دستاوردهای پژوهشهای اخیر در ریاضیات•

(۲)

۱۶. سیستمهای ذراتِ برهمکنش‌کننده

این مبحثِ احتمال، با پیکربندیهای ذراتی که در طول زمان به‌طور تصادفی تحول می‌یابند سروکار دارد. ذره معمولاً طبق قانون خاص خودش حرکت می‌کند، نابود می‌شود، و یا ذرات جدیدی پدید می‌آورد. این قانون فقط به‌حالت سیستم در یک همسایگی ذره بستگی دارد.

مفهوم سیستمهای ذراتِ برهمکنش‌کننده از مطالعهٔ مدل ایزینگ که قبلاً شرح آن آمد، نشأت گرفته است ولی امروز کاربردهای بسیار متنوعی از بررسی نظامهایِ زیستی گرفته تا تصویرپردازی برای پزشکی و مقاصد نظامی، پیدا کرده است. فرایند تماسی[1]، که مدلی بنیادی برای پراکنش جمعیتی از موجودات زنده است، در سال ۱۹۷۴ ارائه شد. برخلاف مدلهای مبتنی بر فرایند شاخه‌ای، این سیستم امکان می‌دهد که در هر واحد سطح فقط تعدادی محدود از عناصر وجود داشته باشند. این قید که از نظر فیزیکی موجه است، بررسی تحلیلی سیستم را خیلی مشکل می‌کند. ویژگیهای بنیادی حالت یک بعدی (رشد خطی سیستم در حالت بقا، و کاهشِ نمایی احتمال بقا در حالت زوال) در اوایل دههٔ اخیر محرز شد ولی ویژگیهای متناظر برای حالت مهم دوبعدی، تازه در همین اواخر اثبات شده است.

اکنون آشکالی از فرایند تماسی در حالتی که ذرات از انواع گوناگون هستند، برای مطالعهٔ رقابتِ گونه‌ها و نظام میزبان-انگل یا شکار-شکارگر به‌کار می‌روند. مدلهای دیگری که ارتباط نزدیکتری با نشت[2] دارند، در مطالعهٔ آتش‌سوزیهای اخیر در جنگلهای یلوستون به‌کار گرفته شده‌اند. مثال سوم، بررسی توزیع و الگوی حرکتیِ کریل[1] های اقیانوس منجمد جنوبی است. سیستم اخیر بسیار جالب توجه است زیرا الگوهایی را روی مقیاسهای مکانی و زمانی چندگانه نمایش می‌دهد. این مثالها فقط سه نمونه از مثالهای روزافزونی هستند که نشان می‌دهند سیستمهای ذرات برهمکنش‌کننده چارچوبی مناسب برای فهم مکانیسمی است که در چند پدیدهٔ اکولوژیک در کار است.

۱۷. آمار مکانی

انگیزهٔ نیرومندی که علاقهٔ پژوهشگران را به بسط نظریه و روشهای آمار مکانی[2] [فضایی] جلب می‌کند، دسته‌ای ازکاربردها از جمله سنجش از دور، برآورد منابع، کشاورزی و جنگل‌داری، اقیانوس‌نگاری، اکولوزی، و دیدبانی[3] محیط زیست است. زنجیره‌ای که این کاربردها را به‌هم می‌پیوندد، عبارت است از مشخص‌سازی و بهره‌گیری از مجاورت[4]. در زیر، برخی از زمینه‌ها و امکانات مهم برای پیشرفتهای آتی را خلاصه‌وار ذکر می‌کنیم.

در سنجش از دور ژئوفیزیکی، داده‌ها معمولاً به‌صورت شبکه‌بندی شده به‌دست می‌آیند. و اثرات جویِ ناخواسته و خطاهای مربوط به مکان و اندازه‌گیری، آنها را مخدوش می‌کند؛ دراین‌مورد اغلب از باندهای چند طول‌موجی استفاده می‌شود که باعث می‌شود داده‌ها چندمتغیری، و غالباً به‌صورت کمیتهای بزرگ، باشد. سؤالهایی که در این زمینه مطرح می‌شود، از این قبیل است: چگونه می‌توان از خطا اجتناب کرد، چگونه می‌توان اطلاعاتی

---

1. contact process 2. percolation

۱. نوعی جانور کوچک، دریایی، از سخت‌پوستان، که خوراک وال بدون دندان است.

2. spatial statistics 3. monitoring 4. proximity

را که از طول موجهای گوناگون به‌دست می‌آید با هم ترکیب کرد، چگونه می‌شود الگوهایی استخراج کرد، کار تا چه حدی خوب پیش می‌رود، داده‌ها را تا چه اندازه می‌توان افزایش داد، و فایدهٔ داده‌های اضافی چیست. ملاحظات مربوط به مجاورت، تأثیر عمیقی بر رهیافتهای آماری برای پاسخگویی به همهٔ این پرسشها خواهد گذاشت.

در روشهایِ حذف اثرات ناخواسته نمی‌توان این واقعیت را از نظر دور داشت که آن اثرات غالباً به‌طور خیلی دقیق مشخص نشده‌اند. شیوه‌های استخراج الگوهای مربوطه از ترکیب اطلاعات چندباندی باید قویاً متکی باشند به مدلهای احتمالی که به‌اندازهٔ کافی برمایه‌اند. از سوی دیگر، براوردهای دقت آماری باید تا حد امکان مستقل از مدل باشند. این الزامات، مسائل مهمی در پیش آماردانان پژوهشگر می‌نهد، که دشواریهای فنی و محاسباتی آنها به‌اندازهٔ قابل ملاحظه‌ای بیشتر از دشواریهای مسائل مربوطه در تحلیل سریهای زمانی یا فرایندهای تصادفی است.

داده‌های مکانیِ مربوط به براورد منابع و دیدبانیِ محیط زیست نوعاً به شکل شبکه‌های منظم به‌دست نمی‌آیند و نسبتاً پراکنده‌اند. اگرچه داده‌های حاصل از حفاری مغزه‌ای ممکن است تفکیک عمودی خوبی داشته باشد، ولی احتمال دارد حفاری بیشتر در مناطق «پرعیار» انجام شده باشد. خواست ما معمولاً براورد کل منابع، توزیع فراوانی منابع، و برنامهٔ معقولی برای بهره‌برداری است که مبتنی بر براوردهای موضعی از منابع، همراه با شاخصهای عدم قطعیت باشد تا معلوم شود در آینده به چه میزان کارهای استخراجی دیگر نیاز هست. مجموعه روشهای آماری اولیه که هنوز هم کاربرد گسترده‌ای دارد، و مبتنی بر مدلسازی فرایند گاوسی است، با توجه به خصلت داده‌های مربوط به منابع که بسیار خطاآمیزند، چندان مناسب نیست و پژوهش بسیار بیشتری لازم است تا روشهایی کشف شوند که با توجه به کیفیات خاص داده‌های منابع، مناسب باشند.

در دیدبانی محیط زیست، نمونه‌گیری ممکن است بیشتر در نواحیِ بسیار آلوده انجام شده باشد. انتخابی‌بودن و تصادفی‌نبودن داده‌های محیط‌زیستی، مشکلات مهمی در مدلسازی و استنباط آماری پیش می‌آورد. به‌علاوه، در هر محل دیدبانی، معمولاً یک سریِ زمانی از داده‌ها وجود دارد که تحت تأثیر تغییرات فصلی است. چیزی که معمولاً می‌خواهیم بدانیم این است که چگونه کیفیت محیط زیست تغییر می‌کند. چون داده‌های ثبت شده معمولاً نمی‌تواند مقدار تغییری [در محیط زیست] را که در میان تغییرات شدید طبیعی قابل اندازه‌گیری باشد برملا سازد، به روشهای مناسبی برای ترکیب اطلاعات گرفته شده از چندین محل دیدبانی نیاز داریم. همچنین افزایش و تجدید آرایش منابع دیدبانی مستلزم پژوهشی در مسائل طرح آماری است که محدودهٔ آن از مسائلِ ایده‌آلی شده و ساده‌ای که جوابهایی برای آنها داریم، بسیار فراتر می‌رود.

