مقاله کلاسیک

مجادله در مبانی آمار•

برادلی افرون*
ترجمهٔ علی عمیدی

۱. مقدمه

به نظر می‌رسد که آمار موضوعی دشوار برای ریاضیدانان است، شاید به دلیل خصوصیت مبهم‌گونه و دامنهٔ گسترده‌اش که باعث می‌شود در عرضهٔ مطالب آن از روش سنتی قضیه-اثبات عدول شود. لذا ممکن است تسلی خاطری باشد که بگوییم آمار برای آماردانان نیز موضوعی مشکل است.

ما اینک، گذشت تقریباً دو سده از شروع مجادله‌ای بر سر مبانی آمار را گرامی می‌داریم. دو گروه اصلی در این ستیز فلسفی، یعنی بیزگراها و فراوانی‌گراها، به تناوب هر کدام برای مدت زمانی بر دیگری تفوق داشته‌اند، ولی در حال حاضر، فراوانی‌گراها از برتری مختصری برخوردارند؛ گروه سوم کوچکتری که بهتر است آنها را فیشریها بنامیم به هر دو گروه دیگر به شدت ایراد می‌گیرند.

آمار، بنا به تعریف، به حالتهای خاص توجهی ندارد. متوسطها خوراک آماردانان‌اند. «متوسط» در اینجا به مفهوم کلی یعنی به معنای هر حکم خلاصه‌ای دربارهٔ جامعهٔ بزرگی از اشیاء است. حکم «متوسط IQی دانشجویان تازه واردِ یک دانشگاه برابر ۱۰۹ است» و همچنین حکم «احتمال اینکه سکه‌ای بیفتد و برآمد آن شیر باشد $\frac{1}{2}$ است.» از این گونه حکمها هستند. مجادله‌هایی که تقسیم‌کنندهٔ دنیای آمارند حول نکتهٔ اساسی زیرند: دقیقاً کدام متوسطها برای به دست‌آوردن استنباط از روی داده‌ها مناسبترند؟ فراوانی‌گراها، بیزگراها، و فیشریها پاسخهای اساساً مختلفی به این سؤال می‌دهند.

این مقاله به جای تشریح اصل موضوعی یا تاریخی دیدگاههای مختلف، با تعدادی مثال به پیش می‌رود. مثالها به خاطر درک عموم، به صورتی ساختگی ساده هستند، ولی خوانندگان مطمئن باشند که این اختلاف دیدگاهها برای داده‌های واقعی هم به همین اندازه صادق‌اند. ذکر این نکته هم مناسب است: این اختلاف دیدگاهها، چه از لحاظ نظری و چه از لحاظ کاربردی لطمه‌ای به آمار نزده‌اند، و در واقع، سهمی در شادابی آن دارند. دستاوردهای جدید مهمی، به ویژه روشهای بیزی تجربی که در بخش ۸ به آنها اشاره می‌شود، مستقیماً از جدال بین دیدگاههای بیزگرا و فراوانی‌گرا حاصل شده‌اند.

۲. توزیع نرمال

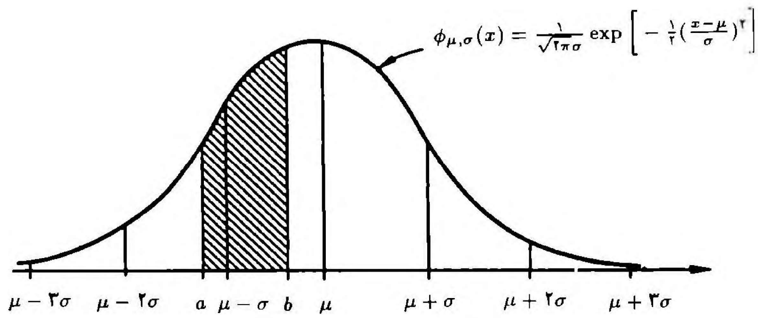
همهٔ مثالهای ما مربوط به توزیع نرمال‌اند، که به دلایل مختلف نقشی اساسی در آمار نظری و کاربردی ایفاء می‌کند. یک متغیر تصادفی نرمال یا گاوسی x، کمیتی است که می‌تواند هر مقداری را روی محور حقیقی اختیار کند، ولی نه با احتمال یکسان. احتمال آنکه x در بازهٔ $[a, b]$ قرار گیرد با مساحت سطح زیر خم معروف زنگدیس گاوس معین می‌شود، یعنی

$$\mathrm{Prob}\{a \leq x \leq b\} = \int_a^b \phi_{\mu,\sigma}(x)dx \qquad (۱.۲)$$

که در آن

$$\phi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp[-\frac{1}{2}(\frac{x-\mu}{\sigma})^2] \qquad (۲.۲)$$

برادلی افرون (Bradley Efron) درجهٔ دکتری خود را در آمار در سال ۱۹۶۴ با راهنمایی روپرت میلر (Rupert Miller) از دانشگاه استنفرد گرفت. زمینهٔ علائق او بیشتر بخشهای آمار نظری و کاربردی ــ به‌خصوص کاربرد روشهای هندسی در مسائل آماری ــ را در بر می‌گیرد.



شکل ۱ توزیع نرمال. کمیت تصادفی $x\sim\mathcal{N}(\mu,\sigma^2)$ با احتمالی برابر با مساحت سطح هاشورزده، در $[a,b]$ رخ می‌دهد. ۶۸٪ احتمال دربازهٔ $[\mu-\sigma,\mu+\sigma]$، ۹۵٪ دربازهٔ $[\mu-2\sigma,\mu+2\sigma]$، و ۹۹٫۷٪ دربازهٔ $[\mu-3\sigma,\mu+3\sigma]$ است.

برای آسانی، چنین متغیر تصادفی را با نماد

$$x\sim\mathcal{N}(\mu,\sigma^2) \qquad (۳.۲)$$

نشان می‌دهیم، که بنا به قرارداد، σ^2 به جای σ، شناسهٔ دوم است.

شکل ۱، نمودار توزیع نرمال را نشان می‌دهد. قلهٔ $\phi_{\mu,\sigma}(x)$ به ازای $x=\mu$ به دست می‌آید. خم به ازای $|x-\mu|>\sigma$ سریعاً به محور xها نزدیک می‌شود. بیشترین مقدار احتمال، ۹۹٫۷٪ ، احتمال قرارگرفتن x در بازهٔ $(\mu-3\sigma,\mu+3\sigma)$ است. می‌توانیم $x\sim\mathcal{N}(\mu,\sigma^2)$ را به صورت $x=\mu+\epsilon$ بنویسیم، که در آن $\epsilon\sim\mathcal{N}(0,\sigma^2)$؛ اضافه‌کردن مقدار ثابت μ، صرفاً $\epsilon\sim\mathcal{N}(0,\sigma^2)$ را، μ واحد انتقال می‌دهد.

پارامتر μ، «میانگین» یا «امید» کمیت تصادفی x است. اگر از نماد «E» برای نمایش امید استفاده کنیم[1]، نتیجه می‌شود که

$$\mu=E\{x\}\equiv\int_{-\infty}^{+\infty}x\phi_{\mu,\sigma}(x)dx \qquad (۴.۲)$$

ممکن است خواننده مایل باشد $E\{g(x)\}$ را برای تابع دلخواه $g(x)$ صرفاً به‌عنوان نماد دیگری از انتگرال $g(x)$ نسبت به $\phi_{\mu,\sigma}(x)dx$ در نظر بگیرد

$$E\{g(x)\}\equiv\int_{-\infty}^{+\infty}g(x)\phi_{\mu,\sigma}(x)dx \qquad (۵.۲)$$

از لحاظ شهودی، $E\{g(x)\}$، متوسط موزون مقادیر ممکن $g(x)$ است، که برحسب احتمالهای $\phi_{\mu,\sigma}(x)dx$ برای بازه‌های بینهایت‌کوچک $[x,x+dx]$، وزندار شده‌اند. به عبارت دیگر، $E\{g(x)\}$، متوسط نظری یک جامعهٔ نامتناهی از مقادیر $g(x)$ است، که x به نسبت $\phi_{\mu,\sigma}(x)$ رخ می‌دهد.

بنابر تقارن، به آسانی دیده می‌شود که μ درواقع متوسط نظری خود x است وقتی که $x\sim\mathcal{N}(\mu,\sigma^2)$. محاسبه‌ای مشکلتر (هر چند برای کسانی که با تابع گاما کار کرده‌اند آسان است) امید $g(x)=(x-\mu)^2$ را به دست می‌دهد

$$E\{(x-\mu)^2\}=\int_{-\infty}^{+\infty}(x-\mu)^2\phi_{\mu,\sigma}(x)dx=\sigma^2 \qquad (۶.۲)$$

پارامتر σ، که «انحراف معیار» نامیده می‌شود، تغییرپذیری x حول مقدار مرکزی μ را، به‌صورتی که شکل ۱ نشان می‌دهد، مقیاس‌بندی می‌کند. متغیر تصادفی $\mathcal{N}(1,10^{-6})$، تحت امتحانهای مکرر، تقریباً تغییرپذیری محسوسی نخواهد داشت، از ۱۰۰۰ تکرار، ۹۹۷ تا در $[0.997, 1.003]$ رخ می‌دهند، زیرا $\sigma=10^{-3}$. یک متغیر تصادفی $\mathcal{N}(1,10^{6})$، اگر به زبان گویای نظریهٔ مخابرات صحبت کنیم، تقریباً تماماً نوفه است و نه سیگنال.

توزیع نرمال، ویژگی بسیار سودمندی دارد که باعث می‌شود کار کردن با تعداد زیادی از مشاهدات، به اندازهٔ کارکردن با یک مشاهدهٔ تک آسان باشد. فرض کنید x_1، x_2، x_3، ...، x_n، n مشاهدهٔ مستقل باشند که هر یک $\mathcal{N}(\mu,\sigma^2)$ است، μ و σ برای همهٔ n تکرار یکی هستند. استقلال بدین معناست که مثلاً مقدار x_1 بر مقادیر دیگر اثری ندارد: مشاهدهٔ $x_1>\mu$، احتمال ۳۴٪ رخداد $x_2\in[\mu,\mu+\sigma]$ و غیره را افزایش یا کاهش نمی‌دهد. یک مثال آشنای (غیرنرمال) متغیرهای مستقل x_1، x_2، ...، مشاهدات متوالی ریختن تاس همگن است.

