Egyetemi doktori (PhD) értekezés tézisei

Határérték tételek véletlen mezőkre

Karácsony Zsolt

Témavezető: Dr. Fazekas István

Debreceni Egyetem
Matematika- és Számítástudományok Doktori Iskola
Debrecen, 2010.

Egyetemi doktori (PhD) értekezés tézisei

Határérték tételek véletlen mezőkre

Karácsony Zsolt

Témavezető: Dr. Fazekas István

Debreceni Egyetem
Matematika- és Számítástudományok Doktori Iskola
Debrecen, 2010.

1. fejezet

Bevezetés

A határérték-tételekhez tartozó témakörök a valószínűségszámítás legfontosabb, legtöbbet kutatott fejezetei közé tartoznak. Dolgozatomban határérték-tételeket bizonyítok különböző feltételek esetén. A könnyebb követhetőség kedvéért a tézisekbeli tétel, definíció számozás megegyezik a disszertációmban használt számozással.

Az értekezés első részében autoregressziós típusú martingál mezőket tanulmányozok. Közismert, hogy a martingáloknak sokfajta általánosítása létezik. A Doob-féle martingál konvergencia tételt többparaméterű esetre Cairoli általánosította (lásd [Cai70]). Vektor értékű martingálokra pedig Chatterji terjesztette ki (lásd [Cha68]). Az ő eredményeit Fazekas István általánosította többparaméterű esetre (lásd [Faz83]).

MacQueen 1973-ban bevezette a lineáris martingál fogalmát ([MQ73]). Megmutatta, hogy az általa vizsgált folyamat sokkal közelebb áll a martingálokhoz, mint a stacionárius folyamatokhoz, ha az együtthatókra bizonyos feltételek teljesülnek. Az általa tekintett speciális esetben a Doob-féle martingál konvergencia tétel is teljesül.

Kétparaméterű esetre való kiterjesztésével Fazekas István foglalkozott a 80-as évek közepén. Kiindulva Fazekas István korábbi eredményeiből ([Faz87], [Faz88]), sikerült kiterjeszteni az autoregressziós típusú martingálokra vonatkozó eredményeket d-paraméterű esetre is. Itt már magának a d-paraméteres autoregressziós típusú martingál mező fogalmának a kialakítása is saját eredmény, nemcsak ezen folyamat alkalmas reprezentációja és a konvergencia tételek igazolása. A bizonyításban új ötleteket kellett felhasználni, mert a korábbi eredményekben használt módszerek nem voltak alkalmazhatóak a d-paraméterű esetben. A 2-nél magasabb dimenziós paramétertér esetén bevezetett új formalizmus a vec operátoron és a Kronecker-szorzaton alapul. Felhasználtam továbbá a Burkholder-egyenlőtlenség általános alakját is, amely Fazekas István cikkében található meg (lásd [Faz05]).

A disszertáció második és harmadik részében határérték-tételeket bizonyítok úgynevezett „sűrűsödő-növekvő” sémára. Az aszimptotikus eredmények általában növekvő tartományra vonatkoznak, miközben a megfigyelési helyek távolsága fix. Véletlen mezők esetén azonban számos esetben adott tartományon belül tekintenek egyre több megfigyelést. Lahiri kezdeményezte a két fenti hozzáállás egyesítését.

A „sűrűsödő-növekvő” tartomány esetén a megfigyeléseink egyre sűrűbbek, miközben a tartomány „mérete” is tart a végtelenhez. Ezt a gondolatot ilyen explicit formában először Lahiri vetette fel, majd egyre többen kezdtek foglalkozni ezzel a témával, hiszen ez a megközelítés hasznos lehet a földtudományok, meteorológia, környezetvédelem, képfeldolgozás stb. területein is. Ezen tudományterületeken számos olyan folyamatot tanulmányoznak, amely térben vagy időben folytonosan változik. Viszont a gyakorlatban nem tudjuk folytonosan megfigyelni a folyamatokat, ezért véges adathalmazokat és diszkrét becsléseket használunk.

Ezekben a fejezetekben gyengén függő valószínűségi mező létezését feltételezzük. Ehhez használjuk az úgynevezett α-keverő tulajdonságot.

Az első központi határeloszlás-tételeket véletlen keverő sorozatokra Rosenblatt bizonyította ([Ros56]) α-keverő esetre, valamint Ibragimov ([Ibr62]) a φ-keverő esetre. A dolgozatom szempontjából fontos előzmény, hogy Ibragimov és Linnik ([IL71]) α-keverő folyamatokra, Guyon ([Guy95]) α-keverő mezőkre, Fazekas és Kukush ([FK00]) pedig α-keverő mezők esetén „sűrűsödő-növekvő” sémára bizonyítanak határérték-tételeket. Fazekas és Kukush nem közöl részletes bizonyítást, ezért jelen munkám második részében rögzítem a bizonyítás lépéseinek részleteit, valamint foglalkozom a tétel p-dimenziós kiterjesztéseivel is.

A dolgozatom harmadik részének témája a regressziós függvény magfüggvényes becslése. A magfüggvényes becsléseket széles körben tanulmányozza a szakirodalom. Nadaraya és Watson ([Nad64], [Wat64]) 1964-ben közölt eredményeit számos cikkben feldolgozták és általánosították. Ezeket az eredményeket többek között Rao ([Rao83]), Devroye és Győrfi ([DG85]), valamint Bosq ([Bos98]) foglalta össze. A magfüggvényes becslések egy fontos tulajdonsága az aszimptotikus normalitás, melyet több cikkben is tanulmányoztak. Például 1972-ben Schuster ([Sch72]) diszkrét esetben bizonyított aszimptotikus normalitást, míg 2001-ben Cai ([Cai01]) folytonos esetben látott be hasonló eredményt. A regressziós függvény magfüggvényes becslésének vizsgálata szorosan kapcsolódik a sűrűségfüggvény becsléséhez (lásd [CL86], [BMP99], [FC06]). Ezen eredményekből kiindulva sikerült a regressziós függvény aszimptotikus normalitását bizonyítani α- keverő mezőkre „sűrűsödő-növekvő” tartományt tekintve (4.1. Tétel). Ezen tételt tekintem a dolgozat legfontosabb eredményének. Dolgozatomból, illetve Fazekas István korábbi eredményeiből ([FC06]) kiderül, hogy ez az eset a diszkrét és a folytonos idejű esetek között helyezkedik el. Pontosabban szólva, a határeloszlás kovariancia struktúrája a diszkrét és a folytonos idejű határeloszlások lineáris kombinációjaként adódik. Dolgozatomat két példával zárom, melyen keresztül szemléltetem a határeloszlást. A numerikus példák jól mutatják az előbb jelzett speciális kovariancia struktúrát.

2. fejezet

Autoregressziós típusú martingál mezők

A disszertáció első részében a lineáris martingálok többparaméteres kiterjesztésével foglalkozom, és fő eredményként majdnem mindenütti konvergenciát bizonyítok. Ez az eredmény speciális esetként tartalmazza Fazekas István eredményeit is (lásd [Faz87], [Faz88]).

Legyen $(\Omega, \mathcal{F}, \mathbb{P})$ valószínűségi mező, $X_{\mathbf{n}}, \mathbf{n} \in \mathbb{N}^d$, valószínűségi változók egy d-paraméterű sorozata, és legyen $\mathcal{F}_{\mathbf{n}} \subseteq \mathcal{F}$ σ-algebra minden $\mathbf{n} \in \mathbb{N}^d$-re.

Tegyük fel, hogy $\mathcal{F}_{\mathbf{m}} \subseteq \mathcal{F}_{\mathbf{n}}$ minden $\mathbf{m} \leq \mathbf{n}$-re. Továbbá, hogy $X_{\mathbf{n}}$ $\mathcal{F}_{\mathbf{n}}$-mérhető és integrálható minden $\mathbf{n} \in \mathbb{N}^d$-re. Azt mondjuk, hogy $(X_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}})$ martingál, ha $\mathbb{E}(X_{\mathbf{n}+\mathbf{k}} \,|\mathcal{F}_{\mathbf{n}}) = X_{\mathbf{n}}$ teljesül m.m., minden $\mathbf{n} \in \mathbb{N}^d$ és $\mathbf{k} \in \mathbb{N}_0^d$ esetén.

Használjuk a következő feltételt. Tetszőleges véges várható értékű η-ra:

$$\mathbb{E}\left(\eta \mid \mathcal{F}_{\mathbf{n}}\right) = \mathbb{E}\Big(\dots\ \mathbb{E}\Big(\eta \mid \mathcal{F}_{\mathbf{n}}^{(i_1)}\Big)\ \dots \mid \mathcal{F}_{\mathbf{n}}^{(i_d)}\Big) \tag{2.3}$$

az $(1, \dots, d)$ minden $(i_1, \dots, i_d)$ permutációja esetén, ahol $\mathcal{F}_{\mathbf{n}}^{(i)} = \sigma\{\mathcal{F}_{\mathbf{l}} : l_i = n_i\}$ minden rögzített $\mathbf{n}$-re és i-re (az $\mathcal{F}_{\mathbf{n}}^{(i)}$ képletben (i) a megfelelő koordináta).

Első eredményem többparaméterű B-értékű martingálok egyenletes konvergenciájáról szól. Ez a tétel a [Faz83] Theorem 4.4. egy változata, melyben az egyenletes konvergenciát nem vizsgálták.

2.6. Tétel. *Legyen B egy valós szeparábilis Banach-tér. Legyen $(X_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}})$, $\mathbf{n} \in \mathbb{N}^d$, B-értékű martingál. B rendelkezzen a Radon-Nikodym tulajdonsággal vagy legyen $X_{\mathbf{n}} = \mathbb{E}(X|\mathcal{F}_{\mathbf{n}})$, $\mathbf{n} \in \mathbb{N}^d$ alakú, valamely $X \in L^1(\mathcal{F}, B)$-re. Tegyük fel, hogy*

$$\sup_{\mathbf{n}} \mathbb{E}\|X_{\mathbf{n}}\|(\log^+ \|X_{\mathbf{n}}\|^{d-1}) < \infty.$$

Ekkor létezik egy olyan A esemény, melyre $\mathbb{P}(A) = 1$ és minden $\omega \in A$-ra teljesül a következő: ha az $\mathbf{n}$ tetszőleges koordinátái konvergálnak a ∞-hez, míg a *maradék koordináták rögzítettek maradnak, akkor $X_{\mathbf{n}}(\omega)$ egyenletesen konvergál. (A határérték egy, a rögzített koordinátáktól függő valószínűségi változó.)*

Ahhoz, hogy leírjuk a vizsgálandó $\xi_{\mathbf{n}}, \mathbf{n} \in \mathbb{N}^d$, véletlen mező struktúráját, szükséges a Kronecker-szorzat (jelölje $\otimes$) és a vec operátor fogalma (lásd [MN88]). Magának a d-paraméteres autoregressziós típusú martingál mező fogalmának a kialakítása is saját eredmény.

2.9. DEFINÍCIÓ. *A $\{\xi_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}}\}$, $\mathbf{n} \in \mathbb{N}^d$, folyamatot autoregresszív típusú martingál mezőnek nevezünk, ha teljesülnek a következő feltételek:*

1. $\xi_{\mathbf{n}}$ $\mathcal{F}_{\mathbf{n}}$-mérhető és integrálható minden $\mathbf{n} \in \mathbb{N}^d$ esetén,

2.

$$\begin{aligned} \mathbb{E}\left(\xi_{\mathbf{n}} | \mathcal{F}^{(j)}_{\mathbf{n}-\mathbf{e}_j}\right) &= a_1^{(j)}(n_j)\xi_{\mathbf{n}-\mathbf{e}_j} \\ &\quad + a_2^{(j)}(n_j)\xi_{\mathbf{n}-2\mathbf{e}_j} + \ldots + a_m^{(j)}(n_j)\xi_{\mathbf{n}-m\cdot\mathbf{e}_j} \end{aligned} \tag{2.10}$$

minden $\mathbf{n}$-re és j-re, ahol m egy rögzített pozitív egész szám, $n_j > m$, $j = 1, \ldots, d$, továbbá $\mathbf{e}_j = (0, \ldots, \underbrace{1}_{j}, \ldots 0) \in \mathbb{N}_0^d$ a j-edik egységvektor, $j = 1, \ldots, d$, és $a_i^{(j)}(n_j)$ nem negatív, nem véletlen együtthatók, amelyekre teljesül, hogy

$$\sum_{i=1}^{m} a_i^{(j)}(n_j) = 1$$

minden $n_j = m+1, m+2, \ldots,$ $j = 1, \ldots, d$ esetén.

Ha az $a_k^{(j)}(l)$ értékek nem függnek l-től, akkor $\xi_{\mathbf{n}}$-t *homogén autoregresszív martingál mező*nek nevezzük.

2.10. ÁLLÍTÁS. *Legyen $(\xi_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}})$ a 2.9. Definícióban bevezetett autoregresszív martingál mező és $X_{\mathbf{n}}$ a $\xi_{\mathbf{n}}$-ből származó tömbértékű véletlen mező. Ekkor*

$$\operatorname{vec}\left[\mathbb{E}\left(X_{\mathbf{n}} \mid \mathcal{F}^{(j)}_{\mathbf{n}-\mathbf{e}_j}\right)\right] = \Big(\underbrace{I \otimes \cdots \otimes I}_{d-j} \otimes A_j^{(n_j)} \otimes \underbrace{I \otimes \cdots \otimes I}_{j-1}\Big) \cdot \operatorname{vec}(X_{\mathbf{n}-\mathbf{e}_j}) \tag{2.11}$$

teljesül minden $\mathbf{n}$-re, ahol $n_j > m$, $j = 1, \ldots, d$, és $A_j^{(l)}$ minden $j = 1, \ldots, d$, $l = m+1, m+2, \ldots$ esetén a következő $m \times m$-es mátrixot jelöli

$$A_j^{(l)} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ a_m^{(j)}(l) & a_{m-1}^{(j)}(l) & \cdots & \cdots & a_1^{(j)}(l) \end{pmatrix}.$$

2.11. Állítás. *Legyen $X_{\mathbf{n}}$ tömbértékű véletlen mező, ami teljesíti a (2.11) egyenlőséget. Feltesszük továbbá, hogy* (2.3) *is teljesül. Ekkor az $(X_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}})$ folyamatra a következő egyenlőség teljesül:*

$$\operatorname{vec}[\mathbb{E}(X_{\mathbf{n+t}}|\mathcal{F}_{\mathbf{n}})] = (A_d(n_d+t_d, n_d) \otimes \cdots \otimes A_2(n_2+t_2, n_2) \otimes A_1(n_1+t_1, n_1)) \cdot \operatorname{vec}(X_{\mathbf{n}}), \tag{2.16}$$

ahol

$$A_j(n_j+t_j, n_j) = A_j^{(n_j+t_j)} \cdot A_j^{(n_j+t_j-1)} \cdots A_j^{(n_j+1)} \tag{2.17}$$

minden $n_j > m$, $j = 1, \ldots, d$ és $\mathbf{n} \in \mathbb{N}^d$, $\mathbf{t} \in \mathbb{N}_0^d$ esetén.

