فلسفهٔ آمار•

(۲)

دنیس لیندلی*

ترجمهٔ حمید پزشک

۱۰. تحلیل داده‌ها

کار آماری، خواه در مکتب فراوانی‌گرا باشد یا بیزی، غالباً ربطی به یک سیستم ریاضی ندارد بلکه در سطحی انجام می‌شود که پیچیدگی کمتری دارد. هنگامی که آماردان با مجموعه‌ای از داده‌های جدید مواجه می‌شود با آنها به نوعی «بازی» می‌کند، بازی‌ای که به آن تحلیل (اکتشافی) داده‌ها می‌گویند. محاسباتی مقدماتی انجام می‌شود و نمودارهای ساده‌ای رسم می‌شوند. برای این «بازی»، ایده‌ها و ابزارهای ارزشمند متعددی مانند بافت نگار و نمودارهای جعبه‌ای عرضه شده‌اند. تحلیل داده‌ها کاری اساسی، مهم و ارزشمند است که با فلسفهٔ آمار کاملاً سازگار است. در اینجا این نظر پذیرفته می‌شود که تحلیل داده‌ها به فرمولبندی الگو کمک می‌کند و فعالیتی مفید قبل از محاسباتِ احتمالاتی لازم برای استنباط است. استدلال ارائه شده در این مقاله نیاز به احتمال را نشان داده است. تحلیل داده‌ها گوشت و پوست برای این اسکلت ریاضی فراهم می‌کند. تنها نکته‌های تازه‌ای که ما به تحلیل مرسوم داده‌ها اضافه می‌کنیم این است که نتایج نهایی آن باید به زبان احتمال بیان شود و علاوه بر داده‌ها باید دربرگیرندهٔ پارامترها نیز باشد. چنانکه در بخش قبل اشاره شد، نتایج تحلیل داده‌ها باید منطقاً سازگار باشند.

مفهوم اساسی در پسِ اندازه‌گیری عدم‌حتمیت، مقایسه با استاندارد است. این‌گونه مقایسه‌ها غالباً دشوار است و باید جانشینی برای آنها یافت. ما برای اندازه‌گیری طول از استاندارد آن یعنی نور کریپتون استفاده نمی‌کنیم بلکه روشهای دیگری را به‌کار می‌گیریم. تحلیل داده‌ها و مفهوم سازگاری منطقی یکی از این جانشینهاست. فرض کنید می‌خواهید یک احتمال تنها را محاسبه کنید؛ آنگاه تنها رهنمودی که باید از آن پیروی کنید این است که این عدد باید بین ۰ و ۱ باشد. اما هنگامی که نیاز دارید احتمال چندین پیشامد یا کمیت وابسته به‌هم را محاسبه کنید، آنگاه نظریهٔ غنی حساب احتمالات در ارزیابیهای شما به کمکتان می‌آید. در مثالی که در پایان بخش ۹ آورده شد، شما ممکن است هر چهار مقدار داده شده در آنجا را به‌دست آورید ولی در نظر گرفتن سازگاری منطقی شما را مجبور می‌کند که حداقل یکی از آنها را اصلاح کنید. سازگاری مانند علم هندسه در اندازه‌گیری فواصل عمل می‌کند؛ یعنی اندازه‌گیریهای مختلف را مجبور به پیروی از یک نظام می‌نماید. این را در تعویض $p(y|x)$ با $p(x|\theta,\alpha)$ و $p(\theta,\alpha)$ دیده‌ایم. حال این موضوع و رابطه‌اش را با تحلیل داده‌ها بررسی می‌کنیم. ابتدا چگالی داده‌ها، $p(x|\theta,\alpha)$، را در نظر می‌گیریم.

در اینجا یک ابزار مفید و آشنا، بافت نگار و نمودارهای جدید مانند نمودار ساقه و برگ است. این ابزارها کمک می‌کنند که مشخص کنیم آیا چگالی نرمال برای توجیه داده‌ها مناسب است و یا به خانوادهٔ غنی‌تری از چگالیها نیاز است. اگر داده‌ها مرکب از دو یا بیش از دو کمیت باشند مثلاً $x=(w,z)$، آنگاه ترسیم نمودار z برحسب w کمک می‌کند که رگرسیون z را روی w و در نتیجه $p(z|w,\theta,\alpha)$ را بررسی کنیم. این ابزارها شامل مفهوم مشاهدات تکراری برای، مثلاً، ساختن بافت‌نگارند. ما به این نکته در بخش ۱۴ که به بحث راجع به تبادل‌پذیری می‌پردازیم، دوباره اشاره خواهیم کرد.

در اینجا موضوعاتی وجود دارد که گاه به آنها توجه نشده است. شما گزاره‌ای احتمالاتی مانند $p(x|\theta,\alpha)$ راجع به یک کمیت x بیان می‌کنید که با توجه به داده‌های در دسترس شما، برای شما حتمیت دارد. به‌علاوه، شما این کار را با تفکر واقعی در قالب تحلیل داده‌ها راجع به x انجام می‌دهید. عجیب است که عدم‌حتمیت (احتمال) را برای تنها کمیّت حتمی موجود به‌کار می‌گیرید. از آن گذشته، فرض کنید که $t(x)$ جنبه‌هایی از داده‌هایی را که شما در نظر گرفته‌اید، مثلاً بافت‌نگار یا رگرسیون را، توضیح دهد. آنگاه نتیجهٔ تحلیل داده‌ها واقعاً عبارت است از $p\{x|\theta,\alpha,t(x)\}$؛ یعنی احتمال را مشروط به $t(x)$ می‌کنید. به‌عنوان مثال شما ممکن است بگویید x نرمال است با میانگین θ و واریانس α، اما این کار فقط هنگامی میسر است که $t(x)$ را

دیده باشید و به‌عبارت دیگر تحلیل داده‌ها را انجام داده باشید. این ممکن است موجب دقت غیرواقعی در محاسبات بعدی شود. یک راه جلوگیری از این وضع این است که الگو را بدون نگاه کردن به داده‌ها بسازیم. در حقیقت این کار برای طرح آزمایش لازم است (بخش ۱۶). ساخت الگو تنها با مشاوره با مشتری امکان‌پذیر است و این کار موجب به‌وجود آمدن الگوهای بزرگ‌تری نسبت به الگوهای مورد استفادهٔ جاری خواهد شد. شاید تحلیل داده‌ها را بتوان به‌عنوان استنباطی تقریبی در نظر گرفت که در آن جنبه‌هایی از مدل بزرگ‌تر که در مدل عملیاتی کوچک مورد نیاز نیستند کنار گذاشته می‌شود.

نکتهٔ دیگر این است که $p(x|\theta,\alpha)$، مثلاً در قالب بافت‌نگار، فقط برای یک مقدار از (θ,α) یعنی مقدار غیرحتمی موجود نمایش داده می‌شود. داده‌ها شامل اطلاعات و شواهدی اندک هستند که حتی اگر $x \sim N(\theta_0, \alpha_0)$، آنگاه در وضعیتهای مشاهده نشده، $x \sim N(\theta,\alpha)$. بنابراین یکی از حالات در ساختن الگوها این است که الگویی بسازید که تا آنجا که توان رایانهٔ شما اجازه می‌دهد بزرگ باشد تا برای حالت غیرنرمال و زمانی که پارامترها می‌توانند هر مقداری را اختیار کنند نیز مناسب باشد. دربارهٔ اندازهٔ الگو در بخش ۱۱ بحث خواهد شد. توجه داشته باشید که مشکلات مطرح‌شده در دو پاراگراف قبل گریبان هر دو گروه فراوانی‌گرا و بیزی را به یک اندازه می‌گیرد.

هنگامی که چگالی برای پارامتر در نظر گرفته می‌شود مسألهٔ ارزیابی کاملاً متفاوت خواهد بود زیرا غالباً تکراری وجود ندارد و ابزارهای آشنای تحلیل داده‌ها دیگر در دسترس نیست. به‌علاوه، در بررسی چگالی داده‌ها الگوهای استاندارد متنوعی در دسترس است، مانند خانوادهٔ نمایی و روشهایی که بر پایهٔ GLIM (الگوی خطی کلی) ساخته می‌شوند. این الگوها در اصل به این علت طراحی شده‌اند که به خاطر دارا بودن ویژگیهای خاصی مانند آماره‌های بسنده با بُعد کوچک مشخص، تحلیل آنها آسان است هر چند دچار این مشکل‌اند که نمی‌تواند داده‌های دورافتاده را به‌راحتی در خود جا دهند. بخشی از این محدودیتها ناشی از محدودیت توان رایانه‌ها بوده است ولی مشکل مهم‌تر این است که در چارچوب رهیافت فراوانی‌گرا اصول فراگیر وجود ندارند و هر الگوی جدید ممکن است نیاز به معرفی ایده‌های جدید داشته باشد. روشهای محاسباتی مدرن مشکل اول را کاهش می‌دهند و روشهای بیزی با استفادهٔ گسترده از حساب احتمال مشکل دوم را به‌کلی از بین می‌برند؛ هدفْ همواره محاسبهٔ $p(\theta|x)$ است. ما در بخش ۱۵ دوباره به این نکته برمی‌گردیم.