فرض کنید

$$\bar{x}\equiv\sum_{i=1}^{n}x_i/n \qquad (۷.۲)$$

متوسط مشاهده‌شدهٔ n متغیر مستقل $\mathcal{N}(\mu,\sigma^2)$ باشد. به آسانی می‌توان نشان داد که

$$\bar{x}\sim\mathcal{N}(\mu,\sigma^2/n) \qquad (۸.۲)$$

توزیع $\bar{x}$ نظیر توزیع x_iی تنهاست، بجز آنکه پارامتر مقیاس آن از σ به $\sigma/\sqrt{n}$ تقلیل یافته است. می‌توانیم، با اختیار nای به‌قدر کافی بزرگ، تغییرپذیری $\bar{x}$ حول μ را به سطحی به دلخواه کوچک، تقلیل دهیم، ولی البته در مسائل واقعی، n محدود است و $\bar{x}$ دارای مؤلفه‌ای غیرقابل تقلیل از تغییرپذیری تصادفی است.

در همهٔ مثالهای ما، σ برای آماردان معلوم فرض خواهد شد. پارامتر مجهول μ، پارامتر مورد توجه است، و هدف، به دست‌آوردن استنباطهایی دربارهٔ مقدار μ بر پایهٔ داده‌های x_1، x_2، ...، x_n است. سر رانلد فیشر در ۱۹۲۵ به این نتیجهٔ اساسی رسید که در این وضعیت، میانگین $\bar{x}$ حاوی همهٔ اطلاعات ممکن دربارهٔ μ است. برای هر مسألهٔ استنباط دربارهٔ μ، دانستن

۱. حرف E اول کلمهٔ Expectation به معنی امید یا میانگین است. ـ م

$\bar{x}$ دقیقاً به اندازهٔ آگاهی از مجموعهٔ همه داده‌های x_1، x_2، ...، x_n مفید است. به زبان امروزی، $\bar{x}$ «آمارهٔ بسنده» برای پارامتر مجهول μ است.

نشان دادن بسندگی در این مورد خاص بسیار آسان است. اگر مقدار مشاهده‌شدهٔ $\bar{x}$ معلوم باشد، محاسبهٔ متعارف احتمال نشان می‌دهد که کمیتهای تصادفی $x_1 - \bar{x}$، $x_2 - \bar{x}$، $x_3 - \bar{x}$، ...، $x_n - \bar{x}$ توزیع توأمی دارند که به هیچ‌وجه به پارامتر مجهول μ بستگی ندارد. به‌عبارت دیگر، آنچه بعد از آگاهی آمار دادن از $\bar{x}$، در داده‌ها باقی می‌ماند حاوی اطلاعی دربارهٔ μ نیست. (این اصلِ ظاهراً ساده از چشم گاوس و لاپلاس دور مانده بود.)

۳. برآورد فراوانی‌گرایانه از میانگین

آماردان ممکن است بخواهد پارامتر مشاهده‌ناپذیر μ را بر پایهٔ داده‌های مشاهده‌شدهٔ x_1، x_2، ...، x_n برآورد کند. «برآورد معمولاً به معنای مطرح کردن حدس $\hat{\mu}(x_1, x_2, x_3, \ldots, x_n)$ وابسته به x_1، x_2، ...، x_n است، با در نظر گرفتن اینکه میزان تاوانی که برای حدس خود خواهید پرداخت تابع صعودی هموارى از خطای برآورد $|\hat{\mu} - \mu|$ است». تابع تاوان معمول، که ما نیز در اینجا به کار می‌بریم، $(\hat{\mu} - \mu)^2$، یعنی تابع زیان توان دوم خطاست که ابتدا گاوس آن را پیشنهاد کرد.

اصل بسندگی فیشر می‌گوید که ما فقط نیاز به در نظر گرفتن قاعده‌های برآوردی داریم که تابع $\bar{x}$ هستند. بدیهی‌ترین نامزد، خود $\bar{x}$ است

$$\hat{\mu}(x_1, x_2, \ldots, x_n) = \bar{x} \qquad (۱.۳)$$

این قاعدهٔ برآورد، برای μ «نااریب» است: مقدار واقعی μ هر چه باشد،

$$E\bar{x} = \mu \qquad (۲.۳)$$

نااریب بودن، همان‌طور که کمی بعد خواهیم دید به هیچ وجه شرطی لازم برای خوب بودن قاعدهٔ برآورد نیست، امّا این تضمین را می‌دهد که آماردان سعی نکرده فرایند برآورد را به سمت مقداری خاص از μ میل دهد و این تضمین از لحاظ شهودی جاذبه‌ای زیاد دارد.

در صورت استفاده از $\hat{\mu} = \bar{x}$، امید تاوان، بنابر (۶.۲) و (۸.۲) عبارت است از

$$E(\hat{\mu} - \mu)^2 = \sigma^2/n \qquad (۳.۳)$$

گاوس نشان داد که بین همهٔ قاعده‌های برآورد نااریب $\hat{\mu}(x_1, x_2, \ldots, x_n)$ که برحسب x_1، x_2، x_3، ...، x_n خطی‌اند، قاعدهٔ $\hat{\mu} = \bar{x}$ به صورت یکنواخت، به ازای هر مقدار μ، مقدار $E(\hat{\mu} - \mu)^2$ را مینیمم می‌کند. در اوایل دههٔ ۱۹۴۰، این حکم تعمیم یافت تا هر برآوردکنندهٔ نااریب خطی یا غیرخطی را شامل شود. برهان این مطلب که وابسته به ایده‌هایی است که فیشر در دههٔ ۱۹۲۰ ارائه کرد، جداگانه به وسیلهٔ کرامر[1] در سوئد و رائو[2] در هند عرضه شد.

اگر موافق تأکید بر ملاک نااریبی باشیم و زیان توان دوم خطا را به کار بریم، به نظر می‌رسد که $\bar{x}$ بهترین برآوردکنندهٔ μ است. برای آماردان مفید

1. H. Cramér 2. C. R. Rao

است که نه تنها یک «برآوردکنندهٔ نقطه‌ای» بلکه دامنه‌ای از مقادیر موجه برای μ را که سازگار با داده‌ها باشد تعیین کند. از (۸.۲) و شکل ۱ درمی‌یابیم که

$$\text{Prob}\{|\bar{x} - \mu| \le 2\sigma/\sqrt{n}\} = 0.95 \qquad (۴.۳)$$

که هم‌ارز حکم

$$\text{Prob}\{\bar{x} - 2\sigma/\sqrt{n} \le \mu \le \bar{x} + 2\sigma/\sqrt{n}\} = 0.95 \qquad (۵.۳)$$

است. بازهٔ $[\bar{x} - 2\sigma/\sqrt{n}, \bar{x} + 2\sigma/\sqrt{n}]$ را «بازهٔ اطمینان ۹۵٪» برای μ می‌نامند. نظریهٔ بازه‌های اطمینان را نیمن[1] در اوایل دههٔ ۱۹۳۰ مطرح کرد. به‌عنوان مثال، با فرض $n = 4$، $\sigma = 1$، داده‌های $x_1 = 1.2$، $x_2 = 0.3$، $x_3 = 0.7$، $x_4 = 0.2$ را به دست آورده‌ایم. در این صورت $\bar{x} = 0.6$ و بازهٔ اطمینان ۹۵٪ برای μ بازهٔ $[-0.4, 1.6]$ است.

همهٔ اینها آن‌قدر بی‌ضرر و سرراست به نظر می‌رسند که خواننده ممکن است تعجب کند که پس مجادله بر سر چیست. واقعیت این است که تمام نتیجه‌هایی که تا اینجا ارائه شدند از لحاظ ماهیت، «فراوانی‌گرایانه» هستند. یعنی وابسته‌اند به متوسطهای نظری نسبت به توزیع $\mathcal{N}(\mu, \sigma^2/n)$ متغیر $\bar{x}$، با μیی که فرض می‌شود یک مقدار واقعی ثابت دارد. نااریبی هم مفهومی فراوانی‌گرایانه است؛ متوسط نظری $\hat{\mu}$، یعنی $E(\hat{\mu})$، با فرض ثابت بودن μ، برابر با μ است. نتایج (۳.۳) و (۵.۳)، و قضیهٔ کرامر-رائو، احکامی فراوانی‌گرایانه هستند. مثلاً، تعبیر مناسب (۵.۳) آن است که بازهٔ $[\bar{x} - 2\sigma/\sqrt{n}, \bar{x} + 2\sigma/\sqrt{n}]$ مقدار واقعی μ را، در دنباله‌ای طولانی از تکرارهای مستقل $\bar{x} \sim \mathcal{N}(\mu, \sigma^2/n)$، با فراوانی ۹۵٪ دربرمی‌گیرد.

کسی شک ندارد که این نتایج درست‌اند. سؤالی که بیزگراها و طرفداران فیشر مطرح کرده‌اند این است که آیا متوسطهای فراوانی‌گرایانه واقعاً به فرایندی استنباطی که دانشمندان در رسیدن از داده‌های خام به مدلهای ریاضی زیربنایی به کار می‌برند مربوط‌اند یا نه. بعداً به نقطه‌نظر بیزیها برمی‌گردیم.

۴. برآورد بیزی میانگین

تا اینجا μ را ثابت، هر چند مجهول، در نظر گرفتیم. فرض کنید که μ خود متغیری تصادفی است که می‌دانیم دارای توزیع نرمال با میانگین m و انحراف معیار s است،

$$\mu \sim \mathcal{N}(m, s^2) \qquad (۱.۴)$$

m و s ثابتهایی هستند که برای آماردان معلوم‌اند. مثلاً، اگر μ، مقدار IQ فردی باشد که به تصادف از جامعهٔ ایالات متحده انتخاب شده است، حکم (۱.۴) با $m = 100$ و $s = 15$ (تقریباً) برقرار است. حدود ۶۸٪ از IQها بین ۸۵ و ۱۱۵، حدود ۹۵٪ بین ۷۰ و ۱۳۰، و الخ، هستند. اطلاعی نظیر (۱.۴)، که به زبان بیزیها، «توزیع پیشین μ» است، طبیعت فرایند برآورد را تغییر می‌دهد.

