معرفی روش انقباض

رامین کاظمی[۱]

تاریخ دریافت: ۱۳۹۸/۱/۲۸
تاریخ پذیرش: ۱۳۹۸/۷/۲۳

چکیده:
هدف این مقاله، معرفی روش انقباض برای تحلیل الگوریتم‌ها است. بر اساس این روش، چندین رده از روابط بازگشتی می‌توانند به‌عنوان حالت‌های خاص چارچوب کلّی بیان شده تحلیل شوند. گام‌های اصلی این فن بر اساس ویژگی‌های انقباض الگوریتم نسبت به مترهای احتمالیِ مناسب پایه‌ریزی می‌شوند. نوعاً توزیع حدی به‌عنوان نقطه ثابت یک عملگر حدی روی ردهٔ توزیع‌های احتمال مشخص‌سازی می‌شود.
واژه‌های کلیدی: الگوریتم‌های بازگشتی، متر احتمال، عملگر حدی، نقطه ثابت، توزیع حدی.

۱ مقدمه

روش‌های اصلی برای تحلیل مجانبی الگوریتم‌ها روش‌های تبدیل (به‌ویژه، توابع مولّد گشتاور) [۳] ، روش مارتینگلی [۴] و روش فرایندهای شاخه‌ای [۲] از طریق تقریب به‌وسیله برون‌گشت‌های براونی هستند. برای یک رده محدودتر از الگوریتم‌های تصادفی روش پایه‌ریزی شده روی تقریب‌های تصادفی نیز در این میان قرار می‌گیرند. روش گشتاورها، یکی از کلاسیک‌ترین روش‌های دست‌یابی توزیع حدی، در مسائل گوناگون در حوزه‌های مختلف است [۱۰]. این روش شامل گام‌های زیر است:

۱- میانگین و واریانس محاسبه می‌شوند.

۲- متغیر تصادفی به‌طور سَره‌ای مقیاسیده می‌شود.

۳- با استفاده از استقراء گشتاورهای بالاتر متغیر تصادفی مقیاسیده محاسبه می‌شوند.

۴- معیار کارلِمان برای تصدیق یکتاییِ قاعده حدی به‌کار گرفته می‌شود.

۵- بنا بر قضیهٔ همگرایی گشتاوری فِرشه-شوهات، همگرایی در توزیع و همگرایی در همه گشتاورها (یا همگرایی در L_p برای هر $p > 0$) نتیجه می‌شود.

روش دیگری موسوم به روش انقباض وجود دارد. اما روش انقباض چیست؟ روش انقباض یکی از مفیدترین روش‌ها برای تحلیل مجانبی الگوریتم‌ها است. این روش تا حدودی شبیه به روش گشتاوری است. روش انقباض برای تحلیل مجانبی یک الگوریتم بازگشتی شامل مراحل زیر است: ابتدا با مطالعه اوّلین گشتاورها مجبور به نرمال‌سازی صحیح الگوریتم هستیم. نسخه نرمال شده مجدداً در یک شکل بازگشتی ارائه می‌شود. سپس اغلب یک معادله نقطه ثابت برای معادله حدی الگوریتم حدس زده می‌شود. این معادله نقطه ثابت حدی توسط یک معادله به شکل $K\mu = \mu$ برای یک عملگر حدی K روی مجموعه همه توزیع‌های احتمال بیان می‌شود. در گام بعدی، مجبور به اثبات وجود یک حل از این معادله نقطه ثابت هستیم. این کار توسط انتخاب یک متر احتمال مناسب روی فضای اندازه‌های احتمال انجام می‌شود به‌طوری‌که K یک نگاشت انقباضی نسبت به این متر است و قضیهٔ نقطه ثابت باناخ به‌کار گرفته می‌شود. در گام آخر، مجبور به اثبات همگرایی الگوریتم (در توزیع) به این نقطه ثابت هستیم. این کار نیز با معرفی دنباله‌های همراه توزیع‌ها که الگوریتم را همانندسازی و همچنین ساختار معادله حدی را

[۱] هیأت علمی گروه آمار، دانشگاه بین المللی امام خمینی، قزوین، ایران

منعکس می‌کنند، انجام می‌شود.

اوّلین مثال که در آن از روش انقباض استفاده شد در مقاله نوشته شده توسط روسلا و در مطالعه الگوریتم مرتب‌سازی سریع بود [۱۱]. در شکل کلّی، این روش به‌طور مستقل توسط روسلا [۱۲] و راشِف و روشندورف [۱۰] بهبود داده شده است. بعد از این، در بسیاری از مطالعه‌های این روش بر روش‌های احتمالی و روش‌های تحلیلی و ترکیبیاتی ترجیح داده شده است.

یِک موضوع کلاسیک نظریهٔ احتمال، قضیهٔ حدی مرکزی برای دنباله‌هایی از مجموع‌هایی از متغیرهای تصادفی وابسته است. این دنباله‌های تجمعی یک ساختار بازگشتی ساده دارند. مترهای احتمال و فن‌های انقباض در این حالت، به‌ویژه در مدرسه روسی احتمال برای اثبات قضیهٔ حدی مرکزی معرفی شده‌اند. در این مدرسه، مترهای احتمال مفید، ویژگی‌های نظم آنها، و ارتباط بین آنها به‌طور جدی مطالعه شده است (برای اطلاعات بیشتر به [۹، ۱۰، ۱۵] مراجعه کنید). این ویژگی‌های نظم در تحلیل الگوریتم‌های بازگشتی کلّی مفید هستند.

۲ گام‌های ترتیبی

فرض کنید $(X_n)_{n\in\mathbb{N}}$ دنباله‌ای از متغیرهای تصادفی باشد که برای هر n، $n \geq n_۰$، در برابری توزیعی

$$X_n \stackrel{d}{=} \sum_{i\in\mathbb{N}} T_{n,i} X^{(i)}_{Z_{n,i}} + C_n \qquad (۱)$$

صدق می‌کند. عدد $n_۰$ یک عدد طبیعی و متغیرهای تصادفی (C_n, Z_n, T_n) و $X_j^{(i)} \in \mathbb{N}$ برای $۰ \leq j \leq n-۱$ مستقل هستند.

از نمادگذاری $Z_n = (Z_{n,۱}, Z_{n,۲}, \dots)$ و $T_n = (T_{n,۱}, T_{n,۲}, \dots)$ استفاده می‌کنیم. متغیرهای تصادفی $Z_{n,i}$ مقادیر خود را در مجموعه $\{۰, ۱, \dots, n-۱\}$ اتخاذ می‌کنند. توزیع X_j برای $۰ \leq j \leq n_۰ - ۱$ داده می‌شود و توزیع $X_j^{(i)}$ همان توزیع X_j است. همچنین، قرار می‌دهیم $X_۰ \equiv ۰$. تحلیل X_n در گام‌های ترتیبی زیر صورت می‌گیرد.

گام اول (امید ریاضی): امید ریاضی $a_n := \mathbb{E}(X_n)$ محاسبه می‌شود. امید ریاضی در معادله بازگشتی

$$a_n = \sum_{i\in\mathbb{N}} \mathbb{E}(T_{n,i}) a_{Z_{n,i}} + c_n \qquad (۲)$$

صدق می‌کند به‌طوری‌که $n_۰ \leq n$ و $c_n = \mathbb{E}(C_n)$. شکل استاندارد معادله به‌صورت

$$a_n = \sum_{j=۰}^{n-۱} \nu_n(j) a_j + c_n \qquad (۳)$$

است. در این‌جا ν_n یک اندازه علامت‌دار روی مجموعه $\{۰, \dots, n-۱\}$ و نه لزوماً یک اندازهٔ احتمال است. در بیشتر مواقع، به رفتار مجانبی a_n در n علاقه‌مند هستیم. در مثال‌های استاندارد، نه جایی‌که $a_n \to ۰$، a_n از مرتبه

$$(\text{ثابت})\ n \ln n + (\text{ثابت})\ n + o(n)$$

یا مرتبهٔ

$$(\text{ثابت})\ n + o(n)$$

و یا از مرتبه $\ln n$ است.

روش‌های متعدد و فن‌های گوناگونی از جمله یک برهان با یک حدس، توابع مولّد و چند معادله جبری، معادلات دیفرانسیلی و معادله تفاضلی در حد، تفسیر احتمالی از طریق فرایندهای تجدید یا مارکف با استفاده از مسئله دریکله و کوشی وجود دارند. یک روش کلّی از نظریه تجدید برای این معالات برخاسته از تحلیل الگوریتم‌ها در روسلا [۱۱] وجود دارد.