تعداد کمی الگوی استاندارد برای چگالی پارامتر وجود دارد که اساساً محدود به چگالیهایی هستند که مزدوج عضو انتخاب‌شده از خانوادهٔ نمایی برای چگالی داده‌ها می‌باشند. شاه‌بیت انتقاد فراوانی‌گراها این است که: «چگالی پیشین را از کجا آوردی؟» این یک طعنهٔ احمقانه نیست؛ مشکلات جدی درکارند ولی بخشی از آنها به‌خاطر عدم‌موفقیت در ارتباط بین نظریه و عمل است. من به کرّات به سؤالات احمقانه‌ای مانند «چگالی پیشین مناسب برای واریانس σ^2 در چگالی نرمال (داده‌ها) چیست؟» برخورد کرده‌ام. این سؤال به این جهت احمقانه است که σ فقط یک حرف از الفبای یونانی است. برای یافتن چگالی پارامتر جامعه باید از الفبا فراتر رفته و واقعیت نهفته در پس σ^2 را بررسی کرد. منظور از واریانس چیست؟ به نظر مشتری این پارامتر چه مقادیری را می‌گیرد؟ به‌یاد آورید که وظیفهٔ آماردان بیان عدم‌حتمیت شما، یعنی مشتری، به زبان احتمالاتی است. شکل معقولی این سؤال می‌تواند این باشد که «عقیدهٔ شما راجع به تغییرات فشار خون در یک مرد میانسال سالم در انگلستان چیست؟» اما حتی در حالتی که به مسائل عملی نظر داریم، احمقانه خواهد بود که مشکلات کاملاً واقعی را که تا حد زیادی حل‌نشده باقی مانده‌اند انکار کنیم. این موضوع مخصوصاً هنگامی صادق است که فضای پارامتری ابعاد زیادی داشته باشد (که اغلب دارد). ما روشهایی برای درک چگالیهای چندمتغیره (چه در مورد داده‌ها و چه در مورد پارامترها) در دست نداریم. فیزیکدانان قوانین نیوتن را به این جهت که بعضی از ایده‌های معرفی‌شده توسط او قابل اندازه‌گیری نبود، منکر نمی‌شدند. بلکه می‌گفتند که این قوانین معنی دارند و هر جا که بتوانیم اندازه‌گیری کنیم به‌کار می‌آیند، پس باید روشهای بهتری برای اندازه‌گیری ابداع کنیم. مشابه این مطلب دربارهٔ احتمال صادق است. محدوده‌ای فراموش‌شده از تحقیقات آماری عبارت است از بیان ایدهٔ چند متغیره بودن برحسب احتمال، که در آن فرض استقلال جهت آسان کردن مطلب آن‌چنان به‌کار می‌رود که غالباً حقیقت زیربا نهاده می‌شود. غالباً فراموش می‌شود که مفهوم استقلال نیز به‌خاطر آنکه با احتمال سروکار دارد مفهومی شرطی است. اینکه بگوییم $(x_1, x_2, \ldots, x_n)$ ــ که تشکیل نمونه‌ای تصادفی می‌دهند ــ مستقل از هم‌اند، وقتی آنها را برای استنباط راجع به x_{n+1} به‌کار می‌گیریم حرف مضحکی است. آنها به شرط θ از هم مستقل‌اند.

گاهی اوقات این بحث مطرح می‌شود که تحلیل داده‌ها هیچ کمکی به ارزیابی یک توزیع برای پارامتر نمی‌کند زیرا این‌کار نیاز به مشاهدهٔ داده‌ها دارد، در صورتی‌که ما به یک توزیع پیشین قبل از مشاهدهٔ داده‌ها نیاز داریم. این حرف مغایر با این واقعیت است که همهٔ ما از داده‌ها استفاده می‌کنیم تا ببینیم بر چه چیزی دلالت دارند و سپس بررسی می‌کنیم که نظرمان نسبت به آن چیز بدون حضور داده‌ها چگونه باشد. شما دنباله‌ای از صفر و یک‌ها را می‌بینید و به تعداد اندکی گردش طولانی توجه می‌کنید. آیا برخلاف آنچه شما پیش‌بینی می‌کردید می‌توان نتیجه گرفت که دنباله به‌جای آنکه تبادل‌پذیر باشد مارکوفی است؟ شما به دلایل وابستگی می‌اندیشید و همین‌که نتیجه گرفتید دنباله تشکیل زنجیر مارکوف می‌دهد، به ارزش آن فکر می‌کنید. هنگامی که مرتبه عددی اول است، اگر شما فقط یک‌ها را مشاهده کرده باشید دیگر دنبال دلیل نمی‌گردید و فقط چیزی عجیبی را که رخ داده است می‌پذیرید.

۱۱. باز هم الگوها

الگو [مدل] در واقع توصیف احتمالاتیِ موقعیت مشتری است که تحلیل داده‌ها و به‌کارگیری دانسته‌های فعلی مشتری به ارزیابی آن کمک می‌کند.

چند مشکل باقی می‌ماند که یکی از آنها اندازهٔ الگو است. آیا لازم است به غیر از x کمیتهای دیگری را به‌عنوان متغیر کمکی در نظر بگیریم؟ آیا لازم است تعداد پارامترها را افزایش دهیم تا الگو انعطاف‌پذیری بیشتری داشته باشد، مثلاً به‌جای یک توزیع نرمال از توزیع tای استودنت استفاده کنیم؟ سه‌ویج هنگامی که گفت الگو باید به اندازهٔ یک فیل باشد، پیشنهاد عاقلانه‌ای کرد. به‌راستی آماردان بیزیِ کمال‌طلب الگویی دارد که همه چیز را در برمی‌گیرد، الگویی که دیدگاه جهانی نامیده شده است. چنین الگویی غیرعملی است و شما باید به دنیای کوچک‌تری که علائق نزدیک شما را در برمی‌گیرد، قانع باشید. اما این دنیای انتخابی باید تا چه حدی کوچک باشد؟ دنیاهای واقعاً کوچک از امتیاز سادگی و امکان به‌دست آوردن نتایج بسیار برخوردارند. اما آنها این عیب را هم دارند که ممکن است شناخت شما را از واقعیت دربرنداشته باشند، به‌طوری که $p(y|x)$ای که بر اساس آنها حساب می‌شود، ممکن است دارای امتیاز منفی زیادی باشد. پس مصالحه‌ای لازم است. اما همیشه باید

بزرگ‌ترین الگویی را که امکانات محاسباتی شما از عهده آن برمی‌آید انتخاب کنید. یک راهبرد موفقیت‌آمیز، این است که الگوی بزرگی انتخاب کنید و از طریق مطالعات مربوط به استواری، معین کنید که کدام جنبه‌های الگو روی نتیجهٔ پایانی شما تأثیر جدی می‌گذارد، یعنی جنبه‌هایی که نمی‌توان آنها را نادیده گرفت؛ و در نتیجه می‌توانید اندازهٔ الگو را کاهش دهید.

روابط بین دنیای کوچک انتخابی و دنیاهای بزرگی که این دنیای کوچک را در برمی‌گیرند، قابل تأمل است. یک کار رایج در انگلستان، چاپ جداول عملکرد مدارس صرفاً با استفاده از نتایج امتحانات دانش‌آموزان است. خیلیها استدلال می‌کنند که این، دنیای فوق‌العاده کوچکی است و عوامل دیگری مانند عملکرد دانش‌آموز در موقع پذیرش به مدرسه را نیز باید در نظر گرفت. بعضی مواقع گفته می‌شود که در رهیافت ما، یک دنیای کوچک‌تر نمی‌تواند در دنیایی بزرگ‌تر جای بگیرد و اگر اولی ناکافی باشد لازم است دوباره از اول شروع کرد. اما این‌طور نیست، این درک اشتباه ناشی از عدم درک ماهیت شرطی احتمال است. برای نمونه، فرض کنید که الگوی شما $x \sim N(\theta, \alpha)$ است. پس در واقع شما $p(x|\theta, \alpha, N, K)$ را توصیف می‌کنید که N دلالت بر نرمال بودن دارد. به عبارت دیگر، با دانستن K و فرض نرمال بودن، x دارای میانگین θ و واریانس α است. اگر با توجه به حضور مشاهدات دورافتاده، گسترش الگو به توزیع tای استودنت مطرح شود یعنی الگو به $x \sim t(\theta, \alpha, \nu)$ با شاخص ν تبدیل شود، آنگاه دو الگو با هم سازگارند، با این تفاوت که اولی این محدودیت یا شرط را دارد که $\nu = \infty$. شما با در نظر گرفتن t، مقادیر بزرگی برای ν در نظر گرفته‌اید. معمولاً در عبور از یک الگوی کوچک‌تر به الگویی بزرگ‌تر، الگوی کوچک‌تر تحت شرایطی شبیه الگوی بزرگ‌تر و محاط در آن است، ولی گاه چنین نیست. برای نمونه شاید در الگویی x نرمال باشد و در الگوی دیگری $\log(x)$ نرمال باشد. یک راه برای برطرف کردن این مشکل، همان‌طوری که باکس و کاکس پیشنهاد کردند[۹]، معرفی یک رشته تبدیلات است که x و $\log(x)$ دو عضو آن باشند. اگر این کار امکان نداشته باشد و شما واقعاً مطمئن نباشید که الگوی M_1 حکمفرماست یا M_2، آنگاه عدم‌حتمیت خود را با احتمال بیان کنید و الگویی تعریف کنید که در آن M_1 دارای احتمال γ و M_2 دارای احتمال $\gamma - 1$ باشد. قسمتی از مسألهٔ استنباط، گذار از γ به $p(M_1|x)$ خواهد بود. این مسأله مورد بحث قرار گرفته[۲۱] و در آن از فرضهای غلط به بهترین وجه پرهیز شده و قابلیت اختلاط مفروض گرفته شده است.

انتقاداتی از الگوهای بزرگ مطرح شده است چون به نظر می‌رسد گهگاه، در مقایسه با الگوهای کوچک‌تر، نتایج بدتری به‌دست می‌دهند. برای نمونه با در نظر گرفتن رگرسیون یک کمیت روی چندین کمیت دیگر مجبورید متغیرهای رگرسور (برگشتی) زیادی را در نظر نگیرید، چون این کار منجر به برازش اشتباه می‌گردد. این جنبهٔ نامطلوب در هنگام استفاده از روشهای فراوانی‌گرا پیش می‌آید. قضیه‌ای در چارچوب بیزی، نشان می‌دهد که این پدیده زمانی که با یک تحلیل منسجم همراه است پیش نمی‌آید، اساساً به این دلیل که [حاصل] بیشینه‌سازی روی یک زیرمجموعه نمی‌تواند از بیشینه‌سازی روی کل مجموعه بیشتر شود. این موضوع با قابلیت اختلاط مربوط است زیرا روش عادی برازش یعنی کمترین مربعات، معادل یک استدلال بیزی است که در آن از یک توزیع پیشین نامناسب، یعنی یک توزیع یکنواخت، بر روی فضاهای پارامترهای رگرسیون استفاده شود. این موضوع وقتی که بعد فضا کم است مشکلی به‌وجود نمی‌آورد. اما با افزایش ابعاد فضا، مشکلات زیاد می‌شوند[۲۶] و بنابراین مسألهٔ برازش اشتباه می‌آید.