آزمونهای استاندارد IQ طوری طرح می‌شوند که اگر فردی را که به تصادف انتخاب کرده‌ایم برای کشف مقدار خاص μی او آزمون کنیم، نمرهٔ*

1. J. Neyman

* نمادهای $\bar{x}$ برای نمرهٔ آزمون و $\sigma/\sqrt{n}$ برای انحراف معیار آن، برای هماهنگی با نمادگذاری قبلی انتخاب شده‌اند، گرچه نمره‌های IQای واقعی، عملاً متوسط n مورد مستقل آزمون نیستند. همان‌طور که در (۴.۲) بیان شد، نرمال بودن کامل، چیزی جز یک حالت ایدآلی نیست که نمره‌های آزمون واقعی، تقریب آن هستند.

کلی آزمون، یعنی $\overline{x}$، نظیر بخش ۳، یک برآوردکنندهٔ نااریب با توزیع نرمال برای μ است

$$\overline{x}|\mu \sim \mathcal{N}(\mu, \sigma^{۲}/n) \qquad (۲.۴)$$

که $\sigma/\sqrt{n}$ حدود ۷٫۵ است. می‌توانیم انتظار داشته باشیم که $\overline{x}$ در ۶۸٪ مواقع حداکثر به فاصلهٔ ۷٫۵ واحد IQ از μ باشد، و نظایر آن. نماد «$\overline{x}|\mu$» تأکید می‌کند که توزیع $\mathcal{N}(\mu, \sigma^{۲}/n)$ برای $\overline{x}$، روی مقدار خاصی که کمیت تصادفی μ اختیار کرده، شرطی است. دلیل این تعویض نماد به‌زودی روشنتر خواهد شد.

قضیهٔ بیز که آن را کشیش برجسته، تامس بیز، در حدود ۱۷۵۰ کشف کرد فرمولی ریاضی برای ترکیب (۱.۴) و (۲.۴) به منظور به دست آوردن توزیع شرطی μ به شرط معلوم بودن $\overline{x}$ است. در این حالت، فرمول بیز به دست می‌دهد

$$\mu|\overline{x} \sim \mathcal{N}(m + C(\overline{x} - m), D) \qquad (۳.۴)$$

که در آن

$$C = \frac{n/\sigma^{۲}}{۱/s^{۲} + n/\sigma^{۲}} \quad \text{و} \quad D = \frac{۱}{۱/s^{۲} + n/\sigma^{۲}} \qquad (۴.۴)$$

مثلاً اگر $\overline{x} = ۱۶۰$ (و $m = ۱۰۰$، $s = ۱۵$، $\sigma/\sqrt{n} = ۷٫۵$)، آنگاه

$$\mu|\overline{x} \sim \mathcal{N}(۱۴۸, (۶٫۷)^{۲}) \qquad (۵.۴)$$

عبارت (۵.۴) یا عبارت کلیتر (۳.۴)، «توزیع پسین μ به ازای مقدار مشاهده‌شدهٔ $\overline{x}$» است. عرضهٔ چنین حکمی در چارچوب بیزی امکان دارد، زیرا ما با این فرض شروع کردیم که خود μ متغیری تصادفی است. در چارچوب بیزی، فرایند متوسط‌گیری وارون می‌شود؛ دادهٔ $\overline{x}$ در مقدار مشاهده شده‌اش، ثابت فرض می‌شود، در حالی که این پارامتر μ است که تغییر می‌کند. مثلاً در (۵.۴) دیدیم که متوسط شرطی μ به شرط $\overline{x} = ۱۶۰$ برابر ۱۴۸ است. اگر به‌تصادف تعداد زیادی از افراد را انتخاب می‌کردیم و از هر کدام، یک آزمون IQ به عمل می‌آوردیم، و زیرمجموعه‌ای از آنها را که نمرهٔ ۱۶۰ دارند، در نظر می‌گرفتیم باید متوسط واقعی IQای این زیرمجموعه ۱۴۸ می‌بود؛ یعنی ۶۸٪ از IQهای واقعی باید در بازهٔ [۱۴۸ − ۶٫۷، ۱۴۸ + ۶٫۷] قرار می‌گرفتند.

در وضعیت بیزی چگونه باید μ را برآورد کنیم؟ استفاده از برآوردکنندهٔ $\mu^*(\overline{x})$، که امید شرطی $(\mu - \mu^*)^{۲}$ را به شرط مقدار مشاهده‌شدهٔ $\overline{x}$ مینیمم می‌کند، طبیعی به نظر می‌رسد. به آسانی از (۳.۴) نتیجه می‌شود که این «برآوردکنندهٔ بیزی» عبارت است از

$$\mu^*(\overline{x}) = m + C(\overline{x} - m) \qquad (۶.۴)$$

که میانگین توزیع پسین μ به فرض معلوم بودن $\overline{x}$ است. اگر $\overline{x}$ مشاهده‌شده برابر ۱۶۰ باشد، برآورد بیزی ۱۴۸ است و نه ۱۶۰. گرچه ما یک آزمون IQی نااریب را به کار برده‌ایم، تعداد IQهای واقعی کوچکتر از ۱۶۰ آنقدر بیشتر از IQهای بزرگتر از ۱۶۰ است که امید خطای برآورد را کاهش می‌دهد و نمرهٔ مشاهده‌شده را به سوی ۱۰۰ اریب می‌کند. شکل ۲ این وضعیت را نمایان می‌کند.

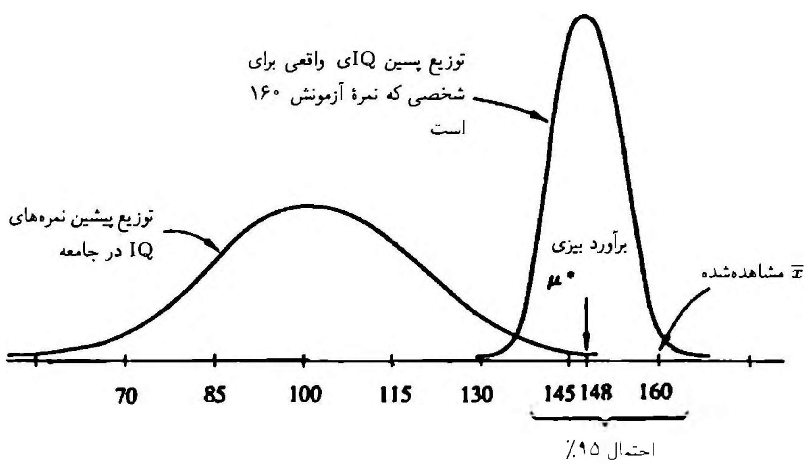
بازه‌های اطمینان، مشابه بیزی آشکاری دارند. از (۳.۴) داریم

$$\mathrm{Prob}\{\mu^*(\overline{x}) - ۲\sqrt{D} \le \mu \le \mu^*(\overline{x}) + ۲\sqrt{D}|\overline{x}\} = ۰٫۹۵ \qquad (۷.۴)$$

نماد $\mathrm{Prob}\{\cdot|\overline{x}\}$ احتمال شرطی مبتنی بر مقدار مشاهده‌شدهٔ $\overline{x}$ را نشان می‌دهد. در مثال IQ،

$$\mathrm{Prob}\{۱۳۴٫۶ \le \mu \le ۱۶۱٫۸|\overline{x} = ۱۶۰\} = ۰٫۹۵$$

هیچ‌کس (بهتر بگوییم، تقریباً هیچ‌کس) با استفاده از روشهای بیزی در وضعیتهایی نظیر مسألهٔ IQ، که برای آن، توزیع پیشین μ به روشنی تعریف



شکل ۲ نمره‌های IQ در کل جامعه دارای توزیع $\mathcal{N}(۱۰۰, (۱۵)^{۲})$ هستند. برآورد می‌شود فردی که به تصادف انتخاب‌شده و در آزمون IQای نااریب نرمال با انحراف معیار ۷٫۵ نمره دارای نمرهٔ ۱۶۰ است، IQای واقعیش برابر با ۱۴۸ است. احتمال اینکه IQای واقعی شخص در بازهٔ [۱۳۴٫۶، ۱۶۱٫۴] باشد ۹۵٪ است.

شده و کاملاً معلوم است، مخالفتی ندارد. نظریهٔ بیز، همان‌گونه که خواهیم دید، مزایای قابل توجهی از لحاظ وضوح و سازگاری دارد. این مزایا حاصل این واقعیت‌اند که متوسطهای بیزی تنها متضمن آن مقدار داده‌ای $\overline{x}$اند که عملاً دیده شده است و نه گردایه‌ای از مقادیر دیگر $\overline{x}$ که به طور نظری ممکن‌اند.

مشکلات و مجادله‌ها به این دلیل رخ می‌دهند که آماردانان بیزی می‌خواهند روشهای بیزی را وقتی به کار برند که هیچ توزیع پیشین آشکاری برای μ وجود ندارد یا حتی فراتر از آن، وقتی‌که واضح است μی مجهول یک ثابت تثبیت‌شدهٔ بدون سرشت تصادفی است. (مثلاً، وقتی μ، ثابتی فیزیکی، نظیر سرعت نور باشد که به‌صورت تجربی برآورده شده است.) امّا مسبب این نهضت بیزی، نوعی انحراف فکری نیست، بلکه پیشینهٔ کاملاً مستند ناسازگاریهای ناخوشایند در رهیافت فراوانی‌گراست.