در طی دههٔ گذشته، گامهای بلندی در جهت ایجاد نظریه و روشهایی برای حل مسائل مکانی با استفاده از ابزارهای آماری برداشته شده است. پس، پژوهشِ موردنیازی که شرح آن در بالا آمد از مبنای محکمی برخوردار است. پیشرفت مهمی که اخیراً در پژوهشهای مربوط به آمار مکانی به‌دست آمده، ابداع و کاربرد مدلهای شبکه‌ای مارکوفی مکانیِ انعطاف‌پذیر برای تصویرپردازی و کاربرد تکنیکهای نیرومند از قبیل الگوریتم مترو پولیس برای اجرای این مدلهاست. از جملهٔ سایر پیشرفتها می‌توان از موارد زیر نام برد: تکنیکهای هموارسازی مکانی که با پیچیدگی موضعی الگوهای مکانی تطبیق داده می‌شود، روشهای کاهش بُعد برای داده‌های مکانی چندمتغیری مبتنی بر محکهای ساختار مکانی، ابداع مدلهای میدانیِ مکانیِ غیرگاوسی که از لحاظ ریاضی قابل تحلیل باشد برای میدانهای پارامتری پیوسته، روشهای درونیابی مکانیِ ناپارامتریِ غیرخطی برای میدانهای مکانی ناپیوسته، اثبات ارتباطات نظری بین روشهای اسپلاین، و پالایه‌های وینر، و مجموعه روشهای مقایسه برای درجه‌بندی تکنیکهای براورد مکانی و تعیین دقت آنها.

۱۸. روشهای آماری برای کیفیت و بهره‌وری

بعد از جنگ دوم جهانی، کارخانه‌هایی که بهره‌وری خود و کیفیت محصولاتشان را بالا برده‌اند، به حیات خود ادامه داده‌اند و کارشان رونق یافته است ولی کارخانه‌هایی که چنین نکرده‌اند، عملکرد خوبی نداشته‌اند و یا به‌کلی از صحنه خارج شده‌اند. مهندسان برای بالابردن کیفیت و بهره‌وری ناچارند از روشهای آماری در تحلیل فرایندهای تولید استفاده کنند.

در این زمینه، چهار مقوله از روشهای آماری بسیار زیاد به‌کار می‌روند: (۱) کنترل فرایند آماری، (۲) طرح آزمایشهای آماری، (۳) اعتمادپذیری، و (۴) نمونه‌گیری برای پذیرش. کنترل فرایند آماری مشتمل است بر روشهایی برای تعیین عملکرد یک فرایند مهندسی در طی زمان به‌خصوص برای تشخیص تغییرات؛ مهمترین روشهای این مقوله، روشهای مبتنی بر نمودارکنترل هستند و این مبحثی است که نیاز به بازنگری دارد. بیشتر روشهایی که امروز در این زمینه مطرح‌اند، چند دههٔ قبل و در دوره‌ای ابداع شده‌اند که محاسبه با دست انجام می‌شد و فرض می‌کردند داده‌ها به‌صورت نرمال توزیع شده باشند، زیرا در غیر این‌صورت محاسبات فوق‌العاده دشوار می‌شد. به‌عنوان مثال، میزان تغییرات در روشهای مبتنی بر نمودارکنترل اغلب بر اساس دامنهٔ داده‌ها اندازه‌گیری می‌شود که محاسبهٔ آن آسان است. این روشها باید از اساس مورد بازنگری قرار گیرند.

طرح آزمایش، مجموعهٔ بسیار مهمی از تکنیکهاست برای مجزا کردن عواملی که می‌توان آنها را تغییر داد تا فرایند بهتر شود؛ بنابراین، برای اینکه یک فرایند مهندسی به‌طور پیوسته بهتر شود، طرحهای آزمایش باید به‌منظور بررسی و آزمودن فرایند به‌طور پیوسته اجرا شوند. بیشتر پژوهشهایی که در این زمینه انجام می‌شود معطوف به‌فهم این موضوع است که سطح متوسط یک پاسخ چه وابستگیی به‌عاملها دارد؛ در این مورد مدلهایی به‌کار می‌روند که در آنها واریانس یا ثابت است و یا، اگر تغییر کند، به‌عنوان یک پارامتر مزاحم در نظر گرفته می‌شود. ولی در بسیاری از فرایندهای مهندسی، واریانس تغییر می‌کند و همانقدر مهم است که تغییر سطح میانگین؛ و مثلاً در طرح «نیرومند» این امر صادق است. در طرح نیرومند این موضوع پذیرفته می‌شود که کیفیت بیشتر محصولات به تعداد زیادی عامل بستگی دارد. کنترل برخی از این عوامل که «پارامترهای کنترل» نامیده می‌شوند ارزان است ولی کنترل عوامل دیگر، موسوم به «پارامترهای نوفه[1]» گران یا حتی غیرممکن است. هدف، طراحی محصولی است که عملکردش در برابر پارامترهای نوفه، غیرحساس یا نیرومند باشد.

روشهایی که در مقولهٔ «اعتمادپذیری» قرار می‌گیرند، برای مطالعهٔ نیم‌عمر مؤلفه‌ها و سیستمها و مربوط ساختن این نیم‌عمرها به‌عواملی که عملکرد را مشخص می‌کنند، به‌کار می‌روند. در این زمینه، لازم است جامعهٔ پژوهشگران

1. noise

روشهایی ابداع کند که مهندسان بوسیلهٔ آنها بتوانند به‌جای تحلیل شاخصهای «زمان تا خرابی»، شاخصهای تباهی[2] را تحلیل کنند که حاوی اطلاع بیشتری است. همچنین در این مبحث لازم است تحقیقات بیشتری در زمینهٔ مدلهای داده‌هایی که از آزمایشهای تسریع شدهٔ زمان تا خرابی به‌دست می‌آیند، انجام شود.

نمونه‌گیری برای پذیرش، عبارت است از روشهایی برای امتحان کردن محموله‌ای از یک کالا، برای اینکه معلوم شود آن محموله قابل پذیرش هست یانه. در بعضی از موارد، هر عدد از کالاهای محموله امتحان می‌شود ولی در غالب موارد، نمونه‌ای از کالاها انتخاب می‌شود و با امتحان‌کردن آن، دربارهٔ کل محموله نتیجه‌گیری می‌شود. روشهای تازه‌ای برای پرداختن به این موضوع لازم به‌نظر می‌رسد. برخی از تحقیقات گذشته نشان داده است که روشهای بیزی راه مناسبی هستند که می‌توان در پیش گرفت.

۱۹. کهادهای گراف

کهادی از یک گراف، گرافی است که از یک زیرگراف گراف اولیه با انقباض یالها به‌دست می‌آید. چند نتیجهٔ مرتبط با هم در مورد کهادهای گراف، که برخی به‌نظریهٔ «محض» گرافها تعلق دارند و برخی به طرح الگوریتمها، در زمرهٔ دستاوردهای جدید به‌شمار می‌روند. با به‌دست آمدن این نتایج، راههای تازه‌ای برای تحقیق گشوده شده و مسائل و فرصتهای تازه‌ای پدیدار گشته است.

در یک مسألهٔ قدیمی و کاملاً حل‌شده در نظریهٔ جریان شبکه‌ای، گرافی مفروض است که برخی از رأسهایش را «مبدأ» و رأسهای دیگرش را «مقصد» می‌نامند. سؤال این است که آیا مثلاً ده مسیر درگراف وجود دارد که هر یک از مبدئی شروع و به مقصدی ختم شود و این مسیرها هم را قطع نکنند؟ چگونه می‌توان برنامه‌ای برای کامپیوتر نوشت که جواب این سؤال را بدهد؟ یک راه، تهیهٔ فهرست همهٔ مسیرها و آزمودن همهٔ ترکیبات ده تا از آنهاست؛ ولی چنین راهی حتی برای یک گراف کوچک خیلی طولانی است. خوشبختانه (چون این مسأله اهمیت زیاد وکاربردهای فراوانی دارد)، الگوریتم دیگری در دست است که غیرمستقیم‌تر ولی بسیار کاراست. بنابراین، مسأله را می‌توان کاملاً حل‌شده دانست.

