $[a,b]$ részhalmazain. Az általános rész fő eredménye a 4.7. és a 4.9. Tétel. Ezekben az I_r folyamatra vonatkozó egyenlőtlenségeket bizonyítunk

$$\sup_{a\le t\le b}|Y_r(t)| \le C\sup_{x\in\mathbb{R}}|K(x)| \tag{5}$$

alakban, ahol $C=C(r,a,b)$ a K folyamattól független konstans.

Ezen általános eredmények alkalmazásaiban a K folyamat az $X_1,\dots,X_n$ változók által meghatározott valamilyen empirikus típusú folyamatot (empirikus vagy elméleti eloszlásfüggvényt, bootstrappelt és/vagy becsült paraméteres empirikus folyamatot) fog jelölni. Mivel a feltevések szerint a minta egy nemnegatív értékű eloszlásból jön, a kapcsolatos empirikus típusú folyamatok kielégítik a K folyamatra tett feltevéseket. Ezáltal a K folyamatra meglevő eredményeket alkalmazva a (5) egyenlőtlenség felhasználásával képesek leszünk konvergenciát és approximációt bizonyítani az I_r folyamatra.

Az empirikus generátor folyamat elemi tulajdonságai

A 4.3. fejezetben megmutatjuk, hogy a γ_n generátor folyamat r. deriváltja

$$\gamma_n^{(r)}(t) = n^{1/2}\big[g_n^{(r)}(t) - g^{(r)}(t)\big] = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\,d\alpha_n(x)$$

alakban írható fel, $r=0{,}1,\dots$, ahol $g_n^{(r)}$ és $g^{(r)}$ az empirikus és az elméleti valószínűségi generátorfüggvény r. deriváltja, továbbá $\alpha_n(x)$, $x\in\mathbb{R}$, az $X_1,\dots,X_n$ minta alapján felírt empirikus folyamat. Emellett definiáljuk az

$$Y_r(t) = \int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\,dB\big(F(x)\big)$$

folyamatot, ahol B a Brown híd. A 4.10. Állításban megmutatjuk, hogy $\gamma_n^{(r)}$ folytonos és az Y_r folyamatnak létezik mintafolytonos modifikációja az $[a,b]$ intervallumain, ahol $[a,b]$ választható úgy, mint

- $[\varepsilon,1-\varepsilon]$ tetszőleges $0<\varepsilon<1/2$ értékkel bármely nemnegatív értékű X változóra;
- $[-\tau,\tau]$ tetszőleges $0<\tau<1$ értékkel, ha X nemnegatív egész értékű változó;
- $[0{,}1]$, ha X nemnegatív egész értékű és $r=0$.

A 4.11. fejezetben belátjuk, hogy $Y_r(t)$, $a\le t\le b$, Gauss folyamat nulla várható értékkel és folytonos kovariancia függvénnyel. Tsirel'son egy, a normális eloszlású változók szuprémumára vonatkozó tételének alkalmazásával bizonyítjuk, hogy a folyamat

$$S_r = \sup_{a\le x\le b}|Y_r(t)|, \qquad S_r^+ = \sup_{a\le x\le b} Y_r(t) \qquad \text{és} \qquad S_r^- = -\inf_{a\le x\le b} Y_r(t) \tag{6}$$

Kolmogorov–Szmirnov típusú funkcionáljainak eloszlásfüggvénye abszolút folytonos a $(0,\infty)$ pozitív félegyenesen, és a kapcsolatos sűrűségfüggvény korlátos minden $[s,\infty)$, $s>0$, részintervallumon. Emellett az Y_r folyamat Karhunen–Loève sorfejtését használva megmutatjuk, hogy az

$$\int_a^b Y_r^2(t)\,dt \tag{7}$$

Cramér–von Mises típusú integrálfunkcionál abszolút folytonos, és sűrűségfüggvénye korlátos a valós egyenesen.