به‌عنوان مثالی از نوع مشکلاتی که فراوانی‌گرایان با آن مواجه می‌شوند، مسألهٔ برآورد IQ را دوباره، ولی بدون فرض آگاهی از توزیع پیشین (۱.۴) برای μ در نظر بگیرید. به عبارت دیگر، فرض کنید که تنها مشاهده $\overline{x} \sim \mathcal{N}(\mu, \sigma^۲/n)$، $\sigma/\sqrt{n} = ۷٫۵$ را داریم و می‌خواهیم μ را برآورد کنیم. با مشاهدهٔ $\overline{x} = ۱۶۰$، نتایج بخش ۳ به ما می‌گویند که μ را با $\hat{\mu} = ۱۶۰$ و با بازهٔ اطمینان ۹۵ درصد $[\hat{\mu} - ۲\sigma/\sqrt{n}, \hat{\mu} + ۲\sigma/\sqrt{n}] = [۱۴۵, ۱۷۵]$ برآورد کنیم.

حال فرض کنید که شخص فراوانی‌گرا نامه‌ای از شرکتی که آزمون IQ را برگزار می‌کند دریافت کند: «در روزی که نمرهٔ $\overline{x} = ۱۶۰$ گزارش شد، ماشین نمره‌دهیِ آزمون بد عمل کرده است هر نمرهٔ $\overline{x}$ کمتر از ۱۰۰، برابر ۱۰۰ گزارش شده است. ماشین برای نمره‌های $\overline{x}$ بعد از ۱۰۰، درست عمل کرده است.»

ممکن است به نظر آید که فرد فراوانی‌گرا دلیلی برای نگرانی در این مورد ندارد، زیرا نمره‌ای که دریافت کرده است، یعنی $\overline{x} = ۱۶۰$، صحیح گزارش شده است. امّا، دلیل استفادهٔ او از $\hat{\mu} = \overline{x}$ برای برآورد μ آن است که $\overline{x}$ بهترین برآوردکنندهٔ ناارِیب است. بد عمل کردن ماشین نمره‌دهی به این معنی است که $\hat{\mu}$ حتی دیگر ناارِیب هم نیست!

اگر مقدار واقعی μ برابر ۱۰۰ باشد، بد عمل کردن ماشین، به نحوی که در نامه توصیف شده است، $E\overline{x} = ۱۰۳$ را با اریبی ۳+ امتیاز تولید می‌کند. برای بازیافت ناارِیبی، آماردان فراوانی‌گرا باید به جای قاعدهٔ برآورد $\hat{\mu} = \overline{x}$ قاعدهٔ $\hat{\mu}' = \overline{x} - \Delta(\overline{x})$ را به کار گیرد، که در آن تابع $\Delta(\overline{x})$ برای حذف اریبیِ حاصل از بد عمل کردن ماشین، انتخاب شده است.

جملهٔ تصحیح‌کنندهٔ $\Delta(\overline{x})$ برای $\overline{x} = ۱۶۰$ خیلی کوچک خواهد بود، امّا اینکه اصلاً تغییری لازم باشد ناراحت‌کننده است. نامهٔ شرکت نمره‌دهی، حاوی اطلاع جدیدی دربارهٔ نمره‌ای که عملاً گزارش شده، یا دربارهٔ IQها به طور کلی، نیست. این نامه فقط مربوط به اتفاق بدی است که ممکن بود رخ داده باشد ولی رخ نداده است. چرا باید استنباطمان را دربارهٔ مقدار واقعی μ عوض کنیم؟ روشهای بیزی از این نقص بری هستند؛ استنباطهایی که از این روشها به دست می‌آیند تنها به مقدار داده‌ای $\overline{x}$ ای که عملاً مشاهده شده است بستگی دارد، زیرا متوسطهای بیزیِ نظیر (۶.۴) و (۷.۴) روی $\overline{x}$ مشاهده‌شده شرطی هستند.

روش تحلیل آماردان بیزی، اگر اطلاعی از توزیع پیشینی مانند (۱.۴) نداشته باشد چگونه است؟ در این مورد از دو رهیافت مختلف استفاده می‌شود. شاخهٔ «ذهن‌گرای» آمار بیزی، قبل ازگردآوری داده‌ها، توزیع احتمال ذهنی آماردان را از طریق یک رشته شرط‌بندی فرضی به پارامتر مجهول μ نسبت می‌دهد. این شرط‌بندیها به صورت «آیا مایل‌اید شرطِ پولی یک به یک روی $۸۵ > \mu$ در مقابل $۸۵ \leq \mu$ ببندید؟ آیا مایل‌اید شرط دو به یک روی $۱۵۰ > \mu$ در مقابل $۱۵۰ \leq \mu$ ببندید؟ ...» هستند. کارهای ساوج[1] و دفینتی[2] نشان می‌دهند که هر فرد کاملاً منطقی باید همواره بتواند با تعمق کافی به یک توزیع پیشین یکتا (برای خودش) از μ، برسد.

رهیافت ذهن‌گرایانه، در مواردی که آماردان (البته معمولاً در همکاری با آزمایشگر) دارای نظرات پیشین هر چند مبهمی دربارهٔ مقدار واقعی μ است و سعی در به هنگام کردن آنها بر پایهٔ دادهٔ $\overline{x}$ مشاهده شده دارد می‌تواند بسیار ثمربخش باشد. چون این رهیافت ذهنی است، در مواردی که عینیت مهم است، مثلاً در انتشار نتایج علمی جدید مجادله‌انگیز، زیاد مورد استفاده قرار نمی‌گیرد.

خط دیگر تفکر بیزی، که می‌توان آن را «بیزگرایی عینی» نامید (که البته معمولاً چنین خوانده نمی‌شود) سعی دارد که در نبود شناخت پیشین، توزیعی پیشین برای μ به دست دهد که هر کسی قبول کند حاکی از نظر پیشین کاملاً بیطرفانه‌ای دربارهٔ μ است. در مسألهٔ IQ چنین توزیع پیشین «همواری» باید صورت $\mu \sim \mathcal{N}(۰, +\infty)$ را اختیار کند که به معنای $\mu \sim \mathcal{N}(۰, s^۲)$ است، که در آن $s^۲$ به بینهایت می‌گراید، از (۳.۴) و (۴.۴) به دست می‌آوریم

$$\mu|\overline{x} \sim \mathcal{N}(\overline{x}, \sigma^۲/n) \qquad (۸.۴)$$

این نتیجه جاذبهٔ زیادی دارد. برآوردکنندهٔ بیزی μ^* برابر با برآوردکنندهٔ فراوانی‌گرای $\hat{\mu} = \overline{x}$ است. بازهٔ احتمال بیزی ۹۵٪ (۷.۴) همان بازهٔ اطمینان فراوانی‌گرای ۹۵٪ (۵.۳) است. به‌علاوه، چون (۸.۴) حکمی بیزی است، نامهٔ شرکت آزمون‌کنندهٔ IQ اثری بر آن ندارد. گویی که از مزایای دو جهان فراوانی‌گرا و بیزی‌گرا برخورداریم.

تلاش فراوانی برای تدوین دیدگاه بیزی عینی انجام شده است. خود بیز این دیدگاه را مطرح کرد (امّا ظاهراً با تردید بسیار ــ مقاله او پس از مرگش با کوششهای یک دوست علاقه‌مند منتشر شد) و لاپلاس آن را بی‌قید و شرط پذیرفت. اعتبار این رهیافت در اوایل ۱۹۰۰ دستخوش تزلزل شد ولی بعدها با کار هارولد جفریز[3] تا حدّی حیات تازه یافته است. یک مشکل این است که توزیع پیشین «هموار» برای μ، ابداً توزیعی هموار برای مثلاً $\mu^۳$ نیست، و اظهار بی‌اطلاعی ما از همواری توزیع پیشین، مربوط به انتخاب تابع موردنظر از پارامتر مجهول است. دربارهٔ مشکل زیان‌بارتری در بخش ۸ بحث شده است: در مسائلی که متضمن برآورد همزمان چندین پارامتر مجهول‌اند، آنچه، توزیع پیشین کاملاً بیطرفانه‌ای به نظر می‌رسد، مستلزم مفروضاتی نامطلوب دربارهٔ پارامترها خواهد بود.

۵. برآورد فیشری میانگین

رانلد فیشر یکی از پایه‌گذاران اصلی نظریهٔ فراوانی‌گرایی بود. با این حال در تمام عمرش منتقد این نظریه بود و غالباً با حرارت زیاد از رهیافت فراوانی‌گرایی متعارف انتقاد می‌کرد. انتقادهای او از همان نوع انتقادهای بیزیها بود: چرا

1. L. J. Savage 2. B. de Finetti 3. Harold Jeffreys

اگر تعداد بینهایت مقدار $\overline{x}$ به تصادف از $\mathcal{N}(\mu,\sigma^2/n)$ با μی ثابت تولید شود باید برای استنباط دربارهٔ آنچه رخ می‌دهد به متوسطهای نظری توجه کنیم؟ ما تنها یک مقدار مشاهده‌شدهٔ $\overline{x}$ در هر مسألهٔ استنباط داریم، و فرایند استنباط باید درست بر این مقدار مشاهده‌شده متمرکز شود.

فیشر با روش بیزی هم مخالف بود شاید بدین سبب که آن نوع از مسائل تحلیل داده‌ها که او در کار ژنتیک و کشاورزی خود با آنها مواجه بود چندان برای نسبت‌دادن توزیعهای پیشین به آنها مناسب نبودند. او با قریحه و ذکاوت خاص خود صورت دیگری از استنباط را به وجود آورد، که نه بیزی بود و نه فراوانی‌گرا.

