

A modern valószínűségszámítás néhány problémája című pályázatunkban a következő kérdésekkel foglalkoztunk:

- a) Olyan mind a valószínűségszámítás mind a számelmélet számára érdekes jelenségek vizsgálatával, ahol a vizsgált valószínűségi változók viselkedésén kívül a (determinisztikus) együtthatók számelméleti tulajdonságai is fontos szerepet játszanak.
- b) Olyan statisztikai problémákkal, ahol néhány rendkívül nagy tag megjelenése miatt a hagyományos statisztikai módszerek nem alkalmazhatóak, ezért a feladatok vizsgálatában új, hasznosabb módszereket kell kidolgozni.
- c) Bizonyos sztochasztikus folyamatoknak Wiener folyamatokkal való jó közelítéséről szóló eredmények általánosításával sztochasztikus folyamatok gazdagabb osztályára.
- d) Normalizált empirikus mértékek szerinti többszörös integrálok és U -statisztikák becslésével és ezen eredmények alkalmazásával nem triviális statisztikai problémák vizsgálatában.

Végül megjegyzem, hogy később csatlakozott csoportunkhoz Pham Ngoc Anh, aki algebrai problémákkal foglalkozik. De az ő algebrai eredményei hasznosak lehetnek a valószínűségszámítás számára is.

Az a) pontban tárgyalt problémák hátterében a következő jelenség áll. Tudjuk, hogy valószínűségi változók tetszőleges (függő) sorozatát elegendően megritkítva, a kapott részsorozat úgy viselkedik, mintha tagjai független, azonos eloszlású (f.a.e.) változók lennének. Ezt hívják az irodalomban részsorozat-elynek. Például a trigonometrikus rendszer elegendően ritka részsorozataira teljesül a centrális határeloszlástétel és az iterált logaritmus tétel. Mivel egy f.a.e. sorozat tagjainak tetszőleges permutációja után is f.a.e. marad, természetes lenne azt várni, hogy hézagos sorok viselkedése is permutáció-invariáns. Ez azonban, mint Fukuyama (2009) megmutatta, nem igaz. Vizsgáltuk, hogy milyen esetekben igaz és milyen esetekben nem igaz a “részsorozat-ely”. A [10] dolgozatban megmutattuk, hogy a trigonometrikus rendszer Hadamard hézagos részsorozataira, azaz a $\sin n_k x$ függvényekre az

$$n_{k+1}/n_k \geq q > 1 \quad k = 1, 2, \dots \quad (1)$$

feltétel mellett igaz a centrális határeloszlás és az iterált logaritmus tétel, de ezek az eredmények nem érvényesek akkor, ha az (1) hézagfeltételt tetszőlegesen gyengítjük, azaz ha

$$n_{k+1}/n_k \geq 1 + \varepsilon_k, \quad \varepsilon_k \rightarrow 0. \quad (2)$$

Másrészt, elég lassan csökkenő ε_k mellett az eredeti (nem permutált) sorozatra még igaz a centrális határeloszlás és iterált logaritmus tétel. A [10]-es dolgozatban megmutattuk azt is, hogy exponenciálisnál lassabban növekvő (n_k) mellett a $\sin n_k x$ sorozatra ezek az eredmények permutáció-invariáns formában is igazak, ha az n_k együtthatók teljesítik a következő számelméleti jellegű feltételt. Az

$$a_1 n_{k_1} + \dots + a_p n_{k_p} = \ell, \quad 1 \leq k_1, \dots, k_p \leq N \quad (3)$$

alakú diofantoszi egyenleteknek “kevés” megoldása van.

Viszont ennek a feltételnek az ellenőrzése, azaz a (3) egyenlet megoldási számának a meghatározása (vagy becslése) az általános esetben rendkívül nehéz feladat. Ezért érdemes külön tekinteni azt a speciális esetet, amikor (n_k) egy úgynevezett Hardy-Littlewood-Pólya sorozat, azaz $p_1^{k_1} \cdots p_r^{k_r}$ alakú számokból áll, ahol $p_1, \dots, p_r$ prímszámok egy tetszőleges fix véges halmaza és a $k_1, \dots, k_r \geq 0$ egész számok tetszőlegesek. A [9] dolgozatban ugyanezzel a problémával foglalkoztunk az $f(n_k x)$ sorozat esetén, ahol f egy sima periodikus függvény, és (n_k) Hadamard hézagos. Megmutattuk, hogy ebben az esetben a permutáció-invariancia szükséges és elégséges feltétele az, hogy a $p = 2$ esetben a (3) diofantoszi egyenlet megoldásainak száma $N \rightarrow \infty$ esetén $O(1)$, egyenletesen $a_1, \dots, a_p$ -ben. Az [5] dolgozatban pedig megmutattuk, hogy egy végtelen rendű diofantoszi feltétel mellett a permutáció-invariancia a trigonometrikus rendszer részsorozataira minden növekedési feltétel mellett fennáll.