اگر ν_n تنها جرمی در نقطه $n-۱$ داشته باشد، آن‌گاه معادله

$$a_n = \nu_n(n-۱) a_{n-۱} + c_n \qquad (۴)$$

ساده‌تر (با تکرار یا از طریق معادلات تفاضلی یا معادلات دیفرانسیل حدی) حل خواهد شد.

گام دوم (نرمال کردن): نرمال کردنِ درست حدس زده می‌شود به‌طوری‌که نسخه نرمال شده Y_n از X_n به یک حد W همگرا باشد. به نسخه

$$Y_n := \frac{X_n - a_n}{b_n}$$

برای $b_n > ۰$ توجه می‌کنیم. قاعده کلّی و دقیقی برای یک انتخاب مناسب b_n نمی‌توان ارائه کرد. یک روش، انتخاب b_n به اندازه ممکن کوچک است به‌طوری‌که نسخه نرمال شده C_n^* از

C_n تعریف شده در زیر به یک حد و در صورت امکان به یک حد ناتباهیده همگرا باشد. اگر واریانس متناهی باشد و قابل محاسبه، در این صورت $b_n=\sqrt{var(X_n)}$ گزینه مناسبی است. مثال‌های نوعی b_n از مرتبه n یا $\sqrt{n}$ و یا $\ln n$ است. معادله پایه (۱) به‌صورت

$$\begin{aligned} Y_n &\overset{d}{=} \sum_{i\in\mathbb{N}} T_{n,i}\frac{b_{Z_{n,i}}}{b_n}Y^{(i)}_{Z_{n,i}} \\ &+ \sum_{i\in\mathbb{N}} \frac{T_{n,i}a_{Z_{n,i}}-\mathbb{E}(T_{n,i})a_{Z_{n,i}}}{b_n}+C_n-\mathbb{E}(C_n) \end{aligned} \tag{۵}$$

قابل بازنویسی است. این معادله را معادله پایه‌ای نرمال شده گویند. این معادله ساختاری مشابه با معادله (۱) با $T_{n,i}$ و C_n جایگزین شده توسط

$$\begin{aligned} T^*_{n,i} &:= T_{n,i}\frac{b_{Z_{n,i}}}{b_n}, \\ C^*_n &:= \sum_{i\in\mathbb{N}} \frac{T_{n,i}a_{Z_{n,i}}-\mathbb{E}(T_{n,i})a_{Z_{n,i}}}{b_n}+C_n-\mathbb{E}(C_n) \end{aligned}$$

دارد. اختلاف اصلی با معادله (۱) این است که همه Y_iها دارای میانگین صفر هستند.

گام سوم (توزیع حدی). یک توزیع حدی W از Y_n حدس زده می‌شود. در بسیاری از وضعیت‌ها، توزیع حدی می‌تواند به‌عنوان نقطه ثابت یک عملگر K حدس زده شود. قضیهٔ نقطه ثابت نقش مهمی در کاربردهای شاخه‌های مختلف ریاضی بازی می‌کند. اصطلاح قضیهٔ نقطه ثابت متریک به آن دسته از نتایج نظری نقطه ثابت اشاره دارد که در شرایط هندسی روی فضاهای اساسی و نگاشت‌ها نقش حیاتی دارند. باناخ ریاضی‌دان لهستانی، نتیجه‌ای بسیار مهم در مورد نگاشت‌های انقباضی ثابت کرد که به‌عنوان اصل انقباض باناخ شناخته می‌شود. این قضیهٔ فنی برای حل مسائل کاربردی گوناگون در علوم ریاضی و مهندسی است.

در معادله نرمال شده (۵)، اگر Y_n به W، $Z_{n,i}$ به ∞ و همچنین (C^*_n,T^*_n) به (C,T) همگرا باشد، آن‌گاه یک معادله حدی به‌صورت

$$\begin{array}{ccccc} Y_n \overset{d}{=} \sum & T^*_{n,i} & Y^{(i)}_{Z_{n,i}} & + & C^*_n \\ \downarrow & \downarrow & \downarrow & & \downarrow \\ W \overset{d}{=} \sum_{i\in\mathbb{N}} & T_i & W_i & + & C, \end{array}$$

داریم. این‌جا (C,T) و W_i برای $i\in\mathbb{N}$ مستقل هستند و همه W_iها همان توزیع W را دارند. توزیع W باید یک حل معادله نقطه ثابت بالا باشد. برخی اوقات، این معادله نقطه ثابت، توزیع حدی را مشخص می‌کند. حال، مایل به فرمول‌بندی این معادله بر حسب توزیع‌ها و یک عملگر هستیم. عملگر K را روی یک زیرمجموعه M از اندازه‌های احتمال به خودش از طریق

$$K(\mu):=\mathcal{L}\left(\sum_{i\in\mathbb{N}}T_iY_i+C\right) \tag{۶}$$

تعریف می‌کنیم. این‌جا (C,T) و Y_i برای هر $i\in\mathbb{N}$ مستقل هستند و Y_iها دارای توزیع μ هستند، و توزیع (C,T) از معادله حدی می‌آید. عملگر K را عملگر حدی می‌نامیم. یک نقطه ثابت K یک اندازه احتمال ν است که در شرط $K(\nu)=\nu$ صدق می‌کند. این شرط معادل با

$$W\overset{d}{=}\sum_{i\in\mathbb{N}}T_iW_i+C \tag{۷}$$

است. توزیع W یک نقطه ثابت است. معادله (۷) را معادله نقطه ثابت و حل آن را یک نقطه ثابت گوییم.

گام چهارم (روش انقباض): این روش ابزاری برای اثبات همگرایی دنباله نرمال شده Y_n به W است؛ ایده پایه یک بحث انقباض روی فضای توزیع‌ها است. نگاشت‌های K_n را برای هر $n\geq n_۰$ روی M^n به M از طریق

$$K_n(\mu_۰,\ldots,\mu_{n-۱}):=\mathcal{L}\left(\sum_{i\in\mathbb{N}}T^*_{n,i}Y^{(i)}_{Z_{n,i}}+C^*_n\right) \tag{۸}$$

تعریف کنید. در این‌جا (C^*_n,Z_n,T^*_n) و متغیرهای $Y^{(i)}_j$ برای $i\in\mathbb{N}$ و $۰\leq j\leq n$ مستقل هستند، $Y^{(i)}_j$ همان توزیع μ_j را دارد و توزیع (C^*_n,Z_n,T^*_n) به‌صورت بالا داده می‌شود.

با شروع از اندازه‌های احتمال $\mu_۰,\ldots,\mu_{n_۰-۱}$، یک دنباله به‌صورت

$$\begin{aligned} &\mu_{n_۰}=K_{n_۰}(\mu_۰,\ldots,\mu_{n_۰-۱}), \\ &\mu_{n_۰+۱}=K_{n_۰+۱}(\mu_۰,\ldots,\mu_{n_۰}), \\ &\vdots \\ &\mu_n=K_n(\mu_۰,\ldots,\mu_{n-۱}) \end{aligned}$$

به دست می‌آید. توزیع μ_n همان توزیع Y_n است هرگاه $\mu_۰,\ldots,\mu_{n_۰-۱}$ توزیع‌های $X_۰,\ldots,X_{n_۰-۱}$ باشند.

روی یک زیرفضای M از اندازه‌های احتمال به متر d توجه کنید (برای سادگی از نماد $d(X,Y) := d(\mathcal{L}(X), \mathcal{L}(Y))$ استفاده می‌شود). فرض کنید v_n دنباله‌ای از توزیع‌ها و همگرا به توزیع معلوم v باشد. معمولاً v_n در یک معادله بازگشتی مشابه با μ_n صدق می‌کند که بر حسب یک عملگر L_n فرمول‌بندی می‌شود. دنباله v_n را دنباله همراه می‌نامند. روش انقباض روشی برای نشان دادن این است که اختلاف بین μ_n و v_n به صفر همگرا است که البته بعد از حدس زدن یک L_n و دنباله v_n اتفاق می‌افتاد. این ثابت می‌کند که توزیع Y_n همگرا به‌توزیع v است. اغلب دنباله v_n یکسان با نقطه ثابت v برای عملگر حدی است. چون با عملگرهای خطی K سروکار داریم مترهای با ویژگی‌های خاص می‌توانند بهتر از دیگر مترها برای نشان دادن اجرای روش انقباض مورد استفاده قرار گیرند. به‌طور مشخص متر واسراشتاین و متر زُلُتارف از این دسته هستند.