اگر مسأله را قدری تغییر دهیم، و فرض کنیم ده مبدأ و ده مقصد وجود دارد و بخواهیم مسیر اول از مبدأ اول تا مقصد اول، مسیر دوم از مبدأ دوم تا مقصد دوم، و ... الی آخر، باشد، آن‌وقت چه می‌شود؟ این مسأله خیلی مشکلتر است. حتی در مورد دو مسیر (دو مبدأ و دو مقصد به‌جای ده تا)، مسأله مشکل است، و مسألهٔ مربوط به سه مسیر تا همین اواخر حل نشده بود. یکی از احکام مهم به‌دست آمده در بارهٔ کهادهای گراف این است که به‌ازای هر تعداد (مثلاً ده) مسیر، الگوریتم کارایی برای حل مسألهٔ ده مسیر وجود دارد (کلمهٔ «کارا» در اینجا یک اصطلاح فنی است و به‌این معنی است که زمان اجرای الگوریتم محدود به یک چندجمله‌ای برحسب تعداد رئوس گراف است.)

دستاورد دیگر، اثباتی است از یک حدس قدیمی متعلق به واگنر، حاکی از اینکه در هر گردآیهٔ نامتناهی از گرافها، گرافی وجود دارد که شامل گراف دیگری است. (یک گراف «شامل» گرافی دیگر است اگر دومی را بتوان به‌وسیلهٔ انقباض یالها از زیرگرافی از اولی به‌دست آورد.) این موضوع در نظریهٔ «محض» گرافها حائز اهمیت است ولی اگر با الگوریتمی که قبلاً ذکر شد ترکیب شود، کاربردهای الگوریتمی نیز دارد. مثلاً فرض کنید می‌خواهیم بدانیم که آیا می‌شود یک گراف خاص را به‌طور فیزیکی (با استفاده از سیم برای ساختن یالها) ساخت به‌نحوی که هیچ مداری ازگراف «گره نخورد» یا نه. هیچ الگوریتم کارایی برای پاسخ دادن به این سؤال پیدا نشده است. ولی از حدس اثبات‌شدهٔ فوق نتیجه می‌شود که الگوریتم کارایی برای مسأله وجود دارد اگرچه کسی تا کنون آن را پیدا نکرده است.

می‌توان حکم مشابهی ارائه کرد که از کلیت بسیار برخوردار است و کاربردهای زیادی در علوم کامپیوتر نظری دارد. فرض کنید می‌خواهیم الگوریتم کارایی برای آزمودن گرافها طرح کنیم که معلوم کند یک گراف، ویژگی خاصی را دارد یا نه. در مورد برخی از ویژگیها، یافتن چنین الگوریتمی غیرممکن است؛ ولی فرض کنید گرافی که خاصیت مورد نظر را دارد، شامل هیچ گرافی بدون آن خاصیت نیست. (مثلاً، این ویژگی که گراف قابل ساختن باشد بدون اینکه گره بخورد، در این شرط صدق می‌کند.) پس الگوریتم کارایی برای آزمودن یک گراف از لحاظ داشتن آن خاصیت وجود دارد، اگرچه ممکن است تا کنون کسی آن را نیافته باشد. باید تأکید کرد که دانستن اینکه الگوریتم کارایی وجود دارد، حتی اگر کسی هنوز آن را پیدا نکرده باشد، خودبه‌خود اطلاع مهم و با معنایی است.

۲۰. اقتصاد ریاضی

گرچه بحث ریاضی دربارهٔ عملکرد بازار در قرن گذشته آغاز شد، نخستین توصیف ریاضی دقیق از مبانی اقتصادی عملکرد بازار در اواخر دههٔ ۱۹۴۰ و اوایل دههٔ ۱۹۵۰ به‌عمل آمد. این توصیف به مدل معروف تعادل عمومی (GE) تحت فرض کامل‌بودن بازارها انجامید یعنی تحت این فرض که همهٔ کالاها را می‌توان در بازارهایی که در یک زمان کالا عرضه می‌کنند و خریدار و فروشنده در این بازارها از اعتبار کامل برخوردارند، خرید و فروش کرد. در حالت تعادل، عرضه با تقاضا برابر است و هر خانوار طبق علائق خود و متناسب با بودجه‌اش عمل می‌کند.

ولی مدل بازارهای کامل، اشکال عمده‌ای دارد. برنامه‌ریزی برای پیشامدها و احتمالات آتی یک قسمت اساسی مسألهٔ تخصیص آماری است و مدل تعادل عمومی را می‌شود به‌این‌صورت تعبیر کرد که در این مدل، زمان و عدم قطعیت بوسیلهٔ طبقه‌بندی کالاها برحسب پیشامد [خرید یا فروش] و زمان منظور می‌شود. ولی تنها یک قید در مورد بودجه، این دیدگاه غیرواقعبینانه را بر این تعبیر تحمیل می‌کند که همهٔ معامله‌ها یکباره انجام می‌شوند. تحقیقات اخیر در زمینهٔ مدلهای بازار ناکامل که این اشکال را رفع می‌کند، به پیشرفت مهمی نائل آمده و مسائل و چشم‌اندازهای تازه‌ای را مطرح کرده است.

در مدل تعادل عمومی با فرض ناکامل‌بودن بازارها (GEI)، معامله‌گران می‌توانند همهٔ کالاهای ممکن را یکباره معامله کنند. برای سادگی، فرض کنید کالاهایی فاسدشدنی داشته باشیم که بتوان آنها را در زمان صفر مصرف کرد و یا تحت هر یک از S حالت ممکن، در زمان یک. به‌علاوه، سرمایهٔ معامله‌گران ممکن است در بعضی حالات زیاد و در بعضی حالات کم باشد. در زمان صفر، معامله‌گران مجازند تعداد محدودی از داراییها را که تحویل آنها به‌صورت ترکیبات مختلفی ازکالاها، بسته به حالت، میسر است معامله کنند. مثلاً ذخیرهٔ انبار یک کارخانه، یک نوع دارایی است که در حالاتی که

2. degradation

برنامه‌های تولید خوب پیش می‌رود، مقدار زیادی کالا و در غیراین‌حالات، مقدار کمی کالا تحویل می‌دهد.

به‌کمک ابزارهایی از توپولوژی دیفرانسیل می‌توان نشان داد که تعادل GEI ویژگیهای بسیارشگفت‌انگیز متعددی دارد. نخست اینکه، درست برخلاف بازارهای کامل، اگر داراییها کمتر از حالتها باشند، آنگاه تعادلهای GEI «نوعاً» ناکارا هستند. ناکارایی این تعادلها نه‌فقط به‌دلیل نبودِ برخی بازارهای دارایی است، بلکه از بازارهایی هم که وجود دارند، به‌درستی استفاده نمی‌شود.

خصیصهٔ تعجب‌آور دیگری که مدل GEI دارد، ویژگیهای خاصی است که تعادل پولی می‌تواند در این مدل داشته باشد. اگر کالایی وجود داشته باشد که هیچ اثری بر مطلوبیت نداشته باشد، و جزو مایهٔ اولیهٔ هیچ معامله‌گری نباشد، دوتا از ویژگیهای پول را دارد. اگر داراییها قابل تحویل به این پول باشند، آنگاه تحت شرایط قضیهٔ ناکارایی، معمولاً $S-1$ بعدِ متمایز برای تخصیص کالا در حالت تعادل وجود دارد. ولی اگر بازارِ دارایی «کامل» باشد، معمولاً فقط تعادلهای منفرد وجود دارد.

کنارگذاشتن فقط یک دارایی، موجب پرش عدم تعین از حالت صفر بعدی به $S-1$ بعدی می‌شود. این بعد، هرچقدر تعداد داراییها کم باشد، ثابت است. بدون استفاده از توپولوژی دیفرانسیل، دلیل قابل قبولی برای پیش‌بینیِ چنین نتیجه‌ای وجود نداشت و مسلماً اثبات آن نیز به‌صورتی قانع‌کننده میسر نبود. توجه کنید که بدون یک کالای پولی معمولاً تعدادی متناهی حالت تعادل وجود دارد. در مدل GEI پول نقش چشمگیری دارد.