Általánosítva a (2.16)-os tulajdonságot, a következő fogalmat kapjuk.

2.12. Definíció. *Egy tömbértékű $(X_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}})$, $\mathbf{n} \in \mathbb{N}^{\mathbf{d}}$, folyamatot A-martingál mezőnek nevezünk, ha teljesülnek a következők:*

1. *$X_{\mathbf{n}}$ $\mathcal{F}_{\mathbf{n}}$-mérhető és integrálható minden $\mathbf{n} \in \mathbb{N}^d$-re,*
2. *a (2.16) egyenlőség teljesül minden $\mathbf{n}, \mathbf{t} \in \mathbb{N}^d$-re, ahol az $A_j(n_j+t_j, n_j)$ mátrixok kielégítik a (2.17)-et. (Minden vizsgált $A_j^{(l_j)}$ $m \times m$ típusú mátrix nem véletlen.)*

2.15. Állítás. *Legyen $X_{\mathbf{n}}$ A-martingál mező, mely teljesíti* (2.3)*-at. Ekkor* $\operatorname{vec}(X_{\mathbf{n}})$ *előállítható a következő alakban:*

$$\operatorname{vec}(X_{\mathbf{n}}) = \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \cdots \sum_{k_d=1}^{n_d} A_d(n_d, k_d) \otimes \cdots \otimes A_1(n_1, k_1) \cdot \operatorname{vec}(\Delta_{\mathbf{k}}),\ \mathbf{k} \in \mathbb{N}^d,\ \mathbf{k} \leq \mathbf{n}, \tag{2.20}$$

ahol $A_j(k_j, k_j) = I$, $j = 1, \ldots, d$, $\Delta_{\mathbf{n}}$ jelöli az $X_{\mathbf{n}}$ A-martingál mező martingál differencia mezőjét.

A következő feltételekre szükségünk lesz a tételeinkben.
Tegyük fel, hogy

$$A_j(i_j+t_j, i_j) \to A_j(\infty, i_j), \text{ ha } t_j \to \infty, \text{ minden } i_j, j \in \mathbb{N} \tag{2.21}$$

és a konvergencia „gyors” a következő értelemben:

$$\|A_j(\infty, i_j) - A_j(i_j+t_j, i_j)\| \leq c_{t_j}^{(j)}, \qquad \forall i_j, j \in \mathbb{N}, \tag{2.22}$$

ahol $\sum_{t_j=1}^{\infty} c_{t_j}^{(j)} < \infty$ minden j-re.

Az $A_j(\infty,k_j) = \lim_{t_j\to\infty} A_j(k_j+t_j,k_j)$ határérték mátrixokról feltesszük, hogy létezik egy olyan pozitív C konstans úgy, hogy

$$\|[A_d(\infty,k_d)\otimes\cdots\otimes A_1(\infty,k_1)]\operatorname{vec}(\Delta_{\mathbf{k}})\| \geq C\|\operatorname{vec}(\Delta_{\mathbf{k}})\|, \tag{2.23}$$

teljesül, minden $\mathbf{k}$-ra.

Továbbá tegyük fel, hogy

$$\begin{aligned}\Bigg\|\sum_{\mathbf{k}_S\leq\mathbf{n}_S} & A_d(\infty,k_d)\otimes A_{d-1}(\infty,k_{d-1})\otimes\cdots\otimes A_1(\infty,k_1)\cdot\operatorname{vec}(\Delta_{\mathbf{k}})\Bigg\| \\ &\geq C\cdot\Bigg\|\sum_{\mathbf{k}_S\leq\mathbf{n}_S} D_{k_d}\otimes D_{k_{d-1}}\otimes\cdots\otimes D_{k_1}\cdot\operatorname{vec}(\Delta_{\mathbf{k}})\Bigg\|,\end{aligned} \tag{2.24}$$

ahol $\mathbf{k}_S$ a $\mathbf{k}\in\mathbb{N}^d$ olyan koordinátáit tartalmazza, amelyek $S\subseteq\{1,\dots,d\}$ elemei, továbbá $D_{k_l}=I$, ha $l\notin S$, illetve $D_{k_l}=A_l(\infty,k_l)$, ha $l\in S$.

2.17. Tétel. *Tegyük fel, hogy az $(X_{\mathbf{n}},\mathcal{F}_{\mathbf{n}})$ A-martingál mező teljesíti a* (2.3), (2.22), (2.23) *és* (2.24) *feltételeket. Ha*

$$\sup_{\mathbf{k}\in\mathbb{N}^d}\mathbb{E}\|\operatorname{vec}(X_{\mathbf{k}})\|\left[\log^+(\|\operatorname{vec}(X_{\mathbf{k}})\|)\right]^{d-1}<\infty, \tag{2.25}$$

akkor $X_{\mathbf{n}}$ konvergens m.m., ha $n_j\to\infty$ minden j esetén. Továbbá, ha $d\geq 2$, akkor $X_{\mathbf{n}}$ konvergens $\mathcal{L}_1$-ben, ha $n_j\to\infty$ minden j esetén.

2.20. Tétel. *Tegyük fel, hogy az $(X_{\mathbf{n}},\mathcal{F}_{\mathbf{n}})$ A-martingál mező teljesíti a* (2.3), (2.22) *és* (2.24) *feltételeket, valamint*

$$\|A_j(i_j,u_j)\|<K<\infty, \tag{2.30}$$

ha $i_j>u_j$, $j=1,\dots,d$. Ha $\sup_{\mathbf{n}}\mathbb{E}\|\operatorname{vec}(X_{\mathbf{n}})\|^{\alpha}<\infty$, ahol $\alpha>1$, akkor $X_{\mathbf{n}}$ konvergens $\mathcal{L}_\alpha$-ban, $\mathbf{n}\to\infty$ esetén.

Végül kimondhatjuk a disszertáció első részének fő eredményét.

2.21. Tétel. *Legyen $(\xi_{\mathbf{n}},\mathcal{F}_{\mathbf{n}})$ homogén autoregresszív martingál mező és tegyük fel, hogy* (2.3) *teljesül. Továbbá tegyük még fel, hogy $a_m^{(j)}>0$, minden $j=1,\dots,d$ esetén és a $\{k: 1\leq k\leq m, a_k^{(j)}>0\}$ számok legnagyobb közös osztója 1.*

a) Ha $\sup_{\mathbf{n}}\mathbb{E}|\xi_{\mathbf{n}}|\left[\log^+|\xi_{\mathbf{n}}|\right]^{d-1}<\infty$, akkor $\xi_{\mathbf{n}}$ konvergens m.m. ha $\mathbf{n}\to\infty$, továbbá ha $d\geq 2$, akkor ξ_n konvergens $\mathcal{L}_1$-ben is.

b) Legyen $\alpha>1$. Ha $\sup_{\mathbf{n}}\mathbb{E}|\xi_{\mathbf{n}}|^{\alpha}<\infty$, akkor $\xi_{\mathbf{n}}$ konvergens $\mathcal{L}_\alpha$-ban (és m.m.), ha $\mathbf{n}\to\infty$.

3. fejezet

Központi határeloszlás-tételek keverő véletlen mezőkre

A független valószínűségi változók sorozataira vonatkozó határérték-tételek függő esetre történő kiterjesztésével számos mű foglalkozott. A gyenge függőség közismert feltételei közül a keverési feltételek fontos szerepet játszanak. Ibragimov és Linnik 1971-ben ([IL71]) központi határeloszlás-tételt igazoltak olyan stacionárius sorozatokra, amelyek bizonyos α-keverő feltételeket teljesítenek. Az eredményüket Bolthausen ([Bol82]) és Guyon ([Guy95]) terjesztette ki α-keverő véletlen mezőkre. Guyon eredményeit Fezekas és Kukush vitte át „sűrűsödő-növekvő" sémára. Fazekas István ([Faz03]) korlátos, Fazekas-Kukush ([FK00]) az egyenletesen integrálható esettel foglalkozott. Ezek a cikkek nem tartalmazzák az említett tételek részletes bizonyításait, hanem csak rövid vázlatot közölnek.

Disszertációm második részében az említett tételek részletes bizonyításával foglalkozom. Az említett bizonyítások tanulsága az, hogy Ibragimov és Linnik ([IL71]), valamint Guyon ([Guy95]) eredeti gondolatmenete csak akkor alkalmazható, ha a véletlen mezőre teljesül egy bizonyos egyenletesen integrálhatósági feltétel. Guyon nem tette fel az egyenletesen integrálhatóságot, sőt nem is írta le a korlátos esettől az általános esethez vezető lépéseket. Ezért nem világos, hogy az eredménye teljesül-e az általános esetben is. Megjegyezzük, hogy Ibragimov és Linnik feltételezte a stacionaritást, tehát az ő bizonyításuk teljes.

A megfigyelések sémája az alábbi. Legyenek $T_1, T_2, \ldots,$ és T_∞ tartományok $\mathbb{R}^d$-ben. Tegyük fel, hogy $T_1 \subset T_2 \subset T_3 \subset \ldots,$ $\bigcup_{i=1}^{\infty} T_i = T_\infty$, és T_i kompakt minden i-re, továbbá T_∞ Lebesgue-mértéke végtelen. Legyen $\{\varepsilon(\mathbf{x}),\ \mathbf{x} \in T_\infty\}$ egy véletlen mező. Az n-edik megfigyelés halmaz (minta) az $\varepsilon(\mathbf{x})$ mező $\mathbf{x}_\mathbf{k} \in T_n$ pontokban felvett értékeiből áll, ahol $\mathbf{k} \in \mathcal{D}_n \subset \mathbb{Z}^d$. Az $\mathbf{x}_\mathbf{k}$ pontok kiválasztása a következő.

Osszuk fel $\mathbb{R}^d$-t az alábbi téglákra

$$\Delta_n(\mathbf{k}) = \prod_{j=1}^{d} \Big(\frac{k_j}{N_{jn}}, \frac{k_j+1}{N_{jn}} \Big],$$

ahol $\mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d$ d-dimenziós egész koordinátájú pont, $\{N_{jn}\}$ pedig pozitív egészek növekvő, nem korlátos sorozata minden rögzített $j = 1, \dots, d$-re. Az n-edik mintavételi helyeket, azaz az $\{\mathbf{x_k},\ \mathbf{k} \in \mathcal{D}_n\}$, halmazt úgy kapjuk, hogy egy $\mathbf{x_k}$ pontot választunk minden nem üres $\Delta_n(\mathbf{k}) \cap T_n$ halmazból. Valójában minden $\mathbf{x_k} = \mathbf{x_k}^{(n)}$ függ n-től is. Azért, hogy elkerüljük a bonyolult jelöléseket, gyakran elhagyjuk az (n) felsőindexet. Tegyük fel, hogy $\lim_{n\to\infty} |\mathcal{D}_n| = \infty$.

Ha a megfigyelések helyei egyre sűrűbbek és sűrűbbek lesznek a tartományok egy növekvő sorozata esetén, akkor ezt a sémát „sűrűsödő-növekvő” sémának nevezzük.

Definiáljuk a diszkrét paraméterű $Y_n(\mathbf{k})$ vektor mezőt a következő módon. Minden $n = 1, 2, \dots,$ és minden $\mathbf{k} \in \mathcal{D}_n$ esetén legyen

$$Y_n(\mathbf{k}) \text{ az } \varepsilon(\mathbf{x_k}^{(n)}) \text{ Borel-mérhető függvénye.} \tag{3.1}$$

Emlékeztetünk az α-keverési együttható definíciójára. Legyen $\mathcal{A}$ és $\mathcal{B}$ két σ-algebra $\mathcal{F}$-ben. Az $\mathcal{A}$ és $\mathcal{B}$ halmazok α-keverési együtthatójának nevezzük az

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup\{|\mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(AB)| \ : \ A \in \mathcal{A}, B \in \mathcal{B}\}$$

értéket. Az $\{\varepsilon(\mathbf{x}) \ : \ \mathbf{x} \in T_\infty\}$ mező α-keverési együtthatója pedig:

$$\alpha(r, u, v) = \sup\{\alpha(\mathcal{F}_{I_1}, \mathcal{F}_{I_2}) \ : \ \varrho(I_1, I_2) \ge r, \ |I_1| \le u, \ |I_2| \le v\},$$

ahol a szuprémum a T_∞ minden I_1 és I_2 véges részhalmazára veendő, továbbá $\mathcal{F}_{I_i} = \sigma\{\varepsilon(\mathbf{x}) \ : \ \mathbf{x} \in I_i\}$, $i = 1, 2$.

Azt mondjuk, hogy az $\{\varepsilon(\mathbf{x})\}$ véletlen mező α-keverő, ha a keverő együtthatók kielégítenek bizonyos feltételeket. Ezen feltételek mindegyike a mező gyenge függőségét jelenti, azaz azt, hogy $\alpha(r, u, v)$ kicsi, ha r nagy.

A tételeinkben felhasználjuk a következő feltételeket:

$$\int_0^\infty s^{d-1} \alpha^{\frac{\tau}{2+\tau}}(s, 1, 1) ds < \infty, \quad \text{ha} \quad 0 < \tau < 1, \tag{3.2}$$

$$\int_0^\infty s^{d-1} \alpha(s, i, j) ds < \infty, \quad \text{ha} \quad i + j \le 4, \tag{3.3}$$

$$\alpha(s, 1, \infty) = \mathrm{o}(s^{-d}), \quad \text{ha} \quad s \to \infty, \tag{3.4}$$

$$\Lambda_n = \mathrm{O}(\lambda_n), \quad \text{ha} \quad n \to \infty, \tag{3.5}$$

ahol

$$\Lambda_n = \max_{1\le j\le d} N_{jn}, \quad \lambda_n = \min_{1\le j\le d} N_{jn}. \tag{3.6}$$

Előljáróban a fő eredményünk által leírt szituációról a következőket jegyezzük meg. Egyrészt arra az esetre koncentrálunk, amikor $\varepsilon(\mathbf{x})$ és $\varepsilon(\mathbf{y})$ nem függetlenek, ha $\mathbf{x}$ és $\mathbf{y}$ közel vannak egymáshoz, ezért a tételünk nem fedi le azon eseteket, amikor $Y_n(\mathbf{k})$-k függetlenek és azonos eloszlásúak. Másrészről, ha $\varepsilon(\mathbf{x})$ egy stacionárius mező folytonos kovariancia függvénnyel és pozitív szórással, akkor a kovariancia közel van egy rögzített pozitív számhoz egy kicsi hipertéglalapon belül. Ezt a szituációt tételezzük fel a háttérben. Emlékeztetőül, $\mathcal{D}_n$ véges halmazok sorozata $\mathbb{Z}^d$-ben, ahol $\lim_{n\to\infty} |\mathcal{D}_n| = \infty$ teljesül.