Foglalkoztunk olyan kérdésekkel is, amelyek arról szólnak, hogy amennyiben a $[0, 1]$ intervallumból kiválasztunk véletlenül egy x pontot, akkor a pozitív egész számok mely (számelméleti tulajdonságú) $n_1, n_2, \dots$ sorozataira igaz, hogy az $\{n_k x\}$ sorozat hasonló valószínűségszámítási törvényeket teljesít, mint a független a $[0, 1]$ intervallumban egyenletes eloszlású $x_1, x_2, \dots$ valószínűségi változók sorozata. Az alábbiakban $\{x\}$ az x szám tört részét jelöli. Ilyen jellegű kérdéseket tárgyalunk az alábbiakban.

Egy véges $(x_1, \dots, x_n) \subset (0, 1)$ sorozat *-diszkrepanciáját a következőképpen definiáljuk:

$$D_N^* = D_N^*(x_1, \dots, x_n) := \sup_{0 \leq a < 1} \left| \frac{\sum_{k=1}^N I_a(x_k)}{N} - a \right|,$$

ahol I_a a $[0, a]$ intervallum indikátorfüggvényét jelenti. Vegyük észre, hogy ez nem más, mint az $(x_1, \dots, x_n)$ minta Kolmogorov-Szmirnov statisztikája. Philipp (1975) egy ismert eredménye szerint, ha (n_k) kielégíti az (1) Hadamard hézagfeltételt, akkor

$$0 < \limsup_{N \rightarrow \infty} \frac{ND_N^*(\{n_k x\}_{1 \leq k \leq N})}{\sqrt{2N \log \log N}} < +\infty \quad \text{a.e.} \quad (4)$$

Ez éppen a Kolmogorov-Szmirnov iterált logaritmus tétel az $\{n_k x\}$ sorozatra. A klasszikus statisztikai eredménnyel szemben azonban az (4)-beli $\limsup$ nem egyenlő $1/2$ -el, és általában az x számtól is függ. A $\sqrt{N}D_N^*$ aszimptotikus viselkedése (amely a klasszikus esetben a Kolmogorov-féle $K(y)$ határeloszláshoz tart) azonban ismeretlen. A [4] dolgozatban ezt a határeloszlást határoztuk meg az $n_k = a^k$ ($a \geq 2$) esetben. Megmutattuk, hogy a megfelelő folyamat gyengén konvergál egy Gauss folyamathoz, amelynek kovarianciája érzékenyen függ az a szám számelméleti tulajdonságaitól. Ha $n_{k+1}/n_k \rightarrow \infty$, akkor a határfolyamat a Brownian bridge. Ekkor tehát az $\{n_k x\}$ változók pontosan úgy viselkednek, mint a független valószínűségi változók.

A [3] dolgozat egy áttekintő cikk e téma 100 évre visszamenő történetéről.

Ugyancsak a valószínűségszámítás és a számelmélet közös témája az a probléma, amelyet a [11] dolgozatban tárgyaltunk. Itt a $\sum_{n=1}^N f(n)X_n$ alakú összegek aszimptotikus

viselkedésének a vizsgálatával foglalkoztunk abban az esetben, amikor $f(n)$ egy additív számelméleti függvény. Ilyen esetekben az együtthatók irregularitása miatt a klasszikus eredmények nem érvényesek, és számos új jelenség lép fel.

A b) pontban megjelölt témakörbe tartoznak a [1], [8], [13], [15] dolgozatokban tárgyalt eredmények. Itt a matematika statisztika paraméteres eljárásait vizsgáltuk végtelen szórású valószínűségi változók esetében. Ilyenkor számos statisztikai próba esetén a normális határeloszlás, illetve a Wiener folyamat helyett ezek stabilis megfelelői lépnek fel. Ez lényeges problémát jelent a gyakorlatban, mivel e folyamatok eloszlását nem tudjuk megadni zárt alakban. Ilyenkor természetes lenne a próba kritikus értékeit bootstrap vagy permutációstatisztikák segítségével meghatározni, de ez a módszer sem működik, mivel, mint azt [1]-ben megmutattuk, végtelen szórás esetén a bootstrap statisztikák határeloszlása véletlen paramétereket tartalmaz. Ennek oka a statisztikai mintákban fellépő nagy mintaelemek jelenléte, amelyek a bootstrap statisztikában nem halnak ki. Ezt a nehézséget a trimming alkalmazásával, azaz a minta megnyírásával küszöbölhetjük ki: az n elemű mintából ilyenkor elhagyva a d_n legnagyobb és legkisebb elemet, ahol $d_n \rightarrow \infty$ és $d_n/n \rightarrow 0$, a maradék mintaelemekhez tartozó statisztika (például a changepoint vizsgálatoknál használt CUSUM statisztika) határfolyamata Gauss-folyamat. Független változók megnyírt összegeit sokan vizsgálták, és számos fontos eredmény született e témakörben. Ugyanakkor több fontos kérdés megválaszolatlan maradt. Például Griffin és Pruitt megmutatták, hogy modulus trimming esetén a centrális határeloszlástétel csak szimmetrikus változók esetében igaz, és az általános eset nyitva maradt. Az [8] dolgozatban erre adunk teljes választ, megmutatván, hogy a nem szimmetrikus esetben a centrális határeloszlás tétel csak véletlen centráló faktor mellett igaz, és ez a megfelelő statisztikai eljárás módosítását kívánja meg. Függő változók megnyírt összegeire pedig nincsenek eredmények, noha az idősorok statisztikájában és a közgazdasági matematikában fellépő lineáris és nemlineáris folyamatok a legtöbb esetben "heavy tail" jellegűek, és sokszor szórásuk sem létezik. [15] dolgozatunk egy centrális határeloszlástételt szolgáltat függő, megnyírt folyamatokra; a vizsgált osztály a stabilis felújításokkal rendelkező autoregresszív folyamatok osztálya. [13] dolgozatunk a téma irodalmát tekinti át.