آماردانان طی سالیان متمادی مجموعه‌ای از الگوهای استاندارد را به‌وجود آورده‌اند؛ بعضی از این الگوها به‌قدری ساده هستند که برای اجرای آنها نرم‌افزارهای رایانه‌ای نیز وجود دارد. این الگوها هنگامی که شکل فراوانی‌گرایانهٔ آنها جرح و تعدیل شود تا تحلیل منسجمی امکان‌پذیر گردد، بی‌شک ارزشمند هستند، اما نباید هیچ‌گاه جای الگوسازی دقیق بر اساس واقعیتهای غیرفرضی را بگیرند. ما در اینجا پیشنهاد مهمی را تکرار می‌کنیم، و آن این است که برای ساختن الگو فکر کن و بقیهٔ کار را به عهده حساب احتمال بگذار. یک مثال، پدیدهٔ نامطلوب «بی‌پاسخی» در بررسیهای آماری است. در اینجا مهم آن است که راجع به سازوکارهایی که به بی‌پاسخی می‌انجامند و به الگو در آوردنِ آنها فکر کنیم. برخی الگوهای موجود در متون آماری بر اساس هیچ نوع درک واقعی از دلایل نقص داده‌ها، بنا نشده‌اند. بنابراین باید به آنها مشکوک بود. دانسته‌های مشتری دربارهٔ واقعیت باید برحسب احتمال الگوسازی شوند.

این پیشنهاد بارها مطرح شده که شایستگی یک الگو بدون مشخص کردن الگوهای رقیب آزمایش شود و روشهایی هم برای انجام این کار به‌وجود آمده‌اند[۸]. ولی ما استدلال می‌کنیم که رد کردن یک الگو امر مطلقی نیست و بد بودن آن فقط در مقایسه با الگویی که از آن بهتر است معنا دارد. دلیلش در ذات احتمال نهفته است، که اساساً یک شاخص مقایسه‌ای است. دنیای بیزی یک دنیای مقایسه‌ای و نسبی است که مطلق در آن وجود ندارد. به این نکته بار دیگر در تحلیل تصمیم در بخش ۱۵ برخواهیم گشت؛ در آنجا تصمیم می‌گیرید کاری را انجام دهید، نه به خاطر آنکه خوب است بلکه به خاطر آنکه بهتر از هر کار دیگری است که به ذهن شما خطور می‌کند. مردمی که در انتخابات از رأی دادن اجتناب می‌کنند چون شرایط هیچ‌کدام از کاندیداها رامطابق خواسته‌های خود نمی‌بینند، به این نکته توجه نمی‌کنند که با وجود محدودیت در کاندیداها، باید آن را که اصلح می‌دانند برگزینند ولو مطلوب نباشد.

۱۲. بهینگی

در بحث خود به این موضع رسیدیم که عدم‌حتمیتهای واقعی باید توسط احتمالات بیان شوند، در الگوی شما گنجانده شوند و آنگاه عملیات لازم طبق قوانین حساب احتمال روی آنها انجام شود. حال به پیامدهایی که عملیات در چارچوب آن حساب روی روشهای آماری دارد (به‌ویژه در تقابل با شیوه‌های فراوانی‌گرا) می‌پردازیم و با این کار بحث در مورد آزمون معنی‌دار بودن و بازهٔ اطمینان بخش ۶ را گسترش می‌دهیم. افرادی که از برآوردها یا آزمونهای بیزی استفاده می‌کنند، گاه اظهار می‌دارند که تنها کاری که در رهیافت بیزی انجام می‌شود افزودن یک توزیع پیشین به الگوی فراوانی‌گرایانه است، یعنی توزیع پیشین فقط به‌عنوان ابزاری برای ساختن یک فرایند، که بعداً در چارچوب فراوانی‌گرا روی آن تحقیق می‌شود معرفی می‌گردد. با این گفته اهمیت توزیع پیشین را که باعث کشف آن فرایند شده، نادیده می‌گیرند. اما این گفته درست نیست، چون قبول کامل دیدگاه بیزی مستلزم تغییری جدی در نحوهٔ تفکر شما راجع به روشهای آماری است.

تلاش زیادی برای استخراج برآوردها و آزمونهای بهینه شده است. این موضوع در قسمت نظری کاملاً مشهود است؛ کتابهای فوق‌العاده غنی لیمن [۲۰a ، ۲۰b]، عمدتاً به روشهایی برای پیدا کردن برآوردها و آزمونهای خوب اختصاص دارند. و باز، به‌طور غیررسمی، در تحلیل داده‌ها دلایلی برای استفاده از یک فرایند به‌جای فرایند دیگر پیش کشیده می‌شود، مثلاً

وقتی به‌حق می‌گویند که در صورت وجود مشاهدات دورافتاده، میانگینهای اصلاح‌شده بهتر از میانگینهای خام‌اند. پس بگذارید به موضوع استنباط بپردازیم یعنی اظهارنظر راجع به یک پارامتر θ با در دست بودن داده‌های x درکنار پارامترهای مزاحم α. آماردان فراوانی‌گرا ممکن است به دنبال بهترین برآورد نقطه‌ای یا بازهٔ اطمینان یا آزمون معنی‌داری برای θ باشد.

یک واقعیت جالب‌توجه، و تا حد زیادی ناشناخته، این است که در چارچوب تفکر بیزی تمام مشکلات بهینگی از بین می‌رود. یعنی کل یک مسأله ناپدید می‌شود. چطور این اتفاق می‌افتد؟ دستورالعملی را در نظر بگیرید که عبارت است از محاسبهٔ $p(\theta|x,K)$ یعنی چگالی پارامتر مورد نظر با توجه به داده‌ها و دانش قبلی. این چگالیْ توصیف کاملی از شناخت فعلی شما از θ است. نکتهٔ دیگری برای گفتن وجود ندارد. این یک برآورد است: تنها برآورد شما، که اگر کل آن روی یک مجموعهٔ H در نظر گرفته شود، تمام اطلاعات شما را دربارهٔ اینکه H درست است یا نه تشکیل می‌دهد. هیچ چیزی بهتر از $p(\theta|x,K)$ وجود ندارد: یکتا و تنها کاندیداست. مورد میانگین اصلاح‌شده را که همین الان به آن اشاره شد، در نظر بگیرید. اگر الگو با نرمال ساده ترکیب شود، آنگاه چگالی θ حول میانگین نمونه یعنی $\bar{x}$، تقریباً نرمال است. اما، فرض کنید که به‌جای توزیع نرمال، توزیع tای استودنت قرار گیرد (با دو پارامتر مزاحم: پراکندگی و درجهٔ آزادی). در این‌صورت، چگالی θ حول $\bar{x}$ متمرکز نخواهد شد، بلکه حول چیزی که اساساً یک میانگین اصلاح شده است، متمرکز می‌شود. به عبارت دیگر، بحث برآورد کردن به ناچار پیش می‌آید و نه به‌خاطر ملاحظات بهینگی.

برآورد یکتای بیزی، توزیع پسین، به پیشین بستگی دارد. پس شباهتهایی بین آماردانی بیزی و آماردانی فراوانی‌گرا که هر دو از یک پیشین برای ساختن برآورد بهینهٔ خود استفاده می‌کنند، وجود دارد. (مجموعهٔ شیوه‌های خوب و مناسب فراوانی‌گرا همان مجموعهٔ بیزی است). فرق واقعی در این است که آماردان فراوانی‌گرا از معیارهایی متفاوت با آماردان بیزی استفاده می‌کند، مثلاً به‌جای استفاده از مفهوم سازگاری منطقی برای قضاوت در مورد کیفیت محصول، از مفهوم نرخ خطا بهره می‌جوید. دربارهٔ این مطلب در بخش ۱۶ بیشتر بحث می‌کنیم.

۱۳. اصل درستنمایی

دیدیم که استنباط پارامتری از طریق محاسبهٔ

$$p(\theta|x) = p(x|\theta)p(\theta) \Big/ \int p(x|\theta)\, p(\theta)\, \mathrm{d}\theta \qquad (۳)$$

به‌دست می‌آید. $p(x|\theta)$ را تابعی از دو کمیت x و θ در نظر بگیرید. $p(\cdot|\theta)$ به‌عنوان تابعی از x، به‌ازای هر مقدار ثابت θ، یک چگالی احتمال است، یعنی مثبت و انتگرال آن نسبت به x برابر ۱ است. به‌عنوان تابعی از θ، به‌ازای هر مقدار ثابت x، مثبت است اما معمولاً انتگرال آن برابر ۱ نیست. $p(x|\cdot)$ درستنمایی متغیر θ به‌ازای x ثابت نامیده می‌شود. از برابری (۳) به‌وضوح دیده می‌شود که تنها کمکی که داده‌ها به استنباط می‌کنند از طریق تابع درستنمایی برای x مشاهده شده می‌باشد. این همان اصل درستنمایی است که می‌گوید مقادیر x، به غیر از آنکه مشاهده‌شده، هیچ نقشی در استنباط ایفا نمی‌کنند. یک منبع مهم برای اثبات این موضوع، کتاب برگر و ولپرت [۴] است.

این موضوع پیامدهای مهم و شناخته شده‌ای در استنباط آماری دارد. هرگاه در استنباطْ عمل انتگرال‌گیری روی مقادیر x انجام گیرد، اصل بالا نقض شده و روش به‌دست آمده، ممکن است منسجم نباشد. برآوردهای نااریب و آزمون معنی‌دار بودن مساحت دُمی جزو عوامل آسیب دیده هستند. پس تابع درستنمایی در آمار بیزی نقش مهم‌تری ایفا می‌کند تا در مکتب فراوانی‌گرا؛ با این حال تابع درستنمایی به تنهایی برای استنباط کافی نیست و لازم است با توجه به توزیع پارامتر تعدیل شود. عدم‌حتمیت باید توسط احتمال بیان شود و نه درستنمایی. قبل از بسط و تفصیل این نکته، لازم است منظور از درستنمایی روشن شود. اگر یک الگو از داده‌های x و پارامترهای (θ,α) تشکیل شده باشد، آنگاه $p(x|\theta,\alpha)$ به‌عنوان تابعی از (θ,α)، به‌ازای هر مقدار ثابت و مشاهده‌شدهٔ x، بدون شک تابع درستنمایی است. با این حال، استنباط در برابری (۳) شامل تمام تابع درستنمایی نیست، بلکه فقط شامل انتگرال آن است:

$$p(x|\theta) = \int p(x|\theta,\alpha)\, p(\alpha|\theta)\, \mathrm{d}\alpha. \qquad (۴)$$

ما این را درستنمایی θ می‌نامیم، اما این نامگذاری همواره قابل قبول نیست. دلیلش کاملاً واضح است. ساختار آن فقط جنبهٔ $p(\alpha|\theta)$ از چگالی پارامتر $p(\theta,\alpha)$ را شامل می‌شود، که فراوانی‌گراها و طرفداران مکتب درستنمایی این دومی را قبول ندارند. در هیچ‌کدام از دو مکتب توافقی عام بر سر اینکه تابع درستنمایی برای پارامتر θی مورد نظر در حضور یک پارامتر مزاحم α از چه چیزی تشکیل می‌شود وجود ندارد. حداقل چندتایی کاندیدا برای این امر معرفی شده است. برای نمونه در معادلهٔ انتگرالی (۴)، از $p(x|\theta,\hat{\alpha})$ هم استفاده می‌شود که در آن $\hat{\alpha}$ مقداری است که باعث می‌شود $p(x|\theta,\alpha)$ نسبت به α یک مقدار ماکسیمم باشد. تعدّد کاندیداها نشان می‌دهد که پیدا کردن تعریفی مناسب و راضی‌کننده که مانع از دخالت نابه‌جای احتمالات در پارامتر شود، غیرممکن است.