3.5. Tétel. *Legyen $\varepsilon(\mathbf{x})$ véletlen mező és legyen $Y_n(\mathbf{k})$ az $\varepsilon(\mathbf{x}_{\mathbf{k}}^{(n)})$ Borel-mérhető függvénye, $\mathbf{k} \in \mathcal{D}_n$. Legyen az $\{|Y_n(\mathbf{k})| : \mathbf{k} \in \mathcal{D}_n, n = 1, 2, \dots\}$ család egyenletesen korlátos. Legyen $S_n = \sum_{\mathbf{k}\in\mathcal{D}_n} Y_n(\mathbf{k})$, $n = 1, 2, \dots$, valamint $\sigma_n^2 = \mathrm{var}(S_n)$. Tegyük fel, hogy (3.3), (3.4) és (3.5) teljesül. Továbbá tegyük fel, hogy*

$$\liminf_{n\to\infty} \frac{\sigma_n^2}{\Lambda_n^d |\mathcal{D}_n|} > 0 \tag{3.15}$$

teljesül. Ekkor $\sigma_n^{-1} S_n \Rightarrow \mathcal{N}(0, 1)$, ha $n \to \infty$.

3.6. Tétel. *Legyen $\{\varepsilon(\mathbf{x}) : \mathbf{x} \in T_\infty\}$ véletlen mező és legyen $Y_n(\mathbf{k})$ az $\varepsilon(\mathbf{x}_{\mathbf{k}}^{(n)})$ Borel-mérhető függvénye, $\mathbf{k} \in \mathcal{D}_n$. Tegyük fel, hogy $\mathbb{E}Y_n(\mathbf{k}) = 0$ minden $\mathbf{k} \in \mathcal{D}_n$, $n = 1, 2, \dots$ esetén. Legyen $S_n = \sum_{\mathbf{k}\in\mathcal{D}_n} Y_n(\mathbf{k})$, $n = 1, 2, \dots$, valamint legyen $\sigma_n^2 = \mathrm{var}(S_n)$. Tegyük fel, hogy létezik olyan $\tau > 0$, hogy teljesül (3.2) és*

$$\{|Y_n(\mathbf{k})|^{2+\tau} \ : \ \mathbf{k} \in \mathcal{D}_n, n = 1, 2, \dots\} \quad \textit{egyenletesen integrálható.} \tag{3.20}$$

Ekkor

$$\limsup_{n\to\infty} \frac{1}{\Lambda_n^d |\mathcal{D}_n|} \sum_{\mathbf{k},\mathbf{l}\in\mathcal{D}_n} |\operatorname{cov}(Y_n(\mathbf{k}), Y_n(\mathbf{l}))| < \infty. \tag{3.21}$$

Ha még a (3.3), (3.4), (3.5), és (3.15) feltételek is teljesülnek, akkor

$$\sigma_n^{-1} S_n \Rightarrow \mathcal{N}(0, 1),$$

ha $n \to \infty$.

Disszertációm második részének végén a 3.5. Tétel és 3.6. Tétel p-dimenziós kiterjesztéseivel foglalkozom. A 3.9. Tétel lényegében [Faz03] Theorem 3.1.

3.9. TÉTEL. *Legyen* $\varepsilon(\mathbf{x})$ *véletlen mező és* $Y_n(\mathbf{k})$ p*-dimenziós véletlen vektor* $\varepsilon(\mathbf{x}_{\mathbf{k}}^{(n)})$ *egy Borel-mérhető függvénye,* $\mathbf{k} \in \mathcal{D}_n$*. Legyen* $\mathbb{E}Y_n(\mathbf{k}) = 0$ *és* $\|Y_n(\mathbf{k})\|$ $\mathbf{k} \in \mathcal{D}_n$, $n = 1, 2, \dots$ *egyenletesen korlátos. Legyen* $S_n = \sum\limits_{\mathbf{k} \in \mathcal{D}_n} Y_n(\mathbf{k})$, $n = 1, 2, \dots,$ $\Sigma_n = \mathrm{var}(S_n)$*. Tegyük fel, hogy teljesül (3.3), (3.4), (3.5) és létezik a*

$$\lim_{n \to \infty} (\Lambda_n^{-d} |\mathcal{D}_n|^{-1} \Sigma_n) = \Sigma$$

határérték. Ekkor $(\Lambda_n^d |\mathcal{D}_n|)^{-\frac{1}{2}} S_n \Rightarrow \mathcal{N}(0, \Sigma)$*, ha* $n \to \infty$*.*

A 3.10. Tétel megegyezik a [FK00] Remark 4.3. megjegyzésével.

3.10. TÉTEL. *Legyen* $\varepsilon(\mathbf{x})$ *egy véletlen mező. Legyen* $Y_n(\mathbf{k})$ *egy centrált* p*-dimenziós véletlen vektor, amely* $\varepsilon(\mathbf{x}_{\mathbf{k}}^{(n)})$*-mérhető minden* $n = 1, 2, \dots,$ *és minden* $\mathbf{k} \in \mathcal{D}_n$ *esetén. Legyen* $S_n = \sum_{\mathbf{k} \in \mathcal{D}_n} Y_n(\mathbf{k})$, $n = 1, 2, \dots,$ $\Sigma_n = \mathrm{var}(S_n)$*. Tegyük fel, hogy teljesül (3.3), (3.4), (3.5) és létezik olyan* $\tau > 0$*, ami kielégíti (3.2)-t és*

$$\{\|Y_n(\mathbf{k})\|^{2+\tau} \ : \ \mathbf{k} \in \mathcal{D}_n, n = 1, 2, \dots\} \qquad \text{egyenletesen integrálható} \ .$$

Ezenkívül tegyük fel, hogy

$$\liminf_{n \to \infty} \lambda_{\min}\big(\Lambda_n^{-d} |\mathcal{D}_n|^{-1} \Sigma_n\big) > 0.$$

Ekkor $\Sigma_n^{-\frac{1}{2}} S_n \Rightarrow \mathcal{N}(0, I_p)$*, ha* $n \to \infty$*.*

4. fejezet

A regressziós függvény becslésének határeloszlása véletlen mezőkre

A magfüggvényes becsléseket széles körben tanulmányozza a szakirodalom. A sűrűségfüggvény magfüggvényes becsléséről Parzen ([Par62]) és Rosenblatt ([Ros56b]) ért el alapvető eredményeket. A regressziós függvény magfüggvényes becsléséről Nadaraya és Watson ([Nad64], [Wat64]) 1964-ben közölt eredményeit számos cikkben feldolgozták és általánosították. Ezeket az eredményeket többek között Rao ([Rao83]), Devroye és Győrfi ([DG85]), valamint Bosq ([Bos98]) foglalta össze. A magfüggvényes becslések egyik fontos tulajdonsága az aszimptotikus normalitás, melyet több cikkben is tanulmányoztak (lásd [Sch72], [Cai01]).

Tekintsünk most egy $(X_{\mathbf{t}}, Y_{\mathbf{t}}), \mathbf{t} \in T_\infty$, erősen stacionárius véletlen mezőt (T_∞ az $\mathbb{R}^d$ tér egy tartománya, $X_{\mathbf{t}}$ és $Y_{\mathbf{t}}$ valós értékű valószínűségi változók.) Célunk az $r(x) = \mathbb{E}\left(\Phi\left(Y_{\mathbf{t}}\right) | X_{\mathbf{t}} = x\right)$ regressziós függvény becslésének határeloszlásának meghatározása, ahol Φ ismert, korlátos és mérhető függvény. Legyen $(X_{\mathbf{t}}, Y_{\mathbf{t}}), \mathbf{t} \in \mathcal{D}_n$ az adathalmazunk, ahol $\mathcal{D}_n$ jelöli a T_n rácspontjait és $T_n \subset \mathbb{R}^d$. Tekintsük a regressziós függvény magfüggvényes becslését

$$r_n(x) = \frac{\sum\limits_{\mathbf{t}\in\mathcal{D}_n} \Phi(Y_{\mathbf{t}}) K\left(\dfrac{x - X_{\mathbf{t}}}{h}\right)}{\sum\limits_{\mathbf{t}\in\mathcal{D}_n} K\left(\dfrac{x - X_{\mathbf{t}}}{h}\right)},$$

ahol K egy magfüggvény (lásd [Nad64], [Wat64]). $(r_n(x_1), \ldots, r_n(x_m))$ együttes aszimptotikus normalitása független megfigyelések esetén jól ismert (lásd [Sch72]).

Az általunk tekintett mintavételezési séma azonban eltér az általánosan használtaktól. A megfigyelésünk helyei egyre sűrűbbek lesznek a tartomány növekedésével. Ezt a jelenséget „sűrűsödő-növekvő” (infill-increasing) sémának nevezték el a szakirodalomban (lásd [LKC99], [Faz03]). Feltesszük, hogy a megfigyelt valószínűségi mező gyengén függő, pontosabban a valószínűségi mezőre teljesül az úgynevezett α-keverő feltétel. Fő eredményünk az $r_n(x)$ aszimptotikus normalitásának bizonyítása. Az eredmény külön érdekessége a szokatlan kovariancia mátrix. Pontosabban szólva $(r_n(x_1), \dots, r_n(x_m))$ vektor aszimptotikus kovariancia mátrixa, egy diagonális mátrix és egy a feltételes kovarianciák integráljait tartalmazó mátrix összegéből áll (lásd 4.1. Tétel).

A vizsgálataink motivációjaként meg kell említeni még a számos helyen alkalmazott mintavételi sémákat is. A mintavételezés történhet véletlen vagy determinisztikus időpontokban. A legtöbb létező eredmény a nem sűrűsödő esetre vonatkozik (lásd [Mas83], [BC93]). A [Bos98] könyvben a mintavételi séma fontossága bemutatásra kerül, de nem említ explicit eredményt regresszió esetén. A [Bos98] 140. oldalán a következő utalás szerepel: „A regressziós és sűrűség becslések egymáshoz hasonlóan viselkednek mintavételi sémák esetén”.

Továbbra is jelölje $|\mathcal{D}|$ egy $\mathcal{D}$ véges halmaz számosságát, valamint $|T|$ a T tartomány térfogatát (Lebesgue-mértékét).

A megfigyelések az alábbi sémát követik. Az egyszerűség kedvéért a d-dimenziós téglák legyenek a megfigyelési tartományok. Legyen $\Lambda > 0$ rögzített. Jelölje $\left(\frac{\mathbb{Z}}{\Lambda}\right)^d$ az $\mathbb{R}^d$-beli Λ-háló pontjait, azaz a hálópontok távolsága $1/\Lambda$:

$$\left(\frac{\mathbb{Z}}{\Lambda}\right)^d = \left\{ \left(\frac{k_1}{\Lambda}, \dots, \frac{k_d}{\Lambda}\right) : (k_1, \dots, k_d) \in \mathbb{Z}^d \right\}.$$

Legyen T korlátos, zárt téglalap $\mathbb{R}^d$-ben, élei párhuzamosak a tengelyekkel. A T-beli Λ-rácspontokat jelölje $\mathcal{D}$, azaz $\mathcal{D} = T \cap (\mathbb{Z}/\Lambda)^d$. A határeloszlás leírásához tekintsük az előző objektumok egy sorozatát, azaz legyenek $T_1, T_2, \dots$ zárt, korlátos $\mathbb{R}^d$-beli téglák egy sorozata. Tegyük fel, hogy $T_1 \subset T_2 \subset T_3 \subset \dots$, $\bigcup_{i=1}^{\infty} T_i = T_\infty$.

Feltesszük még, hogy T_n minden élének hossza egész és tart a ∞-hez, amint $n \to \infty$ (például $T_\infty = \mathbb{R}^d$ vagy $T_\infty = [0, \infty)^d$). Legyen $\{\Lambda_n\}$ a pozitív egész számok egy növekvő sorozata (a nem egész számok esetét lényegében azonos módon kezelhetjük) és a T_n-be eső Λ_n-rácspontok halmaza legyen $\mathcal{D}_n$.

Legyen $\{\xi_{\mathbf{t}} = (X_{\mathbf{t}}, Y_{\mathbf{t}}), \mathbf{t} \in T_\infty\}$ erősen stacionárius kétdimenziós véletlen mező. A megfigyelések n-edik halmaza az $(X_{\mathbf{t}}, Y_{\mathbf{t}})$ véletlen mező minden $\mathbf{t} \in \mathcal{D}_n$ pontban felvett értékeiből áll. Ezekből az adatokból egy becslést konstruálunk a regressziós függvényre. Valójában, minden $\mathbf{t} = \mathbf{t}^{(n)}$ függ n-től de, hogy elkerüljük a bonyolult jelöléseket, elhagyjuk az (n) felső indexet. Feltesszük, hogy $\lim_{n\to\infty} |\mathcal{D}_n| = \infty$.

Ezen megfigyelések helyei egyre sűrűbbek és sűrűbbek lesznek a tartományok egy növekvő sorozata esetén, ezért ezt a sémát „sűrűsödő-növekvő” sémának nevezzük.

Használni fogjuk az előző fejezetben bevezetett α-keverési együtthatót is.