Aszimptotikus eredmények az empirikus generátor folyamatra

Mint azt korábban már említettük, Rémillard és Theodorescu (2000) megmutatta, hogy γ_n eloszlásban konvergál az Y folyamathoz tetszőleges nemnegatív egész értékű változó esetén, de van egy kisebb hiba a bizonyításban. Mindazonáltal az alapötletük érdekes, és a 4.5. fejezetben sikerül kijavítanunk a bizonyítást. Ehez tekintünk egy Ψ függvényt, mely a $D[0,1]$ tér egy D_0 alterén van definiálva, és a $C[0,1]$ téren veszi fel értékeit olyan módon, hogy teljesül a $\gamma_n=\Psi(\beta_n)$ és a $Y=\Psi(B)$ azonosság, ahol β_n egy egyenletes empirikus folyamat és B egy Brown híd. Megmutatjuk, hogy Ψ mérhető a Szkorohod topológiára nézve, és folytonos a $D_0\cap C[0,1]$ halmaz elemein a szuprémum metrikában. Mivel β_n eloszlásban konvergál a B folyamathoz, és a Brown híd egy valószínűséggel a $D_0\cap C[0,1]$ halmazba esik, kapjuk $\Psi(\beta_n)$ gyenge konvergenciáját a $\Psi(B)$ folyamathoz, amint n tart a végtelenbe.

Az empirikus valószínűségi generátor folyamattal kapcsolatban a fő eredményünk a következő erős approximáció. Megjegyezzük, hogy ez a tétel és a következő eredmények nem csak egész értékű, hanem tetszőleges nemnegatív értékű változóra teljesülnek.

4.18. Tétel. *Tekintsük egy tetszőleges nemnegatív értékű X véletlen változó $F(x)$, $x\in\mathbb{R}$, eloszlásfüggvényét. Egy alkalmas valószínűségi mezőn definiálható $X_1,X_2,\dots$ független változó F eloszlásfüggvénnyel, valamint az Y_r folyamat $Y_{r,1},Y_{r,2},\dots$ másolatai olyan módon, hogy*

$$\sup_{a\le t\le b}\big|\gamma_n^{(r)}(t)-Y_{r,n}(t)\big|=\mathcal{O}\big(n^{-1/2}\log n\big)$$

teljesül majdnem biztosan, amint $n\to\infty$.

A tételben szereplő véletlen elemeket a 2. fejezetben ismertetett KMT mező egyenletes eloszlású véletlen változóinak és Brown hídjainak segítségével definiáljuk. Ezután az approximációt az (1) konvergencia és az (5) egyenlőtlenség segítségével bizonyítjuk. Ehhez hasonló módon, Chung iterált logaritmustételéből kapjuk meg a következő iterált logaritmustételt a generátor folyamat deriváltjaira.

4.23. Tétel. *Tetszőleges nemnegatív értékű X változó esetén*

$$\limsup_{n\to\infty}\frac{\sup_{a\le t\le b}|\gamma_n^{(r)}(t)|}{(\log\log n)^{1/2}}\le\frac{C}{2^{1/2}}\qquad m.b.,$$

ahol a $C=C(a,b,r)$ pozitív konstans nem függ X eloszlásától.

A 4.23. Tétel következményeként kapjuk, hogy a g_n empirikus generátorfüggvény r. deriváltja egyenletesen konvergál az elméleti generátorfüggvény r. deriváltjához, továbbá konvergenciasebességet is tudunk bizonyítani. Kapjuk, hogy

$$\sup_{a\le t\le b}\left|g_n^{(r)}(t)-g^{(r)}(t)\right| = \mathcal{O}\Big(n^{-1/2}(\log\log n)^{1/2}\Big)$$

majdnem biztosan, amint $n\to\infty$.

A KMT approximációnak van egy fontos következménye a β_n empirikus folyamat bizonyos funkcionáljaira. Komlós, Major és Tusnády (1975) megmutatta, hogy ha a ψ funkcionál a $D[0,1]$ téren van értelmezve, rendelkezik a Lipschitz tulajdonsággal, valamint $\psi(B)$ abszolút folytonos korlátos sűrűségfüggvénnyel, akkor $\psi(\beta_n)$ eloszlásfüggvénye egyenletesen és $\mathcal{O}(n^{-1/2}\log n)$ sebességgel konvergál $\psi(B)$ eloszlásfüggvényéhez. A bizonyításuk ötleteit és az (5) egyenlőtlenséget felhasználva kapjuk a következő tételt az empirikus valószínűségi generátor folyamat funkcionáljaira.