رابطهٔ $\overline{x}\sim\mathcal{N}(\mu,\sigma^2/n)$ را می‌توان به صورت زیر نوشت

$$\overline{x}=\mu+\epsilon,\qquad \epsilon\sim\mathcal{N}(0,\sigma^2/n)$$ (۱.۵)

ما مشاهدهٔ $\overline{x}$ را با اضافه‌کردن نوفهٔ نرمال $\epsilon\sim\mathcal{N}(0,\sigma^2/n)$، به میانگین μی غیرقابل مشاهده به دست‌می‌آوریم. عبارت (۱.۵) را می‌توان به صورت

$$\mu=\overline{x}-\epsilon$$ (۲.۵)

نیز نوشت. آشکار است، یا حداقل برای فیشر آشکار بود، که در وضعیتی که ما چیزی از پیش دربارهٔ μ نمی‌دانیم، مشاهدهٔ $\overline{x}$ به ما چیزی دربارهٔ ϵ نمی‌گوید. فیشر می‌گفت: در واقع اگر ما بتوانیم از روی $\overline{x}$ اطلاعی دربارهٔ ϵ به دست آوریم، آنگاه مدل (۱.۵) به صورت تنها، جنبهٔ مهمی از وضعیت آماری را نادیده می‌گیرد. ما این بحث را باز هم، به صورت ملموس‌تر، در بخش بعد پیگیری می‌کنیم.

اگر $\epsilon\sim\mathcal{N}(0,\sigma^2/n)$، آنگاه به دلیل تقارن خم زنگدیس حول نقطهٔ مرکزیش $-\epsilon\sim\mathcal{N}(0,\sigma^2/n)$. تعبیر فیشر از (۲.۵) عبارت بود از

$$\mu|\overline{x}\sim\mathcal{N}(\overline{x},\sigma^2/n)$$ (۳.۵)

این شبیه حکم بیزی عینی‌گرایانهٔ (۸.۴) به نظر می‌رسد، ولی بدون توسل به توزیع پیشین μ به‌دست آمده است. حکم بازه‌ای حاصل از (۳.۳) عبارت است از

$$\mathrm{Prob}\{\overline{x}-2\sigma/\sqrt{x}\le\mu\le\overline{x}+2\sigma/\sqrt{n}|\overline{x}\}=0{,}95$$ (۴.۵)

به زبان فیشر این یک حکم احتمالاتی «فیدوسیالی» است.

در بحث «فیدوسیالی»، تصادفی بودن، نه در دادهٔ $\overline{x}$، نظیر محاسبات فراوانی‌گراها، جایی دارد و نه در μ، نظیر محاسبات بیزیها. بلکه در مکانیسمی که μی غیرقابل مشاهده را به $\overline{x}$ مشاهده‌شده تبدیل می‌کند جای می‌گیرد. (در بحث حاضر این مکانیسم، افزودن $\epsilon\sim\mathcal{N}(0,\sigma^2/n)$ به μ است.) احکام فیدوسیالی نظیر (۴.۵) به صورت متوسطهایی روی مکانیسم تبدیل تصادفی به دست می‌آیند.

اوج شکوفایی استدلال فیدوسیالی در دههٔ ۱۹۴۰ بود و از آن زمان به بعد به تدریج از گردونه خارج شده است. اکثر آمارنادان معاصر، البته نه همه، این موضوع را یا صورتی از بیزگرایی عینی به شمار می‌آورند یا یک اشتباه می‌دانند. کاربرد استدلال فیدوسیالی در برآورد همزمان چندین پارامتر، همان‌طور که در بخش ۸ نشان خواهیم داد ممکن است مشکل بزرگی به‌بار آورد.

برای اینکه خواننده زیاد برای فیشر متأثر نشود باید گفت دو اندیشهٔ بدیع دیگرش در مورد متوسط‌گیری، استنباط شرطی و تصادفی‌سازی، هنوز بسیار مورد توجه‌اند و موضوعهای دو بخش آینده‌اند.

۶. استنباط شرطی

به دیدگاه فراوانی‌گرا برمی‌گردیم، امّا این بار از مفهوم «شرطی‌سازی» که فیشر در ۱۹۳۴ معرفی کرد استفاده می‌کنیم. استنباط شرطی منشأ ابهام عمدهٔ دیگری در مجموعهٔ روشهای فراوانی‌گرایانه است که عبارت است از روش انتخاب گردایه‌ای از مقادیر داده‌های نظراً ممکن که برای به دست آوردن استنباطی فراوانی‌گرایانه، متوسط آنها محاسبه می‌شود.

دوباره فرض کنید که متغیرهای نرمال مستقل $x_1, x_2, \ldots, x_n$ را داریم که هر x_i دارای توزیع $\mathcal{N}(\mu,\sigma^2)$ است، امّا قبل از انجام مشاهده، n را با انداختن سکه‌ای همگن به تصادف انتخاب کرده‌ایم

$$n=\begin{cases}10 & \text{با احتمال } \frac{1}{2}\\ 100 & \text{با احتمال } \frac{1}{2}\end{cases}$$ (۱.۶)

هنوز مایلیم μ را بر پایهٔ داده‌های $x_1, x_2, \ldots, x_n$ و n، و با σای که مثل قبل ثابتی معلوم است، برآورد کنیم.

توزیع شرطی $\overline{x}$ وقتی مقدار مشاهده‌شده n معلوم است، مثل (۸.۲) عبارت است از

$$\overline{x}|n\sim\mathcal{N}(\mu,\sigma^2/n)$$ (۲.۶)

متوسط مشاهده‌شده، $\overline{x}$، در این وضعیت، آماره‌ای بسنده نیست. ما هنوز نیاز داریم که بدانیم n برابر با ۱۰ است یا ۱۰۰. بدون این آگاهی باز برآوردکنندهٔ نااریب μ، یعنی $\hat{\mu}=\overline{x}$ را داریم، امّا انحراف معیار $\hat{\mu}$ را نمی‌دانیم.

در چنین وضعیتی امید مربع خطای $\hat{\mu}=\overline{x}$ چیست؟ اگر متوسط (۳.۳) را به ازای دو مقدار n به دست‌آوریم، داریم

$$E(\hat{\mu}-\mu)^2=\frac{1}{2}\frac{\sigma^2}{10}+\frac{1}{2}\frac{\sigma^2}{100}$$ (۳.۶)

فیشر تذکر داده که این محاسبه مضحک است. محاسبهٔ درستی $\hat{\mu}$ به شرط داشتن مقدار مشاهده‌شدهٔ واقعی n به وضوح مناسبتر است، یعنی

$$E\{(\hat{\mu}-\mu)^2|n\}=\begin{cases}\sigma^2/10 & n=10\\ \sigma^2/100 & n=100\end{cases}$$ (۴.۶)

چیز نادرستی در (۳.۶) نیست، بجز آنکه متوسط مربع خطایی که در آن محاسبه می‌شود به هر مقدار خاص n و $\overline{x}$ای که واقعاً مشاهده‌شده نامربوط

است؛ اگر $n = ۱۰۰$، آنگاه نتیجهٔ (۳.۶) در مورد درستی $\hat{\mu}$ خیلی بدبینانه است، در حالی که اگر $n = ۱۰$، خیلی خوش‌بینانه است.

اینها همه ممکن است آن‌قدر بدیهی به نظر آید که به گفتنش نیرزد. کار شگفت فیشر این بود که نشان داد دقیقاً همین وضعیت، با پیچیدگی و ظرافت بیشتر در مسأله‌های دیگر استنباط آماری نیز پیش می‌آید. ما این مطلب را با مثالی شرح می‌دهیم که متضمن برآورد میانگینهای دو توزیع نرمال مختلف، مثلاً $\mu_۱$ و $\mu_۲$ است. برآورد بر پایهٔ برآوردهای نرمال نااریب مستقل هر یک از آنهاست

$$\bar{x}_۱ \sim \mathcal{N}(\mu_۱, ۱), \quad \bar{x}_۲ \sim \mathcal{N}(\mu_۲, ۱) \qquad (۵.۶)$$

$\bar{x}_۱$ و $\bar{x}_۲$ از هم مستقل‌اند. (برای سادگی، فرض کرده‌ایم که $\sigma^۲/n$ برای هر دو برابر ۱ باشد) بردار داده‌ای دوبعدی $(\bar{x}_۱, \bar{x}_۲)$ می‌تواند هر جای صفحه باشد، اما با احتمال زیاد، چندان دورتر از بردار میانگینها، $(\mu_۱, \mu_۲)$، قرار نمی‌گیرد.

اگر اطلاع دیگری در دست نباشد، احتمالاً $(\mu_۱, \mu_۲)$ را با $(\bar{x}_۱, \bar{x}_۲)$ برآورد می‌کنیم. (لکن بخش ۸ را ببینید!). اما، اینک این فرض را اضافه می‌کنیم که می‌دانیم $(\mu_۱, \mu_۲)$ بر دایرهٔ به شعاع ۳ و به مرکز مبدأ قرار دارد.

$$(\mu_۱, \mu_۲) = ۳(\cos\theta, \sin\theta) \quad -\pi \le \theta \le \pi \qquad (۶.۶)$$

مسألهٔ آماری، همان‌طور که در شکل ۳ نشان داده‌ایم، برآورد پارامتر مجهول θ بر پایهٔ $(\bar{x}_۱, \bar{x}_۲)$ است.

فرض کنید مختصات قطبی $(\bar{x}_۱, \bar{x}_۲)$ را با

$$\hat{\theta} \equiv \arctan(\bar{x}_۲/\bar{x}_۱), \quad r \equiv \sqrt{\bar{x}_۱^۲ + \bar{x}_۲^۲} \qquad (۷.۶)$$

نشان دهیم. در این صورت $\hat{\theta}$ برآوردکنندهٔ بدیهی θ است. این برآوردکننده، نااریب است، $E\hat{\theta} = \theta$، و امید مربع خطا عبارت است از

$$E(\hat{\theta} - \theta)^۲ = ۰{,}۱۲ \qquad (۸.۶)$$

(که با انتگرالگیری عددی به دست آمده است؛ در (۸.۶) قرارداد می‌کنیم که $\hat{\theta} - \theta$ به ازای هر مقدار θ، بردی از $-\pi$ تا π دارد، بزرگترین خطای برآورد ممکن وقتی رخ می‌دهد که $(\bar{x}_۱, \bar{x}_۲)$ نسبت به $(\mu_۱, \mu_۲)$ متقاطر باشد. این قرارداد مهم نیست زیرا احتمال $\frac{\pi}{۲} < |\hat{\theta} - \theta|$ تنها ۰٫۰۰۱۴ است.)