Szükségünk lesz a következő feltételre: valamely $1 < a < \infty$ esetén

$$\int_0^\infty s^{2d-1} \alpha^{\frac{a-1}{a}}(s) ds < \infty. \tag{4.1}$$

A $K : \mathbb{R} \to [0, \infty)$ függvényt magfüggvénynek nevezzük, ha K korlátos, folytonos, szimmetrikus sűrűségfüggvény (a Lebesgue-mértékre nézve), melyre

$$\lim_{|u|\to\infty} |u|K(u) = 0, \qquad \int_{-\infty}^{\infty} u^2 K(u) du < \infty. \tag{4.2}$$

Legyen $g(x)$ az $X_{\mathbf{t}}$ (ismeretlen) perem-sűrűségfüggvénye. Feltételezzük, hogy $g(x)$ mindenütt pozitív. Legyen K magfüggvény és legyen $h_n > 0$. Ekkor a $g(x)$ (Parzen-Rosenblatt-féle) magfüggvényes becslése

$$g_n(x) = \frac{1}{|\mathcal{D}_n|} \frac{1}{h_n} \sum_{\mathbf{i} \in \mathcal{D}_n} K\left(\frac{x - X_{\mathbf{i}}}{h_n}\right), \quad x \in \mathbb{R}.$$

Legyen

$$a(x) = \mathbb{E}\left(\Phi^2(Y_{\mathbf{t}}) | X_{\mathbf{t}} = x\right).$$

$\mathbb{R}^d_{\mathbf{0}}$ jelölje az $\mathbb{R}^d \setminus \{\mathbf{0}\}$ halmazt. Legyen $g_{\mathbf{u}}(x, y)$ az $X_{\mathbf{0}}$ és $X_{\mathbf{u}}$ együttes sűrűségfüggvénye, ha $\mathbf{u} \in \mathbb{R}^d_{\mathbf{0}}$ és $x, y \in \mathbb{R}$ és

$$a_{\mathbf{u}}(x, y) = \mathbb{E}\big\{[\Phi(Y_{\mathbf{0}}) - r(X_{\mathbf{0}})]\,[\Phi(Y_{\mathbf{u}}) - r(X_{\mathbf{u}})]\,\big|X_{\mathbf{0}} = x, X_{\mathbf{u}} = y\big\}.$$

Feltesszük, hogy minden rögzített $\mathbf{u}$ esetén

(4.3) $a_{\mathbf{u}}(.,.), g_{\mathbf{u}}(.,.), a(.), r(.), g(.), r'(.), g'(.), r''(.), g''(.)$ korlátos és folytonos

függvények. Továbbá feltesszük, hogy

$$\lim_{n\to\infty} \frac{1}{\Lambda_n^d h_n} = L < \infty, \quad \lim_{n\to\infty} \Lambda_n = \infty \text{ és } \lim_{n\to\infty} h_n = 0, \tag{4.4}$$

valamint

$$\lim_{n\to\infty} |T_n| h_n^4 = 0. \tag{4.5}$$

Arra az esetre koncentrálunk, amikor $\xi_{\mathbf{t}}$ és $\xi_{\mathbf{s}}$ valószínűségi változók függőek, ha $\mathbf{t}$ és $\mathbf{s}$ közel vannak egymáshoz.

Először tekintsük a sűrűségfüggvény g_n becslésének aszimptotikus normalitását.

Legyen $l_{\mathbf{u}}(x, y) = g_{\mathbf{u}}(x, y) - g(x)g(y), \mathbf{u} \in \mathbb{R}^d_{\mathbf{0}}$ és $x, y \in \mathbb{R}$. Jelölje $l_{\mathbf{u}}$ az $l_{\mathbf{u}}(x, y)$ függvényt, mint $l : \mathbb{R}^d_{\mathbf{0}} \to \mathcal{C}(\mathbb{R}^2)$ leképezést, azaz egy olyan függvényt, mely a

$\mathcal{C}(\mathbb{R}^2)$ térből veszi fel értékeit (itt $\mathcal{C}(\mathbb{R}^2)$ az $\mathbb{R}^2$-en értelmezett valós értékű folytonos függvények tere). Legyen

$$\|l_{\mathbf{u}}\| = \sup_{(x,y)\in\mathbb{R}^2} |l_{\mathbf{u}}(x,y)| \tag{4.6}$$

az $l_{\mathbf{u}}$ normája. Legyenek $x_1,\ldots,x_m$ különböző valós számok.
Legyen $\Sigma_l = \left(\int_{\mathbb{R}^d_{\mathbf{0}}} l_{\mathbf{u}}(x_i,x_j)d\mathbf{u}\right)_{1\le i,j\le m}$, legyen továbbá $D^{'}$ diagonális mátrix $Lg(x_i)\int_{-\infty}^{\infty} K^2(u)du$, $i = 1,\ldots,m$ diagonális elemekkel. Vezessük be a $\Sigma^{'} = \Sigma_l + D^{'}$ jelölést.

A. TÉTEL. *([FC06] Theorem 1.) Tegyük fel, hogy $l_{\mathbf{u}}$ (mint $\mathbf{u}$ változójú $l : \mathbb{R}^d_{\mathbf{0}} \to \mathcal{C}(\mathbb{R}^2)$ függvény) Riemann-integrálható minden korlátos zárt d-dimenziós $R \subset \mathbb{R}^d_{\mathbf{0}}$ téglán, továbbá $\|l_{\mathbf{u}}\|$ direkt Riemann-integrálható ($\|l\| : \mathbb{R}^d_{\mathbf{0}} \to \mathbb{R}$ függvényként tekintve). Legyenek $x_1,\ldots,x_m$ egymástól különböző valós számok és tegyük fel, hogy $\Sigma^{'}$ pozitív definit. Tegyük fel, hogy létezik $1 < a < \infty$, melyre teljesül (4.1) és*

$$(h_n)^{-1} \le c|T_n|^{\frac{a^2}{(3a-1)(2a-1)}} \quad \textit{minden } n \textit{ esetén.} \tag{4.7}$$

Ha (4.4) és (4.5) fennáll, akkor

$$\sqrt{\frac{|\mathcal{D}_n|}{\Lambda^d_n}}\{(g_n(x_i)-g(x_i)), i=1,\ldots,m\} \Rightarrow \mathcal{N}(0,\Sigma^{'}), \quad \textit{ha} \quad n\to\infty. \tag{4.8}$$

A fenti előzmények után kimondhatjuk a fő eredményünket.

Legyen $v(x) = a(x) - r^2(x)$. Rögzített m pozitív egész és rögzített, egymástól különböző $x_1,x_2,\ldots,x_m$ valós számok esetén vezessük be a következő jelöléseket.

$$\sigma(x_t,x_s) = \int_{\mathbb{R}^d_{\mathbf{0}}} a_{\mathbf{u}}(x_t,x_s)g_{\mathbf{u}}(x_t,x_s)d\mathbf{u}, \quad t,s=1,\ldots,m, \tag{4.9}$$

$$\Sigma^{(m)} = \left(\frac{\sigma(x_t,x_s)}{g(x_t)g(x_s)}\right)_{1\le t,s\le m}. \tag{4.10}$$

Feltesszük, hogy

$$\lim_{|z|\to\infty} z^3|K(z)| = 0 \tag{4.11}$$

teljesül.

4.1. TÉTEL. *Legyen $(X_{\mathbf{t}},Y_{\mathbf{t}})$, $\mathbf{t}\in T_\infty$, erősen stacionárius kétdimenziós véletlen mező, $r(x) = \mathbb{E}(\Phi(Y_{\mathbf{t}})|X_{\mathbf{t}}=x)$ a regressziós függvény, ahol Φ korlátos, mérhető függvény és K egy magfüggvény. Tegyük fel, hogy az A. Tétel feltételei teljesülnek az*

$l_{\mathbf{u}}$ függvényre, továbbá legyen Σ' pozitív definit. Tegyük fel, hogy az $X_{\mathbf{t}}$ peremsűrűségfüggvénye pozitív, valamint $a_{\mathbf{u}}g_{\mathbf{u}}$ Riemann-integrálható (egy $a \cdot g : \mathbb{R}_{\mathbf{0}}^d \to \mathcal{C}(\mathbb{R}^2)$ függvényként tekintve) minden korlátos zárt d-dimenziós $R \subset \mathbb{R}_{\mathbf{0}}^d$ téglán. Továbbá, $\|a_{\mathbf{u}}g_{\mathbf{u}}\|$ direkt Riemann-integrálható (mint egy $\|a \cdot g\| : \mathbb{R}_{\mathbf{0}}^d \to \mathbb{R}$ függvény), ahol a norma (4.6) szerint definiált. Tegyük fel továbbá, hogy létezik olyan $1 < a < \infty$, hogy (4.1) és (4.7) teljesül. Legyen $\Sigma^{(m)}+D$ mátrix pozitív definit, ahol D egy diagonális mátrix, melynek a diagonális elemei: $Lv(x_i)\int_{-\infty}^{\infty} K^2(t)dt/g(x_i)$, $i = 1, \ldots, m$. Ha a fentieken felül még (4.3), (4.4), (4.5) és (4.11) teljesülnek, akkor

$$\sqrt{\frac{|\mathcal{D}_n|}{\Lambda_n^d}}\{(r_n(x_i) - r(x_i)), i = 1, \ldots, m\} \Rightarrow \mathcal{N}(0, \Sigma), \quad ha \quad n \to \infty,$$

ahol

$$\Sigma = \Sigma^{(m)} + D.$$

4.2. Megjegyzés. Megmutattuk, hogy a 4.1. Tételbeli Σ aszimptotikus kovariancia-mátrix egy diszkrét és egy folytonos esetnek megfelelő aszimptotikus kovariancia-mátrix kombinációja. Schuster igazolta (lásd [Sch72]), hogy (független, azonos eloszlású megfigyelések esetén) $r_n(x_1), \ldots, r_n(x_m)$ aszimptotikusan normális diagonális kovariancia-mátrixszal. Pontosabban $\sqrt{nh_n}(r_n(x_i) - r(x_i)) \Rightarrow \mathcal{N}(0, c_i)$, ahol $c_i = v(x_i)\int_{-\infty}^{\infty} K^2(t)dt/g(x_i)$, ezért a 4.1. Tételben a D diagonális rész megfelel a diszkrét esetbeli határérték kovariancia-mátrixnak.

Dolgozatom utolsó fejezetében szimulációkon keresztül néhány egyszerű példát adok az elméleti állítás bemutatására.

Legyen $X_{\mathbf{u}}, \mathbf{u} \in \mathbb{R}^d$, stacionárius Gauss-féle véletlen mező nulla várható értékkel és $\rho_{\mathbf{u}}$ kovariancia függvénnyel. A következő példában ugyanazt az $X_{\mathbf{u}}$ véletlen mezőt vizsgáljuk, amelyeket Fazekas és Chuprunov már tanulmányoztak ([FC06], [Faz07]). Legyen

$$\Phi(Y_{\mathbf{u}}) = 10\sin(X_{\mathbf{u}}) + 100 + \delta_{\mathbf{u}},$$

ahol $\delta_{\mathbf{u}} = \widetilde{X_{\mathbf{u}}}$, és $\widetilde{X_{\mathbf{u}}}$ olyan stacionárius véletlen mező, hogy $\widetilde{X_{\mathbf{u}}}$ és $X_{\mathbf{u}}$ eloszlása megegyezik, valamint legyenek $X_{\mathbf{u}}$ és $\widetilde{X_{\mathbf{u}}}$ függetlenek.

4.7. Példa. Tekintsünk egy $X(u), u \in \mathbb{R}$, Gauss-folyamatot nulla várható értékkel és $\rho_u = e^{-|u|}, u \in \mathbb{R}$ kovariancia függvénnyel. Ezt a folyamatot figyeljük meg a $T = [0, t]$ tartomány $1/\Lambda$-hálópontjaiban, ahol $\Lambda = 40$ és $t = 60$. Azaz a minta $z_1 = X(1/40), \ldots, z_s = X(2400/40)$, ahol $s = 2400$. Ennek a minta vektornak a kovariancia-mátrixa $(\rho^{|i-j|})_{i,j=1}^s$, ahol $\rho = e^{-1/\Lambda}$. Ezért az adatok a szimulációhoz könnyen generálhatóak. Valóban, legyenek $y_1, \ldots, y_s$ független, azonos eloszlású standard normális valószínűségi változók és legyen

$$z_i = \rho^{i-1}y_1 + \sqrt{1-\rho^2}\sum_{j=2}^{i}\rho^{i-j}y_j, i = 1, \ldots, s.$$

Ezeket az adatokat felhasználva, meghatároztuk az r_n regressziós függvény becslését az $x_1 = -0.5$, $x_2 = -0.25$, $x_3 = 0$, $x_4 = 0.25$, és $x_5 = 0.5$ pontokban. Sávszélességnek két értéket használtunk, $h_1 = 0.025$-öt és $h_2 = 0.005$-öt. Továbbá a K magfüggvénynek a standard normális eloszlás sűrűségfüggvényét használtuk.

A szimulációkat MATLAB programcsomag segítségével hajtottuk végre. Az eljárást 5000-szer ismételtük. Mindkét sávszélesség (h_1 és h_2) használata esetén az adathalmazok megegyeztek. A regressziós függvény elméleti értékeit és becsléseinek átlagát az alábbi 4.1. táblázat mutatja. Megállapíthatjuk, hogy mindkét sávszélesség esetén az elméleti érték és a közelítő értékek átlaga közel vannak egymáshoz.

4.1. táblázat. A regressziós függvény elméleti értékei és becsléseinek átlaga a 4.7. Példa adataira.

x	-0.5	-0.25	0	0.25	0.5
$r(x)$	95.2057	97.5260	100.0000	102.4740	104.7943
$\overline{r_n(x)}$, ha $h_1 = 0.025$	95.2039	97.5220	99.9953	102.4684	104.7929
$\overline{r_n(x)}$, ha $h_2 = 0.005$	95.1970	97.5229	99.9939	102.4707	104.7976

A $\sqrt{\frac{|\mathcal{D}|}{\Lambda}}(r_n(x_1) - r(x_1), \ldots, r_n(x_5) - r(x_5))$ standardizált becslésekre (a standardizáló faktor $\sqrt{\frac{|\mathcal{D}|}{\Lambda}} = 7.7459$) kiszámítottuk a Σ_1 (h_1 sávszélesség esetén) és Σ_2 (h_2 sávszélesség esetén) empirikus kovariancia-mátrixokat. A Σ_1 és Σ_2 mátrixok csak a főátlókban térnek el jelentősen, többi elemük majdnem megegyezik.

Határozzuk most meg a 4.1. Tételben leírt kovariancia-mátrix D diagonális mátrixának elemeit. Esetünkben, a D_k diagonális mátrix elemei h_k ($k = 1, 2$-re) sávszélesség esetén a következők:

$$\frac{1}{\Lambda}\frac{1}{h_k}v(x_i)\frac{1}{g(x_i)}\int_{-\infty}^{\infty} K^2(u)du = \frac{1}{40}\frac{1}{h_k} \cdot 1 \cdot \frac{1}{g(x_i)}\frac{1}{2\sqrt{\pi}}.$$

A „sűrűsödő-növekvő" eset miatt a kovariancia-mátrixban csak a főátlóban lehet különbség különböző sávszélesség használata esetén. A 4.2. táblázatban bemutatjuk, hogy az empirikus kovariancia-mátrixok főátlóbeli elemei különbsége ($diag(\Sigma_2 - \Sigma_1)$) és az elméleti kovariancia-mátrixok különbsége ($diag(D_2 - D_1)$, hogyan viszonyul egymáshoz.