4.20. Tétel. *Tekintsünk tetszőleges nemnegatív értékű X változót, és legyen ψ olyan funkcionál a $C[a,b]$ téren, mely rendelkezik a Lipschitz tulajdonsággal, tehát*

$$\left|\psi(h_1)-\psi(h_2)\right| \le M \sup_{a\le u\le b}\left|h_1(u)-h_2(u)\right|, \qquad h_1,h_2\in C[a,b],$$

egy M pozitív konstanssal. Ha az $\psi(Y_r)$ változó eloszlásfüggvénye abszolút folytonos korlátos sűrűségfüggvénnyel egy $[s,\infty)$, $s\in\mathbb{R}$, intervallumon, akkor

$$\sup_{x\ge s}\left|P\big(\psi(\gamma_n^{(r)})\le x\big)-P\big(\psi(Y_r)\le x\big)\right| = \mathcal{O}\big(n^{-1/2}\log n\big), \qquad n\to\infty.$$

Jegyezzük meg, hogy (6) szerint ez az eredmény alkalmazható a generátor folyamat szuprémum funkcionáljaira. Továbbá, néhány ponton változtatva a bizonyítás menetén megmutatjuk, hogy a $\gamma_n^{(r)}(t)$, $a\le t\le b$, folyamat négyzetintegráljának eloszlásfüggvénye konvergál a (7) változó eloszlásfüggvényéhez.

Finkelstein (1971) egyik ismert és szép eredménye, hogy a β_n empirikus folyamat legfeljebb 0 valószínűséggel konvergál a $D[0,1]$ térben, amint n tart a végtelenbe. Ez azért van, mert a folyamat relatív kompakt C_β limeszhalmazzal, ahol C_β azon $h\in D[0,1]$ függvények halmaza, melyek abszolút folytonosak, eltűnnek a 0 és az 1 pontban, és a Lebesgue mértékre vett $h'(u)$, $0\le u\le 1$, Radon–Nikodym deriváltjukra

$$\int_0^1 \big(h'(u)\big)^2\,du \le 1.$$

A következő eredmény az állítja, hogy az empirikus valószínűségi generátor folyamatnak hasonló az aszimptotikus viselkedése.

4.25. Tétel. *Tetszőleges nemnegatív értékű X változó és $r=0,1,\ldots$ egész esetén a $\gamma_n^{(r)}(t)$, $a\le t\le b$, folyamat relatív kompakt a $C[a,b]$ térben, és a limeszpontok halmaza*

$$C_r[a,b]=\left\{\int_{\mathbb{R}} x(x-1)\cdots(x-r+1)t^{x-r}\,dh\big(F(x)\big),\, a\le t\le b : h\in C_\beta\right\}.$$

A bootstrap generátor folyamat és konfidenciasávok

A 4.8. fejezetben az empirikus valószínűségi generátor folyamat bootstrap változatát vizsgáljuk. Tekintsünk független nemnegatív értékű $X_1, X_2, \dots$ változókat $F(x)$, $x \in \mathbb{R}$, eloszlásfüggvénnyel, továbbá $X^*_{1,n}, \dots, X^*_{m_n,n}$ Efron típusú, tehát nemparaméteres bootstrap mintát az $X_1, \dots, X_n$ megfigyelésekre nézve. Definiáljuk az empirikus valószínűségi generátorfüggvény bootstrap változatát a

$$g^*_{m_n,n}(t) = \frac{1}{n}\sum_{i=1}^{m_n} t^{X^*_{i,n}}, \qquad a \le t \le b,$$

formulával, és tekintsük a

$$\gamma^*_{m_n,n}(t) = n^{1/2}\big[g^*_{m_n,n}(t) - g_n(t)\big], \qquad a \le t \le b,$$

bootstrap empirikus valószínűségi generátor folyamatot. Alkalmazva Csörgő és Mason (1989) a klasszikus bootstrap empirikus folyamatra vonatkozó approximációs tételét a következő eredményt kapjuk.