این ایده‌های ریاضی کاربرد دیگری هم در مدل GEI دارند. مدتها تصور می‌شد که شاید یک مدل GEI وجود نداشته باشد که به مفهومی «نیرومند» باشد. ترفندی که برای تحلیل صحیح مسأله به‌کار می‌رفت، پذیرفتن این موضوع بود که محدودیت بودجهٔ GEI را می‌توان برحسب حیطهٔ بازده‌های مالی در تمام S حالت ـ و بنابراین برحسب یک خمینهٔ گراسمان که با استفاده از فضای حالات ساخته شود ـ بیان کرد. استدلالهایی مبتنی بر نظریهٔ درجه نشان می‌دهند که دستگاه معادلاتی که به‌وسیلهٔ تعادل GEI بر این خمینهٔ گراسمان تعریف می‌شود، نوعاً جواب دارد.

۲۱. الگوریتمها و معماریهای موازی

پیشرفتهای چشمگیر در تکنولوژی تولید کامپیوتر، تأثیر چشمگیری هم بر نحوهٔ استفادهٔ ما از کامپیوتر داشته است. مثلاً، امروز قویترین کامپیوترها در واقع مرکب از تعداد زیادی پردازندهٔ کوچک (یعنی تراشه) هستند که با هم جمع شده و یک ماشین موازی پدید آورده‌اند. در چنین ماشین موازی، پردازنده‌ها معمولاً ماشینهای ترتیبی سنتی هستند که با هم برای حل یک مسألهٔ بزرگ، کار می‌کنند. قاعدتاً انتظار داریم که با گردهم‌آوردن N پردازنده برای انجام دادن یک کار خاص، آن کار N بار سریعتر از وقتی که فقط یک پردازنده کار می‌کند، انجام شود. اگرچه از لحاظ شهودی به‌نظر می‌رسد که N پردازنده باید بتوانند یک مسأله را N بار سریعتر از یک پردازندهٔ تنها حل کنند، این امر همیشه امکان‌پذیر نیست. در واقع، ممکن است طرح الگوریتمهایی موازی که در ماشینهایِ با N پردازنده، N بار سریعتر از ماشینهای تک‌پردازنده اجرا شوند بسیار مشکل باشد.

برای اینکه ببینیم وقتی می‌خواهیم یک الگوریتم ترتیبی را به‌صورت موازی درآوریم چه مشکلاتی ممکن است پیش آید، مثالی بسیار مقدماتی می‌آوریم که عبارت است ازکار «پیش‌پاافتادهٔ» جمع کردن دو عدد N رقمی.

جمع دو این عدد در N مرحلهٔ پیاپی مسلماً آسان است ولی آیا می‌توان این‌کار را در یک مرحله به‌وسیلهٔ N پردازندهٔ موازی انجام داد؟ واقعیت این است که نمی‌توان چنین کاری کرد و حتی بدتر از این، در اولین نگاه به‌نظر می‌رسد که مسألهٔ جمع را اصلاً نمی‌توان با استفاده از N پردازنده سریعتر از یک پردازنده حل کرد. بنابراین، مثال جمع در نگاه اول خیلی ناامید کننده به‌نظر می‌آید زیرا اگر عمل جمع را نتوان به‌طور موازی انجام داد، چه امیدی به‌حل موازی مسائل دیگر می‌توان داشت؟

خوشبختانه، امروز به‌دست‌آوردن الگوریتمهای موازیِ سریع برای بسیاری از مسائل محاسباتی در علوم و مهندسی امکان‌پذیر است. مثلاً در مورد عمل سادهٔ جمع، الگوریتمی وجود دارد که در صورت استفاده از $N/\log(N)$ پردازنده، فقط $\log(N)$ مرحله دارد؛ اگرچه همان‌طور که دیدیم، وجود چنین الگوریتمی به‌هیچ‌وجه واضح و بدیهی نیست. طی پنج سال گذشته، پیشرفت چشمگیری درکشف روشهای جالب و بسیارکارا برای حل برخی مسائل مهم که ذاتاً ترتیبی به‌نظر می‌رسند، به‌دست آمده است. از جملهٔ این روشها، الگوریتمهایی برای حساب، محاسبات ماتریسی، عملیات روی چندجمله‌ایها، معادله‌های دیفرانسیل، همبندی گرافها، انقباض و محاسبهٔ درخت، جورسازی[1] گرافها، مسائل مجموعهٔ مستقل، برنامه‌ریزی خطی، هندسهٔ محاسباتی، جورسازی رشته‌ها، و برنامه‌ریزی پویا را می‌توان نام برد.

همچنین پیشرفت زیادی در زمینهٔ مسائل نهفته در طراحی ماشینهای موازی حاصل شده است. مثلاً، انواع متعددی از شبکهٔ مخابراتی برای وصل کردن پردازنده‌های یک ماشین موازی به‌یکدیگر ابداع شده و الگوریتمهای سریعی برای هدایت داده‌های صحیح به مکان صحیح در زمان صحیح به‌دست آمده است. این‌کار تا حد زیادی جنبهٔ ریاضی دارد و مبتنی است بر استفادهٔ فراوان از ترکیبیات، تحلیل احتمالاتی، و جبر. درواقع، معمولاً در کامپیوترهای موازی از شبکه‌های ارتباطی و الگوریتمهای مسیردهی که مبتنی بر ترکیبیات هستند استفاده می‌شود. پس در این زمینه هم، این دستاوردهای اساسی، امکانات زیادی برای پیشرفتهای بعدی فراهم کرده است.

۲۲. الگوریتمهای تصادفی(شده)

دانشمندان کامپیوتر طی ۱۵ سال گذشته به‌مزایایِ فراوانِ الگوریتمهایی که در حین اجرا، به‌اصطلاح «تاس می‌ریزند» [یعنی مبتنی بر انتخاب تصادفی عدد، داده، یا رویداد هستند] پی برده‌اند. برای کارهای بسیار متنوعی، از آزمایشِ اول‌بودن یک عدد گرفته تا تخصیص منابع در سیستمهای کامپیوتر پیچیده، ساده‌ترین و کاراترین الگوریتمهایی که در حال حاضر می‌شناسیم، الگوریتمهای تصادفی‌شده هستند. بنابراین، افزایش آگاهی در زمینهٔ این‌گونه الگوریتمها، مسأله‌ای است که پژوهشگران را به مبارزه می‌طلبد و زایندهٔ امکاناتی در آینده است.

تقریباً از همان آغاز عصر کامپیوتر، مولدهای اعداد تصادفی، به‌کمک تکنیکهای نمونه‌گیری پیچیده‌ای که مجموعاً روش مونت‌کارلو نامیده می‌شوند، در شبیه‌سازیِ سیستمهای پیچیده‌ای که شامل صف‌بندی و سایر پدیده‌های تصادفی هستند، و نیز در براورد انتگرالهای چند بعدی و سایر کمیتهای ریاضی به‌کار می‌رفته‌اند.

عامل مهمی که توجه متخصصان کامپیوتر را به‌کاربردهای گسترده‌تر

1. matching

«تصادفی‌سازی» جلب کرد، کشف دو الگوریتم تصادفی کارا برای آزمایش اول‌بودن یک عدد بود، که در سال ۱۹۷۵ روی داد. هر دو الگوریتم مبتنی بر نشانه‌های [شواهد] مرکب‌بودن عددند. مثال ساده‌ای از مفهوم «نشانه» از قضیه‌ای به‌دست می‌آید که منسوب به‌فرماست. این قضیه می‌گوید که اگر n یک، عدد اول باشد، به‌ازای هر عدد صحیحی چون m که مضربی از n نباشد، $m^{(n-1)} - ۱$ مضربی از n است. اگر این محاسبه برای عددی مثل m انجام شود و نتیجه‌ای که قضیهٔ فرما پیش‌بینی می‌کند به‌دست نیاید، آنگاه n مرکب است (یعنی اول نیست)؛ دراین‌صورت، m را نشانه‌ای از مرکب بودن n می‌خوانند. آزمونهای ذکر شده مبتنی بر انواعی از نشانه‌ها هستند که قدری پیچیده‌تر از این نوع‌اند. کارایی این آزمونها از قضایایی ناشی می‌شود که نشان می‌دهند اگر n مرکب باشد، آنگاه بیشتر اعداد صحیح بین ۱ و $n - ۱$، نشانهٔ این امر خواهند بود. یک جنبهٔ جالب توجه این آزمونها این است که نشانه‌ای از اول‌بودن به‌دست نمی‌دهند، ولی این نقطهٔ ضعف با تعریف نشانهٔ اول‌بودن، به‌جای مرکب‌بودن، رفع شد، یعنی با نشان‌دادن اینکه اگر n عدد اولی باشد، بیشتر اعدادی که به‌تصادف انتخاب می‌شوند، نشانه‌ای از این واقعیت خواهند بود. الگوریتمهای تصادفی متعدد دیگری هم بر اساس فراوانی نشانه‌ها وجود دارد.