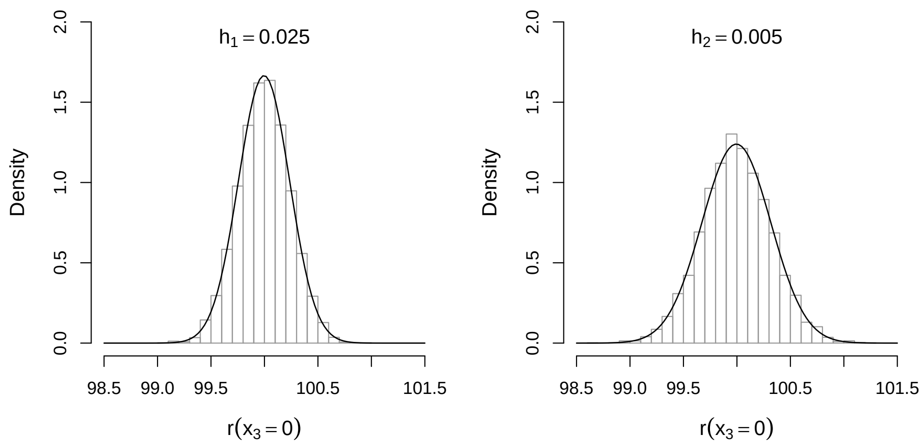
Az eredmények azt mutatják, hogy a 4.1. Tételben szereplő D diagonális mátrix jól magyarázza a határérték kovariancia-mátrix függőséget a sávszélességtől, mivel az arányok közel vannak egyhez.

Végül, a 4.1. ábra az $r(x_3 = 0)$ pontbeli becslések relatív gyakoriságát mutatja $h_1 = 0.025$ (bal oldali ábra) és $h_2 = 0.005$ (jobb oldali ábra) sávszélességek esetén.

4.2. táblázat. Az empirikus és az elméleti kovariancia-mátrixok főátlóbeli elemei különbségének aránya a 4.7. Példa adataira.

x	-0.5	-0.25	0	0.25	0.5
$\frac{diag(\Sigma_2-\Sigma_1)}{diag(D_2-D_1)}$	1.0428	1.0316	0.9812	1.0397	1.0465

A hisztogramokra ráhelyeztük azt a normális eloszlást, melynek a várható értékét és szórását a mintából becsültük. A 4.1. Tételben kimondott regressziós becslés közelítő normális eloszlása látható ezekben az ábrákban. A különböző sávszélességek különböző normális eloszlásokat eredményeznek.



4.1. ábra. Az $r(x_3 = 0)$ pontbeli becslések relatív gyakorisága $h_1 = 0.025$ (bal ábra) és $h_2 = 0.005$ (jobb ábra) sávszélességek esetén, valamint a hozzájuk tartozó becsült normális eloszlások sűrűségfüggvényei.

PhD Thesis

Limit theorems for random fields

Zsolt Karácsony

Supervisor: Dr. István Fazekas

University of Debrecen
Doctoral School of Mathematics and Computational Sciences
Debrecen, 2010.

Chapter 1

Introduction

This Ph.D thesis contains new results in the field of limit theorems of probability theory and statistics. It consists of three parts, which can be treated separately. In this summary the numeration of our theorems and definitions are the same as those used in the dissertation.

In the first part of the dissertation we deal with autoregressive type martingale fields. This is in this part a generalization of d-index martingales is studied.

A d-index process is called an autoregressive martingale field if it satisfies certain autoregressive type stochastic difference equations. An almost sure convergence theorem is proved for autoregressive martingale fields.

There are several extensions of the notion of a martingale. The so called linear martingales were studied in [MQ73], [Hey80] and [Faz87]. The notion of a linear martingale was extended to the two indexes case in [Faz88]. We mention that a lot of papers are devoted to the study of multiindex martingales (e.g. [Cai70], [Faz83]). It is well-known that the almost sure (a.s.) convergence of a multiindex sequence (in particular a martingale) requires stronger conditions than that of a single index sequence. The a.s. convergence of multiindex martingales is described in [Cai70].

We extend the notion of a linear martingale to the multiindex case. The construction of the notion of d-index autoregressive type martingale fields is also our own result for $d > 2$. Then we obtain an a.s. convergence result for it (Theorem 2.21.). This theorem contains previous results of [Faz87] and [Faz88] as special cases.

In the second part of our dissertation we deal with central limit theorems for mixing random fields.

Ibragimov and Linnik ([IL71]) proved a central limit theorem for stationary sequences satisfying certain α-mixing conditions. Bolthausen ([Bol82]) and Guyon ([Guy95]) extended it to α-mixing random fields. Fazekas ([Faz03]) and Fazekas and Kukush ([FK00]) presented so called infill-increasing versions of Guyon's result for the bounded and the uniformly integrable cases, respectively. These papers do not contain the proofs of the theorems mentioned, just a sketch of the proof is given. In the second part of our dissertation detailed proofs for the above mentioned theorems are given. The importance of the detailed proof is the following. It turns out that the original method by Ibragimov and Linnik ([IL71]) and Guyon ([Guy95]) can be applied if the random field satisfies a certain uniform integrability condition. Guyon ([Guy95]) does not assume uniform integrability but he does not describe the step from the bounded case to the general case. Therefore we do not know if his result is valid in the general case. We mention that Ibragimov and Linnik ([IL71]) assumed stationarity so their proof is complete.

In the third part of our dissertation we deal with asymptotic normality of kernel type regression estimators for random fields.

Kernel type regression estimators have been widely studied in the literature. The original results by Nadaraya ([Nad64]) and Watson ([Wat64]) have been extended in several papers, and they are summarized for example in [Rao83], [DG85], and [Bos98]. One important issue for kernel type regression estimators is their asymptotic normality, which has been studied in several papers, like in [Sch72] and [Cai01].

The asymptotic normality of the Nadaraya-Watson regression estimator is studied for α-mixing random fields. The infill-increasing setting is considered, that is when the locations of observations become dense in an increasing sequence of domains. This setting fills the gap between continuous and discrete models. In the infill-increasing case the asymptotic normality of the Nadaraya-Watson estimator holds, but with an unusual asymptotic covariance structure. It turns out that this covariance structure is a combination of the covariance structures that we observe in the discrete and in the continuous case (Theorem 4.1.).

Chapter 2

Autoregressive type martingale fields

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $X_{\mathbf{n}}, \mathbf{n} \in \mathbb{N}^d$, a d-index sequence of random variables, and let $\mathcal{F}_{\mathbf{n}} \subseteq \mathcal{F}$ be σ-algebras for all $\mathbf{n} \in \mathbb{N}^d$.

Recall the notion of a martingale. Suppose that $\mathcal{F}_{\mathbf{m}} \subseteq \mathcal{F}_{\mathbf{n}}$ for every $\mathbf{m} \leq \mathbf{n}$. Assume that $X_{\mathbf{n}}$ is $\mathcal{F}_{\mathbf{n}}$-measurable and integrable for every $\mathbf{n} \in \mathbb{N}^d$. We say that $(X_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}})$ is a martingale if $E(X_{\mathbf{n}+\mathbf{k}} \,|\mathcal{F}_{\mathbf{n}}) = X_{\mathbf{n}}$ a.s., for all $\mathbf{n} \in \mathbb{N}^d$ and $\mathbf{k} \in \mathbb{N}_0^d$.

We shall use the following condition. For any η with finite expectation

$$E\left(\eta|\mathcal{F}_{\mathbf{n}}\right) = E\left(\ldots E\left(\eta|\mathcal{F}_{\mathbf{n}}^{(i_1)}\right)\ldots|\mathcal{F}_{\mathbf{n}}^{(i_d)}\right) \tag{2.3}$$

for any permutation $(i_1, \ldots, i_d)$ of $(1, \ldots, d)$, where $\mathcal{F}_{\mathbf{n}}^{(i)} = \sigma\{\mathcal{F}_{\mathbf{l}} : l_i = n_i\}$ for any fixed $\mathbf{n}$ and i (in the notation $\mathcal{F}_{\mathbf{n}}^{(i)}$ superscript (i) shows the appropriate coordinate).

In order to prove our result we have to use a new martingale convergence theorem (Theorem 2.6.). Theorem 2.6. is a uniform a.s. convergence result for Banach space valued multiindex martingales.

More precisely, we have the following theorem.

Theorem 2.6. *Let B be a real separable Banach space. Let $(X_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}}), \mathbf{n} \in \mathbb{N}^d$, be a B-valued martingale. Let B have Radon-Nikodym property or let $X_{\mathbf{n}}$ be of the form $X_{\mathbf{n}} = \mathbb{E}(X|\mathcal{F}_{\mathbf{n}}), \mathbf{n} \in \mathbb{N}^d$, for an $X \in L^1(\mathcal{F}, B)$. Assume that*

$$\sup_{\mathbf{n}} \mathbb{E}\|X_{\mathbf{n}}\|(\log^+ \|X_{\mathbf{n}}\|^{d-1}) < \infty.$$

Then there exists an event A with $\mathbb{P}(A) = 1$ such that for $\omega \in A$ we have: if arbitrary coordinates of $\mathbf{n}$ converge to ∞ while the remaining coordinates remain *fixed then $X_{\mathbf{n}}(\omega)$ converges uniformly. (The limit is a random variable depending on the coordinates remaining fixed.)*

To describe the structure of the random field $\xi_{\mathbf{n}}, \mathbf{n} \in \mathbb{N}^d$, we shall use the Kronecker product (denoted by $\otimes$) and the vec operation (see, e.g. [MN88]). The construction of the notion of d-index autoregressive type martingale fields is also our own result.

Definition 2.9. *The process $\{\xi_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}}\}$, $\mathbf{n} \in \mathbb{N}^d$, is called an autoregressive martingale field if $\xi_{\mathbf{n}}$ is $\mathcal{F}_{\mathbf{n}}$-measurable and integrable for every $\mathbf{n} \in \mathbb{N}^d$,*

$$\begin{aligned} \mathbb{E}\left(\xi_{\mathbf{n}} | \mathcal{F}^{(j)}_{\mathbf{n}-\mathbf{e}_j}\right) &= a_1^{(j)}(n_j)\xi_{\mathbf{n}-\mathbf{e}_j} \\ &\quad + a_2^{(j)}(n_j)\xi_{\mathbf{n}-2\mathbf{e}_j} + \ldots + a_m^{(j)}(n_j)\xi_{\mathbf{n}-m\cdot\mathbf{e}_j} \end{aligned} \tag{2.10}$$

for every $\mathbf{n}$ and j, with $n_j > m$, $j = 1, \ldots, d$, where m is a fixed positive integer, $\mathbf{e}_j = (0, \ldots, \underbrace{1}_{jth}, \ldots 0) \in \mathbb{N}_0^d$ is the jth unit vector, $j = 1, \ldots, d$, and $a_i^{(j)}(n_j)$ are non-negative non-random coefficients with $\sum_{i=1}^m a_i^{(j)}(n_j) = 1$ for every $n_j = m+1, m+2, \ldots,\ j = 1, \ldots, d$.
If the coefficients $a_k^{(j)}(l)$ do not depend on l, then $\xi_{\mathbf{n}}$ is called a homogeneous autoregressive martingale field.

Let $\xi_{\mathbf{n}}$, $\mathbf{n} \in \mathbb{N}^d$, be a d-index random field. Using $\xi_{\mathbf{n}}$, we shall construct another random field $X_{\mathbf{n}}$. The values of this new field are d-index arrays. For any fixed $m \in \mathbb{N}$ and $\mathbf{n} \in \mathbb{N}^d$ (with $n_i \geq m, i = 1, \ldots, d$) $X_{\mathbf{n}}$ denotes the elements of the random field $\xi_{\mathbf{k}}$ with indices being in a hypercube of size $\underbrace{m \times m \times \ldots \times m}_{d}$.

Proposition 2.10. *Let $(\xi_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}})$ be the autoregressive martingale field introduced in Definition 2.9. Let $X_{\mathbf{n}}$ be the array valued random field corresponding to $\xi_{\mathbf{n}}$. Then*

$$\operatorname{vec}\left[\mathbb{E}\left(X_{\mathbf{n}} \mid \mathcal{F}^{(j)}_{\mathbf{n}-\mathbf{e}_j}\right)\right] = \Big(\underbrace{I \otimes \cdots \otimes I}_{d-j} \otimes A_j^{(n_j)} \otimes \underbrace{I \otimes \cdots \otimes I}_{j-1}\Big) \cdot \operatorname{vec}(X_{\mathbf{n}-\mathbf{e}_j}) \tag{2.11}$$

for every $\mathbf{n}$ with $n_j > m$, $j = 1, \ldots, d$, where $A_j^{(l)}$ denotes the following $m \times m$ matrix

$$A_j^{(l)} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ a_m^{(j)}(l) & a_{m-1}^{(j)}(l) & \cdots & \cdots & a_1^{(j)}(l) \end{pmatrix},$$

for every $j = 1, \ldots, d$ and $l = m+1, m+2, \ldots$.

Proposition 2.11. Let $X_{\mathbf{n}}$ be an array-valued random field satisfying (2.11). Assume that (2.3) is valid. Then for the process $(X_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}})$ the equation

$$\begin{aligned} \operatorname{vec}[E(X_{\mathbf{n+t}}|\mathcal{F}_{\mathbf{n}})] &= [A_d(n_d+t_d, n_d) \otimes \cdots \\ &\qquad \otimes A_2(n_2+t_2, n_2) \otimes A_1(n_1+t_1, n_1)] \operatorname{vec}(X_{\mathbf{n}}) \end{aligned} \tag{2.16}$$

holds, where

$$A_j(n_j+t_j, n_j) = A_j^{(n_j+t_j)} A_j^{(n_j+t_j-1)} \cdots A_j^{(n_j+1)} \tag{2.17}$$

for every $n_j > m,\ j = 1, \ldots, d$ and $\mathbf{n} \in \mathbb{N}^d$, $\mathbf{t} \in \mathbb{N}_0^d$.

Above and in the following $A_j(n_j, n_j) = I$ (the unit matrix).

Generalizing property (2.16), we get the following notion.

Definition 2.12. An array-valued process $(X_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}})$, $\mathbf{n} \in \mathbb{N}^{\mathbf{d}}$, is called an *A-martingale field* if

1) $X_{\mathbf{n}}$ is $\mathcal{F}_{\mathbf{n}}$-measurable and integrable for every $\mathbf{n} \in \mathbb{N}^d$,
2) equation (2.16) is satisfied for every $\mathbf{n}, \mathbf{t} \in \mathbb{N}^d$, where the matrices $A_j(n_j + t_j, n_j)$ are given by (2.17). (All matrices $A_j^{(l_j)}$ considered are nonrandom and of type $m \times m$.)