A [7], [19] dolgozatok tartoznak a c) pontban megjelölt témakörbe. Ezekben a Komlós-Major-Tusnády approximációtételt terjesztettük ki függő valószínűségi változókra. [19]-ben megmutattuk, hogy ha (X_n) egy Bernoulli shift sorozat, azaz teljesül rá az

$$X_n = f(\dots, \varepsilon_{n-1}, \varepsilon_n, \varepsilon_{n+1}, \dots), \quad n = \pm 1, \pm 2, \dots$$

előállítás, ahol (ε_k) független egyforma eloszlású valószínűségi változók sorozata, és f egy olyan mérhető függvény, amely elég gyors polinomiális rendben cseng le az ε_k változói mentén $k \rightarrow \infty$ mellett, akkor az $EX_0 = 0$, $E|X_0|^p < \infty$ feltétel esetén teljesül az

$$\sum_{k=1}^n X_k = W(\sigma^2 n) + o(n^{1/p}) \quad \text{a.s.} \quad (5)$$

approximáció valamely $\sigma^2 \geq 0$ mellett, ahol W egy Wiener-folyamat. Ez az eredmény olyan esetekben ad jó normális approximációt valószínűségi változók sorozatára, ami-

kor a Komlós–Major–Tusnády eredmény nem alkalmazható. A tétel bizonyítása az (X_n) sorozat egy triadikus felbontásán alapul, és igen bonyolult. A [19] dolgozatban megmutattuk, hogy e tételre egy igen egyszerű bizonyítás adható, ha az (5) reláció jobboldalán a $W(n)$ Wiener folyamat mellett egy további $W_2(\tau_n)$ tagot is megengedünk, ahol W_2 is egy Wiener folyamat, és $\tau_n = O(n^\gamma)$ valamely $0 < \gamma < 1$ kitevővel. Mint [19]-ban megmutattuk, a második Wiener folyamat annak kisebb skálafaktora miatt a tétel alkalmazásait nem zavarja, noha a W_2 Wiener folyamat nem független a W Wiener folyamattól.

A d) témakörbe tartoznak a [2] és [16] illetve bizonyos mértékig a [20] munkák. A [2] dolgozat áttekintő jellegű, míg [16] egy a Springer Mathematical Lecture Note-ban megjelent munka, ahol egy a valószínűségszámításban fontos elmélet részletesen ki van dolgozva. A Lecture Note-ban vizsgált fő kérdés a következő. Tekintsük független, egyforma eloszlású valószínűségi változók $\xi_1, \dots, \xi_n$ sorozatát valamilyen μ eloszlással, és valós értékű k -változós függvényeknek egy szép tulajdonságokkal rendelkező $\mathcal{F}$ családját. Adjunk jó becslést a

$$\sup_{f \in \mathcal{F}} \int' f(x_1, \dots, x_k) \mu_n(dx_1) \dots \mu_n(dx_k) \quad (6)$$

véletlen integrálok szuprémumának a farokeloszlására, ahol μ_n a $\xi_1, \dots, \xi_n$ sorozat normalizált empirikus mértéke, és a vessző az $\int'$ -ban azt jelenti, hogy az átlókat kihagyjuk az integrálási tartományból. A [16] Lecture Note több korábbi cikk eredményének a feldolgozását tartalmazza, de maga a könyv több, mint a meglevő eredmények feldolgozása. A könyv számos eredménynek új bizonyítását tartalmazza, és ezek a bizonyítások jobban elmagyarázzák a problémák és eredmények hátterét.