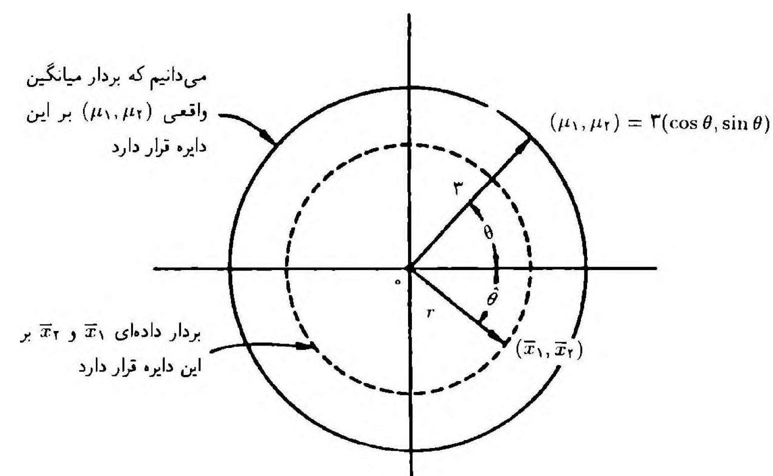
واقعیت ناآشکاری که فیشر متذکر شد آن است که r همان نقشی را دارد که n در مثالهای (۱.۶) تا (۴.۶) داشت.

(الف) توزیع r به مقدار واقعی θ بستگی ندارد. (برای خوانندگان آشنا با چگالی نرمال دو متغیره، این مطلب از تقارن دایره‌ای توزیع (۵.۶) متعلق به جفت $(\bar{x}_۱, \bar{x}_۲)$ حول $(\mu_۱, \mu_۲)$ نتیجه می‌شود.)

(ب) اگر r کوچک باشد، آنگاه $\hat{\theta}$ درستیی کمتر از آنچه (۸.۶) نشان می‌دهد دارد، در حالی که اگر r بزرگ باشد، $\hat{\theta}$ درستیی بزرگتر از آنچه (۸.۶) نشان می‌دهد دارد. جدول ۱ امید مربع خطای شرطی $E\{(\hat{\theta} - \theta)^۲|r\}$ را به صورت تابعی از r ارائه می‌کند.

در اصطلاح فیشر، r یک آمارهٔ «کمکی» است. به دلیل ویژگی (الف) این آماره مستقیماً اطلاعی دربارهٔ θ نمی‌دهد، ولی مقدارش، درستی $\hat{\theta}$ را تعیین می‌کند. اینک به نظر واضح می‌آید که باید ارزیابی خودمان از درستی $\hat{\theta}$ را روی مقدار مشاهده‌شدهٔ r شرطی کنیم. اگر $r = ۲$، نظیر شکل ۳، آنگاه $E\{(\hat{\theta} - \theta)^۲|r\} = ۰{,}۱۸$ بیشتر از امید غیرشرطی $E(\hat{\theta} - \theta)^۲ = ۰{,}۱۲$ به دقت $\hat{\theta}$ مرتبط است.

بسیاری از مسائل آماری واقعی دارای این ویژگی‌اند که برخی از مقادیر داده‌ها به وضوح آگاهی‌بخش‌تر از بقیه هستند. در این‌گونه مسائل، شرطی



جدول ۱ امید مربع خطای شرطی برآورد در مسألهٔ دایره، یعنی $E\{(\hat{\theta}-\theta)^{۲}|r\}$ به صورت تابعی از آمارهٔ کمکی $r=\sqrt{x_{۱}^{۲}+x_{۲}^{۲}}$. درستی $\hat{\theta}$ وقتی r افزایش می‌یابد اصلاح می‌شود. فیشر استدلال کرد که $E\{(\hat{\theta}-\theta)^{۲}|r\}$ معیار مناسبتری از درستی $\hat{\theta}$ است تا امید غیرشرطی $E(\hat{\theta}-\theta)^{۲}$.

r	۱٫۵	۲	۲٫۵	۳	۳٫۵	۴	۴٫۵	۵	مقدار غیرشرطی $E(\hat{\theta}-\theta)^{۲}$
$E\{(\hat{\theta}-\theta)^{۲}\|r\}$	۰٫۲۶	۰٫۱۸	۰٫۱۴	۰٫۱۲	۰٫۱۰	۰٫۰۹	۰٫۰۸	۰٫۰۷	۰٫۱۲

کردن، به طور شهودی راه صحیح حل مسأله است، امّا وضعیتهای معدودی، به وضوح ساختاری مثل مسألهٔ دایره را دارند. گاهی بیش از یک آمارهٔ کمکی وجود دارد، و یک مقدار داده، بسته به اینکه کدام آمارهٔ کمکی شرطی شود، برآوردهای درستی مختلفی را نتیجه می‌دهد. بیشتر اوقات آمارهٔ کمکی وجود ندارد، ولی آماره‌های کمکی تقریبی مختلف به ذهن القا می‌شوند. آنچه مثال دایره آشکار می‌کند آن است که حکمهای فراوانی‌گرایانه نظیر (۸.۶) ممکن است راست ولی نامربوط باشند. دیدگاه فیشر این بود که متوسط نظری $(\hat{\theta}-\theta)^{۲}$ نباید روی همهٔ مقادیر داده‌ای ممکن محاسبه شود، بلکه باید تنها روی آن مقادیری محاسبه شود که حاوی مقدار یکسانی از اطلاعات دربارهٔ θ باشند. تاکنون صورت‌بندی این حکم به صورتی رضایتبخش میسر نشده است.

بیزگرا قبول دارد که شرطی کردن عقیده‌ای درباره درستی $\hat{\theta}$ روی مقدار مشاهده‌شدهٔ r کاری صحیح است، امّا می‌پرسد که چرا فراتر نرویم و شرطی‌کردن را روی مقدار مشاهده‌شدهٔ خود $(\bar{x}_{۱},\bar{x}_{۲})$ انجام ندهیم. این کار در چارچوب فراوانی‌گرایی غیرممکن است، زیرا اگر مجموعه‌ای را که متوسط‌گیری روی آن انجام می‌شود به یک نقطهٔ داده‌ای تقلیل دهیم، برای میانگین‌گیری چیزی باقی نمی‌ماند. استنباطهای بیزی همیشه روی نقطهٔ داده‌ای واقعاً مشاهده شده شرطی هستند. در مسألهٔ دایره، توزیع پیشین هموار طبیعی، توزیع یکنواختی روی $\theta\in[-\pi,\pi]$ است. با این توزیع پیشین نتیجه می‌شود که $E\{(\theta-\hat{\theta})^{۲}|(\bar{x}_{۱},\bar{x}_{۲})\}$، همان‌طور که در جدول ۱ نشان داده‌ایم برابر با $E\{(\hat{\theta}-\theta)^{۲}|r=\sqrt{\bar{x}_{۱}^{۲}+\bar{x}_{۲}^{۲}}\}$ است، پس در این حالت خاص، دیدگاههای بیزگرای عینی و فراوانی‌گرای شرطی همانندند. (توجه کنید که در اولین امید، «θ» کمیتی تصادفی است، در حالی که در دومی، این «$\hat{\theta}$» است که تغییر می‌کند.)

۷. تصادفی‌سازی

تصادفی‌سازی هم صورت دیگری از متوسط‌گیری استنباطی است که فیشر معرفی کرده است. برای اینکه بحث دربارهٔ این موضوع ساده شود، باید مسائل آماری را از چارچوب نظریهٔ برآورد به چارچوب «آزمون فرض» منتقل کنیم. داده‌ها اینک به صورت $۲n$ مشاهدهٔ نرمال مستقل $x_{۱}, x_{۲}, \ldots, x_n, y_{۱}, y_{۲}, \ldots, y_n$ از توزیعهای

$$x_i\sim\mathcal{N}(\mu_{۱},\sigma^{۲}),\ y_i\sim\mathcal{N}(\mu_{۲},\sigma^{۲}),\ i=۱,۲,\ldots,n \qquad (۱.۷)$$

با مقدار معلوم σ و مقادیر مجهول $\mu_{۱}$ و $\mu_{۲}$ هستند. می‌خواهیم این «فرض صفر» $\mu_{۲}=\mu_{۱}$ را در برابر «فرض مقابل» $\mu_{۲}>\mu_{۱}$ که اغلب به صورت

$$(۲.۷) \qquad H:\mu_{۲}=\mu_{۱} \quad \text{در برابر} \quad A:\mu_{۲}>\mu_{۱}$$

نوشته می‌شود، آزمون کنیم. (برای مقاصد ما، $\mu_{۲}<\mu_{۱}$ غیرممکن فرض می‌شود.)

در آزمون فرض، فرض صفر H معمولاً نقش «مخالف‌خوان مصلحتی» را دارد که آزمایشگر سعی می‌کند نظر او را رد کند. مثلاً، xها ممکن است پاسخهای مربوط به یک داروی قدیم و yها پاسخهای مربوط به یک داروی جدید باشند که آزمایشگر امید به بهتر بودن آن دارد. چون انگیزهٔ قوی برای بی‌اعتبار کردن H وجود دارد، روشهای آماری محافظه‌کارانه‌ای ارائه‌شده‌اند که شواهدی در سطح نسبتاً بالا برای اعلام بی‌اعتباری H طلب می‌کنند. نظریهٔ فراوانی‌گرا که در آزمون فرض، نظریه‌ای غالب است لازم می‌داند که احتمال به غلط رد کردن H به نفع A، وقتی H درست است، همواره کمتر از سطحی پایین، معمولاً ۰٫۰۵ باشد. آزمونی که این محک را در نظر می‌گیرد آزمون در «سطح ۰٫۰۵» برای آزمون کردن H در برابر A نامیده می‌شود.