Proposition 2.15. Assume (2.3). For the A-martingale field $X_{\mathbf{n}}$, we have the representation:

$$\operatorname{vec}(X_{\mathbf{n}}) = \sum_{k_1=1}^{n_1} \sum_{k_2=1}^{n_2} \cdots \sum_{k_d=1}^{n_d} [A_d(n_d, k_d) \otimes \cdots \otimes A_1(n_1, k_1)] \operatorname{vec}(\Delta_{\mathbf{k}}), \tag{2.20}$$

where $A_j(k_j, k_j) = I,\ j = 1, \ldots, d$.

The following conditions will be used in our theorems. Suppose that

$$A_j(i_j + t_j, i_j) \to A_j(\infty, i_j), \text{ as } t_j \to \infty, \text{ for every } i_j, j \in \mathbb{N} \tag{2.21}$$

and that the convergence is 'fast' in the following sense:

$$\|A_j(\infty, i_j) - A_j(i_j + t_j, i_j)\| \le c_{t_j}^{(j)}, \qquad \forall i_j, j \in \mathbb{N}, \tag{2.22}$$

where $\sum_{t_j=1}^{\infty} c_{t_j}^{(j)} < \infty$ for every j.

For the limit matrices $A_j(\infty, k_j) = \lim_{t_j \to \infty} A_j(k_j + t_j, k_j)$, we assume that there exists a positive number C such that

$$\|[A_d(\infty, k_d) \otimes \cdots \otimes A_1(\infty, k_1)] \operatorname{vec}(\Delta_{\mathbf{k}})\| \ge C \|\operatorname{vec}(\Delta_{\mathbf{k}})\|, \tag{2.23}$$

for every $\mathbf{k} = (k_1, \ldots, k_d)$.

For arbitrary $S \subseteq \{1, \ldots, d\}$ and arbitrary $\mathbf{n}$

$$\left\| \sum_{\mathbf{k}_S \leq \mathbf{n}_S} A_d(\infty, k_d) \otimes A_{d-1}(\infty, k_{d-1}) \otimes \cdots \otimes A_1(\infty, k_1) \cdot \operatorname{vec}(\Delta_{\mathbf{k}}) \right\| \\ \geq C \cdot \left\| \sum_{\mathbf{k}_S \leq \mathbf{n}_S} D_{k_d} \otimes D_{k_{d-1}} \otimes \cdots \otimes D_{k_1} \cdot \operatorname{vec}(\Delta_{\mathbf{k}}) \right\|, \tag{2.24}$$

where the matrix $D_{k_l} = A_l(\infty, k_l)$ if $l \in S$ and $D_{k_l} = I$ if $l \notin S$.

Theorem 2.17. Assume that the A-martingale field $(X_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}})$, $\mathbf{n} \in \mathbb{N}^d$, satisfies (2.3), (2.22), (2.23) and (2.24). If

$$\sup_{\mathbf{k} \in \mathbb{N}^d} E\| \operatorname{vec}(X_{\mathbf{k}})\| \left[\log^+(\| \operatorname{vec}(X_{\mathbf{k}})\|)\right]^{d-1} < \infty, \tag{2.25}$$

then $X_{\mathbf{n}}$ converges a.s. as $n_j \to \infty$ for all j. If, moreover, $d \geq 2$ then $X_{\mathbf{n}}$ converges in $\mathcal{L}_1$, as $n_j \to \infty$ for all j.

Theorem 2.20. Suppose that for the A-martingale field $(X_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}}), \mathbf{n} \in \mathbb{N}^d$, condition (2.3), (2.22) and (2.24) hold, and

$$\|A_j(i_j, u_j)\| < K < \infty \tag{2.30}$$

if $i_j > u_j,\ j = 1, \ldots, d$. If $\sup_{\mathbf{n}} E\| \operatorname{vec}(X_{\mathbf{n}})\|^{\alpha} < \infty$, where $\alpha > 1$, then $X_{\mathbf{n}}$ converges in $\mathcal{L}_\alpha$, as $\mathbf{n} \to \infty$.

Finally, the main result is the following.

Theorem 2.21. *Let* $(\xi_{\mathbf{n}}, \mathcal{F}_{\mathbf{n}}), \mathbf{n} \in \mathbb{N}^d$, *be a homogeneous autoregressive martingale field and suppose that* (2.3) *is satisfied. Assume, for each* $j = 1, \ldots, d$, $a_m^{(j)} > 0$, *and the greatest common divisor of* $\{k : 1 \leq k \leq m, a_k^{(j)} > 0\}$ *is equal to 1.*

a) If $\sup_{\mathbf{n}} \mathbb{E}|\xi_{\mathbf{n}}| \left[\log^+ |\xi_{\mathbf{n}}|\right]^{d-1} < \infty$, *then* $\xi_{\mathbf{n}}$ *converges a.s, if moreover,* $d \geq 2$, *then* ξ_n *converges in* $\mathcal{L}_1$, *as* $\mathbf{n} \to \infty$.

b) Let be $\alpha > 1$. *If* $\sup_{\mathbf{n}} \mathbb{E}|\xi_{\mathbf{n}}|^{\alpha} < \infty$, *then* $\xi_{\mathbf{n}}$ *converges in* $\mathcal{L}_\alpha$ *(and a.s.), as* $\mathbf{n} \to \infty$.

Chapter 3

A central limit theorems for mixing random fields

Ibragimov and Linnik ([IL71]) proved a central limit theorem for stationary sequences satisfying certain α-mixing conditions. Bolthausen ([Bol82]) and Guyon ([Guy95]) extended it to α-mixing random fields. Fazekas ([Faz03]) and Fazekas and Kukush ([FK00]) presented so called infill-increasing versions of Guyon's result for the bounded and the uniformly integrable cases, respectively. These papers do not contain the proofs of the theorems mentioned, just a sketch of the proof is given. In the second part of our dissertation detailed proofs for the above mentioned theorems are given. The importance of the detailed proof is the following. It turns out that the original proof by Ibragimov and Linnik ([IL71]) and Guyon ([Guy95]) can be applied if the random field satisfies a certain uniform integrability condition. Guyon ([Guy95]) does not assume uniform integrability but he does not describe the step from the bounded case to the general case. Therefore we do not know if his result is valid in the general case. We mention that Ibragimov and Linnik ([IL71]) assumed stationarity so their proof is complete.

The scheme of observations is the following. Let $T_1, T_2, \ldots,$ and T_∞ be domains in $\mathbb{R}^d$. Suppose that $T_1 \subset T_2 \subset T_3 \subset \ldots,$ $\bigcup_{i=1}^{\infty} T_i = T_\infty$. Assume that T_i is compact for each i, T_∞ is of infinite Lebesgue measure. Let $\{\varepsilon(\mathbf{x}),\ \mathbf{x} \in T_\infty\}$ be a random field. The n-th set of observations consists of values of the random field $\varepsilon(\mathbf{x})$ taken at points $\mathbf{x}_{\mathbf{k}} \in T_n$, where $\mathbf{k} \in \mathcal{D}_n \subset \mathbb{Z}^d$. The choice of points $\mathbf{x}_{\mathbf{k}}$ is the following. Divide $\mathbb{R}^d$ into hyperrectangles

$$\Delta_n(\mathbf{k}) = \prod_{j=1}^{d} \Big(\frac{k_j}{N_{jn}}, \frac{k_j+1}{N_{jn}} \Big],$$

where $\mathbf{k} = (k_1, \ldots, k_d) \in \mathbb{Z}^d$ is a d-dimensional integer lattice point and $\{N_{jn}\}$ is an increasing and unbounded sequence of positive integers for each $j = 1, \dots, d$. Now, select the n-th data sites $\mathbf{x_k}$, $\mathbf{k} \in \mathcal{D}_n$, by choosing an arbitrary point $\mathbf{x_k}$ from each $\Delta_n(\mathbf{k}) \cap T_n$ which is non-empty. Actually, each $\mathbf{x_k} = \mathbf{x_k}^{(n)}$ depends on n but to avoid complicated notation we often omit superscript (n). We suppose that $\lim_{n\to\infty} |\mathcal{D}_n| = \infty$.

As the locations of the observations become more and more dense in an increasing sequence of domains, we call our setup infill-increasing.

We need the notion of α-mixing (see e.g. Doukhan [Dou94], Guyon [Guy95]). Let $\mathcal{A}$ and $\mathcal{B}$ be two σ-algebras in $\mathcal{F}$. The α-mixing coefficient of $\mathcal{A}$ and $\mathcal{B}$ is

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup\{|\mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(AB)| \ : \ A \in \mathcal{A}, B \in \mathcal{B}\}.$$

The α-mixing coefficient of $\{\varepsilon(\mathbf{x}) \ : \ \mathbf{x} \in T_\infty\}$ is

$$\alpha(r, u, v) = \sup\{\alpha(\mathcal{F}_{I_1}, \mathcal{F}_{I_2}) \ : \ \varrho(I_1, I_2) \geq r, \ |I_1| \leq u, \ |I_2| \leq v\},$$

where I_1 and I_2 are finite subsets in T_∞, $\mathcal{F}_{I_i} = \sigma\{\varepsilon(\mathbf{x}) \ : \ \mathbf{x} \in I_i\}$, $i = 1, 2$. We list the conditions that will be used in our theorems.

$$\int_0^\infty s^{d-1} \alpha^{\frac{\tau}{2+\tau}}(s, 1, 1) ds < \infty, \quad \text{for some} \quad 0 < \tau < 1. \tag{3.2}$$

$$\int_0^\infty s^{d-1} \alpha(s, i, j) ds < \infty, \quad \text{for} \quad i + j \leq 4. \tag{3.3}$$

$$\alpha(s, 1, \infty) = \mathrm{o}(s^{-d}), \quad \text{as} \quad s \to \infty. \tag{3.4}$$

$$\Lambda_n = \mathrm{O}(\lambda_n), \quad \text{as} \quad n \to \infty, \tag{3.5}$$

where

$$\Lambda_n = \max_{1 \leq j \leq d} N_{jn}, \quad \lambda_n = \min_{1 \leq j \leq d} N_{jn}. \tag{3.6}$$

Actually (3.4) means that $\lim_{n\to\infty} \alpha(s_n, 1, k_n) s_n^d = 0$ if $s_n \to \infty$ and $k_n \to \infty$.

In the following theorems we concentrate on the case when $\xi_\mathbf{t}$ and $\xi_\mathbf{s}$ are dependent if $\mathbf{t}$ and $\mathbf{s}$ are close to each other.

Theorem 3.5. *Let $\varepsilon(\mathbf{x})$ be a random field and let $Y_n(\mathbf{k})$ be a Borel measurable function of $\varepsilon(\mathbf{x}_\mathbf{k}^{(n)}), \mathbf{k} \in \mathcal{D}_n$. Suppose that $\mathbb{E} Y_n(\mathbf{k}) = 0$ and $|Y_n(\mathbf{k})|$ $\mathbf{k} \in \mathcal{D}_n$, $n = 1, 2, \dots$ are uniformly bounded. Let $S_n = \sum_{\mathbf{k}\in\mathcal{D}_n} Y_n(\mathbf{k})$, $n = 1, 2, \dots$, $\sigma_n^2 =$ var(S_n). Suppose that the conditions (3.3), (3.4) and (3.5) are satisfied. Assume that*

$$\liminf_{n\to\infty} \frac{\sigma_n^2}{\Lambda_n^d |\mathcal{D}_n|} > 0, \tag{3.15}$$

hold. Then $\sigma_n^{-1} S_n \Rightarrow \mathcal{N}(0,1)$*, as* $n \to \infty$*.*

Theorem 3.6. *Let* $\varepsilon(\mathbf{x})$ *be a random field and let* $Y_n(\mathbf{k})$ *be a Borel-measurable function of* $\varepsilon(\mathbf{x}_{\mathbf{k}}^{(n)}), \mathbf{k} \in \mathcal{D}_n$*. Suppose that* $\mathbb{E}Y_n(\mathbf{k}) = 0$ *for* $\mathbf{k} \in \mathcal{D}_n$*,* $n = 1, 2, \ldots$*. Let* $S_n = \sum_{\mathbf{k}\in\mathcal{D}_n} Y_n(\mathbf{k})$*,* $n = 1, 2, \ldots$*,* $\sigma_n^2 = \operatorname{var}(S_n)$*. Suppose that there exists a* $\tau > 0$ *such that (3.2) is satisfied and*

$$\{|Y_n(\mathbf{k})|^{2+\tau} \ : \ \mathbf{k} \in \mathcal{D}_n, n = 1, 2, \ldots\} \quad \text{are uniformly integrable.}$$

Then

$$\limsup_{n\to\infty} \frac{1}{\Lambda_n^d |\mathcal{D}_n|} \sum\nolimits_{\mathbf{k},\mathbf{l}\in\mathcal{D}_n} |\operatorname{cov}(Y_n(\mathbf{k}), Y_n(\mathbf{l}))| < \infty.$$

If additionally, conditions (3.3), (3.4), (3.5), and (3.15) are satisfied, then $\sigma_n^{-1} S_n \Rightarrow \mathcal{N}(0,1)$*, as* $n \to \infty$*.*

Now we turn to p-dimensional extensions of Theorem 3.5. and Theorem 3.6. Our Theorem 3.9. is the same as Theorem 3.1. of Fazekas ([Faz03]).

Theorem 3.9. *Let* $\varepsilon(\mathbf{x})$ *be a random field and let the* p*-dimensional random vector* $Y_n(\mathbf{k})$ *be a Borel measurable function of* $\varepsilon(\mathbf{x}_{\mathbf{k}}^{(n)}), \mathbf{k} \in \mathcal{D}_n$*. Suppose that* $\mathbb{E}Y_n(\mathbf{k}) = 0$ *and* $\|Y_n(\mathbf{k})\|$ $\mathbf{k} \in \mathcal{D}_n$*,* $n = 1, 2, \ldots$ *are uniformly bounded. Let* $S_n = \sum_{\mathbf{k}\in\mathcal{D}_n} Y_n(\mathbf{k})$*,* $n = 1, 2, \ldots$*,* $\Sigma_n = \operatorname{var}(S_n)$*. Suppose that conditions (3.3), (3.4) and (3.5) are satisfied. Assume that the limit*

$$\lim_{n\to\infty} (\Lambda_n^{-d} |\mathcal{D}_n|^{-1} \Sigma_n) = \Sigma$$

exists. Then $(\Lambda_n^d |\mathcal{D}_n|)^{-\frac{1}{2}} S_n \Rightarrow \mathcal{N}(0, \Sigma)$*, as* $n \to \infty$*.*

Our Theorem 3.10. is the same as Remark 4.3. in Fazekas and Kukush ([FK00]).