۳ متر واسراشتاین

فرض کنید $f : [a,b] \to \mathbb{R}$ تابعی تعریف شده روی بازه $[a,b]$ باشد به‌طوری‌که انتگرال $\int_a^b |f(x)|^p dx$ متناهی است. برای هر $p \geq ۱$، p-نُرم به‌صورت

$$||f||_p = \left(\int_a^b |f(x)|^p dx \right)^{\frac{1}{p}}$$

تعریف می‌شود. متر ℓ_p با $۱ \leq p < \infty$ که متر مالوس [۱] نیز نامیده می‌شود به‌صورت

$$\ell_p(\mu, v) = \inf \{ \|X - Y\|_p \}$$

تعریف می‌شود. اینفیمم روی همه متغیرهای تصادفی X با توزیع μ و متغیرهای تصادفی Y با توزیع v روی فضای احتمال غیراتمی گرفته می‌شود. علامت $\|\cdot\|_p$ بیان‌کنندهٔ نرم L_p معمول است. نمادگذاری مرسوم برای این متر، ℓ_p است. اینفیمم به‌صورت

$$\ell_p(\mu, v) = \|F_\mu^{-۱}(U) - F_v^{-۱}(U)\|_p$$

است. این‌جا U به‌طور یکنواخت روی بازه‌ی $[۰, ۱]$ توزیع می‌شود و F_μ تابع توزیع μ است. متر ℓ_p دارای ویژگی‌های

$$\ell_p(aX, aY) \leq |a| \ell_p(X, Y),$$
$$\ell_p\left(\sum_i X_i, \sum_i Y_i \right) \leq \sum_i \ell_p(X_i, Y_i),$$
$$\ell_p(AX, AY) \leq \|A\|_p \ell_p(X, Y)$$

است که در آن $a \in \mathbb{R}$ و برای هر $i \in \mathbb{N}$، (X_i, Y_i) مستقل هستند و A مستقل از (X, Y) است. فرض کنید $\mathcal{M}_p$ فضای همه توزیع‌های μ با گشتاورهای قدر مطلق متناهی باشد، یعنی

$$\int |x|^p \mu(\mathrm{d}x) < \infty.$$

فضای $(\mathcal{M}_p, \ell_p)$ یک فضای متریک تفکیک‌پذیر کامل است [۱]. در ادامه، به حالت $p = ۲$ توجه می‌کنیم.

قضیهٔ ۱.۳. *[۱۰] فرض کنید $\mu, v \in \mathcal{M}_۲$ دو اندازهٔ احتمال با*

$$\int x \mu(\mathrm{d}x) = ۰ = \int x v(\mathrm{d}x)$$

باشند. در این صورت،

$$\ell_۲(K(\mu), K(v)) \leq \sqrt{\mathbb{E}\left(\sum_j T_j^۲ \right) \ell_۲(\mu, v)}. \qquad (۹)$$

نابرابری (۹) که به لم انقباض مشهور است، نقش اساسی در روش انقباض بازی می‌کند. قضیهٔ زیر که به قضیهٔ نقطه ثابت باناخ مشهور است، وجود و یکتایی حل معادله نقطه ثابت را تضمین می‌کند.

قضیهٔ ۲.۳. *(قضیهٔ نقطه ثابت باناخ). فرض کنید (E, d) یک فضای متریک کامل باشد. فرض کنید $K : E \longmapsto E$ یک انقباض اکید با ثابت انقباض $k < ۱$ باشد. در این صورت، دنباله $K^n(e)$ برای $n \in \mathbb{N}$ به‌طور نمایی با مرتبه k^n به نقطه ثابت یکتای K برای هر $e \in E$ همگرا است.*

قضیهٔ نقطه ثابت باناخ به قضیهٔ بنیادی زیر در فضای متر $\ell_۲$ منجر می‌شود.

قضیهٔ ۳.۳. *فرض کنید $\mathbb{E}(C) = ۰$، $\mathbb{E}(C^۲) < \infty$ و $\sum_i \mathbb{E}(T_i^۲) < ۱$. در این صورت، دنباله $K^n(\mu)$ برای هر $n \in \mathbb{N}$*

و $\mu \in \mathcal{M}^{\circ}$ به‌طور نمایی در متر $\ell_{۲}$ به نقطه ثابت یکتای K در $\mathcal{M}^{\circ}$ همگرا است.

اثبات. به [۱۴] مراجعه کنید. □

مثال ساده زیر را در نظر بگیرید. فرض کنید $(X_n)_{n\geq ۱}$ دنباله‌ای از متغیرهای تصادفی و $\frac{X_n-\mathbb{E}(X_n)}{s_n}$ نسخه نرمال شده آن باشد. یادآور می‌شویم که $s_n^{۲}$ دقیقاً واریانس نیست. به‌طور مشخص فرض کنید به هر روشی توانسته‌ایم $\mathbb{E}(X_n)$ را به دست آوریم ولی واریانس X_n تعیین نشده است. آیا می‌توان تقریب مناسبی برای واریانس ارائه کرد؟ عموماً $s_n^{۲}$ به‌عنوان مقیاسی مشخص پیشنهاد می‌شود که مرتبه آن توان دوم مرتبه میانگین است. برای مثال اگر $\mathbb{E}(X_n)\sim \log n$، آنگاه می‌توان $s_n=\sqrt{\log n}$ اختیار کرد. دلیل این انتخاب واضح است [۱۴] . اما پرسش اصلی این است که آیا این امکان وجود دارد که متغیری تصادفی چون X پیدا کرد که در یک معادله نقطه ثابت صدق کند که جوابی یکتا دارد، امید آن صفر است و واریانس آن قابل محاسبه است به‌طوری‌که $\frac{X_n-\mathbb{E}(X)}{s_n}\xrightarrow{d} X$. برای یک لحظه فرض کنید واریانس X تعیین شود. در این صورت،

$$var(X_n)\sim s_n^{۲}var(X).$$

روسلا [۱۱] اوّلین بار روش انقباض را در مطالعهٔ تعداد مقایسه‌های مورد نیاز به‌وسیله الگوریتم مرتب‌سازی سریع برای مرتب کردن یک فهرست از n عدد درون ترتیب‌های طبیعی آنها به‌کار گرفت. در ادامه، به این مسئله و چگونگی حل آن اشاره می‌شود. فرض کنید توانسته‌ایم نشان دهیم متغیر تعریف شده در بالا در رابطهٔ هم‌توزیعی زیر صدق می‌کند (صرف نظر از چگونگی آن): برای هر $n\geq ۲$،

$$Y_n \stackrel{d}{=} Y_{I_n}+Y^{*}_{n-۱-I_n}+n-۱,$$

که در آن $Y_{\circ}=Y_{۱}=\circ$، $Y_{۲}=۱$ و I_n دارای توزیع یکنواخت گسسته روی مجموعه $\{\circ,۱,۲,\ldots,n-۱\}$ است. همچنین، $Y_j\stackrel{d}{=}Y_j^{*}$ و I_n، Y_j و Y_j^{*} برای $۱\leq j\leq n$ مستقل هستند.

به‌وضوح،

$$\begin{aligned}\mathbb{E}(Y_n)&=\mathbb{E}(Y_{I_n})+\mathbb{E}(Y^{*}_{n-۱-I_n})+n-۱\\&=\frac{۱}{n}\sum_{j=۱}^{n}\mathbb{E}(Y_{j+۱})+\frac{۱}{n}\sum_{j=۱}^{n}\mathbb{E}(Y_{n-j})+n-۱\\&=۲(n+۱)\sum_{h=۱}^{n+۱}\frac{۱}{h}-۴(n+۱)+۲\\&=۲n\log n+n(۲\gamma-۴)+۲\log n+۲\gamma+۱\\&\quad+O\Big(\frac{\log n}{n}\Big),\end{aligned}$$

زیرا

$$H_n=\sum_{h=۱}^{n}\frac{۱}{h}=\log n+\gamma-\frac{۱}{۲n}+O(n^{-۲}),$$

که در آن γ ثابت اویلر است. حال، دنباله متغیرهای تصادفی $\frac{Y_n-\mathbb{E}(Y_n)}{n}$ را در نظر بگیرید. ادعا می‌کنیم که

$$\frac{Y_n-\mathbb{E}(Y_n)}{n}\xrightarrow{d} X,$$

که در آن متغیر تصادفی X در معادله نقطه ثابت

$$X\stackrel{d}{=}UX+(۱-U)X^{*}+b(U) \qquad (۱۰)$$

صدق می‌کند به‌طوری‌که $U\sim U(\circ,۱)$، $X\stackrel{d}{=}X^{*}$ و X، X^{*} و U مستقل هستند و

$$b(x)=۲x\log x+۲(۱-x)\log(۱-x)+۱.$$

قبل از اثبات این ادعا توجه کنید که

$$\begin{aligned}\mathbb{E}(X)&=\mathbb{E}(UX+(۱-U)X^{*}+b(U))\\&=\circ\end{aligned}$$