دلیل آنکه درستنمایی به تنهایی کافی نیست، این است که درستنمایی بر خلاف احتمالْ جمع‌پذیر نیست. اگر A و B دو مجموعهٔ مجزا باشند، آنگاه داریم $p(A\cup B) = p(A) + p(B)$ در حالی‌که تساوی $l(A\cup B) = l(A) + l(B)$ برای یک تابع درستنمایی $l(\cdot)$ درست نیست. چون ویژگیهایی که به‌عنوان اصول موضوع در ارائهٔ یک استنباط به‌کار می‌روند (مثلاً در کارِ سه‌ویج) به مجموع‌پذیری می‌انجامد، هر گونه نقضی ممکن است باعث نقض اصول موضوع گردد. این نقض در مورد اصل درستنمایی اتفاق می‌افتد. در بخش ۵ مثالی دربارهٔ رنگ پوست و جنسیت آوردیم که برحسب مفهومِ غیردقیقِ محتمل‌تر بودن یک پیشامد نسبت به پیشامد دیگر بیان شده بود. در واقع اگر کلمهٔ «محتمل» به معنیِ فنیِ تعریف‌شده در اینجا به‌کار رود باز هم آن مثال درست است. درستنمایی یک عنصر ضروری در دستورالعمل استنباط است. اما تنها عنصر ضروری نیست.

توجه داشته باشید که اصل درستنمایی فقط در مورد استنباط صادق است، یعنی در مورد محاسبهٔ بعد از مشاهدهٔ داده‌ها. قبل از این مرحله، مثلاً در بعضی از جنبه‌های انتخاب الگو، به‌طور کلی در طرح آزمایش یا در تحلیل تصمیم، در نظر گرفتن مقادیر ممکن متعددی از داده‌ها ضروری است. (به بخش ۱۶ رجوع شود).

۱۴. مفاهیم فراوانی‌گرایانه

از دههٔ ۱۹۲۰ به بعد رهیافت فراوانی‌گرایانه بر آمار غلبه داشته و بر اساس هر معیار معقولی موفق هم بوده است. با این حال دیده‌ایم که این رهیافتْ تعارضاتی جدی با یک دیدگاه منسجم دارد. چطور چنین چیزی امکان دارد؟ توضیح ما این است: یک ویژگی وجود دارد که هر دو دیدگاه بیزی و فراوانی‌گرا در آن مشترک‌اند و آنها را بیشتر از همهٔ عواملی که تا به حال در اینجا بیان شده، به‌هم پیوند می‌دهد. این ویژگی عبارت است از مفهوم تبادل‌پذیری[1]. دنباله‌ای چون $(x_1, x_2, \ldots, x_n)$ از کمیتهای غیرحتمی با در نظر گرفتن شرایط K، برای شما تبادل‌پذیر است اگر توزیع احتمال توأم شما، به شرط K، به‌ازای جایگشتی از اندیسها ناوردا باشد. به‌عنوان مثال، $p(x_1 = 3, x_2 = 5|K) = p(x_2 = 3, x_1 = 5|K)$ روی جایگشت ۱ و ۲ ناورداست. یک دنبالهٔ نامتناهی تبادل‌پذیر است اگر هر زیر دنبالهٔ متناهی آن، این‌گونه باشد. به نقش «شما» و K به این علت اشاره شد که تأکید شود تبادل‌پذیری یک قضاوت ذهنی و شخصی است و اگر شرایط تغییر یابند، شما هم ممکن است عقیده‌تان را تغییر دهید.

اگر تشخیص دهید که یک دنباله (بینهایت بار) تبادل‌پذیر است، آنگاه ساختمان احتمال شما برای آن دنباله هم‌ارز با معرفی یک پارامتر، مثلاً ψ، است به‌نحوی که به شرط ψ اعضای دنباله به‌طور مستقل و یکسان توزیع شده‌اند (IID). چون این پارامترِ غیرحتمی است، شما یک توزیع احتمال برای آن خواهید داشت. این نتیجه از آنِ دوفینتی [۱۶a، ۱۶b] است. معمولاً ψ شامل عناصر (θ, α) است که در آن θ پارامتر مورد نظر و α مزاحم است. بنابراین، تبادل‌پذیریْ ساختار به‌کاررفته در بالا را تحمیل می‌کند و به‌علاوه داده‌های x شکل خاص مؤلفه‌های IID را دارند. همچنین ψ بستگی به خواص فراوانی دنبالهٔ مزبور دارد. بنابراین، در حالت ساده‌ای که دنبالهٔ برنولی است و x_i مقدار ۰ یا ۱ را می‌گیرد، ψ نسبت حدی x_iهایی است که مقدار ۱ را دارند. بنابراین، آماردانی بیزی که تبادل‌پذیری را تشخیص می‌دهد، در حقیقت همین قضاوت را در مورد داده‌ها به‌عنوان یک آماردان فراوانی‌گرا انجام می‌دهد، با این فرق که مقدار احتمال برای پارامتر را نیز تعیین می‌کند.

در قرن بیستم، مفهوم مشاهدات IID بر علم آمار سلطه داشته است. حتی وقتی که نامناسب بودن آن کاملاً روشن است، مثلاً در بررسی سری زمانی، از آن به‌عنوان الگوی پایه استفاده می‌شود. برای نمونه ممکن است $x_n - \theta x_{n-1}$، به‌عنوان مقداری از θ، IID فرض شود، و منجر به یک فرایند اتورگرسیو [خود برگشتی] خطی مرتبهٔ اول شود. در محدودهٔ فرض IID، ایده‌ها و مفاهیم مرتبط با فراوانی مناسب‌اند، حتی بعضی از این ایده‌ها در محدودهٔ قانون بیزی نیز مناسب و بجا هستند، به‌طوری که نظریه‌ای به‌وجود آمده که عدم‌حتمیت و احتمال را مبتنی بر فراوانی می‌داند. بعضی از متون آماری فقط با داده‌های IID سروکار دارند و بنابراین فعالیتهای آماری را محدود می‌کنند. مثالهای آنها از علوم تجربی گرفته می‌شوند که در آنجا تکرار پایهٔ کار است، و نه از قانون که در آن تکرار وجود ندارد. با این حال فراوانی کفایت نمی‌کند، چون معمولاً تکراری برای پارامترها وجود ندارد، پارامترها مقادیر یکتا و نامعلومی دارند. بنابراین سردرگمی بین فراوانی و احتمال، امکان استفاده از احتمال برای عدم‌حتمیت پارامتر را از فراوانی‌گرایان گرفته، و در نتیجه برای آنها لازم بوده مفاهیمی ناسازگار [با دیدگاه فراوانی‌گرا] مانند بازهٔ اطمینان را تعریف کنند.

1. exchangeability

استفاده از مفاهیم فراوانی در خارج از محدودهٔ تبادل‌پذیری باعث به‌وجود آمدن مشکل دیگری می‌شود. فراوانی‌گرایان معمولاً اظهارات خود را با گفتن اینکه «در درازمدت» حق با آنهاست، توجیه می‌کنند، که جواب منطقی به این گفته این است که «کدام درازمدت؟». مثلاً یک بازهٔ اطمینان مقدار واقعی را به نسبت $\alpha - 1$ بار در درازمدت دربر خواهد گرفت. برای اینکه معنی این مطلب را درک کنیم، لازم است که مورد خاص داده‌های x را در دنباله‌ای از مجموعه‌های داده‌های مشابه جای دهیم، اما این سؤال پیش می‌آید که، کدام دنباله؟ کدام داده‌ها مشابه هستند؟ مثال کلاسیک این موضوع مجموعه‌ای از داده‌هاست مرکب از r پیروزی در n آزمایش که تشخیص داده می‌شود یک دنبالهٔ برنولی است. در این دنباله آیا ما n را ثابت نگه می‌داریم یا r را و یا جنبه‌ای دیگر از داده‌های مشاهده‌شده را؟ این موضوع مهم است. آماردانان بیزی جوابی برای حالتِ مورد نظر به‌دست می‌دهند. در حالی‌که برای آماردانان فراوانی‌گرا معمولاً لازم است آن حالت را در دنباله‌ای از حالتها قرار دهند.

محدودیت احتمال به فراوانی، ممکن است باعث بروز اشتباه در کار شود. در اینجا مثالی دربارهٔ تعیینِ ثابتهای فیزیکی مانند ثابت گرانشی، G، می‌آوریم. معمول و منطقی است که فرض کنیم اندازه‌گیریهای انجام‌شده در یک محل و زمان قابل تبادل و نااریب هستند و هر کدام همان امید ریاضی G را دارند. منطقی است که از میانگین آنها به‌عنوان برآورد جاری G استفاده کنیم. شاید لازم باشد داده‌های دورافتاده را قبل از محاسبهٔ میانگین کنار بگذاریم. عدم‌حتمیتی که همراه این برآورد است، با محاسبهٔ s^2 به‌دست می‌آید که s^2 برابر میانگین مربع انحرافها از میانگین است و خطای معیار $s/\sqrt{n}$ است که در آن n تعداد دفعاتی است که اندازه‌گیری انجام‌شده یا به‌عبارت دیگر حجم نمونه است. این عمل باعث تعیین حدود اطمینان برای G می‌شود. تجربه نشان می‌دهد که معمولاً برآوردهای جدیدتر در خارج از حدود اطمینان قبلی قرار دارند. به بیان دیگر، فاصله بیش از حد کوتاه است. یک قاعدهٔ امتیازدهی برای برآوردکنندگان G، موجب جریمه‌ای بزرگ می‌شود. دلیلش این است که اندازه‌گیریها واقعاً اریب هستند. اما چون مقدار اریبی پذیرایِ ایده‌های مبتنی بر فراوانی نیست، نادیده گرفته می‌شود. در رهیافت بیزی یک توزیع برای این اریبی وجود دارد و برای G از یک پیشین که خودش پسینِ برآورد قبلی است، استفاده می‌شود که احتمالاً با توجه به فرایند اندازه‌گیریْ تعدیل و ترمیم شده است. معمولاً خطاهای معیار بسیار کوچک‌اند، چون فقط مؤلفهٔ تبادل‌پذیر عدم‌حتمیت در نظر گرفته می‌شود. اشتباهات مشابه ممکن است در پیش‌بینی تعداد آدمهایی که در آینده به بیماری ایدز مبتلا خواهند شد، رخ دهند. در این پیش‌بینی‌ها ممکن است تغییرات در رفتار شخصی و سیاستهای دولت، تغییراتی که نمی‌توانند در تحلیل فراوانی‌گرایانه وارد شوند، نادیده گرفته شوند.