A Lecture Note egyrészt annak becslésével foglalkozik, hogy a (6) formulában felírt véletlen integrálok szuprémuma milyen valószínűséggel vesz fel viszonylag nagy értékeket, másrészt annak megmutatásával, hogy ezen probléma megoldása más vizsgálatokban is hasznos lehet. Vannak olyan a matematikai statisztikában fontos határeloszlástételek, amelyek bizonyítása az előbb megfogalmazott feladat megoldásán alapul. Bizonyos nem paraméteres maximum likelihood feladatok megoldása a vizsgált statisztika alkalmas Taylor sorfejtéshez hasonló közelítésén alapul, és a bizonyítás fő nehézsége annak megmutatása, hogy a számolásban megjelenő maradéktagok elhanyagolhatóan kicsik. Valójában egy olyan hibatag van, amely nehezen becsülhető, és ennek vizsgálatában a (6) formulában felírt kifejezés farokeloszlásának a becslésére van szükségünk, azaz a [16] jegyzet fő problémájának a megoldására.

A (6) kifejezés eloszlásának vizsgálata akkor is nehéz, ha az $\mathcal{F}$ függvényosztály egyetlen függvényből áll. Ekkor a (6) véletlen integrál momentumait ki tudjuk számolni egy úgynevezett diagram formula segítségével. Ennek a diagram formulának a bizonyítása önmagában is érdekes és fontos eredmény, amely más vizsgálatokban is hasznos.

Az igazán nehéz probléma az, amikor nem csak egy integrálnak, hanem integrálok szuprémumának a farokeloszlására akarunk jó becslést adni. Ennek a problémának a vizsgálata a modern valószínűségszámítás egy fontos módszerének, az úgynevezett

Vapnik–Červonenkis osztályok elméletének az alkalmazását és az ott kidolgozott módszerek finomítását igényelte. Ezenkívül számos egyéb technikai problémát is meg kellett oldani.

Hasonló problémákkal foglalkozott több cikkben és egy könyvben is a mai valószínűségszámítás egyik leginkább megbecsült alakja, Michel Talagrand. Érdeemes e könyv eredményeit összehasonlítani Talagrand eredményeivel. Természetes módon felmerül a kérdés, hogy melyik eredmény élesebb. Az én válaszom erre a kérdésre kissé meglepő lehet. A két eredmény nem összehasonlítható. Mindkét esetben éles eredményeket kapunk lényegében ugyanarra a problémára. De ezek az eredmények más feltételek mellett érvényesek. Az eredményekben megjelenő különböző feltételeknek komoly, nem technikai jellegű oka van. Talagrand azt az esetet vizsgálta, amikor a tekintett véletlen integrálok nagyon hasonlítanak a Gauss esetre, egy olyan esetre, amely gyakran előfordul, és jól vizsgálható. Talagrand megtalálta azokat a lehető leggyengébb feltételeket, amelyek lehetővé tesznek egy jó Gauss típusú közelítést, és ilyen módon kapott éles eredményeket. A [16] könyv viszont azzal az esettel foglalkozik, amikor a Gauss típusú közelítés nem működik, és a jó becslések más okok miatt érvényesek. Vannak olyan feladatok, amelyekben Talagrand eredményei a hasznosak, és vannak olyanok, amelyekben a [16] könyv eredményei. Megjegyzem, hogy a [16] könyv vizsgálatának háttérben lévő statisztikai problémák vizsgálatában Talagrand eredményei nem alkalmazhatók.

A [20] Lecture Note egy régebbi Lecture Note új, javított kiadása. Úgy gondoltam érdemes ezt a Lecture Note-ot újra kiadni, mert az ott tárgyalt téma (nem centrális határeloszlástételek Gauss mezők nem lineáris funkcionáljaira) és az alkalmazott módszerek (több-változós Wiener–Itô integrálok elmélete) továbbra is aktuális, és a jegyzet eredeti változata nehezen olvasható. A Springer kiadó elfogadta az érveimet, és újra kiadta ezt a könyvet. Ez egyrészt tipográfiai új (az első változat megjelenése idején még nem létezett a $\text{\TeX}$ program), de a matematikai részben is sok javítás, finomítás van, amelyek reményeim szerint ezt a könyvet jobbá tették.

Csatlakozott csoportunkhoz Pham Ngoc Anh, aki a [14] dolgozattal vett részt ebben az OTKA kutatásban. Ez algebrai jellegű eredmény. De az ilyen mátrixokról szóló eredmények hasznosak lehetnek a Markov láncok vizsgálatában is. Sőt, az ebben a dolgozatban vizsgált Virasoro és Heisenberg algebrák fontos szerepet játszanak bizonyos statisztikus fizikai problémák vizsgálatában, és ez is egy kapocs ezen OTKA témája és Pham Ngoc Anh kutatása között.

Major Péter