و

$$var(X)=\int_{\circ}^{۱}b^{۲}(x)\mathrm{d}x=۷-\frac{۲}{۳}\pi^{۲}$$

بنا بر این،

$$var(Y_n)\sim n^{۲}var(X)=\Big(۷-\frac{۲}{۳}\pi^{۲}\Big)n^{۲}.$$

برای اثبات ادعا، فرض کنید $\mathcal{M}_{۲}$ معرف فضای اندازه‌ها روی $\mathbb{R}$ با گشتاور دوم متناهی و میانگین صفر باشد. در این صورت متر واسراشتاین یا (متر $L_{۲}$) که با $\ell_{۲}$ نشان داده می‌شود به‌صورت

$$\ell_{۲}(\mu,v)=\inf\|X-Y\|_{۲}$$

تعریف می‌شود که در آن $\|\cdot\|_۲$ معرف $L_۲$ -نرم است و اینفیمم روی همه متغیرهای تصادفی X با قانون $\mu = \mathcal{L}(X)$ و همه Yها با قانون $v = \mathcal{L}(Y)$ گرفته می‌شود. روی فضای $(\mathcal{M}_۲, \ell_۲)$، دنباله μ_n به μ همگراست در $\mathcal{M}_۲$ اگر و تنها اگر μ_n در توزیع به μ همگرا باشد و اگر گشتاورهای دوم μ_n به گشتاور دوم μ همگرا باشد [۱] . اگر قرار دهیم $X_n = \frac{Y_n - \mathbb{E}(Y_n)}{n}$، آنگاه

$$X_n \stackrel{d}{=} X_{I_n}\frac{I_n}{n} + X^*_{n-۱-I_n}\frac{n-۱-I_n}{n} + b_n(I_n) \quad n \geq ۲ \tag{۱۱}$$

که در آن $X_۰ = X_۱ = ۰$ و $X_j \stackrel{d}{=} X_j^*$ و I_n، X_j و X_j^* مستقل هستند. دقت می‌کنیم که $U = \frac{I_n}{n}$ دارای توزیع یکنواخت روی بازه $(۰,۱)$ است. به‌علاوه،

$$\begin{aligned} b_n(j) &= \frac{n-۱}{n} + \frac{۱}{n}(\mathbb{E}(Y_j) + \mathbb{E}(Y_{n-۱-j}) - \mathbb{E}(Y_n)) \\ &= ۱ + ۲\frac{j}{n}\log\frac{j}{n} + ۲\left(۱-\frac{j}{n}\right)\log\left(۱-\frac{j}{n}\right) \\ &\quad + O\left(\frac{\log n}{n}\right). \end{aligned}$$

بنا بر این، اگر X_n دارای یک توزیع حدی به X باشد، آنگاه باید در رابطهٔ (۱۰) صدق کند. اوّلین گام نشان دادن این است که رابطهٔ (۱۰) دارای یک حل یکتا با $\mathbb{E}(X) = ۰$ است. فرض کنید T، $T : \mathcal{M}_۲ \to \mathcal{M}_۲$، نگاشتی تعریف شده به‌صورت

$$T(\mu) = \mathcal{L}(UX + (۱-U)X^* + b(U))$$

باشد که در آن X, X^*, U مستقل هستند $\mu = \mathcal{L}(X) = \mathcal{L}(X^*)$، و $U \sim f(u) = ۱,\ ۰ < u < ۱$. در این صورت، T یک انقباض نسبت به متر واسراشتاین $\ell_۲$ است و بنا بر این یک نقطه ثابت $T(\mu) = \mu$ با $\mu \in \mathcal{M}_۲$ است. برای این فرض کنید $\mu, \nu \in \mathcal{M}_۲$ و اینکه $\mathcal{L}(X) = \mathcal{L}(X^*) = \mu$، $\mathcal{L}(Y) = \mathcal{L}(Y^*) = \nu$ به‌طوری‌که U, X, X^* و U, Y, Y^* مستقل هستند. در این صورت،

$$\begin{aligned} T(\mu) &= \mathcal{L}(UX + (۱-U)X^* + b(U)), \\ T(\nu) &= \mathcal{L}(UY + (۱-U)Y^* + b(U)). \end{aligned}$$

در نتیجه،

$$\begin{aligned} &\ell_۲^۲(T(\mu), T(\nu)) \\ \leq\ & \|UX + (۱-U)X^* - UY - (۱-U)Y^*\|_۲^۲ \\ =\ & (\mathbb{E}(U^۲) + \mathbb{E}(۱-U)^۲)\mathbb{E}(X-Y)^۲ \\ =\ & \frac{۲}{۳}\mathbb{E}(X-Y)^۲, \end{aligned}$$

زیرا $\mathbb{E}(U^۲) = ۱/۳$ و $\mathbb{E}(U) = ۱/۲$. با گرفتن اینفیمم روی همه X و Yهای ممکن،

$$\ell_۲(T(\mu), T(\nu)) \leq \sqrt{\frac{۲}{۳}}\ell_۲(\mu, \nu),$$

که ادعا را ثابت می‌کند. گام بعدی نشان دادن این مطلب است که X_n واقعاً به X همگرا است. یادآوری می‌کنیم که $X_n \xrightarrow{d} X$ بر این دلالت دارد که $\ell_۲(\mathcal{L}(X_n), \mathcal{L}(X)) \to ۰$. فرض کنید $\nu_n = \mathcal{L}(X_n)$، X و X^* مستقل باشند و $\mathcal{L}(X) = \mathcal{L}(X^*) = \mu$، که در آن μ نقطه ثابت $T(\mu) = \mu$ است. نسخه‌های X_j, X_j^* که برای $۱ \leq j \leq n-۱$ مستقل هستند را با

$$\begin{aligned} var(X_j - X) &= \ell_۲^۲(\nu_j - \mu), \\ var(X_j^* - X^*) &= \ell_۲^۲(\nu_j - \mu) \end{aligned}$$

انتخاب کنید. برای $x \in \left(\frac{j}{n}, \frac{j+۱}{n}\right]$ قرار دهید $V_x = X_j$ و $V_x^* = X_j^*$. در این صورت، برای U مستقل از X_j و X_j^*،

$$\nu_n = \mathcal{L}\left(\frac{\lceil nU \rceil - ۱}{n}V_U + \frac{n - \lceil nU \rceil}{n}V^*_{۱-U} + b_n(\lceil nU \rceil)\right).$$

همچنین،

$$\sup_{۰ \leq x \leq ۱} |b_n(\lceil nx \rceil) - b(x)| = \mathcal{O}\left(\frac{\log n}{n}\right). \tag{۱۲}$$

بنا بر تعریف متر واسراشتاین،

$$\begin{aligned}
\ell_2^2(v_n,\mu) &\le \mathbb{E}\Big(\frac{\lceil nU\rceil}{n}V_U+\frac{n-\lceil nU\rceil}{n}V^*_{1-U}+b_n(\lceil nU\rceil)\\
&- UX-(1-U)X^*-b(U)\Big)\\
&= \mathbb{E}\Big(\frac{\lceil nU\rceil-1}{n}V_U-UX\Big)^2\\
&+ \mathbb{E}\Big(\frac{n-\lceil nU\rceil}{n}V^*_{1-U}-(1-U)X^*\Big)^2\\
&+ \mathbb{E}(b_n(\lceil nU\rceil)-b(U))^2\\
&+ 2\mathbb{E}\Big(\Big(\frac{\lceil nU\rceil-1}{n}V_U-UX\Big)\\
&\times (b_n(\lceil nU\rceil)-b(U))\Big)\\
&+ 2\mathbb{E}\Big(\Big(\frac{n-\lceil nU\rceil}{n}V^*_{1-U}-(1-U)X^*\Big)\\
&\times (b_n(\lceil nU\rceil)-b(U))\Big)\\
&+ 2\mathbb{E}\Big(\Big(\frac{\lceil nU\rceil-1}{n}V_U-UX\Big)\\
&\times \Big(\frac{n-\lceil nU\rceil}{n}V^*_{1-U}-(1-U)X^*\Big)\Big).
\end{aligned}$$