۱۵. تحلیل تصمیم

اشاره شد که علم آمار با جمع‌آوری و نمایش داده‌ها شروع شد و بعد گسترش یافت تا پردازش و بررسی داده‌ها و فرایندی را که امروز استنباط می‌نامیم، دربربگیرد. مرحله‌ای فراتر از این هم وجود دارد، یعنی می‌توانیم برای رسیدن به یک تصمیم و یک حرکت آغازین از داده‌ها و استنباطهایی که از آنها حاصل می‌شود استفاده کنیم. به نظر من آماردانان سهمی حقیقی در تحلیل تصمیم دارند و باید جمع‌آوری داده‌ها و استنباطهای خود را طوری گسترش دهند تا

شامل عمل نیز بشود. روشهای رمزی و سه‌ویچ نشان داده که چگونه می‌توان مفاهیم پایه‌ای را از طریق تحلیل تصمیم به نمایش گذاشت. این گسترش برای دربرگرفتن عمل، زمانی بهتر فهمیده می‌شود که بپرسیم هدف از یک استنباط که شامل حساب کردن $p(y|x)$ برای داده‌های آینده، y، به شرط داده‌های قبلی، x است، چیست؟ مثالی دربارهٔ پزشکی ذکر کردیم که داده‌هایی راجع به دارویی در دست داشت و می‌خواست حدس بزند چه اتفاقی می‌افتد اگر آن دارو را به مریضِ خاصی بدهد. این مثال شامل یک تصمیم است، یعنی تصمیم راجع به اینکه کدام دارو تجویز شود؛ تجویز دارویی دیگر ممکن است منجر به یک استنباط دیگر در مورد y شود. ما از روش رمزی پیروی می‌کنیم و می‌گوییم استنباط فقط زمانی ارزش دارد که بتوانیم از آن برای حرکت آغازین استفاده کنیم. دانسته‌های جزئی که قابل استفاده نیستند، ارزش خیلی کمی دارند. $p(\theta|x)$، حتی اگر در شکل پارامتری‌اش باشد، تنها زمانی ارزش دارد که بتوان آن را با اعمالی آمیخت که با θی نامعلوم مربوط‌اند. مارکس درست می‌گفت: مسأله این نیست که فقط دنیا را بشناسیم (استنباط)، بلکه این است که آن را تغییر دهیم (عمل). حال بگذارید ببینیم چگونه می‌توان این کار را از دیدگاه بیزی انجام داد.

ساختاری که سه‌ویچ و دیگران آن را به‌کار بردند، تنظیم فهرستی از تصمیمهای ممکنِ d است که می‌توانیم اتخاذ کنیم. عدم‌حتمیت در یک مقدار (یا پارامتر) قرار دارد. جفت (d,θ) یک نتیجه تلقی می‌شود که نشان می‌دهد اگر پارامتر مقدار θ را داشته باشد و شما تصمیم d را بگیرید چه اتفاقی می‌افتد. دیده‌ایم که چرا لازم است عدم‌حتمیت θ توسط یک توزیع احتمال $p(\theta)$ بیان شود. این عمل به سطح معلومات خود شما، که در نماد نشان داده نمی‌شود بستگی دارد. ممکن است به تصمیم هم‌بستگی داشته باشد، مثلاً زمانی که تصمیم عبارت است از سرمایه‌گذاری کردن یا نکردن در تبلیغات، و θ نشان‌دهنده فروش سال آینده است. بنابراین می‌نویسیم $p(\theta|d)$. استدلال ادامه می‌یابد تا نشان داده شود که مزایای (d,θ) را می‌توان توسط یک عدد حقیقی $u(d,\theta)$ بیان کرد که به‌عنوان مطلوبیت نتیجه تلقی می‌شود. زمانی یک نتیجه به دیگری ترجیح داده می‌شود که از مطلوبیت بیشتری برخوردار باشد. اگر این مطلوبیتها عاقلانه بنا شوند، بهترین تصمیم آن است که مقدار مطلوبیت مورد انتظارِ شما یعنی

$$\int u(d,\theta)\, p(\theta|d)\, \mathrm{d}\theta$$

را به حداکثر برساند. اضافه کردن تابع مطلوبیت و ترکیب آن با توصیف احتمالاتی عدم‌حتمیت به حل مسألهٔ تصمیم می‌انجامد. مطلوبیت باید با دقت بیان شود زیرا صرفاً شاخصی از ارزش نیست بلکه شاخصی از ارزش در مقیاس احتمال است. اگر بهترین نتیجه دارای مطلوبیت ۱ و بدترین نتیجه دارای مطلوبیت ۰ باشد، آنگاه نتیجهٔ (d,θ) دارای مطلوبیت $u(d,\theta)$ است به شرط اینکه شما فرقی بین (الف) نتیجهٔ حتمی و (ب) یک شانس $u(d,\theta)$ برای بهترین (و $1-u(d,\theta)$ برای بدترین) قائل نباشید (به جنبهٔ ذهنی توجه داشته باشید). این ساختار احتمال است که به امید ریاضی امکان می‌دهد به‌عنوان تنها معیار مربوط به انتخاب تصمیم مورد استفاده قرار گیرد. مطلوبیت تمام جنبه‌های نتیجه را دربر می‌گیرد. برای نمونه، اگر یک برآمد در بازی قمار بردی به مبلغ ۱۰۰ پوند باشد، مطلوبیت آن نه تنها جنبهٔ پولی بلکه هیجان قمارکردن را نیز دربر دارد. بعضی تحلیلها که فقط بر اساس پول پایه‌ریزی شده‌اند، به خاطر محدودیت در دیدگاه مطلوبیت، ناقص هستند.

توجه داشته باشید: همان‌طور که $p(\theta)$ عدم‌حتمیت آماردان نیست، بلکه عدم‌حتمیت مشتری است، مطلوبیت هم از آن تصمیم گیرنده است. نقش آماردان این است که علائق مشتری را به شکل یک تابع مطلوبیت بیان کند، همان‌طور که عدم‌حتمیت او را با احتمال بیان می‌کند. همچنین توجه داشته باشید که در تحلیل فرض می‌شود فقط یک تصمیم‌گیرنده وجود دارد که همان «شما»ی متن ماست، هر چند ممکن است این «شما» چندین نفر باشند که یک گروه را تشکیل می‌دهند و تصمیمی دسته‌جمعی می‌گیرند. هیچ‌کدام از بحثهای مطرح‌شده در اینجا شامل موردی که دو یا چند تصمیم‌گیرنده هدف مشترکی نداشته باشند یا احیاناً با هم تقابل داشته باشند، نیست. این یک محدودیت مهم برای ماکسیمم امید ریاضی مطلوبیت است.

یکی از موضوعات مورد علاقهٔ آماردانان حداقل از زمانی که فیشر اثر درخشان خود را عرضه کرد [۱۷]، طرح آزمایشهاست. این مسأله، یک مسألهٔ تصمیم‌گیری است و در اصولی که همین الان شرح داده شد، به‌خوبی می‌گنجد. فرض کنید e عضوی از رده‌ای از آزمایشهای ممکن باشد که باید از بین آنها یکی انتخاب شود و x نمایانگر داده‌هایی باشد که ممکن است از چنین آزمایشی حاصل شوند. این آزمایشها قاعدتاً هدفی را دنبال می‌کنند که با انتخاب یک تصمیم (نهایی) d بیان می‌شود. طبق معمول، عدم‌حتمیت با θ نمایش داده می‌شود. (تحلیل مشابهی هم وقتی که استنباط در مورد داده‌های آینده، y، است به‌کار می‌رود) نتیجهٔ نهایی آزمایش و عمل عبارت است از (e,x,d,θ) که به آن یک مطلوبیت $u(e,x,d,\theta)$ نسبت می‌دهید. پس امید ریاضی مطلوبیت برابر است با

$$\int u(e,x,d,\theta)\, p(\theta|e,x,d)\, \mathrm{d}\theta \qquad (۵)$$

که عدم‌حتمیت موجود، مشروط به تمام عوامل دیگر است. تصمیم بهینه آن dای است که عبارت (۵) را ماکسیمم می‌کند. مقدار ماکسیمم حاصل با $\overline{u}(e,x)$ نمایش داده می‌شود. امید ریاضی آن برابر است با

$$\int \overline{u}(e,x)\, p(x|e)\, \mathrm{d}x. \qquad (۶)$$

از آنجایی که x تنها عامل غیرحتمی در این مرحله است، عدم‌حتمیت به‌وضوح وابسته به آزمایش e است. حاصل ماکسیمم سازی نهایی عبارت (۶) طرح آزمایش بهینه است.

به سادگیِ اصولی که در اینجا به‌کار برده می‌شوند توجه کنید، هر چند عملیات تکنیکی لازم ممکن است دشوار باشد. سلسله‌ای از مراحل طی می‌شود که به تناوب عبارت‌اند از گرفتن امید ریاضی روی کمیتهای غیرحتمی و ماکسیمم‌سازی روی تصمیمهایِ ممکن. هر عدم‌حتمیت باید مشروط به دانسته‌ها در زمان مورد نظر محاسبه شود. مطلوبیت به برآمد نهایی منسوب می‌شود و سایر مطلوبیتها(ی مورد انتظار) مانند $\overline{u}(e,x)$، از آن استخراج می‌شوند. این فرایندْ یک چارچوب رسمی برای طرح آزمایشها به‌وجود می‌آورد.