Theorem 3.10. *Let* $\varepsilon(\mathbf{x})$ *be a random field. For each* $n = 1, 2, \ldots$*, and for each* $\mathbf{k} \in \mathcal{D}_n$ *let* $Y_n(\mathbf{k})$ *be a centered* p*-dimensional random vector that is* $\varepsilon(\mathbf{x}_{\mathbf{k}}^{(n)})$*-measurable. Let* $S_n = \sum_{\mathbf{k}\in\mathcal{D}_n} Y_n(\mathbf{k})$*,* $n = 1, 2, \ldots$*,* $\Sigma_n = \operatorname{var}(S_n)$*. Assume that conditions (3.3), (3.4), and (3.5) are satisfied. Moreover, assume that there exists a* $\tau > 0$ *such that (3.2) is satisfied, and*

$$\{\|Y_n(\mathbf{k})\|^{2+\tau} \ : \ \mathbf{k} \in \mathcal{D}_n, n = 1, 2, \ldots\} \quad \text{are uniformly integrable.}$$

Assume that

$$\liminf_{n\to\infty} \lambda_{\min}\big(\Lambda_n^{-d} |\mathcal{D}_n|^{-1} \Sigma_n\big) > 0.$$

Then $\Sigma_n^{-\frac{1}{2}} S_n \Rightarrow \mathcal{N}(0, I_p)$*, as* $n \to \infty$*.*

Chapter 4

Asymptotic normality of kernel type regression estimators for random fields

Kernel type regression estimators have been widely studied in the literature. The original results by Nadaraya ([Nad64]) and Watson ([Wat64]) have been extended in several papers, and they are summarized for example in [Rao83], [DG85], and [Bos98]. One important issue for kernel type regression estimators is their asymptotic normality, which has been studied in several papers, like in [Sch72] and [Cai01].

We consider $(X_{\mathbf{t}}, Y_{\mathbf{t}}), \mathbf{t} \in T_\infty$, to be a strictly stationary random field. (Here T_∞ is a domain in $\mathbb{R}^d$, $X_{\mathbf{t}}$ and $Y_{\mathbf{t}}$ are real-valued.) We want to estimate the regression function $r(x) = \mathbb{E}\left(\Phi\left(Y_{\mathbf{t}}\right) | X_{\mathbf{t}} = x\right)$, where Φ is a known bounded measurable function. The data set is $(X_{\mathbf{t}}, Y_{\mathbf{t}}), \mathbf{t} \in \mathcal{D}_n$. We consider the well-known kernel type regression estimator

$$r_n(x) = \frac{\sum_{\mathbf{t}\in\mathcal{D}_n} \Phi(Y_{\mathbf{t}}) K\left(\frac{x-X_{\mathbf{t}}}{h}\right)}{\sum_{\mathbf{t}\in\mathcal{D}_n} K\left(\frac{x-X_{\mathbf{t}}}{h}\right)},$$

where K is a kernel function (see [Nad64], [Wat64]). However, our sampling scheme is unusual. The locations of observations become dense in an increasing sequence of domains. It is called the infill-increasing setting, see, for example, [LKC99] and [Faz03]. We suppose that the observed random field is weakly dependent, more precisely, the random field satisfies a certain α-mixing condition. The main result is that $r_n(x)$ is asymptotically normal with an unusual covariance structure. That is, the asymptotic covariance matrix of $(r_n(x_1), \ldots, r_n(x_m))$ is the sum of a diagonal matrix and a matrix containing integrals of the conditional covariances, see Theorem 4.1.

Concerning the motivation of our studies we have to refer to the sampling schemes. Continuous time processes can be observed at deterministic or random time. Most of the existing results concern the non infill case (see [Mas83], [BC93]). In [Bos98] the importance of the sampling schemes is expressed, however no explicit result is mentioned for regression. In [Bos98], p. 140 only the following hint is given: "regression and density estimators behave alike when sampled data are available". Actually, for kernel type density estimators there are several results for infill-increasing type sampling schemes. We refer to [Bos98], pp. 118-127 and [BP03].

$|\mathcal{D}|$ denotes the cardinality of the finite set $\mathcal{D}$ and at the same time $|T|$ denotes the volume of the domain T.

The scheme of observations is the following. For simplicity we restrict ourselves to rectangles as domains for the observations. Let $\Lambda > 0$ be fixed. By $\left(\frac{\mathbb{Z}}{\Lambda}\right)^d$ we denote the Λ-lattice points in $\mathbb{R}^d$, i.e. lattice points with distance $\frac{1}{\Lambda}$:

$$\left(\frac{\mathbb{Z}}{\Lambda}\right)^d = \left\{ \left(\frac{k_1}{\Lambda}, \ldots, \frac{k_d}{\Lambda}\right) : (k_1, \ldots, k_d) \in \mathbb{Z}^d \right\}.$$

T will be a bounded, closed rectangle in $\mathbb{R}^d$ with edges parallel to the axes, and $\mathcal{D}$ will denote the Λ-lattice points belonging to T, i.e. $\mathcal{D} = T \cap (\mathbb{Z}/\Lambda)^d$. For describing the limit distribution we consider a sequence of the previous objects. I.e. let $T_1, T_2, \ldots$ be bounded, closed rectangles in $\mathbb{R}^d$, suppose that $T_1 \subset T_2 \subset T_3 \subset \ldots$, $\bigcup_{i=1}^{\infty} T_i = T_\infty$.

We assume that the length of each edge of T_n is an integer and converges to ∞, as $n \to \infty$ (e.g. $T_\infty = \mathbb{R}^d$ or $T_\infty = [0, \infty)^d$). Let $\{\Lambda_n\}$ be an increasing sequence of positive integers (the non-integer case is essentially the same) and let $\mathcal{D}_n$ be the Λ_n-lattice points belonging to T_n.

Let $\{\xi_{\mathbf{t}} = (X_{\mathbf{t}}, Y_{\mathbf{t}}), \mathbf{t} \in T_\infty\}$ be a strictly stationary two-dimensional random field. The n-th set of observations involves the values of the random field $(X_{\mathbf{t}}, Y_{\mathbf{t}})$ taken at each point $\mathbf{k} \in \mathcal{D}_n$. We shall construct the estimator from the data $(X_{\mathbf{k}}, Y_{\mathbf{k}}), \mathbf{k} \in \mathcal{D}_n$. Actually, each $\mathbf{k} = \mathbf{k}^{(n)}$ depends on n, but to avoid complicated notation we omit the superscript (n). By our assumptions, $\lim_{n\to\infty} |\mathcal{D}_n| = \infty$.

As the locations of the observations become more and more dense in an increasing sequence of domains, we call our setup infill-increasing.

We list the conditions that will be used in our theorems.

Let

$$\int_0^\infty s^{2d-1} \alpha^{\frac{a-1}{a}}(s) ds < \infty, \quad \text{for some} \quad 1 < a < \infty. \tag{4.1}$$

A function $K : \mathbb{R} \to [0, \infty)$ will be called a kernel if K is a bounded, continuous, symmetric density function (with respect to the Lebesgue measure),

$$\lim_{|u|\to\infty} |u|K(u) = 0, \qquad \int_{-\infty}^{\infty} u^2 K(u)du < \infty. \tag{4.2}$$

Let $g(x)$ be the (unknown) marginal density function of $X_{\mathbf{t}}$. We assume that $g(x)$ is positive. Let K be a kernel and let $h_n > 0$, then the kernel-type (or Parzen-Rosenblatt-type) estimator of g is

$$g_n(x) = \frac{1}{|\mathcal{D}_n|}\frac{1}{h_n}\sum_{\mathbf{i}\in\mathcal{D}_n} K\left(\frac{x - X_{\mathbf{i}}}{h_n}\right), \quad x \in \mathbb{R}.$$

Let

$$a(x) = E\left(\Phi^2(Y_{\mathbf{t}})|X_{\mathbf{t}} = x\right).$$

Denote by $\mathbb{R}^d_{\mathbf{0}}$ the set $\mathbb{R}^d \setminus \{\mathbf{0}\}$. Let $g_{\mathbf{u}}(x, y)$ be the joint density function of $X_{\mathbf{0}}$ and $X_{\mathbf{u}}$, if $\mathbf{u} \in \mathbb{R}^d_{\mathbf{0}}$ and $x, y \in \mathbb{R}$. Let

$$a_{\mathbf{u}}(x, y) = E\big\{[\Phi(Y_{\mathbf{0}}) - r(X_{\mathbf{0}})]\,[\Phi(Y_{\mathbf{u}}) - r(X_{\mathbf{u}})]\,\big|X_{\mathbf{0}} = x, X_{\mathbf{u}} = y\big\}.$$

We shall assume that for each fixed $\mathbf{u}$ the functions

(4.3) $a_{\mathbf{u}}(.,.), g_{\mathbf{u}}(.,.), a(.), r(.), g(.), r'(.), g'(.), r''(.), g''(.)$ are bounded and continuous.

Furthermore we shall suppose that

$$\lim_{n\to\infty}\frac{1}{\Lambda_n^d h_n} = L < \infty, \ \lim_{n\to\infty}\Lambda_n = \infty \text{ and } \lim_{n\to\infty} h_n = 0 \tag{4.4}$$

and

$$\lim_{n\to\infty} |T_n| h_n^4 = 0. \tag{4.5}$$

We concentrate on the case when $\xi_{\mathbf{t}}$ and $\xi_{\mathbf{s}}$ are dependent if $\mathbf{t}$ and $\mathbf{s}$ are close to each other.

First recall the asymptotic normality of the density estimator g_n.

Let $l_{\mathbf{u}}(x, y) = g_{\mathbf{u}}(x, y) - g(x)g(y), \mathbf{u} \in \mathbb{R}^d_{\mathbf{0}}$ and $x, y \in \mathbb{R}$. Let $l_{\mathbf{u}}$ denote $l_{\mathbf{u}}(x, y)$ as a function $l : \mathbb{R}^d_{\mathbf{0}} \to \mathcal{C}(\mathbb{R}^2)$, i.e. a function with values in $\mathcal{C}(\mathbb{R}^2)$ the space of continuous real-valued functions over $\mathbb{R}^2$. Let

$$\|l_{\mathbf{u}}\| = \sup_{(x,y)\in\mathbb{R}^2} |l_{\mathbf{u}}(x, y)|$$

be the norm of $l_{\mathbf{u}}$.

Let $x_1, \dots, x_m$ be given distinct real numbers.

Let $\Sigma_l = \left(\int_{\mathbb{R}^d_{\mathbf{0}}} l_{\mathbf{u}}(x_i, x_j)d\mathbf{u}\right)_{1\le i,j\le m}$ and let D' be a diagonal matrix with diagonal elements $Lg(x_i)\int_{-\infty}^{\infty} K^2(u)du$, $i = 1, \dots, m$. Let $\Sigma' = \Sigma_l + D'$.

Theorem A. *(Theorem 1. in [FC06].) Assume that* $l_{\mathbf{u}}$ *is Riemann integrable (as a function* $l : \mathbb{R}_{\mathbf{0}}^d \to \mathcal{C}(\mathbb{R}^2)$*) on each bounded closed d-dimensional rectangle* $R \subset \mathbb{R}_{\mathbf{0}}^d$*, moreover* $\|l_{\mathbf{u}}\|$ *is directly Riemann integrable (as a function* $\|l\| : \mathbb{R}_{\mathbf{0}}^d \to \mathbb{R}$*). Let* $x_1, \ldots, x_m$ *be given distinct real numbers and assume that* $\Sigma^{'}$ *is positive definite. Suppose that there exists* $1 < a < \infty$ *such that (4.1) is satisfied and*

$$(h_n)^{-1} \leq c|T_n|^{\frac{a^2}{(3a-1)(2a-1)}} \quad \text{for each } n. \tag{4.7}$$

If (4.4) and (4.5) are satisfied then

$$\sqrt{\frac{|\mathcal{D}_n|}{\Lambda_n^d}}\{(g_n(x_i) - g(x_i)), i = 1, \ldots, m\} \Rightarrow \mathcal{N}(0, \Sigma^{'}) \quad \text{as} \quad n \to \infty.$$

Now we can state our main result. Let $v(x) = a(x) - r^2(x)$.

For a fixed positive integer m and fixed distinct real numbers $x_1, x_2, \ldots, x_m$ we introduce the notation

$$\sigma(x_t, x_s) = \int_{\mathbb{R}_{\mathbf{0}}^d} a_{\mathbf{u}}(x_t, x_s) g_{\mathbf{u}}(x_t, x_s) d\mathbf{u}, \quad t, s = 1, \ldots, m, \tag{4.9}$$

$$\Sigma^{(m)} = \left(\frac{\sigma(x_t, x_s)}{g(x_t) g(x_s)} \right)_{1 \leq t, s \leq m}. \tag{4.10}$$

We assume that

$$\lim_{|z| \to \infty} z^3 |K(z)| = 0. \tag{4.11}$$

Theorem 4.1. *Let* $(X_{\mathbf{t}}, Y_{\mathbf{t}}), \mathbf{t} \in T_\infty$*, be a strictly stationary two-dimensional random field and let* $r(x) = \mathbb{E}\left(\Phi\left(Y_{\mathbf{t}}\right) | X_{\mathbf{t}} = x\right)$ *be the regression function, where* Φ *is a bounded measurable function. Let* K *be a kernel. Assume that the conditions of Theorem A. on the function* $l_{\mathbf{u}}$ *are satisfied, and that* $\Sigma^{'}$ *is positive definite. Furthermore, assume that the marginal density function of* $X_{\mathbf{t}}$ *is positive, and that* $a_{\mathbf{u}} g_{\mathbf{u}}$ *is Riemann integrable (as a function* $a \cdot g : \mathbb{R}_{\mathbf{0}}^d \to \mathcal{C}(\mathbb{R}^2)$*) on each bounded closed d-dimensional rectangle* $R \subset \mathbb{R}_{\mathbf{0}}^d$*. Moreover,* $\|a_{\mathbf{u}} g_{\mathbf{u}}\|$ *is directly Riemann integrable (as a function* $\|a \cdot g\| : \mathbb{R}_{\mathbf{0}}^d \to \mathbb{R}$*). Suppose there exists* $1 < a < \infty$ *such that (4.1) and (4.7) are satisfied. Assume that the matrix* $\Sigma^{(m)} + D$ *is positive definite where* D *is a diagonal matrix with diagonal elements* $Lv(x_i) \int_{-\infty}^{\infty} K^2(t) dt / g(x_i)$*,* $i = 1, \ldots, m$*. If the conditions (4.3), (4.4), (4.5) and (4.11) hold then*

$$\sqrt{\frac{|\mathcal{D}_n|}{\Lambda_n^d}}\{(r_n(x_i) - r(x_i)), i = 1, \ldots, m\} \Rightarrow \mathcal{N}(0, \Sigma), \quad \text{as} \quad n \to \infty,$$

where

$$\Sigma = \Sigma^{(m)} + D.$$

Remark 4.2. *We see that the asymptotic covariance matrix* Σ *in Theorem 4.1. is a combination of the asymptotic covariance matrices in the discrete and the continuous cases. In [Sch72] it is shown that (for independent identically distributed observations)* $r_n(x_1), \dots, r_n(x_m)$ *is asymptotically normal with diagonal covariance matrix. In particular,* $\sqrt{nh_n}(r_n(x_i) - r(x_i)) \Rightarrow \mathcal{N}(0, c_i)$, *where* $c_i = v(x_i) \int_{-\infty}^{\infty} K^2(t)dt/g(x_i)$. *Therefore, in Theorem 4.1. the diagonal part* D *corresponds to the limiting covariance matrix in the discrete case.*

In the last section we present simple examples that give numerical evidence for the phenomena described in Theorem 4.1. The results show that the diagonal matrix D of Theorem 4.1. explains well the dependence of the limit covariance matrix on the bandwidth.