از طرفی،

$$\begin{aligned}
&\mathbb{E}\Big(\frac{\lceil nU\rceil-1}{n}V_U-UX\Big)^2\\
&=\mathbb{E}\Big(\sum_{j=1}^{n} I_{\left(\frac{j-1}{n},\frac{j}{n}\right]}(U)\Big(\frac{j-1}{n}X_{j-1}-UX\Big)\Big)^2\\
&=\frac{1}{n}\sum_{j=1}^{n}\frac{(j-1)^2}{n^2}\ell_2^2(v_{j-1},\mu)\\
&+O\Big(\frac{1}{n^2}\sum_{j=1}^{n}\frac{j-1}{n}\ell_2(v_{j-1},\mu)\Big)+O\Big(\frac{1}{n^2}\Big).
\end{aligned}$$

برای عبارت دوم نیز به‌طریق مشابه خواهد بود. عبارت سوم با استفاده از رابطهٔ (۱۲) به‌صورت

$$\mathbb{E}(b_n(\lceil nU\rceil)-b(U))^2=O\Big(\frac{(\log n)^2}{n^2}\Big),$$

برآورد می‌شود. بنا بر فرض استقلال و این‌که $\mathbb{E}(X_j)=\mathbb{E}(X)=0$ داریم:

$$\begin{aligned}
&\mathbb{E}\Big(\Big(\frac{\lceil nU\rceil-1}{n}V_U-UX\Big)\\
&\times \Big(\frac{n-\lceil nU\rceil}{n}V^*_{1-U}-(1-U)X^*\Big)\Big)\\
&= O\Big(\frac{1}{n^2}\sum_{j=1}^{n}\frac{j-1}{n}\ell_2(v_{j-1},\mu)\Big)+O\Big(\frac{1}{n^2}\Big).
\end{aligned}$$

در نهایت برای عبارت پنجم (و به‌طور مشابه برای عبارت ششم)،

$$\begin{aligned}
&\mathbb{E}\Big(\Big(\frac{\lceil nU\rceil-1}{n}V_U-UX\Big)\\
&\times (b_n(\lceil nU\rceil)-b(U))\Big)\\
&= O\Big(\frac{\log n}{n^2}\sum_{j=1}^{n}\frac{j-1}{n}\ell_2(v_{j-1},\mu)\Big)+O\Big(\frac{\log n}{n^2}\Big).
\end{aligned}$$

بنا بر این، با انتخاب $a_j=\ell_2^2(v_j,\mu)$،

$$\begin{aligned}
a_n &= \frac{2}{n}\sum_{j=1}^{n}\Big(\frac{j-1}{n}\Big)^2 a_{j-1}\\
&+ O\Big(\frac{\log n}{n^2}\Big)\sum_{j=1}^{n}\frac{j-1}{n}\sqrt{a_{j-1}}+O\Big(\frac{(\log n)^2}{n^2}\Big),
\end{aligned}$$

و لذا برای مقادیر مثبت C_1 و C_2،

$$a_n\le\frac{2}{3}\max_{0\le j\le n-1}a_j+C_1\frac{\log n}{n}\max_{0\le j\le n-1}\sqrt{a_j}+C_2\frac{\log^2 n}{n^2}.$$

به آسانی می‌توان نشان داد که $a_n\longrightarrow 0$. بنا بر این،

$$\lim_{n\to\infty}\ell_2(\mathcal{L}(X_n),\mathcal{L}(X))=0.$$

برای جزئیات بیشتر به [۱۱] مراجعه کنید.

۱.۳ روش انقباض روی L_2

روشی که در بخش ۳ معرفی شد را می‌توان به چندین روش تعمیم داد. برای رسیدن به این هدف، یک رده از دنباله‌های بازگشتی از توزیع‌ها که می‌توانند توسط ویژگی‌های فضای L_2 تحلیل شوند، مشخص‌سازی می‌شوند. به‌منظور ارائه یک حالت کلّی‌تر به بردارهای تصادفی در $\mathbb{R}^d$ توجه می‌کنیم.

فرض کنید یک دنباله از بردارهای d-بُعدی $(Y_n)_{n\ge 1}$ در رابطهٔ بازگشتی

$$Y_n \overset{d}{=} \sum_{r=1}^{K}A_r(n)Y^{(r)}_{I_r^{(n)}}+b_n,\qquad n\ge n_0. \tag{۱۳}$$

صدق می‌کنند که در آن

$$(Y_n^{(1)}),\ldots,(Y_n^{(k)})$$

و

$$(A_1(n),\ldots,A_K(n),b_n,I^{(n)})$$

مستقل هستند، $A_{۱}(n)$، ...، $A_K(n)$ ماتریس‌های تصادفی $d \times d$، b_n یک بردار d-بُعدی تصادفی، $I^{(n)}$ برداری از صحیح‌های تصادفی $I_r^{(n)}$ متعلق به $\{۰, ۱, ۲ \dots, n\}$ و $(Y_n^{(۱)})$، ...، $(Y_n^{(K)})$ دارای توزیع (Y_n) هستند. برای $n_۰ \geq ۱$، بردارهای $Y_۰$، ...، $Y_{n_۰-۱}$، بردارهای تصادفی آغازین هستند. تعداد $K \geq ۱$ تعیینی است. بردار تصادفی (Y_n) در رابطهٔ (۱۳) توسط

$$X_n = (Y_n - M_n)C_n^{-۱/۲}, \qquad (۱۴)$$

که در آن $M_n \in \mathbb{R}^d$ و C_n یک ماتریس $d \times d$ معین-مثبت متقارن است، بازمقیاسیده می‌شود. اگر دو گشتاور اول Y_n متناهی باشند، آنگاه M_n و C_n نوعاً به‌ترتیب از مرتبه امید ریاضی و ماتریس کوواریانس Y_n هستند. از رابطهٔ بازگشتی (۱۳) برای X_n رابطهٔ بازگشتی بهبودیافته

$$X_n \overset{d}{=} \sum_{r=۱}^{K} A_r^{(n)} X_{I_r^{(n)}}^{(n)} + b^{(n)}, \qquad n \geq n_۰ \qquad (۱۵)$$

با

$$A_r^{(n)} = C_n^{-۱/۲} A_r(n) C_{I_r^{(n)}}^{۱/۲}, \qquad (۱۶)$$

$$b^n = C_n^{-۱/۲}\left(b_n - M_n + \sum_{r=۱}^{K} A_r(n) M_{I_r^{(n)}}\right) \qquad (۱۷)$$

به دست می‌آیند که استقلال‌ها مانند رابطهٔ (۱۳) است. روش انقباض به‌صورت زیر دنبال می‌شود.

فرض می‌کنیم که X_n توسط (۱۵) مشخص‌سازی می‌شود. در این صورت، همگرایی ضرایب

$$A_r^{(n)} \xrightarrow{d} A_r^*, \quad b^{(n)} \xrightarrow{d} b^*, \qquad (۱۸)$$

برای همگرایی توزیع کمّیت‌های (X_n) به یک حد X دلالت می‌کنند. توزیع حدی $\mathcal{L}(X)$، توسط یک معادله نقطه ثابت، مشخص‌سازی می‌شود که از رابطهٔ بازگشتی بهبودیافته با قرار دادن $n \to \infty$ به دست می‌آید:

$$X \overset{d}{=} \sum_{r=۱}^{K} A_r^* X^{(r)} + b^*. \qquad (۱۹)$$

در اینجا $(A_۱^*, \dots, A_K^*, b^*)$ و $X^{(۱)}$، ...، $X^{(K)}$ مستقل هستند و برای هر $r = ۱, \dots, K$،

$$X^{(r)} \overset{d}{=} X.$$

برای بازفرمول‌بندی ویژگی‌های نقطه ثابت، فضای همه اندازه‌های احتمال روی $\mathbb{R}^d$ را با $\mathcal{M}^d$ نشان می‌دهیم و قرار می‌دهیم:

$$T : \mathcal{M}^d \longrightarrow \mathcal{M}^d, \mu \longmapsto \mathcal{L}\left(\sum_{r=۱}^{K} A_r^* Z^{(r)} + b^*\right), \qquad (۲۰)$$

که در آن $(A_۱^*, \dots, A_K^*, b^*)$ و $Z^{(۱)}$، ...، $Z^{(K)}$ مستقل هستند و برای $r = ۱, \dots, K$، $\mathcal{L}(Z^{(r)}) = \mu$. در این صورت، X یک حل نقطه ثابت (۱۹) است اگر و تنها اگر توزیع $\mathcal{L}(X)$ یک نقطه ثابت نگاشت T باشد.