به نظر می‌رسد یک نقد معقول از روشی که کلیات آن همین الان بیان شد، این است: بسیاری از آزمایشها با یک تصمیم نهایی در ذهن اجرا نمی‌شوند، بلکه فقط برای جمع‌آوری اطلاعات در مورد θ اجرا می‌شوند. می‌توان این جنبه از موضوع را با گسترش تعبیر یک تصمیم، ملحوظ کرد. اطلاعات راجع به θ وابسته به عدم‌حتمیت شما در مورد θ است که طبق معمول با احتمال بیان می‌شود. پس فرض کنید تصمیم d در مورد انتخاب چگالی مربوط،

در اینجا $p(\theta|e,x)$، باشد. بعد می‌توان یک تابع مطلوبیت ساخت. بیشتر اوقات منطقی است که u را جمع‌پذیر در نظر بگیریم، یعنی

$$u(e,x,d,\theta)=u(e,x)+u(d,\theta) \qquad (۷)$$

اولین جمله شامل هزینهٔ آزمایش است و دومی شامل نتایج نهایی. به ارتباط بین $u(d,\theta)$ و قواعد امتیازدهی پیشنهادشده در بخش ۹ توجه کنید. در اینجا $u(d,\theta_0)$ ممکن است امتیاز مثبتی فرض شود که منسوب به تصمیم d برای اعلام $p(\theta|e,x)$ است هنگامی که پارامتر مقدار θ_0 را دارد. شاخص معمول برای اندازه‌گیری اطلاعات به‌دست آمده از $p(\theta)$، فرمول شانن:

$$\int p(\theta)\log\{p(\theta)\}\mathrm{d}\theta$$

است.

نیمن و دیگران زبان مورد استفاده در تحلیل تصمیم را در ارتباط با آزمودن فرض، آنجا که صحبت از تصمیم‌گیری برای پذیرش یا رد فرض H است، به‌کار برده‌اند. مواردی وجود دارد که مجازیم قبول یا رد کردن را یک عمل در نظر بگیریم، مانند رد کردن یک گروه از اشیا. در مقابل موارد دیگری هم وجود دارد که می‌توانیم $p(H|x,K)$ را به‌عنوان استنباطی در مورد H با داده‌های x و دانسته‌های K، به‌دست آوریم. همان‌طور که در پاراگراف قبل فرض شد، شکل دوم را می‌توان یک تصمیم در نظر گرفت. هر دو شکل معتبر هستند و مصارف گوناگون نیز دارند. فلسفهٔ ما هر دو دیدگاه را دربر می‌گیرد و کاملاً به خود شما بستگی دارد که چگونه واقعیتِ پیش رویِ خود را الگوسازی کنید. یک ویژگی مهم پارادایم بیزی، آن است که می‌تواند وضعیتهای بسیار گوناگون را به کمک معدودی اصل بنیادی دربر بگیرد.

بعضی از نویسندگان متون آماری در بحث راجع به آزمودن فرض اظهار کرده‌اند که وضعیتهای مختلف زیادی وجود دارد که بعضی از آنها ممکن است واقعاً شامل عمل باشند و بعضی دیگر صرفاً استنباطی باشند. موارد دیگری که توضیح داده شدند، مانند درستنمایی، به موقعیتهای پیچیده‌ای ختم می‌شوند. بعضی از مردم عاشق پیچیدگی هستند، چون پیچیدگی باعث پنهان ماندن خطاها و نارساییها می‌شود.

۱۶. (باز هم دربارهٔ) اصل درستنمایی

در بخش ۱۳ دیدیم که چگونه اصل درستنمایی می‌تواند پایه‌ای برای استنباط باشد؛ ولی بسیاری از نظریه‌های فراوانی‌گرا این موضوع را انکار می‌کنند. اصل درستنمایی هنگامی که طرح آزمایش جزو تحلیل تصمیم است، اساساً به‌خاطر انتگرال‌گیری نسبت به x در عبارت (۶) دیگر صادق نیست. در اولین مرحله، وقتی که تصمیم می‌گیرید کدام آزمایش را اجرا کنید، داده‌ها به شرط هر آزمایشی که انتخاب کنید، برای شما غیرحتمی‌اند. این عدم‌حتمیت توسط $p(x|e)$ بیان می‌شود و با گرفتن امید ریاضی در عبارت (۶) حذف می‌شود. در اجرای استنباط یا در تصمیم‌گیری نهایی، مقدار x را می‌دانید چون داده‌ها در دسترس‌اند. پس لازم نیست مقدار داده‌های دیگر را در نظر بگیرید. تنها چیزی که نیاز دارید درستنمایی است. هنگامی که با مسألهٔ طرح آزمایش سروکار دارید، داده‌ها مسلماً در دست نیستند و تمام امکانات باید در نظر گرفته شوند. این تقابل بین داده‌های پیشین و پسین، اهمیت شرایط را وقتی با عدم‌حتمیت مواجه‌اید، نشان می‌دهد. احتمال تابعی از دو متغیر است نه یک متغیر.

در زیر نشان می‌دهیم که چگونه بررسی آزمایش می‌تواند شامل شکلی از انتگرال‌گیری نسبت به x که مورد استفادهٔ فراوانی‌گرایان است، یعنی همان نرخ خطا، باشد. $d^*(e,x)$ را تصمیمی در نظر بگیرید که مطلوبیت مورد انتظار (۵) را ماکسیمم می‌کند. امید ریاضی نسبت به x، عبارت (۵)، را می‌توان به‌صورت زیر نوشت

$$\iint u\{e,x,d^*(e,x),\theta)\}p\{\theta|e,x,d^*(e,x)\}\mathrm{d}\theta p(x|e)\,\mathrm{d}x.$$

حال

$$p\{\theta|e,x,d^*(e,x)\}=p(\theta|e,x)$$

و چون اضافه کردن d^*، تابعی از e و x، هیچ شرط جدیدی را به مسأله اضافه نمی‌کند، احتمال دوم برابر است با

$$p(x|e,\theta)\,p(\theta|e)/p(x|e).$$

با قرار دادن این کمیت در امید ریاضی و جابه‌جا کردن ترتیب انتگرال‌گیری داریم

$$\iint u\{e,x,d^*(e,x),\theta)\}p(x|e,\theta)\,\mathrm{d}x\,p(\theta|e)\,\mathrm{d}\theta$$

که انتگرال داخلی همان شکل انتگرال‌گیریِ فراوانی‌گرایانه نسبت به x را نشان می‌دهد. برای یک آزمایش ثابت، با مطلوبیتی که به‌طور مستقیم شامل x نیست، انتگرال مربوط برابر است با

$$\int u\{d^*(e,x),\theta\}p(x|e,\theta)\,\mathrm{d}x.$$

به‌ازای دو تصمیم و مطلوبیت ۰ – ۱، بلافاصله $\int p(x|e,\theta)\,\mathrm{d}x$ روی زیرمجموعه‌ای از فضای نمونه و هر دو نوع خطای آشنا را در اختیار داریم.

وجود نرخهای خطا باعث یک نوع سردرگمی می‌شود چون معمولاً این مقادیر را کمیتهایی در نظر می‌گیرند که باید کنترل شوند، و بنابراین جایگاه عمده‌ای در تحلیل تصمیم دارند. در حالی که بررسی اصلی ما بر اساس ساختار مطلوبیت است. هنگامی که ساختار مطلوبیت اِعمال شود خطاها خودبه‌خود کنترل می‌شوند. اما در نظر گرفتن خطاهای گوناگون ممکن است منجر به یک رشته تغییرات ناخواسته در ساختار مطلوبیت شود. دیدگاه بیزی این است که مطلوبیتها و نه خطاها، ناورداهای تحلیل هستند. برای نمونه، طرح یک آزمایش برای دستیابی به نرخ خطای از پیش تعیین‌شده ممکن است نامعقول باشد. به‌جای خطا، مطلوبیت باید مشخص شود.

۱۷. مخاطره

مخاطره واژه‌ای است که تا به حال از آن استفاده نکرده‌ایم. این اصطلاح را داکورث[۱۳] چنین تعریف کرده است: امکانِ قرار گرفتن در معرض رویدادهایی غیرحتمی که وقوع آنها پیامدهایی نامطلوب خواهد داشت. با آنکه داکورث مانند اغلب آماردانان بیشتر بر زیان و نامطلوب بودن تأکید دارد تا بر سود و مطلوبیت، تعریف او دو عنصر اصلی را در آنچه ما تحلیل تصمیم می‌نامیم، یعنی عدم‌حتمیت و مطلوبیت را، در نظر می‌گیرد. تفاوت در واقع لفظی است. در نتیجه مخاطره به دو متغیر وابسته است و بحث مبنایی ما در بخش ۳ وابسته به جدایی این دو متغیر برحسب عدم‌حتمیت و ارزش

است. با این حال، روش معمول که داکورث هم از آن پیروی کرده این است که شاخص مخاطره را با یک تک‌عدد بیان کنند و این جدایی را انکار کنند. بنابراین، مقدار مخاطرهٔ مربوط به پرواز ۱۰۰۰ مایلی برابر ۱٫۷ برحسب واحد مناسب است. این گفته بر اساس دلیل زیر قابل دفاع است.

تصمیم بهینه، مقدار امید ریاضی مطلوبیت را به ماکسیمم می‌رساند، که برای داده‌های x متناسب است با

$$\int u(d,\theta)\, p(\theta)\, p(x|\theta)\, \mathrm{d}\theta$$

و می‌توان آن را به‌صورت درستنماییِ وزندار زیر نوشت

$$\int w(d,\theta)\, p(x|\theta)\, \mathrm{d}\theta$$

که در آن $w(d,\theta) = u(d,\theta)\, p(\theta)$. این تحلیل، برای داده‌های مفروض و بنابراین برای یک درستنمایی داده شده، به مطلوبیت و احتمال، یعنی دو پایهٔ اساسی فلسفهٔ ما، به‌طور جداگانه وابسته نیست، بلکه فقط به حاصلضرب این دو وابسته است. به عبارت دیگر، اگر شما یک انسان عاقل را در حالتی مشاهده کنید که عملی انجام می‌دهد (نه در حالتی که افکار خود را بیان می‌کند) نمی‌توانید، بر اساس اعمال مشاهده‌شده، دو عامل فکر و عمل را از هم تشخیص دهید، تنها چیزی که می‌توانید مشخص کنید تابع وزن است.

با این حال، چندین دلیل برای جداسازی مطلوبیت از احتمال وجود دارد. مهم‌ترین دلیل نیاز به استنباط است، یعنی درک صحیح دنیا بدون رجوع به عمل. فلسفه می‌گوید که این نیاز از طریق ساختار احتمالاتی شما برای دنیا برآورده می‌شود. در حیطهٔ استنباط، عملیات کاملاً در محدودهٔ حساب احتمال انجام می‌گیرد، بنابراین از مطلوبیت جدا می‌شود. کسانی معتقدند که استنباط به شکل متداول در علوم محض، هنگامی که از کاربردهایش در فناوری جدا شود، نامطلوب است. بدون شک موضوع مهم این است که استنباط باید در قالبی مناسب برای تصمیم‌گیری باشد و نه عملی که جدا از کاربردهاست. دیدیم که چگونه استنباط بیزی کاملاً برای این منظور مناسب است. در بخش ۱۹ می‌بینیم که چگونه بعضی از جنبه‌های حقوق استنباط را از تصمیم جدا می‌کنند.