با داده‌هایی مثل (۱.۷)، به نظر طبیعی می‌رسد که $\bar{x}=\sum_{۱}^{n}x_i/n$ و $\bar{y}=\sum_{۱}^{n}y_i/n$ را حساب کنیم و H را به نفع A رد کنیم اگر

$$(۳.۷) \qquad \bar{y}-\bar{x}>c$$

ثابت c به قسمی انتخاب می‌شود که اگر H راست باشد آنگاه $\mathrm{Prob}\{\bar{y}-\bar{x}>c\}=۰٫۰۵$. محاسبات احتمالاتی متعارف نشان می‌دهد که انتخاب صحیح $c=۲٫۳۲۶\sigma/\sqrt{n}$ است. نظریهٔ آزمون اپتیمال که نیمن و پیرسن آن را در حدود ۱۹۳۰ ارائه دادند نشان می‌دهد که (۳.۷) در واقع بهترین آزمون فرض H در برابر A در سطح ۰٫۰۵ است، بدین معنا که اگر A واقعاً راست باشد آنگاه احتمال رد H به نفع A، ماکسیمم است.

xها و yهایی که مشاهده می‌کنیم درواقع اندازه‌های مربوط به نوعی از واحدهای آزمایش، مثلاً دانشجویان تازه وارد یا موشهای سفید یا مبتلایان به سردرد، هستند. فرض کنید این واحدها را با $U_{۱}, U_{۲}, \ldots, U_{۲n}$ نشان دهیم. موقعیت برای تصادفی‌سازی وقتی پیش می‌آید که آزمایشی داشته باشیم که در آن بتوانیم از قبل تصمیم بگیریم که کدام n واحد، xها و کدام n واحد، yها هستند. اگر دقیق نباشیم می‌توانیم درست اولین n واحدی را که به‌دست می‌آوریم تیمار x و n واحد آخر را تیمار y بگیریم. این آغاز

مصیبت است! اولین n فرد مبتلا به سردرد ممکن است آنهایی باشند که سردردهای شدیدی دارند، اولین n موش ممکن است آنهایی باشند که با حیوانات سنگینتر در قفس هستند و الخ. در آزمایشی که بدون دقت انجام شده است ممکن است احتمال به غلط ردکردن فرض صفر، به دلیل چنین عاملهای کنترل‌نشده‌ای، خیلی بیش از ۰٫۰۵ باشد.

فیشر در اثر بسیار مهم و مؤثرش دربارهٔ طرح آزمایش، استدلال کرد که انتخاب واحدهای آزمایش باید با تصادفی‌سازی صورت گیرد. یعنی، تخصیص n واحد به گروه تیمار x و n واحد به گروه تیمار y برای هر یک از $(2n!)/(n!)^2$ تخصیص، با احتمال یکسان انجام شود. برای اجرای فرایند تصادفی‌سازی یک ابزار مولد اعداد تصادفی به کار می‌رود.

فیشر خاطرنشان کرد که بررسیهای مبتنی بر تصادفی‌سازی احتمالاً از آن نوع اریبی‌های آزمایشی که در بالا از آن بحث شد بری هستند. مثلاً فرض کنید که نوعی «همتغییر» وجود دارد که به واحدهای آزمایش مربوط است. منظور ما از این اصطلاح، کمیتی است که تصور می‌رود بر مشاهدهٔ مربوط به هر واحد، صرف‌نظر از اینکه چه تیماری داده شده است، اثر می‌گذارد. مثلاً وزن ممکن است همتغییری مهم برای موش سفید باشد. موش سنگین وزن ممکن است کمتر از موش سبک وزن به یک محرک پاسخ دهد. اگر n به صورتی معقول بزرگ، مثلاً برابر ۱۰، باشد بسیار نامحتمل است که در آزمایشِ مبتنی بر تصادفی‌سازی، همهٔ موشهای سنگین در گروه x و همهٔ موشهای سبک در گروه y قرار گیرند. این حکم در مورد هر همتغییریِ صادق است، چه بدانیم و چه ندانیم که بر پاسخ اثر می‌گذارد، و حتی اگر از وجود آن آگاه نباشیم.

هیچ‌کدام از این مطالب ارتباطی با متوسط‌گیری ندارد. ارتباط، از پیشنهاد بعدی فیشر حاصل می‌شود: که متوسطهای نظری را روی همهٔ توزیعهای نرمال مفروض محاسبه نکنیم، بلکه روی خود فرایند تصادفی‌سازی حساب کنیم. فرض کنید که اگر همهٔ $2n$ واحد آزمایش، تیمار x را دریافت کنند، مشاهدات، X_1، X_2، ...، X_{2n} باشند، که X_i مشاهدهٔ مربوط به واحد U_i است. حروف بزرگ نشان می‌دهند که اینها مشاهدات فرضی‌اند و الزاماً داده‌های مشاهده‌شده نیستند. تحت فرض صفر H، تیمار y همان تیمار x است و بنابراین می‌توانیم فرض کنیم که همهٔ $2n$ واحد، تیمار x را دریافت می‌کنند. در این حالت، داده‌های مشاهده‌شدهٔ x_1، x_2، ...، x_n، y_1، y_2، ...، y_n با مقادیر نظری X_1، X_2، ...، X_{2n} یکی هستند. فرض کنید $\mathscr{S}(x)$ مجموعهٔ اندیسهای آن واحدهایی باشد که واقعاً به تیمار x تخصیص یافته‌اند و $\mathscr{S}(y)$، مجموعهٔ اندیسهایی باشد که به تیمار y تخصیص یافته‌اند. در این صورت اگر H درست باشد،

$$\bar{x} = \sum_{i\in\mathscr{S}(x)} X_i/n,\quad \bar{y} = \sum_{i\in\mathscr{S}(y)} X_i/n \qquad (۴.۷)$$

اگر بررسی بر اساس تصادفی‌سازی انجام شود، آنگاه $\bar{x}$ صرفاً میانگین n مقدار X است که به تصادف انتخاب شده‌اند و $\bar{y}$ متوسط بقیهٔ n تا Xهاست.

آزمون تصادفی‌سازی (یا «جایگشتی») H مشابه با (۳.۷)، به صورت زیر طرح می‌شود:

(الف) اگر داده‌های مشاهده‌شدهٔ x_1، x_2، ...، x_n، y_1، y_2، ...، y_n مفروض باشند، تعریف می‌کنیم $u_1 \equiv x_1$، $u_2 \equiv x_2$، ...، $u_{x+1} \equiv y_1$، ...، $u_{2n} \equiv y_n$. (توجه کنید که اگر H درست باشد، uها با Xهای پاراگراف قبل مطابقت دارند.)

(ب) برای هر افراز $\mathscr{P} \equiv \{\mathscr{S}_1, \mathscr{S}_2\}$ از $\{1, 2, \ldots, 2n\}$ به دو زیرمجموعهٔ مجزا با اندازهٔ n، مقدار زیر را حساب کنید

$$(\bar{y} - \bar{x})_{\mathscr{P}} \equiv \sum_{i\in\mathscr{S}_2} u_i/n - \sum_{i\in\mathscr{S}_1} u_i/n \qquad (۵.۷)$$

(پ) همهٔ $(2n!)/(n!)^2$ مقدار $(\bar{y} - \bar{x})_{\mathscr{P}}$ را به ترتیب صعودی فهرست کنید.

(ت) فرض H را به نفع A رد کنید اگر مقدار $\bar{y} - \bar{x}$ی که عملاً مشاهده شده است در قسمت ۵٪ بالایی فهرست باشد.

در آزمون تصادفی‌سازی، ۵٪ احتمال دارد که H به غلط رد شود، و این احتمال ۰٫۰۵ اشاره بر متوسطی دارد که روی همهٔ $(2n!)/(n!)^2$ تخصیص تصادفی انواع تیمارها به واحدهای آزمایش محاسبه می‌شود. آزمون، باز به صورت «H به نفع A رد می‌شود اگر $\bar{y} - \bar{x} > c$» است بجز آنکه c دیگر برابر با مقدار ثابت ۲٫۳۲۶$\sigma/\sqrt{n}$ نیست. در عوض c تابعی از مجموعهٔ مقادیر $\{u_1, u_2, \ldots, u_{2n}\}$ است که در (الف) ساخته شد. برای هر مجموعهٔ $\{u_1, u_2, \ldots, u_{2n}\}$، c به طوری انتخاب می‌شود که در (ت) صدق کند.

آزمون تصادفی‌سازی یک مزیت عمده بر آزمون (۳.۷) دارد. احتمال ۰٫۰۵ ردکردن H به غلط، تحت هر فرض صفری که می‌گوید $2n$ مقدار x و y از یک توزیع احتمال، نرمال یا غیر آن، تولید شده‌اند معتبر باقی می‌ماند. درواقع اصلاً نیازی به فرض تصادفی‌بودن مشاهدات نیست. می‌توانیم فرض صفر را تنها این بگیریم که هر واحد U_i دارای پاسخ تثبیت‌شدهٔ X_iی مربوط به آن است، بدون توجه به آنکه تیمار x یا y منظور شده باشد. این حکم اخیر مجدداً تأکید می‌کند که آزمون تصادفی‌سازی باید متضمن یک صورت غیرفراوانی‌گرایانهٔ متوسط‌گیری باشد.

تصادفی‌سازی، یا حداقل استنباط مبتنی بر تصادفی‌سازی، به نظر آماردان بیزی مردود می‌آید. بیزگرای واقعی باید شرطی کردن را روی تخصیص $\{\mathscr{S}(x), \mathscr{S}(y)\}$ی واحدها به تیمارهایی که عملاً به کار رفته است انجام دهد، زیرا این، بخشی از داده‌های موجود است، و نه متوسطی روی همهٔ افرازهای ممکن. (به نظر می‌رسد که استدلالهای فیشر دربارهٔ آمارهٔ کمکی دقیقاً اشاره‌ای در همین راستاست که مستقیماً مخالف با تصادفی‌سازی است!)