نگاشتی از نوع (۲۰) اغلب نقاط ثابت یکتا در فضای همه توزیع‌های احتمال ندارد و مشخص‌سازی مجموعه همه نقاط ثابت یک مسئله مهم و باز است. به‌منظور یک مطالعهٔ دقیق‌تر زیرمجموعه‌هایی از $\mathcal{M}^d$ به‌صورت زیر تعریف می‌شوند:

$$\mathcal{M}_s^d = \{\mu \in \mathcal{M}^d : \| \mu \|_s < \infty\}, \qquad s > ۰ \qquad (۲۱)$$

(۲۲)

$$\mathcal{M}s^d(M) = \{\mu \in \mathcal{M}_s^d : \mathbb{E}(\mu) = M\}, \qquad s \geq ۱$$

(۲۳)

$$\mathcal{M}s^d(M, C) = \{\mu \in \mathcal{M}_s^d(M) : cov(\mu) = C\}, \quad s \geq ۲$$

که در آن $M \in \mathbb{R}^d$ و C یک ماتریس معین مثبت و متقارن $d \times d$ است و $||\mu||_s$، $\mathbb{E}(\mu)$ و $cov(\mu)$ به‌ترتیب بیان‌کنندهٔ s-امین گشتاور قدر مطلق، امید ریاضی و ماتریس کوواریانس یک بردار تصادفی با توزیع μ هستند.

هسته این روش اعطای یک زیرفضای مناسب $\mathcal{M}^* \in \mathcal{M}^d$ روی یکی از مجموعه‌های ارائه شده در (۲۱) تا (۲۳) با یک متر کامل δ است به‌طوری‌که قید T به $\mathcal{M}^*$ یک انقباض روی فضای متریک $(\mathcal{M}^*, \delta)$ به معنای اشاره شده در قضیهٔ نقطه ثابت باناخ است. این نشان می‌دهد که یک نقطه ثابت موجود $\mathcal{L}(X)$ از T در $\mathcal{M}^*$ یکتا می‌شود.

در گام دوم هدف نشان دادن همگرایی کمّیت‌های بازمقیاسیدهٔ $\mathcal{L}(X_n)$ به $\mathcal{L}(X)$ در متر δ است. یعنی $\delta(\mathcal{L}(X_n), \mathcal{L}(X)) \to ۰$ وقتی $n \to \infty$ که روی یک همگرایی مناسب ضرایب بیان شده در (۱۸) پایه‌ریزی می‌شود. اگر δ

به‌گونه‌ای انتخاب شود که همگرایی در δ بر همگرایی ضعیف دلالت کند، آن‌گاه همگرایی در توزیع نتیجه می‌شود.

نکتهٔ حیاتی انتخاب یک متر سَره برای نشان دادن این مطلب است که نگاشت T یک انقباض است. یک انتخاب طبیعی، متر ℓ_p برای $p > 0$ است که به‌صورت

$$\ell_p(\mu, v) := \inf\{\| X - Y \|_p; \mathcal{L}(X) = \mu, \mathcal{L}(Y) = v\}$$

برای $\mu, v \in \mathcal{M}_p^d$ تعریف می‌شود که در آن

$$||X||_p = (\mathbb{E}(||X||^p))^{\min\{\frac{1}{p},1\}},$$

بیان‌کنندهٔ نرم L_p برای یک بردار تصادفی X و $||X||$ بیان‌کنندهٔ نُرم اقلیدسی آن است. متر واستراشتاین ℓ_2 یک حالت خاص و البته مهم‌ترین حالت است. به‌علاوه فضای $(\mathcal{M}_p^d, \ell_p)$ برای $p > 0$ به‌خوبی $(\mathcal{M}_p^d(M), \ell_p)$ برای $M \in \mathbb{R}^d$ و $p \geq 1$ فضای متریک کامل است و همگرایی در ℓ_p معادل همگرایی ضعیف و همچنین همگرایی گشتاور قدر مطلق pام است.

برای $\mu, v \in \mathcal{M}_p^d$ همیشه بردارهای X و Y روی یک فضای احتمال توأم با $\mathcal{L}(X) = \mu$ و $\mathcal{L}(Y) = v$ و $\ell(\mu, v) = ||X - Y||_p$ وجود دارند به‌طوری‌که این بردارها مفصل‌های بهینهٔ μ و v نامیده می‌شوند. برای این مطلب و ویژگی‌های متر ℓ_p به [۶] مراجعه کنید.

به‌منظور دست‌یابی به ویژگی‌های انقباض نگاشت (۲۰) نُرم عملگر یک ماتریس مربع A را با

$$||A||_{\mathsf{op}} = \sup_{||x||=1} ||Ax||$$

و ترانهادهٔ آن را با A^T و ثابت لیپ‌شیتس T را در صورت متناهی بودن با

$$\mathsf{lip}(T) = \inf_{\substack{\mu, v \in M^* \\ \mu \neq v}} \frac{\ell_p(T(\mu), T(v))}{\ell_p(\mu, v)}$$

نشان می‌دهیم. در حالت $p = 2$ ثابت لیپ‌شیتس به آسانی برآورد می‌شود.

لم ۴.۳. *فرض کنید $(A_1^*, \dots, A_K^*, b^*)$ یک بردار L_2-انتگرال‌پذیر از ماتریس‌های $d \times d$ تصادفی A_1^*، ... ، A_K^* و یک بردار d-بُعدی تصادفی b^* با $\mathbb{E}(b^*) = 0$ باشد و فرض کنید T همان باشد که در رابطهٔ (۲۰) معرفی شد. در این صورت، قید T به $\mathcal{M}_2^d(0)$ لیپ‌شیتس پیوسته در ℓ_2 است و برای ثابت لیپ‌شیتس $\mathsf{lip}(T)$ داریم:*

$$\mathsf{lip}(T) \leq \left\| \mathbb{E}\left(\sum_{r=1}^{K} (A_r^*)^T A_r^* \right) \right\|_{\mathsf{op}}^{\frac{1}{2}}.$$

گام بعدی روش انقباض نشان دادن همگرایی در ℓ_2 برای دنباله $(\mathcal{L}(X_n))$ ارائه شده در (۱۵) است. قضیهٔ زیر شرط‌های کافی برای این موضوع را نشان می‌دهد. برای برهان و جزئیات آن به قضیهٔ ۱.۳ در [۱۳] و قضیهٔ ۱.۴ در [۷] مراجعه کنید.

قضیهٔ ۵.۳. *فرض کنید (X_n) دنباله‌ای از بردارهای تصادفی d-بُعدی انتگرال‌پذیر در L^2 باشد که در شرط (۲۰) صدق می‌کند و در آن ماتریس‌های $d \times d$ تصادفی و L_2-انتگرال‌پذیر $A_r^{(n)}$ برای $1 \leq r \leq K$ و بردار تصادفی L_2-انتگرال‌پذیر $b^{(n)}$ و بردار مقادیر صحیح تصادفی $I^{(n)} = (I_1^{(n)}, \dots, I_K^{(n)})$ با $I_r^{(n)} \in \{0, 1, \dots, n\}$ معرفی شدند. فرض کنید که برای $\ell \geq 0$ و $r = 1, \dots, K$،*

$$(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}) \xrightarrow{L_2} (A_1^*, \dots, A_K^*, b^*), \qquad (۲۴)$$

$$\mathbb{E}\left(\sum_{r=1}^{K} \| (A_r^*)^T A_r^* \|_{\mathsf{op}} \right) < 1, \qquad (۲۵)$$

$$\mathbb{E}\left(I_{\{I_r^{(n)} \leq \ell\}} \|(A_r^{(n)})^T A_r^{(n)}\|_{\mathsf{op}} \right) \to 0, \ n \to \infty. \qquad (۲۶)$$

در این صورت، وقتی $n \to \infty$،

$$\ell_2(\mathcal{L}(X_n), \mathcal{L}(X)) \to 0,$$

که در آن $\mathcal{L}(X)$ نقطه ثابت نگاشت T تعریف شده در (۲۰) است که در $\mathcal{M}_2^d(0)$ یکتا است.