یک دلیل دیگر برای جداسازی احتمال و مطلوبیت از همدیگر، در علاقهٔ مردم به ارتباط با یکدیگر، و در واقع ارتباط بین «شماها»ی مختلف، نهفته است. برای نمونه پرواز ۱۰۰۰ مایلی را که در بالا مطرح شد در نظر بگیرید. قسمتی از محاسبات به نرخ حوادث مشاهده شدهٔ قبلی در مورد هواپیما بستگی دارد و قسمت دیگر به عواقب پرواز هواپیما وابسته است. شما ممکن است عکس‌العملهای متفاوتی در برابر این دو عامل از خود نشان دهید. مثلاً می‌دانیم که برای افراد پیر، ساعتها نشستن در صندلیهای ناراحت ممکن است خطر بروز مشکلاتی در جریان خون را افزایش دهد. بنابراین ممکن است ارزیابی شما از نرخ حوادث متفاوت با ارزیابی‌ای باشد که صرفاً از نرخ حوادث مربوط به هواپیما به‌دست می‌آید. در مقابل، یک مدیرعامل سالم میانسال که در شرایط راحت‌تری در قسمت درجهٔ یک نشسته است، ممکن است آمار حوادث را بپذیرد، اما مثلاً به خاطر اهمیت ملاقاتی که می‌خواهد در این سفر انجام دهد نظر متفاوتی دربارهٔ مطلوبیت داشته باشد. از این ملاحظات چنین برمی‌آید که نرخ حوادث و عواقب یک حادثه را باید از یکدیگر تفکیک کرد، چون ممکن است بتوان از یکی استفاده کرد و از دیگری نه، در حالی‌که استفاده از تابع وزن به تنهایی دشوار خواهد بود.

۱۸. علوم

کارل پیرسون گفته است: «وحدت علوم فقط در روش آن است نه در محتوایش» [۲۳]. درست نیست که بگوییم فیزیک جزو علوم است ولی ادبیات نیست. مواقعی هست که فیزیکدان مانند هنرمند پرشی به دنیای تخیل می‌کند. تحلیل تعداد کلمات یک متن می‌تواند به شناسایی نویسندهٔ یک متن ادبیِ فاقد امضا کمک کند. روش علمی در فیزیک به‌وضوح مهم‌تر است تا در ادبیات، اما این امکان بالقوه را دارد که در هر رشته‌ای بتوان از آن استفاده کرد.

این روش شامل چیست؟ متون بیشماری در پاسخ به این پرسش نوشته شده است و گستاخانه خواهد بود که من ادعا کنم جواب را پیدا کرده‌ام. اما اعتقاد راسخ دارم که آماردانان در مطالعات عمیق خود دربارهٔ گردآوری و تحلیل داده‌ها، شاید هم ندانسته، جواب را به‌دست آورده‌اند و جواب در فلسفه‌ای که در اینجا ارائه شد، مستتر است. آزمایش کردن که باعث تولید داده‌ها می‌شود، یک عنصر اصلی در روش علمی است، پس ارتباط بین آمار و علوم تجربی تعجب‌آور نیست. از این دیدگاه روش علمی عبارت است از بیان نظر شما دربارهٔ دنیای غیرحتمی شما به زبان احتمال، از طریق اجرای آزمایشهایی برای به‌دست آوردن داده‌ها و استفاده از آن داده‌ها برای روزآمد کردن احتمال و در نتیجه، روزآمد کردن نظرتان نسبت به دنیا. اگرچه در این مورد معمولاً بر آمار بیزی تأکید می‌شود، اما قانون حاصلضرب یعنی حذف پارامترهای مزاحم همه‌جا حاضر با اضافه کردن قاعدهٔ ۲ (جمع)، نیز عملاً مهم است. همان‌طور که دیدیم، طرح آزمایش هم پذیرایِ بررسی آماری است. روش علمی شامل دنباله‌ای از مراحل است که به تناوب بین استدلال و آزمایش کردن، تغییر می‌کنند. همان‌طور که در بخش ۸ بیان شد، هر دانشمند یک «شما» است با عقاید خاص خود که این عقاید از طریق گردآوری داده‌ها به سوی هماهنگی پیش می‌روند. رسیدن به این اجماع یا اتفاق‌آرا اساس علوم عینی است.

این ایراد را به این دیدگاه ساده گرفته‌اند که دانشمندان به روشی که در پاراگراف قبل بیان شد، عمل نمی‌کنند، حتی می‌توان گفت که آزمونهای معنی‌دار بودن مساحت دُمی را انجام می‌دهند. جواب این اعتراض آن است که فلسفهٔ ما تجویزی است نه توصیفی. هدف ما بیانِ شیوهٔ عمل دانشمندان نیست بلکه بیان این است که اگر می‌دانستند چگونه باید عمل کنند چه می‌کردند. حساب احتمال جواب این «چگونه» را فراهم می‌کند. نقص مهمی که رویِ این «چگونه» اثر می‌گذارد، نبودِ روشهای خوب برای ارزیابی احتمالات در زمانی است که هیچ فرض تبادل‌پذیری برای راهنمایی شما در دسترس نیست. معمولاً گفته می‌شود که این فرض برای تعیین توزیع پیشین شماست، اما در واقع گسترده‌تر از آن است. بعضی از حملاتی که به علوم می‌شود در واقع حمله به چگونگی رفتار دانشمندان است. بیشتر این اعتراضات بجا هستند ولی این نوع حملات اگر متوجه استانداردهای تجویز شده برای عمل (نه شیوهٔ عمل رایج) بود نمی‌توانست این‌قدر کوبنده باشد. دانشمندان انسان هستند و شرایط بیرونی بر کار آنها تأثیر می‌گذارد. می‌توان امیدوار بود که دانشمندی که برای یک شرکت چند ملیتی کار می‌کند یا دانشمندی که در استخدام سازمان محیط زیست است، تنها از لحاظ احتمالاتی که برای پدیده‌ها قائل می‌شوند با هم متفاوت باشند و افکار خود را بر آن اساس روزآمد کنند. ولی گمان می‌رود که مسائل دیگری هم در میان باشد. امید من این است که رهیافت بیزی به آشکار ساختن هر گونه ناآرایی و یا اشتباه در هر کدام از استدلالهای مبلغان نظریه‌ها کمک کند.

۱۹. حقوق جنایی

این بخش را به دو دلیل در این مقاله می‌آورم: اول به دلیل علاقهٔ شخصی خود به مباحث دادگاهی، و دوم به علت اعتقادی که این علاقه پدید آورده است، اعتقاد به اینکه بعضی از جنبه‌های مهم حقوق جنایی پذیرای روش علمیِ مورد بحث در بخش قبل است. این جنبه‌ها به فرایند دادرسی مربوط می‌شوند که در آن دربارهٔ گناهکار بودن متهم عدم‌حتمیت وجود دارد، عدم حتمیتی که بعداً توسط داده‌هایی به شکل شواهد، تعدیل می‌شود و امید آن است که اتفاق‌نظری در مورد گناهکار بودن یا نبودن متهم به‌دست آید. این مطلب به‌وضوح در الگویی که در اینجا مطرح شده، می‌گنجد. امتیاز کشف حقیقت به‌طور انحصاری از آنِ وکلا نیست و این کار قرنها توسط دانشمندان با موفقیت انجام می‌شده است. جنبه‌هایی از حقوق هست، مانند نوشتن قانون، که روش علمی نمی‌تواند چیزی به آن اضافه کند. ولی دادگاهها فقط با جرم سروکار ندارند، آنها باید حکم هم صادر کنند. حقوقدانان هم مانند ما این دو مورد را از هم جدا کرده‌اند. این دو مورد را می‌توان به‌ترتیب به‌عنوان استنباط (تشخیص جرم) و تصمیم‌گیری (صادر کردن حکم) قلمداد کرد.

متهم در دادگاه یا واقعاً گناهکار است (G) یا گناهکار نیست ($\sim G$). گناهکار بودن متهم حتمی نیست، بنابراین باید برحسب احتمال $p(G)$ بیان شود. (دانسته‌های قبلی در این نمادگذاری منظور نمی‌شود.) داده‌ها به‌صورت شواهد E به‌دست می‌آیند و احتمال روزآمد می‌شود. چون فقط دو فرض وجود دارد: G یا $\sim G$، کار کردن با بختها راحت‌تر است:

$$o(G) = p(G)/p(\sim G).$$

با استفاده از قضیهٔ بیز داریم

$$o(G|E) = \frac{p(E|G)}{p(E|\sim G)}\, o(G)$$

که شامل ضرب بخت اولیه در نسبت درستنمایی است. معمولاً شواهد شامل پارامترهای مزاحم هستند. اما اصولاً می‌توان پارامترهای مزاحم را به روش معمول یعنی با استفاده از قانون جمع حذف کرد. چون ممکن است چندین راه برای ارتکاب جرم، به‌غیر از راهی که متهم پیموده است، وجود داشته باشد، پارامترهای مزاحم بیشتر مواقع وارد عبارت $p(E|\sim G)$ خواهند شد. با ادامهٔ دادرسی، شواهد بیشتری مطرح می‌شوند و با ضرب پی‌درپی در نسبت درستنمایی بالاخره بخت نهایی مشخص می‌شود. مشکلی که وجود دارد این است که به فرض هر کدام از G یا $\sim G$، اجزای پی‌درپی مدارک ممکن است مستقل از هم نباشند.

تا به حال کاربرد این روش در مورد شواهد علمی مانند لکهٔ خون و DNA با موفقیت همراه بوده است[۱]. قابلیت کاربرد آن به‌طور کلی وابسته به روشهای مناسب در تعیین احتمال است. این روش از این امکان بالقوه برخوردار است که به دادگاه در تلفیق انواع مختلف شواهد کمک می‌کند، چون همان‌طور که در بخش ۳ گفته شد، مزیت اصلی اندازه‌گیری در توانایی آن برای تلفیق چندین مورد عدم‌حتمیت و تبدیل آنها به یک مورد است.