معلوم شده است که تکنیکهای تصادفی در ساختن الگوریتم برای کارهایی از قبیل مرتب‌کردن، جستجو، و هندسهٔ محاسباتی نیز بسیار کارا هستند. مسألهٔ تهیهٔ فهرست تمام تلاقیهای تعدادی پاره‌خط واقع در صفحه، مثال ساده‌ای در این زمینه است. الگوریتم نقوی نسبتاً ساده‌ای وجود دارد که در آن، پاره‌خطها یکی پس از دیگری در نظرگرفته می‌شوند و تلاقیهای هر پاره‌خط جدید با همهٔ خطهای قبلی گزارش می‌شود. اگر پاره‌خطها اتفاقاً به‌ترتیبی بسیار نامناسب در نظرگرفته شوند، زمان اجرای این الگوریتم فوق‌العاده زیاد خواهد بود. ولی می‌توان نشان داد که اگر پاره‌خطها به‌ترتیبی تصادفی پردازش شوند، به‌احتمال بسیار زیاد، الگوریتم خیلی سریع خواهد بود.

به‌علاوه، تصادفی‌سازی نقش فوق‌العاده مهمی در طراحی سیستمهای کامپیوتریِ توزیع‌شده دارند، یعنی سیستمهایی که در آنها تعداد زیادی کامپیوترِ پراکنده و دور از هم که با رابطهای مناسبی به‌هم مربوط شده‌اند، به‌صورت اجزاء یک سیستم واحد با هم کار می‌کنند. در این سیستمها باید برای این مسأله که هر یک از کامپیوترها با رابطها ممکن است از کار بیفتند و یا در یک زمان، سرعتهای متفاوتی داشته باشند، تدبیری اندیشیده شود، و اطمینان حاصل شود که زمان صرف‌شده برای ارتباط بین پردازنده‌ها نامعقول نیست. روشهای تصادفی به‌خصوص در محول‌کردن کارهای محاسباتی به‌هر یک از پردازنده‌ها و انتخاب مسیرهایی برای ارتباط که داده‌ها در طول آنها جریان یابند بسیار کارا هستند. می‌توان نشان داد که در وضعیتها و حالات بسیار متنوعی، احالهٔ تصادفی کارها به‌پردازنده‌ها و تخصیص تصادفی داده‌ها به پیمانه‌های حافظه، همراه با مسیریابی تصادفی پیامها، به‌احتمال زیاد نتیجه‌ای تقریباً بهینه به‌همراه خواهد داشت.

همهٔ کاربردهایی که برای تصادفی‌سازی برشمردیم، وابسته به این فرض است که الگوریتمها، یا برنامه‌های کامپیوتری، به رشته‌ای از بیتهای تصادفی مستقل دسترسی دارند. معمولاً کامپیوتر از اعداد شبه‌تصادفی استفاده می‌کند که به‌وسیلهٔ فرایندی تکرّری و کاملاً تعیّنی از یک عدد اولیه، به‌نام بذر یا نطفه، تولید می‌شوند. این مولدها معمولاً در معرض آزمونهای آماری خاصی قرار می‌گیرند تا مسلّم شود که رشته‌هایی از اعداد که از آنها به‌دست می‌آیند، برخی از خاصیتهای دنباله‌های تصادفی را دارند هر چند که با فرایندی کاملاً تعیّنی تولید می‌شوند. در حال حاضر، پژوهش عمیقی در ویژگیهای مولدهای اعداد شبه‌تصادفی در جریان است. هدف این پژوهش، نشان‌دادن این نکته است که مادامی که بذر تصادفی باشد، محصول آن را نمی‌توان به‌وسیلهٔ هیچ آزمون محاسباتی با زمان چندجمله‌ای، از یک دنبالهٔ کاملاً تصادفی تمییز داد.

و بالاخره، بیشتر تحقیقات نظری جدید معطوف به‌رابطهٔ بین مولدهای شبه‌تصادفی و مفهوم تابع یکطرفه است که مفهوم اساسی در رمزنگاری است. تابع یکطرفه تابعی است که محاسبهٔ آن آسان ولی وارون‌کردنش مشکل است. نشان داده‌اند که هر تابع یکطرفه را می‌توان برای ساختن یک مولد اعداد شبه‌تصادفی که توجیه دقیقی داشته باشد به‌کاربرد. متأسفانه، پژوهشگران نظریهٔ پیچیدگی محاسبه هنوز حتی وجود توابع یکطرفه را معلوم نکرده‌اند. این یکی از مسائل مهم متعددی است که باقی مانده و باید حل شود.

۲۳. الگوریتم چندقطبی سریع

درزمینهٔ الگوریتمهایِ مربوط به‌مسائلی از قبیل باریکه‌های ذرات در فیزیک پلاسما، آکوستیک زیرآبی، مدلسازی مولکولی، و حتی جنبه‌های بسیار مهمی از آئرودینامیک، امکانات زیادی برای پیشرفت وجود دارد. یک نوع محاسبهٔ بنیادی که در این کاربردها اهمیت اساسی دارد، محاسبهٔ برهمکنشها در سیستمی بس‌ذره‌ای است. این برهمکنشها اغلب بلند برد هستند، بنابراین همهٔ جفتهای ذرات را باید در نظرگرفت. به‌خاطر این قید، محاسبهٔ مستقیم همهٔ برهمکنشها در سیستمی متشکل از N جسم مستلزم عملیاتی است که تعداد آنها از مرتبهٔ $N^۲$ است. این نوع محاسبه را محاسبهٔ پتانسیل N جسمی می‌نامیم.

تلاشهایی به‌منظورکاهش پیچیدگی محاسباتی مسألهٔ پتانسیل N جسمی به‌عمل آمده است. قدیمیترین روشی که در این زمینه به‌کار می‌رود، روش ذره در سلول (PIC) است که مستلزم عملیاتی است که تعداد آنها از مرتبهٔ $N \log(N)$ است. متأسفانه، این روشها نیز مستلزم شبکه‌ای است که تفکیک محدودی را به‌دست می‌دهد و وقتی توزیع ذره نایکنواخت باشد، ناکاراست. یکی از روشهای جدیدتر، روش «حل‌کننده‌های سلسله مراتبی[1]» است که روشهایی بی‌شبکه برای شبیه‌سازیهای بس‌ذره‌ای هستند و پیچیدگی محاسباتی آنها نیز از مرتبهٔ $N \log(N)$ براورد شده است.

در روش چندقطبی سریع (FMM) که اخیراً پیدا شده است، مقدار کار لازم برای محاسبهٔ همهٔ برهمکنشهای دوبه‌دو با خطایی در حدود خطای گردکردن، بدون توجه به توزیع ذره، فقط متناسب با N است. این روش همچون روش حل‌کننده‌های سلسله مراتبی، یک الگوریتم «تجزیه و حل» است که مبتنی بر ایدهٔ تجمع ذرات روی مقیاسهای طولیِ مکانیِ گوناگون است؛ و در حقیقت مبتنی است بر ترکیبی از فیزیک (انبساطهای چندقطبی)، ریاضیات (نظریهٔ تقریب)، و علم کامپیوتر، و کاربرد آن رو به فزونی است.

تکنیکهایی که در این زمینه ابداع می‌شوند، کاربردهای صنعتی بلاواسطه‌ای دارند. نتیجهٔ استفادهٔ سرراست از شبیه‌سازیهای ذرات، قابل ملاحظه است و تقریباً بی‌واسطه به‌دست می‌آید. این نوع شبیه‌سازیها در مدلسازیهای ادوات ریز موجی با باریکهٔ الکترونیِ پرتوان (مانند ژیروترونها و لیزرهای آزادالکترون)، باریکه‌های ذرات، ادوات گداخت کنترل‌شده، و نظایر اینها، به‌کار می‌رود.