شرط (۲۴) بدین معنا است که همگرایی ضرایب معرفی شده در (۱۸) باید در L_2 برقرار باشد. برای این، مجاز به قید $(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)})$ مطابق با $(A_1^*, \dots, A_K^*, b^*)$ روی فضای احتمال توأم هستیم. یعنی (۲۴) بدین معنا است که

$$\ell_2(\mathcal{L}(A_1^{(n)}, \dots, A_K^{(n)}, b^{(n)}), \mathcal{L}(A_1^*, \dots, A_K^*, b^*)) \to 0.$$

شرط (۲۵) بنا بر نابرابری یِنسون قوی‌تر از شرط انقباض

$$\left\| \mathbb{E}\left(\sum_{r=1}^{K} (A_r^*)^T A_r^* \right) \right\|_{\mathsf{op}} < 1 \qquad (۲۷)$$

در لم ۴.۳ است. شرط (۲۶) یک شرط فنی است که معمولاً برای دست‌یابی در کاربرد آسان است. در برخی مواقع بررسی همگرایی $X_n \xrightarrow{d} X$ به‌طور مستقیم صورت می‌گیرد.

برای کاربرد قضیهٔ ۵.۳ برای دنباله‌های بازگشتی (Y_n) در (۱۳) توجه به نکات زیر ضروری است. برای مقیاس‌بندی در (۱۴) مجبور به انتخاب $M_n = \mathbb{E}(Y_n)$ به‌منظور تعیین شرط‌های $\mathbb{E}(X_n) = ۰$ و $\mathbb{E}(b^{(n)}) = ۰$ هستیم. چون $b^{(n)}$ شامل کمّیت‌های M_n است نیازمند به دست آوردن یک حد برای $b^{(n)}$ هستیم. این دلالت بر این دارد که برای کاربرد قضیهٔ ۵.۳ یک بسط مجانبی میانگین $\mathbb{E}(Y_n)$ باید مشخص شود. در مقابل ماتریس کوواریانس $cov(Y_n)$ باید در بسط مجانبی مرتبه اول آن حدس زده شود. چون همگرایی در $\ell_۲$ بر همگرایی گشتاور دوم دلالت دارد، قضیهٔ ۵.۳ به‌طور خودکار بر یک بسط مجانبی ماتریس کوواریانس $cov(Y_n)$ دلالت می‌کند.

مجدداً مثال ارائه شده توسط روسلا در بخش ۳ را در نظر بگیرید. قرار دهید:

$$
\begin{aligned}
d &= ۱,\\
K &= ۲,\\
A_۱^{(n)} &= A_۲^{(n)} = ۱,\\
I_۱^{(n)} &\sim DU\{۰, \ldots, n-۱\},\\
I_۲^{(n)} &= n - ۱ - I_۱^{(n)},\\
b_n &= n - ۱.
\end{aligned}
$$

یادآوری می‌کنیم که مجبور به‌کار در فضای $\mathcal{M}_۲(۰)$ هستیم. یعنی مجبور به داشتن اطلاع روی اوّلین گشتاور به‌منظور تقلیل (۱۳) به (۱۵) هستیم. با نرمال کردن توسط $۱/n$، داریم:

$$X_n = \frac{Y_n - \mathbb{E}(Y_n)}{n}.$$

با بحث مشابه در قسمت ۳، به رابطهٔ بازگشتی (۱۱) می‌رسیم که به‌شکل (۱۵) است. معادله حدی برای X توسط (۱۰) داده می‌شود. بنا بر این،

$$
\begin{aligned}
A_۱^* &= U,\\
A_۲^* &= ۱ - U,\\
b^* &= ۱ + ۲U \log U + ۲(۱-U)\log(۱-U).
\end{aligned}
$$

شرایط قضیهٔ ۵.۳ می‌توانند به‌طور مستقیم بررسی شوند. برای مثال، ثابت لیپ‌شیتس می‌تواند توسط

$$
\begin{aligned}
\left(\mathbb{E}\|(A_۱^*)^۲\|_{op} + \mathbb{E}\|(A_۲^*)^۲\|_{op}\right)^{۱/۲} &= \left(۲\int_۰^۱ u^۲ du\right)^{۱/۲}\\
&= \sqrt{\frac{۲}{۳}}
\end{aligned}
$$

برآورد شود.

حال، فرض کنید دو دنباله (W_n) و (Y_n) در برابری‌های توزیعی زیر صدق کنند:

$$
\begin{aligned}
W_n &\stackrel{d}{=} W_{I_n}^{(۱)} + (n + I_n)Y_{I_n} + W_{n-۱-I_n}^{(۲)} + (I_n + ۱)Y_{n-I_n-۱}\\
&\quad + ۲I^{(n)}(n - ۱ - I^{(n)}) + n - ۱,\\
Y_n &\stackrel{d}{=} Y_{I_n}^{(۱)} + Y_{n-۱-I_n}^{(۲)} + n - ۱.
\end{aligned}
$$

این بدان معنا است که

$$
\begin{aligned}
d &= ۲,\\
K &= ۲,\\
I_۱^{(n)} &\sim DU\{۰, \ldots, n-۱\},\\
I_۲^{(n)} &= n - ۱ - I_۱^{(n)},\\
A_۱^{(n)} &= \begin{pmatrix} ۱ & n - I_۱^{(n)} \\ ۰ & ۱ \end{pmatrix},\\
A_۲^{(n)} &= \begin{pmatrix} ۱ & n - I_۲^{(n)} \\ ۰ & ۱ \end{pmatrix},\\
b_n &= \begin{pmatrix} ۲I_۱^{(n)} I_۲^{(n)} + n - ۱ \\ n - ۱ \end{pmatrix}.
\end{aligned}
$$

بنا بر این،

$$
\begin{aligned}
\mathbb{E}(W_n) &= ۲n^۲ H_n - ۶n^۲ + ۸nH_n - ۱۰n + ۶H_n\\
&= ۲n^۲ \log n + (۲\gamma - ۶)n^۲ + o(n^۲).
\end{aligned}
$$

اگر به بردار نرمال شده

$$X_n = \left(\frac{W_n - \mathbb{E}(W_n)}{n^۲}, \frac{Y_n - \mathbb{E}(Y_n)}{n} \right)$$

توجه کنیم، آنگاه

$$A_۱^* = \begin{pmatrix} (۱-U)^۲ & U(۱-U) \\ ۰ & ۱-U \end{pmatrix},$$

$$A_۲^* = \begin{pmatrix} (۱-U)^۲ & U(۱-U) \\ ۰ & ۱-U \end{pmatrix},$$

$$b^* = \begin{pmatrix} ۶U(۱-U) + ۲U\log U + ۲(۱-U)\log(۱-U) \\ ۱ + ۲U\log U + ۲(۱-U)\log(۱-U) \end{pmatrix}$$

که در آن U به‌طور یکنواخت روی بازهٔ $[۰,۱]$ توزیع می‌شود. متغیر W_n شاخص وینر و Y_n طول مسیر داخلی درخت‌های جستجوی دودویی هستند (برای جزئیات [۷] را ببینید).

استفاده از $L_۲$ نسبتاً ساده است اما چندین محدودیت در استفاده از آن وجود دارد. برای مثال، اگر

$$\mathbb{E}\Big(\sum_{r=۱}^{K} (A_r^*)^۲ \Big) = ۱, \tag{۲۸}$$

آنگاه $L_۲$ را نمی‌توان به‌کار گرفت. یعنی، شرط (۲۷) و به‌ترتیب (۲۵) برقرار نیستند. به‌ویژه، کاربردهای فراوانی وجود دارد که به معادله حدی

$$X \overset{d}{=} \sum_{r=۱}^{K} A_r^* X^{(r)} \tag{۲۹}$$

با

$$\sum_{r=۱}^{K} (A_r^*)^۲ = ۱ \tag{۳۰}$$

منجر می‌شود. از ویژگی پیچش توزیع نرمال نتیجه می‌شود که متغیرهای تصادفی نرمال با قاعده $N(۰,\sigma^۲)$ حل‌های معادله حدی (۲۹) تحت قید (۳۰) هستند. بنا بر این، به نظر می‌رسد که نمی‌توان یک قضیهٔ حدی مرکزی را حتی اگر وجود داشته باشد ثابت کرد [۱۰]. چون معادله حدی، نقطه ثابت یکتا ندارد بدون شک یک متر برای این‌که T یک انقباض باشد نیز وجود ندارد. مشکل اصلی دیگر روش انقباض، ظهور معادله‌های حدی تباهیده است. برای مثال، اگر رابطهٔ بازگشتی

$$Y_n \overset{d}{=} Y_{I_n} + b_n, \quad n \geq n_۰$$

را بازمقیاسیده کنیم ممکن است به معادله حدی تباهیده‌ای به‌صورت $X \overset{d}{=} X$ برسیم. این معادله حدی هیچ اطلاعی در خصوص توزیع حدی ندارد و مفهوم روش انقباض نیازمند توسیعی برای چنین مواردی است.