4.26. Tétel. *Tekintsük egy tetszőleges nemnegatív értékű véletlen változó $F(x)$, $x \in \mathbb{R}$, eloszlásfüggvényét, és tegyük fel, hogy létezik C_1 és C_2 pozitív konstans, hogy*

$$C_1 < m_n/n < C_2, \qquad n = 1,2,\dots$$

Ekkor egy megfelelően gazdag valószínűségi mezőn definiálhatunk $X_1, X_2, \dots$ független véletlen változókat F eloszlásfüggvénnyel, és $X^*_{1,n}, \dots, X^*_{m_n,n}$, $n = 1,2,\dots$, bootstrap mintaelemeket, és az Y_r folyamat $Y^*_{r,1}, Y^*_{r,2}, \dots$ reprezentásait olyan módon, hogy

$$\sup_{a \le t \le b}\big|\gamma^{*(r)}_{m_n,n}(t) - Y^*_{r,m_n}(t)\big| = \mathcal{O}\big(\max\{l(m_n), l(n)\}\big),$$

ahol $l(n) = n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}$.

A tétel egy érdekes alkalmazása, hogy egy $X_1, \dots, X_n$ minta alapján aszimptotikusan korrekt konfidenciasávot konstruálhatunk az ismeretlen g valószínűségi generátor függvényre. Ez azt jelenti, hogy egy adott $0 < \alpha < 1$ mellett definiálhatunk egy $c_n(\alpha)$, $n = 1,2,\dots$, sorozatot olyan módon, hogy

$$P\Big(g_n(t) - c_n(\alpha) \le g(t) \le g_n(t) + c_n(\alpha), a \le t \le b\Big) \to 1 - \alpha$$

teljesüljön, amint az n mintaméret tart a végtelenbe.

Becsült paraméteres valószínűségi generátor folyamatok

A 2. fejezetben ismertettünk egy bootstrap algoritmus, mellyel tesztelhetjük független és azonos eloszlású változók illeszkedését egy $\mathcal{F} = \{F(x,\theta), x \in \mathbb{R}, \theta \in \Theta \subseteq \mathbb{R}^d\}$ paraméteres eloszláscsaládhoz. Sajnos léteznek olyan családok, melyek $F(x,\theta)$ paraméteres eloszlásfüggvénye nem írható fel egyszerű alakban, és ematt nehéz az $\hat{\alpha}_n$ becsült paraméteres empirikus folyamatot és a kapcsolatos bootstrap folyamatokat alkalmazni. Ezzel szemben, sok esetben a valószínűségi generátorfüggvény kezelhető. Ez az észrevétel vezérel minket arra, hogy definiáljuk a valószínűségi generátor folyamat becsült paraméteres változatát.

Tegyük fel, hogy $\mathcal{F}$ csak nemnegatív értékű eloszlásokat tartalmaz, és tekintsük a

$$g(t,\theta) = \int_{\mathbb{R}} t^x \, dF(x,\theta)\,, \qquad a' \leq t \leq b'\,, \quad \theta \in \Theta\,,$$

paraméteres generátorfüggvényt, ahol a' és b' alkalmas konstans. Továbbá, tekintsünk független $X_1, X_2, \ldots$ változókat közös $F(x,\theta_0)$, $x \in \mathbb{R}$, eloszlásfüggvénnyel, ahol $\theta_0 \in \Theta$, és legyen $\hat{\theta}_n$ a θ_0 paraméter becslése az első n megfigyelést alapul véve. Ha g_n jelöli az empirikus generátorfüggvényt, akkor a becsült paraméteres empirikus valószínűségi generátor folyamat a

$$\hat{\gamma}_n(t) = n^{1/2}\big[g_n(t) - g(t,\hat{\theta}_n)\big] = \int_{\mathbb{R}} t^x \, d\hat{\alpha}_n(x)\,, \qquad a' \leq t \leq b'\,,$$

formulával definiálható. Emellett, a (4) formula $G(x)$, $x \in \mathbb{R}$, Gauss folyamatával legyen

$$\hat{Y}(t) = \int_{\mathbb{R}} t^x \, dG(x)\,, \qquad a' \leq t \leq b'\,.$$

Tekintsük Burke, Csörgő, Csörgő és Révész (1979) approximációját az $\hat{\alpha}_n$ becsült paraméteres empirikus folyamatra, melyet a (3) formulában ismertettünk. Ennek alkalmazásával megkonstruálhatjuk az X_i sorozat egy reprezentánsát, valamint az $\hat{Y}$ folyamat $\hat{Y}_1, \hat{Y}_2, \ldots$ másolatait olyan módon, hogy teljesüljön a