1. hierarchical solvers

یک کاربرد صنعتی بلاواسطهٔ دیگر، در دینامیک مولکولی است که تکنیکی است برای مطالعهٔ خواص مایعات (و سایر حالتهای ماده) بوسیلهٔ شبیه‌سازی کامپیوتری. وقتی مواضع و سرعتهای اولیه برای تعدادی از ذرات نماینده مشخص شده باشد، مسیرهای آنها با انتگرالگیری از قانون دوم حرکت نیوتن معلوم می‌شود. در شبیه‌سازیهای اولیه، فقط سیالات غیرقطبی موردنظر بود. دراین‌مورد، میزان محاسبه در هر مرحلهٔ زمانی متناسب با تعداد ذرات N است. در سیالات قطبی وضع کاملاً فرق می‌کند؛ یک جملهٔ کولنی به‌تابع پتانسیل افزوده می‌شود و همهٔ برهمکنشهای دوبه‌دو باید به‌حساب آیند، یعنی با یک محاسبهٔ متعارف N جسمی سروکار داریم. روش FMM، شبیه‌سازیِ سیستمهای شیمیایی بسیار بزرگتری، نسبت به سیستمهای قبلی، را امکان‌پذیر می‌سازد. مثلاً بررسی یک مولکول پروتئین در آب، مستلزم درنظرگرفتن حرکت دهها هزار اتم در دهها هزار مرحلهٔ زمانی است. ادامهٔ مطالعات مشروح در زمینهٔ سینماتیک شیمیایی، احتمالاً در درازمدت فواید و دستاوردهای واقعی خواهد داشت.

۲۴. روشهای [مبتنی بر] نقطهٔ درونی برای برنامه‌ریزی خطی

برای بسیاری ازمسائل تخصیص منبع می‌توان مدلهایی از نوع مسألهٔ «برنامه‌ریزی خطی» ساخت. در این نوع مسأله، تابعی خطی را روی رئوس یک چندوجهی n بعدی بهینه می‌سازند. الگوریتم سنتی سادکی برای حل این مسأله، که بر اساس حرکت کردن روی مرز چندوجهی است، از زمان کشف آن در چهل سال پیش تاکنون ارزش و تأثیر زیادی داشته است، ولی یک اشکال نظریِ مهم دارد و آن اینکه، مدت زمان اجرای آن در حالات نامساعد ممکن است به‌صورت تابعی نمایی از تعداد متغیرها افزایش یابد. الگوریتم بسیار مطلوبتر، دست‌کم از لحاظ نظری، الگوریتمی است که زمان اجرایش «در بدترین حالت»، محدود به چندجمله‌ای باشد.

اولین نمونه از چنین الگوریتم مطلوبی، روش بیضیواری بود که در ۱۹۷۶ کشف شد. زمان اجرای این‌الگوریتم، $O(n^4L^2)$ بود که n تعداد متغیرها و L شاخصی از تعداد بیتهای لازم برای توصیف ورودی است. خاصیت مطلوب دیگری که این الگوریتم دارد، این است که می‌توان آن را در مورد «برنامه‌ریزی محدب» کلیتر بکاربرد؛ به‌علاوه، در این الگوریتم، توصیف کامل جسم محدبی که بهینه‌سازی روی آن انجام می‌گیرد لازم نیست و فقط یک زیرروال «جعبهٔ سیاه» لازم است که در مورد هر نقطهٔ مفروض، معین می‌کند که آن نقطه در چندوجهی هست یا نه، و اگر نبود، آن‌رصفحه‌ای مشخص می‌کند که نقطه را از چندوجهی جدا می‌سازد. پس روش بیضیواری، نسبت به‌روش سادکی، برای دستهٔ بسیار بزرگتری از مسائل قابل کاربرد است و وجودش به پیدایش بسیاری الگوریتمهای بازمان چندجمله‌ای برای مسائلی که قبلاً حل نشده بودند انجامیده است. ولی پژوهشگران به‌زودی دریافتند که برای برنامه‌ریزی خطی، بهترشدن کرانِ زمان اجرای الگوریتم در بدترین حالت، ربطی به بهتر اجراشدن آن در عمل ندارد.

بنابراین، به‌نظر می‌رسید که طرح الگوریتمهای برنامه‌ریزی با زمانِ چندجمله‌ای، از لحاظ نظری کار ظریفی است ولی فایدهٔ عملی ندارد. این طرز فکر در سال ۱۹۸۴ با ارائهٔ نوع جدیدی از الگوریتم برنامه‌ریزی خطی با زمان چندجمله‌ای، تغییر کرد.این نوع الگوریتم مبتنی بر استفادهٔ هوشمندانه‌ای از یک ایدهٔ قدیمی است، ایده‌ای که مدتها قبل از آن غیرعملی به‌حساب آمده و کنار گذاشته شده بود. این ایده عبارت بود از اینکه در جستجو برای یک رأس بهینه، از درون چندوجهی عبور کنیم حال آنکه در روش سادکی حرکت فقط در روی چندوجهی انجام می‌شد. دشواریِ کار در مورد چنین رهیافت «نقطهٔ درونی»، پیداکردن جهت و مسافت صحیحی است که در هر مرحله باید طی شود تا رهیافت ثمربخش باشد. راه رفع این مشکل، استفاده از تبدیلهای تصویری و یک تابع پتانسیل لگاریتمی برای هدایتِ جستجو است؛ با این راه‌حل، زمان اجرای $O(n^{3.5}L^2)$ به‌دست می‌آید. ولی موضوع اصلی، بهبود نظریِ زمان اجرای روش بیضیواری نبود. نکتهٔ مهمتر این است که به‌نظر می‌رسد این الگوریتم (و چندین صورت مختلف آن)، وقتی با تکنیکهای مناسبِ مبتنی بر ماتریسهای تُنُک اجرا شود، قادر به رقابت با روش سادکی است. علاوه بر آن، الگوریتم گفته شده در موارد بزرگ یاتبهگون، مزایای قابل ملاحظه‌ای از لحاظ زمان اجرا دارد. در واقع، این الگوریتم کاربردهای عملی مهمی در طرح شبکه‌های مخابراتی بزرگ و در حل مسائل بزرگ‌مقیاس مربوط به برنامه‌ریزی اجستیک و زمانبندی یافته است.

از زمانی که نخستین گزارشها از کاربردپذیریِ بالقوهٔ این رهیافت انتشار یافته، پژوهشهای بسیار در زمینهٔ روشهای نقطهٔ درونی انجام شده است. کندوکاو گسترده‌ای در روابط بین این الگوریتم والگوریتمهای پیشین صورت گرفته است. مثلاً این الگوریتم را می‌توان نوعی‌الگوریتم «تابع مانع[1] لگاریتمی» یا حتی کاربردی از روش نیوتن (در فضایی که به‌طرز مقتضی تبدیل شده باشد) دانست. این الگوریتم در حد، وقتی طول مرحله بینهایت کوچک باشد، میدانی برداری در درون چندوجهی پدید می‌آورد که همهٔ مسیرهای حدیش به رأس بهینه می‌رسند. در این زمینه، الگوریتم را می‌توان به‌مثابهٔ دنبال‌کردنِ تقریبیِ چنین مسیری، با برداشتن گامهای کوتاه در روی خطهای مماس، در نظرگرفت. این خود انواعی از الگوریتم را مطرح می‌سازد که در آنها، در طول خمهایی که معرف تقریبهایِ سری توانی بالامرتبه‌ترِ میدان‌برداری هستند، گام بر می‌داریم. انواع دیگر این الگوریتم، مبتنی بر طی‌کردنِ تقریبیِ مسیری خاص هستند که «مسیر مرکزی» نامیده می‌شود. انواع اخیر به‌زمانهای اجرایِ بهتر و بهتری برای بدترین حالت انجامیده است، و بهترین مورد درحال‌حاضر، با زمانِ اجرای $O(n^{2.5}L^2)$، مبتنی بر استفادهٔ هوشمندانه از پیشرفتهای اخیر در زمینهٔ ضرب ماتریسی سریع است.