۴ متر زُلُتارف

به‌منظور بی‌پاسخ نگذاشتن مسائلی از معادلات حدی که حل یکتایی ندارند مجبور به انتخاب یک زیرفضای سَره هستیم. برای مثال، اگر با $\mathcal{M}_s^d(M,C)$ به‌جای $\mathcal{M}_s^d(M)$ کار کنیم، آنگاه کوواریانس ثابت است و همچنین معادله $X \overset{d}{=} \sum_{r=۱}^{K} A_r^* X^{(r)}$ تحت قید $\sum_{r=۱}^{K} (A_r^*)^۲ = ۱$ یک حد یکتا خواهد داشت. به‌علاوه، مجبور به استفاده از متر دیگری که باعث می‌شود که نگاشت T یک انقباض روی زیرفضا داشته باشد، هستیم. این موضوع توسط نِینینگر و روشندورف [۸] بررسی شد. آنها از متر زُلُتارف ξ_s روی بردارهای تصادفی X و Y در $\mathbb{R}^d$ که به‌صورت

$$\xi_s(X,Y) = \sup_{f \in \mathcal{F}_s} |\mathbb{E}(f(X) - f(Y))|$$

تعریف می‌شود، استفاده کردند که در آن $s = m + \alpha$ با $۰ < \alpha \leq ۱$ است، $m \in N_۰$ و

$$\mathcal{F}_s = \{ f \in C^m(\mathbb{R}^d, \mathbb{R}) : \parallel f^{(m)}(x) - f^{(m)}(y) \parallel \leq \parallel x - y \parallel^\alpha \},$$

بیانگر فضای توابع mبار مشتق‌پذیر پیوسته از $\mathbb{R}^d$ به $\mathbb{R}$ است به‌طوری‌که m-امین مشتق پیوسته هولدر از مرتبهٔ α است. از نماد کوتاه

$$\xi_s(\mathcal{L}(X), \mathcal{L}(Y)) = \xi_s(X,Y)$$

استفاده می‌کنیم. داریم که $\xi_s(X,Y) < \infty$ هرگاه همه گشتاورهای آمیخته مرتبه‌های ۱، ... ، m از X و Y برابر باشند و s-امین گشتاورهای قدر مطلق X و Y متناهی باشند. به‌علاوه $(\mathcal{M}_s^d, \xi_s)$ برای $۰ < s \leq ۱$، $(\mathcal{M}_s^d(M), \xi_s)$ برای $۱ < s \leq ۲$ و $(\mathcal{M}_s^d(M,C), \xi_s)$ برای $۲ < s \leq ۳$ فضاهای متریک کامل با

$M \in \mathbb{R}^d$ و ماتریس $d \times d$ معین-مثبت و متقارن C هستند. ثابت لیپ‌شیتس نگاشت T می‌تواند توسط

$$\mathsf{lip}(T) \leq \mathbb{E}\Big(\sum_{r=۱}^{k} \parallel A_r^* \parallel_{op}^s \Big)$$

برآورد شود. برای مثال، اگر $\sum_{r=۱}^{K} (A_r^*)^۲ = ۱$، آن‌گاه

$$T : \mathcal{M}_s^۱(\circ, ۱) \to \mathcal{M}_s^۱(\circ, ۱)$$

یک انقباض برای هر $۲ < s \leq ۳$ است. یک توسیع مستقیم از قضیهٔ ۵.۳ به متر زُلُتارف با کاربردهای بسیاری در آمار و الگوریتم‌های بازگشتی و موضوعات مرتبط وجود دارد که منتهی به قضایای حدی مرکزی تحت شرایط کلّی‌تر می‌شود. حتی کار کردن در یک فضای باناخ مناسب نیز امکان‌پذیر است. اوّلین مسئله‌ای که به‌کمک متر زُلُتارف مورد بررسی قرار گرفت به شرح زیر است: فرض کنید $L_\circ = L_۱ = ۱$ و

$$L_n \stackrel{d}{=} ۱ + L_{I_n}^{(۱)} + L_{n-I_n}^{(۲)},$$

که در آن $I_n \sim \mathsf{Bin}(n,p)$ و $L_k^{(۱)} \stackrel{d}{=} L_k^{(۲)}$. همچون گذشته شرط استقلال برقرار است. ثابت شد که تحت شرایط خاصی،

$$Y_n = \frac{L_n - \mathbb{E}(L_n)}{\sqrt{n}} \to N(\circ, \sigma^۲),$$

که در آن $\sigma^۲ = \mathbb{E}(Y_n^۲)$. برای جزئیات محاسبه‌ای به [۱۰] مراجعه کنید.

۵ نتیجه‌گیری

در این مقاله، یکی از مهم‌ترین روش‌های تحلیل روابط بازگشتی تصادفی که اغلب در تحلیل الگوریتم‌ها و ساختارهای بازگشتی رخ می‌دهد، معرفی شد. معادله بازگشتی زیر را با شریط استقلال بیان‌شده در طول مقاله و شرایط آغازین دلخواه در نظر بگیرید:

$$X_n \stackrel{d}{=} X_{I_n}^{(۱)} + X_{n-I_n}^{(۲)} + T_n.$$

بخش غیروابسته به متغیر اصلیِ تحت بررسی که در معادله بازگشتی به‌صورت T_n ظاهر می‌شود تابع عوارض نامیده می‌شود. تابع عوارض یا به I_n وابسته است یا وابسته نیست. در مثال حل‌شده توسط روسلا این تابع برابر با $n - ۱$ و در مثال بخش ۴ برابر با ۱ است. اگر تابع عوارض به I_n وابسته باشد تحلیل پیچیده‌تر خواهد شد. هوانگ و نِینینگر [۵] این موضوع را مورد بررسی دقیق قرار داده‌اند. برای مثال‌های گوناگون در حوزه‌های مختلف [۵، ۸، ۱۰] را ببینید.

مراجع

[1] Bickel, P., and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, **9**, 1196-1217.

[2] Gutjahr, W. and Pflug, G. Ch. (1992). The asymptotic contour process of a binary tree is a Brownian excursion, *Stochastic Processes and their Applications*, **41(1)**, 69-89.

[3] Hofri, M. (1987). Probabilistic Analysis of Algorithms, Springer Verlag, New York.

[4] Maejima, M., and Rachev, S. (1987). An ideal metric and the rate of convergence to a self-similar process, *Annals of Probability*, **14**, 708-727.

[5] Hwang, H-K., and Neininger, R. (2002). Phase change of limit laws in the quicksort recurrence under varying toll functions, *SIAM Journal on Computing (SICOMP)*, **31(6)**, 1687-1722.

[6] Major, P. (1978). On the invariance principle for sums of independent identically distributed random variables. *Journal of Multivariate Analysis*, **8(4)**, 487-517.

[7] Neininger, R. (2001). On a multivariate contraction method for random recursive structures with applications to Quicksort. *Random Structures Algorithms*, **19(3-4)**, 498-524.

[8] Neininger, R., and Ruschendorf, L. (2004). A general limit theorem for recursive algorithms and combinatorial structures. *Annals of Applied Probability*, **14(1)**, 378-418.

[9] Rachev, S. (1991). *Probability Metrics and the Stability of Stochastic Models*, Wiley, New York.

[10] Rachev, S. and Rschendorf, L. (1995). Probability metrics and recursive algorithms, *Advances in Applied Probability*, **27**, 770-799

[11] Roesler, U. (1991). A limit theorem for "Quicksort", *RAIRO-Theoretical Informatics and Applications*, **25(1)**, 85-100.

[12] Roesler, U. (1992). A fixed point theorem for distributions, *Stochastic Processes and their Applications*, **37**, 95-214.

[13] Roesler, U. (2001). The analysis of stochastic divide and conquer algorithms, *Algorithmica*, **29(1-2)**, 238-261

[14] Roesler, U., and Rüschendorf, L. (2001). The contraction method for recursive algorithms, *Algorithmica*, **29**, 3-33.

[15] Zolotarev, V. M. (1997). *Modern Theory of Summation of Random Variables*, VSP, Utrecht.