علم حقوق با فلسفهٔ جداسازی استنباط از تصمیم موافق است و حتی اجازه می‌دهد که شواهد مختلف بر روی این دو فرایند تأثیر بگذارند. برای نمونه ممکن است محکومیتهای قبلی یک متهم در صادر کردن حکم تأثیر بگذارد (تصمیم‌گیری)، اما همیشه روی تعیین جرم (استنباط) تأثیر نمی‌گذارد. تحلیل مطلوبیت مورد انتظار شامل قضیه‌ای است به این مضمون که انتظار می‌رود اطلاعات رایگان همواره مقدار مطلوبیت را افزایش دهد. از این مطلب چنین برمی‌آید که تنها دلیل برای نپذیرفتن شواهد باید بر اساس هزینهٔ آن باشد [۱۵].

بخشی از فرایند دادرسی که به تشخیص گناهکار بودن یا بی‌گناهی متهم می‌انجامد، در نظریهٔ ما جای خود را به محاسبهٔ بخت $o(G|E)$ می‌دهد که در آن E کل شواهد پذیرفته شده می‌باشد. در این نظریه هیأت منصفه نباید اظهار نظر قطعی دربارهٔ گناهکار بودن یا بی‌گناهی متهم بکند، بلکه باید بخت نهایی یا احتمال گناهکار بودن را از نظر خودش بیان کند. با این کار، حداقل، امکان گفتگویی انعطاف‌پذیر و روشنگر را فراهم می‌کند. مهم‌تر اینکه، اطلاعاتی را که قاضی برای صدور حکم به آنها نیاز دارد در اختیارش می‌گذارد. اگر d تصمیمی ممکن در مورد زندانی یا جریمه شدن مجرم باشد، آنگاه مطلوبیت d برابر است با

$$u(d,G)p(G|E) + u(d,\sim G)p(\sim G|E).$$

حکم بهینه آن dای است که این امید ریاضی را به مقدار ماکسیمم خود برساند. مطلوبیتها در اینجا بازتاب ارزیابی جامعه از مزایای حکمهای مختلف برای مجرم و جدی بودن قضاوتهای اشتباه است. ما فاصلهٔ زیادی تا اجرای این ایده‌ها داریم، اما حتی حالا هم این ایده‌ها می‌توانند ما را در انتخاب شیوه‌های منطقی و دوری از شیوه‌های نامعقول راهنمایی کنند.

۲۰. نتیجه‌گیری

فلسفهٔ آماری که در اینجا بیان شد دارای سه اصل اساسی است: اول اینکه عدم حتمیت باید برحسب احتمال بیان شود. دوم اینکه ارزش پیامدها باید به‌وسیلهٔ مطلوبیت بیان شود، و سوم اینکه برای گرفتن تصمیم بهینه، احتمالها و مطلوبیتها باید از طریق محاسبهٔ امید ریاضی مطلوبیت و ماکسیمم کردن آن، با هم ترکیب شوند. اگر این اصول پذیرفته شد، آنگاه اولین قدم آماردان تشکیل یک الگوی احتمال است که تمام علائق و عدم‌حتمیت مشتری را دربر داشته باشد. الگوی مورد نظر، داده‌ها و هر پارامتری را که به نظر لازم برسد دربر خواهد داشت. هنگامی که این کار انجام شد، استنباطِ لازم با استفاده از فنون حساب احتمال به‌دست می‌آید. اگر نیاز به تصمیم‌گیری هم باشد، الگو باید گسترش یابد تا شامل مطلوبیتها شود و به دنبال آن یک عمل مکانیکی دیگر یعنی ماکسیمم کردن امید ریاضیِ مطلوبیت لازم است. یک جنبهٔ جذاب این موضوع آن است که کل فرایند به‌خوبی تعریف شده است و نیازی به فرضهای موردی نیست. با وجود این، نیاز قابل‌ملاحظه‌ای به تقریب‌زدن داریم. اجرای این طرح در مورد دنیای بزرگ موجود، غیرممکن است. ضروری است از یک دنیای کوچک‌تر که ساده‌سازیهایی وارد کار می‌کند ولی اغلب موجب تحریفاتی هم می‌شود استفاده کنیم. حتی در محاسبات هم به تقریب عددی نیاز است. هر دوی این موضوعها در متون آماری، خواه فراوانی‌گرا یا بیزی، بررسی شده‌اند و پیشرفت قابل‌ملاحظه‌ای هم به‌دست آمده است. مشکل واقعی در ساختن الگو پیش می‌آید. روشهای ارزشمندِ بسیاری برای ساختن الگو معرفی شده‌اند، اما به خاطر تأکید رهبافت فراوانی‌گرا بر کارهای قبلی، ما نحوهٔ تعیین احتمالات ــ چگونگی بیان عدم‌قطعیتهای خود به شکل درخواستی ــ را دقیقاً نمی‌دانیم. حال که در آستانهٔ قرن جدید قرار داریم، به نظر من مهم‌ترین موضوع تحقیقاتی آماری، تعیین روشهای منطقی برای ارزیابی احتمال است. این کار نیاز به همکاری روانشناسان تجربی که استعداد

ریاضی داشته باشند و مقدار زیادی کار آزمایشی دارد. یکی از همکاران این موضوع را به‌صورتی ساده اما با کمی مبالغه بیان کرده است: «در آمار هیچ مشکلی باقی نمانده است، مگر ارزیابی احتمال». جالب این است که متخصصان حرفه‌ای احتمال هیچ چیزی راجع به ارزیابی نمی‌دانند و علاقه‌ای هم به دانستن آن ندارند.

قبول موضعی که رئوس آن در این مقاله بیان شد، حیطهٔ عمل آماردان را گسترش می‌دهد و تصمیم‌گیری نیز، علاوه بر جمع‌آوری داده‌ها، ساختن الگو و استنباط، در آن حیطه قرار می‌گیرد. اما باعث محدودیتی نیز در کار آماردان می‌شود که این محدودیتها هنوز کاملاً شناخته نشده‌اند. آماردانان در خانهٔ خودْ صاحبخانه نیستند. وظیفهٔ آنها کمک به مشتری برای رویارویی با عدم‌حتمیتی است که در برابرش قرار گرفته است. «شما» در تجزیه و تحلیل همان مشتری است، نه آماردان. مجله‌های ما و حتی شاید هم کارهای ما، بیش از حد از احتیاجات مشتری دور هستند. در این امر من هم مثل دیگران گناهکارم. اما نظریه‌پرداز دست‌کم روشها را به‌وجود آورده است. وظیفهٔ شما آن است که این روشها را به مصرف درست برسانید.

مراجع

1. Aitken, C. G. G. and Stoney, D. A. (1991) *The Use of Statistics in Forensic Science*. Chichester: Horwood.
2. Bartholomew, D. J. (1988) Probability, statistics and theology (with discussion). *J. R. Statist. Soc.* A, **151**, 137-178.
3. Berger, J. O. and Delampady, M. (1987) Testing precise hypotheses (with discussion). *Statist. Sci.*, **2**, 317-352.
4. Berger, J. O. and Wolpert, R. L. (1988) *The Likelihood Principle*. Hayward: Institute of Mathematical Statistics.
5. Bernardo, J. M. (1999) Nested hypothesis testing: the Bayesian reference criterion. In *Bayesian Statistics 6* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith). Oxford: Clarendon.
6. Bernardo, J. M., Berger, J. O., Dawid, A, P. and Smith, A. F. M. (eds) (1999) *Bayesian Statistics 6*. Oxford: Clarendon.
7. Bernardo, J. M. and Smith, A. F. M. (1994) *Bayesian Theory*. Chichester: Wiley.
8. Box, G. E. P. (1980) Sampling and Bayes' inference in scientific modelling (with discussion). *J. R. Statist. Soc.* A, **143**, 383-430.
9. Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations (with discussion). *J. R. Statist. Soc.* B, **26**, 211-252.
10. Dawid, A. P., Stone, M. and Zidek, J. V. (1973) Marginalization paradoxes in Bayesian and structural inference (with discussion). *J. R. Statist. Soc.* B, **35**, 189-233.
11. DeGroot, M. H. (1970) *Optimal Statistical Decisions*. New York: McGraw-Hill.
12. Draper, D. (1995) Assessment and propagation of model uncertainty (with discussion). *J. R. Statist. Soc.* B, **57**, 45-98.
13. Duckworth, F. (1998) The quantification of risk. *RSS News*, **26**, no. 2, 10-12.
14. Edwards, W. L., Lindman, H. and Savage, L. J. (1963) Bayesian statistical inference for psychological research. *Psychol. Rev*, **70**, 193-242.
15. Eggleston, R. (1983) *Evidence, Proof and Probability*. London: Weidenfeld and Nicolson.
16a. de Finetti, B. (1974) *Theory of Probability*, vol. 1. Chichester: Wiley.
16b. ___ (1975) *Theory of Probability*, vol. 2. Chichester: Wiley.
17. Fisher, R. A. (1935) *The Design of Experiments*. Edinburgh: Oliver and Boyd.
18. Healy, M. J. R. (1969) Rao's paradox concerning multivariate tests of significance. *Biometrics*, **25**, 411-413.
19. Jeffreys, H. (1961) *Theory of Probability*. Oxford: Clarendon.
20a. Lehmann, E. L. (1983) *Theory of Point Estimation*. New York: Wiley.
20b. ___ (1986) *Testing Statistical Hypotheses*. New York: Wiley.
21. O'Hagan, A. (1995) Fractional Bayes factors for model comparison (with discussion). *J. R. Statist. Soc.* B, **57**, 99-138.
22. Onions, C. T. (ed.) (1956) *The Shorter English Dictionary*. Oxford: Clarendon.
23. Pearson, K. (1892) *The Grammar of Science*. London: Black.
24. Ramsey, F. P. (1926) Truth and probability. In *The Foundations of Mathematics and Other Logical Essays* (ed. R. B. Braithwaite), pp. 156-198. London: Kegan Paul.
25a. Savage, L. J. (1954) *The Foundations of Statistics*. New York: Wiley.
25b. ___ (1977) The shifting foundations of statistics. In *Logic, Laws and Life: Some Philosophical Complications* (ed. R. G. Colodny), pp. 3-18. Pittsburgh: Pittsburgh University Press.
26. Stein, C. (1956) Inadmissibility of the usual estimation of the mean of a multivariate normal distribution. In *Proc. 3rd Berkeley Symp. Mathematical Statistics and Probability* (eds J. Neyman and E. L. Scott), vol. 1, pp. 197-206. Berkeley: University of California Press.
27. Walley, P. (1991) *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall.

******

• Dennis V. Lindley, "The philosophy of statistics", *The Statistician*, (3) **49** (2000) 293-319.

به علت طولانی بودن مقاله، ترجمهٔ آن به دو قسمت تفکیک شد و قسمت اول در شمارهٔ گذشته آمده است.

* دنیس لیندلی، انگلستان

thombayes@aol.com