الگوریتمها و بصیرتهای جدید همچنان با آهنگی سریع پدید می‌آیند وگرچه به‌نظر نمی‌رسد که این جریان به‌جوابهایی با زمان چندجمله‌ای برای مسائل NPـ تمام دشوارتر بینجامد، امید زیادی هست که رهیافتِ [مبتنی بر] نقطهٔ درونی، حوزهٔ مسائلی را که می‌شود پاسخهای مفیدی برایشان یافت، گسترش دهد.

۲۵. برنامه‌ریزی خطی تصادفی

مدلهای تعیّنی برای مسائل برنامه‌ریزی خطی و راه‌حلهای آنها، از ۴۰ سال پیش که برای نخستین بار مطرح شدند تاکنون پابه‌پای پیشرفتهای خارق‌العادهٔ کامپیوتر به‌پیش رفته‌اند. درزمان‌حاضر، حکومتها و مؤسسات صنعتی هر روز با حل مدلهایی سروکار دارند که در آنها عدم قطعیتی درکار نیست، یعنی به حل مسائلی برای طرح‌ریزی و زمانبندیِ هدفها می‌پردازند که تعیّنیِ خطی

1. barrier

و ریاضی هستند و گاه شامل هزارها متغیر همراه با یک تابع هدف خطی یا غیرخطی و هزاران قید به‌صورت نابرابری‌اند. پیش‌فرض این کار مثلاً این است که اطلاعات ما در بارهٔ انواع تکنولوژی که در آینده در دسترس خواهد بود، قطعی و دقیق است. در نتیجه، جوابهای حاصل از مدلهای تعیّنی ناقص است زیرا در این جوابها، پیشامدها و احتمالات آتی به‌درستی منظور نشده است. گرچه به‌آسانی می‌توان مدلهای ریاضی را مجدداً صورتبندی کرد تا عدم قطعیت در آنها ملحوظ شود، ولی دراین‌صورت اندازهٔ دستگاههای ریاضیی که باید حل شود در اکثر کاربردهای عملی فوق‌العاده زیاد خواهد بود. در حل مدلهای تصادفی، قسمت مشکل کار عبارت بوده (و هست) از محاسبهٔ امکانات.

بنابراین، علی‌رغم پیشرفتی که حاصل شده، هنوز دسته‌ای از مسائل «تصمیم» که اهمیت بسیار دارند، حل نشده‌اند، یعنی مسائل مربوط به یافتن جواب «بهینه» برای برنامه‌های ریاضی و خطیِ تصادفی. در اینجا «تصادفی» به‌معنی نبودِ قطعیت ویقین درمورد مسائلی ازقبیل دردسترس‌بودن منابع، رقابت خارجی، یا تأثیرات اغتشاشات سیاسی و احتمالی است. چون موضوع این‌گونه مسائل، تخصیص بهینهٔ منابع کمیاب بدون داشتن اطلاعات قطعی است، ملحوظ‌کردن عدم قطعیت در صورتبندی مسأله از اهمیت اساسی برخوردار است. اگر چنین مسائلی را بتوان در حالت کلی حل کرد، توانایی ما در طرح‌ریزی و زمانبندی برنامه‌های مربوط به شرایط پیچیده بسیار افزایش خواهد یافت.

درحال‌حاضر در ایالات متحده وکشورهای دیگر تلاش شدیدی برای حل دسته‌های خاصی از برنامه‌های تصادفیِ خطی و غیرخطی در جریان است. پیشرفتهای مهم اخیر در زمینهٔ ساخت‌افزار کامپیوتر، به‌خصوص پدیدآمدن کامپیوترهای بزرگ چندپردازنده‌ای و پردازنده‌های موازی، محرک، وانگیزه‌ای برای این تلاش بوده است. امید می‌رود که ترکیبی از سه نوع تکنیک ــ روشهای اصلاح شده برای تجزیهٔ ریاضی مسائل بزرگ‌مقیاس به‌مسائل کوچکتر، تکنیکهای مهم نمونه‌گیری مونت‌کارلو، و استفاده از پردازنده‌های موازی ــ پیشرفتهای مهمی در آیندهٔ نسبتاً نزدیک به‌بار آورد.

۲۶. کاربردهای آمار در [مطالعهٔ] ساختار DNA

هر زنجیرهٔ DNA را می‌توان به‌صورت رشته‌ای از بازها، A، C، G، T، نشان داد که اطلاعات مربوط به رشد وکارکرد یک موجود زنده را در بردارد. بنابراین، تلاش زیادی برای تعیین توالیهای DNA به‌عمل آمده است. در سال ۱۹۷۶ روشهایی برای تعیین سریع توالی پیدا شد و نتیجهٔ آن، افزایش سرسام‌آور اطلاعات کمّی ژنتیکی بود. امروز بیش از ۲۵ میلیون نوکلئوتید، متعلق به انواع بسیارگوناگونی از موجودات زنده، در قطعاتی با میانگین طول ۱۰۰۰، مشخص شده است. بهسازیِ روشهای تعیین توالی همچنان ادامه دارد، و اکتشافات وابسته به آن در زیست‌شناسیِ پایه، مبهوت‌کننده است.

دو نوع نقشه برای DNA رسم می‌شود: نقشه‌های ژنتیکی و نقشه‌های فیزیکی. هر دو نوع نقشه بااستفاده از آنزیمهای محدودکننده‌ای که توالی DNA را با الگوهای خاصی می‌برند و آن را به قطعاتی تجزیه می‌کنند، تهیه می‌شوند. طولهای این قطعه‌ها را می‌توان با قدری خطای آزمایشیِ ذاتی اندازه گرفت. در سال ۱۹۸۰ این فکر مطرح شد که تغییرات اندکی در توالی DNA ممکن است قطعاتی با طولهای متفاوت پدید آورد که می‌شود آنها را به‌عنوان «نشانگر» به‌کارگرفت و تقریباً با جایگاههای ویژه روی کروموزومهای ویژه مطابقت دارند. مقدار اطلاعاتی که متعاقباً به‌دست آمد، مسائل آماری جدیدی مطرح کرده است.

نقشه‌های فیزیکی، جایگاههای نسبی قطعاتی از DNA را که هویت مشخص و قابلیت دودمان‌سازی دارند، به‌دست می‌دهند. در دسترس‌بودن یک نقشهٔ فیزیکی، تعیین کامل توالی یک زنجیرهٔ DNA را آسانتر می‌کند. اگر نقشهٔ جایگاههای مجموعه‌ای از دودمانها[1] ـ که هر یک طولی در حدود دهها هزار نوکلئوتید دارد ـ در اختیار باشد، چند آزمایشگاه می‌توانند با همکاری هم و به‌طور همزمان به تعیین توالیِ تک‌تک دودمانها بپردازند. می‌توان انتظار داشت که تحلیل آماری پارامترهای طرح، از قبیل تعداد آنزیمهای بُرنده و اینکه چه آنزیمهایی انتخاب شوند، فوایدی در فرایند نقشه‌برداری داشته باشد. فرایند تعیین فیزیکی محل دودمانها روی ژنوم، با شناخت پارامترهای طرح و منابع تغییری که ذاتیِ فرایند است، بسیار آسان می‌شود.

به‌دست‌آوردن اطلاعات مربوط به توالی DNA فقط نخستین گام در زیست‌شناسی مولکولی جدید است. این توالی بعداً مورد بررسی گسترده قرار می‌گیرد تا ارتباط آن با اطلاعاتی که قبلاً دربارهٔ توالیهای DNA به‌دست‌آمده روشن شود. چون در جریان تکامل، ویژگیهای مهمِ زیست‌شناختی، از جمله خصوصیات توالی، محفوظ می‌ماند، این بررسیها ممکن است در فهم کارکرد بخشهای ویژه‌ای از توالی بسیار مهم باشد. اغلب استفاده از کامپیوتر برای اجرای الگوریتمهای پیچیده لازم می‌آید؛ تحلیل ریاضی الگوریتمها، هم برای تعبیر نامبهم و آموزنده‌ای از نتایج و هم برای کسب اطمینان از اینکه الگوریتمی که به‌صورت برنامه درآمده عملیاتش را با سرعت کافی انجام می‌دهد، لازم است.