Irodalomjegyzék/References

[BP03] D. Blanke, B. Pumo, (2002), Optimal sampling for density estimation in continuous time. *J. Time Ser. Anal.* **24** (2003), no.1, 1–23.

[Bol82] E. Bolthausen, (1982), On the central limit theorem for stationary mixing random fields. *Ann. Probability*, **10**, 1047–1050.

[Bos98] D. Bosq, (1998), *Nonparametric Statistics for Stochastic Processes.* Springer, New York - Berlin - Heidelberg.

[BC93] D. Bosq, N. Cheze, (1993), Erreur quadratique asymptotique optimale de l'estimateur non paramétrique de la régression pour des observations discrétisées d'un processus stationnaire á temps continu. *C. R. Acad. Sci. Paris* **317** (I), no. 9, 891–894.

[BMP99] D. Bosq, F. Merlevède, M. Peligrad (1999), Asymptotic normality for density kernel estimators in discrete and continuous time *Journal of Multivariate Analysis* **68**, 78–95.

[Cai01] Z. Cai, (2001), Weighted Nadaraya-Watson regression estimation. *Statistics & Probability Letters,* **51**, 307–318.

[Cai70] R. Cairoli, (1970), Une inégalité pour martingales à indices multiples et ses applications, *Séminaire de Probabilités IV, Université de Strasbourg, Lecture Notes in Math.* **124**, 1–27, Springer-Verlag, Berlin.

[CL86] J.V. Castellana, M.R. Leadbetter (1986), On smoothed probability density estimation for stationary processes *Stochastic Processes and their Applications,* **21**, 179–193.

[Cha68] S.D. Chatterji, (1968), Martingale convergence and the Radon-Nikodym theorem in Banach spaces, *Math. Scand.* **22**, 21–41.

[Cre91] N. A. C. Cressie, (1991), *Statistics for Spatial Data.* Wiley, New York.

[DG85] L. Devroye, L. Györfi, (1985), *Nonparametric density estimation. The L_1 view.* Wiley, New York.

[Dou94] P. Doukhan, (1994), *Mixing. Properties and Examples.* Lecture Notes in Statistics **85**, Springer, New York.

[Faz83] I. Fazekas, (1983), Convergence of vector-valued martingales with multidimensional indices. *Publ. Math. Debrecen*, **30**, 157–164.

[Faz87] I. Fazekas, (1987), On the convergence of linear martingales. *Publ. Math. Debrecen*, **34**, 99–104.

[Faz88] I. Fazekas, (1988), On the convergence of regression type martingal fields. *Proc. of the 5th Pannonian Symp., Reidel, Dordrecht*, 43–52.

[Faz03] I. Fazekas, (2003), Limit theorems for the empirical distribution function in the spatial case. *Statistics & Probability Letters*, **62**, 251–262.

[Faz05] I. Fazekas, (2005), Burkholder's inequality for multiindex martingales. *Annales Mathematicae et Informaticae*, **32**, 45–51.

[Faz07] I. Fazekas, (2007), Central limit theorems for kernel type density estimators. *Proceedings of the 7th International Conference on Applied Informatics,* Eger, (Vol. 1), 209–219.

[FC04] I. Fazekas, A. Chuprunov, (2004), A central limit theorem for random fields. *Acta Mathematica Academiae Paedagogicae Nyiregyhaziensis,* **20** (1), 93–104, www.emis.de/journals/AMAPN.

[FC06] I. Fazekas, A. Chuprunov, (2006), Asymptotic normality of kernel type density estimators for random fields. *Stat. Inf. Stoch. Proc.* **9**, 161–178.

[FK00] I. Fazekas, A. G. Kukush, (2000), Infill asymptotics inside increasing domains for the least squares estimator in linear models. *Stat. Inf. Stoch. Proc.*, **3**, 199–223.

[FK02] I. Fazekas, A. G. Kukush, (2002), A central limit theorem for mixing random fields and its statistical applications. *Limit theorems in probability and statistics (Balatonlelle, 1999)*, Vol. **II**, pp. 59–75, János Bolyai Math. Soc., Budapest.

[FKT00] I. Fazekas, A.G. Kukush, T. Tómács, (2000), On the Rosenthal inequality for mixing fields. *Ukrain. Mat. Zh.*, **52** (2), 266–276., translation in *Ukrainian Math. J.*, **52** (2), 305–318.

[Guy95] X. Guyon, (1995), *Random Fields on a Network. Modeling, Statistics, and Applications.* Springer, New York.

[Hey80] C. C. Heyde, (1980), On a probabilistic analogue of the Fibonacci sequence, *J. Appl. Probab.*, **17**, 1079–1082.

[Ibr62] I.A. Ibragimov, (1962), Some limit theorems for stationary processes. *Theory Probab. Appl.*, **7**, 349–382.

[IL71] I.A. Ibragimov, Yu. V. Linnik, (1971), *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

[Lah96] S. N. Lahiri, (1996), On inconsistency of estimators based on spatial data under infill asymptotics., *Sankhya*, **58**, Ser. A, 403–417.

[LKC99] S.N. Lahiri, M.S. Kaiser, N. Cressie, N. J. Hsu, (1999), Prediction of spatial cumulative distribution functions using subsampling. *J. Amer. Statist. Assoc.* **94** (445), 86–110.

[MQ73] J. B. MacQueen, (1973), A linear extension of the martingale convergence theorem, *Ann. Probab.*, **1**, 263–271.

[MN88] J. R. Magnus, H. Neudecker, (1988), *Matrix Differential Calculus*, John Wiley and Sons, Chichester - New York - Weinheim - Brisbane - Singapore - Toronto.

[Mas83] E. Masry, (1983), Probability density estimation from sampled data. *IEEE Transactions in information theory*, vol. **IT-29**, no. 5, 696–709.

[Nad64] E.A. Nadaraya, (1964), On estimating regression. *Theor. Probability Appl.,* **9**, 141–142.

[Par62] E. Parzen, (1962), On estimation of a probability density and mode.*Ann. Math. Statist,* **33**, 1065–1076.

[Rao83] B.L.S. Prakasa Rao, (1983), *Nonparametric Functional Estimation.* Academic Press, INC. London.

[Ros56] M. Rosenblatt, (1956), A central limit theorem and a string mixing condition. *Proc. Nat. Ac. Sc. USA.,***42**, 43–47.

[Ros56b] M. Rosenblatt, (1956), Remarks on some nonparametric estimates of a density *Ann. Math. Statist,* **27**, 832–835.

[Sch72] E.F. Schuster, (1972), Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. *The Annals of Mathematical Statistics,* **43** (1), 84–88.

[Wat64] G.S. Watson, (1964), Smooth regression analysis. *Sankhya. Ser. A* **26**, 359–372.

A szerző publikációi/Publications

Referált folyóiratban megjelent cikkek:

1. Karácsony Zs., *Autoregressive type martingale fields*, AMAPN 22 (2006), 101 – 111.
   `http://www.emis.de/journals/AMAPN/vol22_1/11.html`

2. Karácsony Zs., *A central limit theorem for mixing random fields*, Miskolc Mathematical Notes 7 (2006), Number 2, 147 – 160.

3. Karácsony Zs., Filzmoser P., *Asymptotic normality of kernel type regression estimators for random fields*, Jurnal of Statistical Planning and Inference 140 (2010), 872 – 886.

Közlésre benyújtott cikkek:

1. Karácsony Zs., Libor Zs., *Longest runs in coin tossing. Comparison of recursive formulae, asymptotic theorems, computer simulations*, Beküldve az "Acta Univ. Sapientiae, Mathematica"-ba.

2. Fazekas I., Karácsony Zs., Libor Zs., *A leghosszabb szériák vizsgálata*, Beküldve az "Alkalmazott Matematikai Lapok"-ba.

Konferenciakiadványban megjelent cikkek:

1. Karácsony Zs., *Longest runs in coin tossing.* MicroCAD 2010 International Scientific Conference, vol H., Miskolci Egyetem, Miskolc (2010), 33 – 38.

2. Fazekas I., Karácsony Zs., Libor Zs., *Interesting questions in coin-tossing.* 33rd International Congress of Teachers of Mathematics, Physics and IT, Budapest (2009), 119 – 125.

3. Karácsony Zs., *Leghosszabb szériák vizsgálata.* Doktoranduszok fóruma, Miskolci Egyetem, Gépészmérnöki és Informatikai Kar, Miskolc (2009), 71 – 76.

4. Karácsony Zs., *Certain properties of the Nadaraya-Watson estimator*, MicroCAD 2009 International Scientific Conference, vol G., Miskolci Egyetem, Miskolc (2009), 29 – 34.

5. Karácsony Zs., *A regressziós függvény becslésének határeloszlása*, Doktoranduszok fóruma, Miskolci Egyetem, Gépészmérnöki és Informatikai Kar, Miskolc (2008), 29 – 34.

6. Karácsony Zs., *Asymptotic normality of kernel type regression estimators for random fields*, MicroCAD 2008 International Scientific Conference, vol G., Miskolci Egyetem, Miskolc (2008), 31 – 36.

7. Karácsony Zs., *Centrális határeloszlás tételek véletlen mezőkre*, Doktoranduszok fóruma, Miskolci Egyetem, Gépészmérnöki és Informatikai Kar, Miskolc (2006), 78 – 83.

8. Karácsony Zs., *A central limit theorem for mixing random fields*, MicroCAD 2006 International Scientific Conference, vol G., Miskolci Egyetem, Miskolc (2006), 59 – 65.

9. Karácsony Zs., *Autoregressive type martingale fields*, MicroCAD 2005 International Scientific Conference, vol G., Miskolci Egyetem, Miskolc (2005), 79 – 85.

10. Karácsony Zs., *Autoregresszív típusú martingál mezők*, Doktoranduszok fóruma, Miskolci Egyetem, Gépészmérnöki Kar, Miskolc (2005), 87 – 93.

11. Karácsony Zs., *Neurális hálók alkalmazásai*, Doktoranduszok fóruma, Miskolci Egyetem, Gépészmérnöki Kar, Miskolc (2004), 126 – 131.

A szerző konferencia-előadásai/List of conference talks

1. Karácsony Zs., *Longest runs in coin tossing.* MicroCAD 2010 International Scientific Conference, Miskolci Egyetem, Miskolc, 2010. március 18 – 19.

2. Karácsony Zs., Libor Zs., *Longest runs in coin tossing. Recursive formulae, asymptotic theorems, computer simulations*, 8th ICAI, Eger, 2010. január 27 – 30.

3. Karácsony Zs., *Asymptotic normality of kernel type regression estimators for random fields*, 14th Young Statisticians Meeting, Basovizza (Trieste), Olazország, 2009. október 17 – 18.

4. Karácsony Zs., *Certain properties of the Nadaraya-Watson estimator*, 16th European Young Statisticians Meeting, Bukarest, Románia, 2009. augusztus 24 – 28.

5. Karácsony Zs., *Asymptotic normality of kernel type regression estimators for random fields*, International conference Probability and Statistics with Application, Debreceni Egyetem, 2009. június 8 – 12.

6. Fazekas I., Karácsony Zs., Libor Zs., *Longest runs in coin tossing. Recursive formulae, asymptotic theorems, computer simulations*, International conference Probability and Statistics with Application (poszter), Debreceni Egyetem, 2009. június 8 – 12.

7. Karácsony Zs., *Asymptotic normality of kernel type regression estimators for random fields*, XXVII International Seminar on Stability Problems for Stochastic Models, Zakopane, Lengyelország, 2009. május 31 - június 5.

8. Karácsony Zs., *Certain properties of the Nadaraya-Watson estimator*, MicroCAD 2009 International Scientific Conference, Miskolci Egyetem, Miskolc, 2009. március 19 – 20.

9. Karácsony Zs., *A regressziós függvény becslésének határeloszlása*, Doktoranduszok fóruma, Miskolci Egyetem, Gépészmérnöki és Informatikai Kar, Miskolc, 2008. november 13.

10. Karácsony Zs., *Asymptotic normality of kernel type regression estimators for random fields*, MicroCAD 2008 International Scientific Conference, Miskolci Egyetem, Miskolc, 2008. március 20 – 21.

11. Karácsony Zs., *Centrális határeloszlás tételek véletlen mezőkre*, Doktoranduszok fóruma, Miskolci Egyetem, Gépészmérnöki Kar, Miskolc, 2006. november. 7.

12. Karácsony Zs., *A central limit theorem for mixing random fields*, MicroCAD 2006 International Scientific Conference, Miskolci Egyetem, Miskolc, 2006. március 16 – 17.

13. Karácsony Zs., *Autoregresszív típusú martingál mezők*, Doktoranduszok fóruma, Miskolci Egyetem, Gépészmérnöki Kar, Miskolc, 2005. november. 9.

14. Karácsony Zs., *Autoregressive type martingale fields*, MicroCAD 2005 International Scientific Conference, Miskolci Egyetem, Miskolc, 2005. március 10 – 11.

15. Karácsony Zs., *Neurális hálók alkalmazásai*, Doktoranduszok fóruma, Miskolci Egyetem, Gépészmérnöki Kar, Miskolc, 2004. november. 9.

16. Karácsony Zs., *A neurális hálók alapjai és alkalmazásai*, A Debreceni Egyetem Hatvani István Szakkollégiuma Országos Hallgatói Konferenciája, Debrecen, 2003. április 24 – 25.