$$\sup_{0 \leq t < 1} \big|\hat{\gamma}_n(t) - \hat{Y}_n(t)\big| \xrightarrow{P} 0\,, \qquad n \to \infty\,,$$

gyenge approximáció. Ez az eredmény a 4.29. Tétel, mely a (3) approximáció feltételei, valamint bizonyos, a $\nabla_\theta F(x,\theta)$ függvény aszimptotikus viselkedésére vonatkozó további feltevések teljesülése esetén áll fenn.

Fontos észrevenni, hogy a $\hat{\gamma}_n$ folyamat gyakorlati alkalmazása egy újabb nehézséghez vezethet. Mint azt már korábban elmondtuk, a legtöbb esetben egy $\psi(\hat{\alpha}_n)$ teszt statisztika kritikus értékeit nem lehet elméleti számítások útján meghatározni, és ezen statisztikák általában nem is eloszlásmentesek. Mivel az $\hat{Y}$ folyamatot egy G szerinti integrállal definiáljuk, ugyanez a probléma $\psi(\hat{\gamma}_n)$ alakú funkcionálokra is felléphet. Egy lehetséges megoldásként bevezetjük a $\hat{\gamma}_n$ becsült paraméteres empirikus valószínűségi generátor folyamat bootstrappelt változatait, és a 3.2. és a 3.3. Tétel approximációit használva belátjuk, hogy a folyamatok konvergálnak eloszlásban. Következményként kapjuk, hogy egy nemnegatív megfigyelésekből álló $X_1, \ldots, X_n$ minta illeszkedése az $\mathcal{F}$ családhoz tesztelhető a korábban ismertetett bootstrap algoritmussal azzal a módosítással, hogy az $\hat{\alpha}_n$ empirikus folyamat és bootstrap változatai helyett a $\hat{\gamma}_n$ folyamatot és bootstrap változatait kell alkalmazni.

Hivatkozások

Babu, G.J. és Rao, C.R. (2004) Goodness-of-fit tests when parameters are estimated. *Sankhyā: The Indian Journal of Statistics* **66** 63–74.

Burke, M.D., Csörgő, M., Csörgő, S. és Révész, P. (1979) Approximations of the empirical process when parameters are estimated. *The Annals of Probability* **7** 790–810.

Csörgő, M., Csörgő, S., Horváth, L. és Mason, D.M. (1986) Supnorm convergence of the empirical process indexed by functions and applications. *Probability and Mathematical Statistics* **7** 13–26.

Csörgő, S. (1981) Limit behaviour of the empirical characteristic function. *The Annals of Probability* **9** 130–144.

Csörgő, S. és Mason, D.M. (1989) Bootstraping empirical functions. *The Annals of Statistics* **17** 1447–1471.

Durbin, J. (1973) Weak convergence of the sample distribution, when parameters are estimated. *The Annals of Statistics* **1** 279–290.

Efron, B. (1979) Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* **7** 1–26.

Finkelstein, H. (1971) The law of the iterated logarithm for empirical processes. *The Annals of Mathematical Statistics* **42** 607–615.

Komlós, J., Major, P. és Tusnády, G. (1975) An approximation of partial sums of independent R.V.'s, and the sample D.F.I. *Z. Wahrscheinlichkeitstheorie und verw. Gebiete* **32** 111–131.

Rémillard, B. és Theodorescu, R. (2000) Inference based on the empirical probability generating function for mixtures of Poisson distributions. *Statistics & Decisions* **18** 349–366.

Szűcs, G. (2005) Approximations of empirical probability generating processes. *Statistics & Decisions* **23** 67–80.

Szűcs, G. (2008) Parametric bootstrap tests for discrete distributions. *Metrika* **67** 63–81.

Szűcs, G. (20??) Non-parametric bootstrap tests for parametric distribution families. *Acta Scientarium Mathematicarum*, accepted.

Stute, W., Manteiga, W.G. és Quindimil, M.P. (1993) Bootstrap based goodness-of-fit-tests. *Metrica* **40** 243–256.