توانایی ما در شناخت DNA و توالیهای پروتئین باعث پیشرفت در مطالعهٔ تکامل مولکولی شده است. اکنون نمونه‌گیری از توالی مربوط به یک ژن در جامعه‌ای مشخص، دارد امکان‌پذیر می‌شود. این امر، مسائل جدید زیادی را مطرح می‌کند. تکامل مولکولی در درازمدت و کوتاه‌مدت چگونه انجام می‌شود؟ با استفاده از مدلهای فرایند تصادفی برای تکامل مولکولی، هم ترسیم درختهای تکامل و هم تعیین آهنگهای تکامل انجام‌پذیر است. برخی از مهمترین کارهایی که در این زمینه انجام می‌شود، با تعیین ناحیه‌های کدگذاری پروتئین یا ژنهای موجود در DNA آغاز می‌گردد. تعیین جایگاه یک ژن ۶۰۰ حرفی که به‌صورت قطعاتی کوچک در طول ۱۰۰۰۰ یا ۲۰۰۰۰ حرف DNA پراکنده است، کاری طاقت‌فرسا ولی ضروری است و به تحلیل ترکیبیاتی و آماریِ پیچیده‌ای نیاز دارد.

علم زیست‌شناسی مولکولی به‌سرعت در حال پیشرفت است. میزان داده‌های مربوط به توالی DNA و پروتئین در آینده، زمینه را برای رشد این مبحث از لحاظ ریاضی مهیا می‌کند. ماهیت این علم چنان است که ارتباط نزدیک ریاضیدانان و زیست‌شناسان را ضروری می‌سازد. هر دو رشته مسلماً از چنین همکاری سود خواهند برد.

۲۷. آمار زیستی و اپیدمیولوژی

اپیدمیولوژی با توزیع بیماری در جوامع انسانی و عوامل مؤثر در آن سروکار دارد و موضوعات متنوعی را از قبیل تفاوت میزان بروز بیماری در نواحی جغرافیایی گوناگون، استانداردهای قرارگرفتن‌درمعرضِ تابش در محل کار، و

1. clones

ارزیابی تأثیر واکسیناسیون با استفاده از آزمایشهای میدانی تصادفی، در بر می‌گیرد.

در بیشتر تحقیقاتی که در اپیدمیولوژی بیماریهای مزمن انجام می‌گیرد، دو طرح مطالعاتی متمایز یعنی «گروهان[1]»ی و «مورد-شاهد[2]»ی به‌کار می‌رود. در بررسیهای گروهانی، میزانِ قرارگرفتن درمعرضِ [عاملی که مظنون به بیماری‌زایی است] و متغیرهای کمکی مربوطه برای گروه مشخصی از افراد سالم اندازه‌گیری می‌شود و افراد گروه در طی زمان تحت‌نظر قرار می‌گیرند تا معلوم شود که کدام‌یک از آنها به بیماری موردنظر مبتلا می‌شوند. یا بر اثر آن می‌میرند. روشها و مفاهیم تحلیل بقا، به‌خصوص مدل رگرسیون خطرهای متناسب، تأثیر زیادی در بررسی آماری داده‌های گروهانی داشته‌اند؛ و چارچوب ریاضی دقیقی برای فرموابندی مفاهیم اساسی اپیدمیولوژی یعنی میزان بروز و مخاطرهٔ نسبی فراهم کرده‌اند. تعبیر جدیدی از تکنیکهای قدیمی اپیدمیولوژی به زبان نظریهٔ کلاسیک آمار ارائه شده، و راه برای ابداع روشهای انعطاف‌پذیرتر و نیرومندتر در تحلیل داده‌ها هموار شده است.

در بررسیهای «مورد-شاهد»ی با نمونه‌هایی ازافراد بیمار و تندرست سروکار داریم که سابقهٔ درمعرض‌قرارگرفتنِ آنها معلوم است. ثابت‌شدن این موضوع که نسبتِ مربوط به درمعرض‌قرارگرفتگی که از روی داده‌های بیمار-شاهدی محاسبه می‌شود، تقریباً برابر با نسبت میزانهای بروز بیماری بین درمعرض‌قرارگرفتگان و قرارنگرفتگان است، اهمیت بسیار زیادی در تثبیت اعتبار عملی این طرح داشته است. اخیراً متخصصان آمارزیستی و اقتصادسنجی، مستقل از هم، روشهایی ابداع کرده‌اند برای تحلیل داده‌های بیمار - شاهدی و سایر داده‌ها در حالتی که نمونه بر اساس پیامدی که بیش از همه موردنظر است، انتخاب شود. لازم است این روشها گسترش بیشتری یابند تا در طرحهای طبقه‌بندی‌شدهٔ کلیتری به‌کار آیند و نیز در مواردی که فقط اطلاعاتی جزئی در مورد تعداد زیادی از آزمودنیهای نمونه‌گیری‌شده داریم، مفید واقع شوند.

همچنین تحقیقات بیشتری در زمینهٔ روشهای تحلیل داده‌های اپیدمیولوژیک در حالتی که برآمدها وابسته‌اند، لازم است؛ چنین حالتی مثلاً در مطالعات «آینده‌نگر» که اندازه‌گیریهای مکرری روی یک شخص واحد در طی زمان انجام می‌شود و یا در مطالعات ژنتیکیِ الگوهای بیماری در خانواده‌ها، پیش می‌آید. روشهای بهتری نیز برای تعیین مقدار خطاهای اندازه‌گیری و برای تصحیح اثر این خطاها در ضعیف جلوه‌دادن ارتباطِ بین «درمعرض‌قرارگرفتن» و بیماری، لازم است. پیشرفتهای جدید در محاسبات آماری و نظریهٔ آماریِ تحلیلِ شبیه‌درستنمایی بر اساس معادلات براوردکنندهٔ تعمیم‌یافته، نویدبخش پیشرفتهای سریع در این زمینه است.

و بالاخره، ایدز که مهمترین بیماریِ همه‌گیر عصر ماست، مسائل مبرمی در برابر متخصصان آمارزیستی و اپیدمیولوژی قرار داده است. یکی از این مسأله‌ها، براوردکردن میزان آلودگی به ویروس ایدز در گذشته، حال، و آینده، به‌منظور تعیین تعداد مبتلایان به ایدز در آینده است. مسألهٔ دیگر، به‌دست‌آوردن اطلاعات بیشتر در بارهٔ الگوهای انتقال ویروس ایدز در داخل گروههایی از افراد مستعد به ابتلا و نیز در بین این گروههاست تا برنامه‌ریزی برای سازماندهی بهینهٔ منابع جامعه به‌منظور جلوگیری از آلودگی بیشتر امکان‌پذیر شود. جمع‌آوری داده‌های مربوط به ایدز به‌ندرت سرراست و قاعده‌مند است؛ در اغلب موارد بعضی از اطلاعات اساسی دردست نیست و بنابراین، نتایج بررسیها به‌ندرت تن به تحلیلِ سرراست می‌دهند. از دیدگاه ریاضی، مسألهٔ براوردکردن میزانهای آلودگی به ویروس ایدز با استفاده از داده‌های مربوط به‌میزانهای بروز ایدز و توزیع مدت زمانِ از آلودگی تا تشخیص می‌تواند نوع نامتعارفی از مسألهٔ بازپیچش[1] به‌حساب آید. این مسأله با مسائلی ارتباط دارد که در مبحث تصویرپردازی مطرح می‌شوند، مبحثی که پیشرفت سریعی در آن صورت گرفته است. ولی فقدان یا کمبود داده‌های کلیدی باعث می‌شود که این مسأله فوق‌العاده دشوارتر باشد.

* * * * * *

• این مقاله، بخش دوم از ترجمهٔ یکی از پیوستهای گزارش معروف به گزارش دیوید II است که بخش اول آن را در شمارهٔ گذشته (شمارهٔ ۱ و ۲، سال ۵) خواندید. در اینجا از متخصصانی در زمینه‌های مختلف، به‌خصوص خانم دکتر الهی و آقایان دکتر صادقی، دکتر عمیدی، مهندس کیانی، و دکتر مرندی، که برای ترجمهٔ برخی از عبارتها و اصطلاحات این مقاله با آنها مشورت شده است، سپاسگزاری می‌شود.

1. cohort 2. case - control

1. deconvolution