یک جنبهٔ تصادفی‌سازی، هم فراوانی‌گرایان و هم‌بیزگرایان را نگران می‌کند. فرض کنید تنها به دلیل بخت بد، در فرایند تصادفی‌سازی، همهٔ موشهای سنگین وزن به تیمار x و همهٔ موشهای سبک وزن به تیمار y تخصیص یابند. آیا باز هم می‌توانیم از آزمون تصادفی‌سازی در سطح ۰٫۰۵ برای رد H به نفع A استفاده کنیم؟ به نظر می‌رسد که جواب به وضوح منفی است، امّا مشخص کردن راه اجتناب از چنین دامهایی مشکل است. با نگرش به مطلب از راهی دیگر، فرض کنید وزنهای w_1، w_2، ...، w_{2n} موشها را قبل از شروع آزمایش بدانیم. تحت پذیره‌های معقول فراوانی‌گرا، بهترین راه یکتای $\{\mathscr{S}(x), \mathscr{S}(y)\}$ برای تخصیص موشها به تیمارها به منظور آزمون تیمار x در برابر تیمار y وجود دارد که به صورت بهینه، تخصیصهای وزن به دو

جدول ۲ احتمال آنکه $||\boldsymbol{\mu}|| \geq ||\overline{\mathbf{x}}||$، همیشه بزرگتر از ۵۰ر۰ است. برای حالت $k = ۱۰$، احتمالها به ازای مقادیر نه چندان بزرگ و نه چندان کوچک $||\boldsymbol{\mu}||$ خیلی بیش از ۵۰ر۰اند.

$\|\|\boldsymbol{\mu}\|\|^۲$	۰	۶	۱۲	۱۸	۲۴	۳۰	۴۰	۶۰
$\text{Prob}\{\|\|\overline{\mathbf{x}}\|\| > \|\|\boldsymbol{\mu}\|\|\}$	۱٫۰۰	۰٫۹۶۷	۰٫۹۰۴	۰٫۸۵۷	۰٫۸۲۲	۰٫۷۹۵	۰٫۷۶۲	۰٫۷۱۹

گروه را یکسان می‌کند. آماردانانی که در مکتب فیشری کار آموخته‌اند چنین «طرحهای آزمایش بهینه» را به دلیل حذف شدن عنصر تصادفی‌سازی، مشکل می‌پذیرند.

۸. پدیدهٔ استاین

خواننده ممکن است توجه کرده باشد که بحث و مجادله‌ای که تا اینجا شرح داده شد، بیشتر علمی بود تا عملی. همهٔ نحله‌های فلسفی توافق دارند که با عدم شناخت پیشین، بازهٔ $[\overline{x} - ۲\sigma/\sqrt{n}, \overline{x} + ۲\sigma/\sqrt{n}]$ یک بازهٔ ۹۵٪ برای μ است، عدم توافق، دربارهٔ معنای «۹۵٪» است. این وضعیت وقتی به صورت بدی درمی‌آید که برآورد همزمان پارامترهای زیادی را در نظر بگیریم.

فرض کنید که باید چند میانگین نرمال $\mu_۱$، $\mu_۲$، $\ldots$، μ_k را برآورد کنیم که برای هر یک از آنها یک برآورد نرمال ناریب مستقل را به صورت

$$\overline{x}_i \sim \mathcal{N}(\mu_i, ۱) \qquad i = ۱, ۲, \ldots, k \qquad (۱.۸)$$

مشاهده می‌کنیم. (برای راحتی، باز هم واریانس $\sigma^۲/n$ را برابر ۱ گرفته‌ایم.) وقتی چندین پارامتر برای برآوردکردن وجود دارند، مشابه طبیعی زیان مربع خطا، مربع فاصلهٔ اقلیدسی است. برای ساده کردن نمادگذاری، فرض کنید $\overline{\mathbf{x}} = (\overline{x}_۱, \overline{x}_۲, \ldots, \overline{x}_k)$ بردار متوسطهای مشاهده شده باشد، $\boldsymbol{\mu} = (\mu_۱, \mu_۲, \ldots, \mu_k)$ بردار واقعی میانگینهاست، و $\hat{\boldsymbol{\mu}} = (\hat{\mu}_۱, \hat{\mu}_۲, \ldots, \hat{\mu}_k)$ بردار برآوردها باشد. در این صورت مربع خطا، یعنی تاوان برآورد بد، عبارت است از

$$||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}||^۲ = \sum_{i=۱}^{k} (\hat{\mu}_i - \mu_i)^۲ \qquad (۲.۸)$$

قبل از ادامهٔ مسألهٔ برآورد کردن $\boldsymbol{\mu}$ بر پایهٔ $\mathbf{x}$، یک حکم مقدماتی ولی مهم را تذکر می‌دهیم. این حکم، که خوانندگان آشنا با توزیع نرمال چند متغیره می‌توانند آن را در یک سطر ثابت کنند، آن است که برای هر بردار پارامتر $\boldsymbol{\mu}$، داریم

$$\text{Prob}\{||\overline{\mathbf{x}}|| > ||\boldsymbol{\mu}||\} > ۰٫۵۰ \qquad (۳.۸)$$

یعنی، بردار داده‌ای $\overline{\mathbf{x}}$ **گرایش** به این دارد که بیشتر از بردار پارامتر $\boldsymbol{\mu}$، هر چه باشد $\boldsymbol{\mu}$، از مبدأ دور باشد. جدول ۲ نشان می‌دهد که به ازای $k = ۱۰$، برای مقادیر نه چندان بزرگ و نه چندان کوچک $||\boldsymbol{\mu}||$، احتمال خیلی بزرگتر از ۵۰ر۰ است.

فرض کنید که $k = ۱۰$، بردار داده‌ای $\overline{\mathbf{x}}$ را با مربع طول $||\mathbf{x}||^۲ = ۱۲$ مشاهده کنیم. فرض کنید که اطلاع قبلی نیز دربارهٔ $\boldsymbol{\mu}$ نداریم. با ملاحظهٔ جدول ۲، به نظر می‌رسد که احتمال $||\boldsymbol{\mu}||^۲ < ۱۲$ بسیار زیاد است. برای $||\boldsymbol{\mu}||^۲$ در بازهٔ $[۰, ۴۰]$ که اگر $||\mathbf{x}||^۲ = ۱۲$، روی دادن آن تقریباً قطعی است، بیش از ۷۵٪ مواقع داریم $||\boldsymbol{\mu}|| < ||\mathbf{x}||$. امّا، این یک «هفتاد و پنج درصدِ» فراوانی‌گرایانه است که با $\boldsymbol{\mu}$ی تثبیت‌شده و $\overline{\mathbf{x}}$ی که به تصادف مطابق (۱.۸) تغییر می‌کند حساب شده است. استدلالی مشابه استدلال بیزگرایانهٔ عینی که در بخش ۴ معرفی شد نتایج کاملاً متفاوتی به دست می‌دهد.

با فرض ناآگاهی کامل از توزیع پیشین بردار پارامتر $\boldsymbol{\mu}$، استفاده از توزیع پیشین هموار به صورت $\mu_i \sim \mathcal{N}(۰, \infty)$ (یعنی $\mu_i \sim \mathcal{N}(۰, s^۲)$ با $s^۲ \to \infty$) که μ_iها به ازای $i = ۱, ۲, \ldots, k$ مستقل‌اند، طبیعی به نظر می‌رسد. این فرض، به توزیع پسین (۴.۸) برای هر پارامتر μ_i، یعنی

$$\mu_i | \overline{x}_i \sim \mathcal{N}(\overline{x}_i, ۱) \qquad (۴.۸)$$

که به ازای $i = ۱, ۲, \ldots, k$، مستقل از هم هستند، منجر می‌شود. البته این، حکمی بیزی با $\overline{x}_i$های تثبیت‌شده در مقادیر مشاهده شده است و μ_iها به‌طور تصادفی مطابق (۴.۸) تغییر می‌کنند. تعویض نامهای کمیتهای تثبیت‌شده و تصادفی در جدول ۲، نتیجه می‌دهد که

$$\text{Prob}\{||\boldsymbol{\mu}|| > ||\overline{\mathbf{x}}|| \,\big|\, ||\overline{\mathbf{x}}||^۲ = ۱۲\} = ۰٫۹۰۴ \qquad (۵.۸)$$

اینک به نظر می‌رسد که احتمال $||\boldsymbol{\mu}|| > ||\overline{\mathbf{x}}||$ زیاد است. در واقع برای هر بردار داده‌ای مشاهده‌شدهٔ $\overline{\mathbf{x}}$

$$\text{Prob}\{||\boldsymbol{\mu}|| > ||\overline{\mathbf{x}}|| \,\big|\, \overline{\mathbf{x}}\} > ۰٫۵۰ \qquad (۶.۸)$$

استدلال فیدوسیالی فیشر در بخش ۵ هم به (۴.۸) تا (۶.۸) منجر می‌شود.

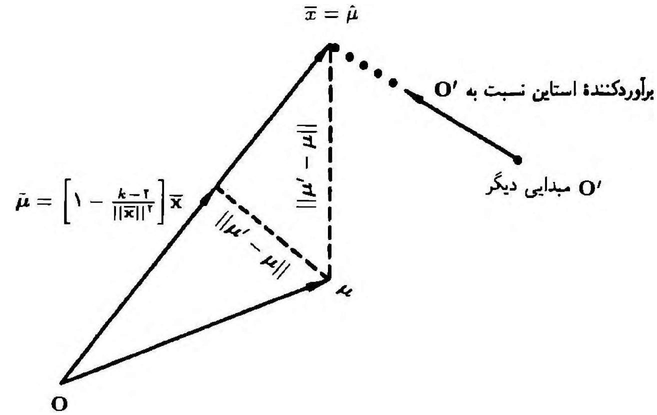
معادله‌های (۳.۸) و (۶.۸) تناقضی آشکار بین دیدگاههای فراوانی‌گرا و بیزگرا را نشان می‌دهند. کدام صحیح است؟ استدلالی بسیار شگفت‌آور و متقاعدکننده به نفع محاسبهٔ فراوانی‌گرای (۳.۸) وجود دارد. این استدلال را چارلز استاین در اواسط دههٔ ۱۹۵۰ عرضه کرده است و به برآورد $\boldsymbol{\mu}$ بر پایهٔ بردار داده‌ای $\overline{\mathbf{x}}$ (یا هم‌ارز آن، برآورد پارامترهای $\mu_۱$، $\mu_۲$، $\ldots$، μ_k بر پایهٔ $\overline{x}_۱$، $\overline{x}_۲$، $\ldots$، $\overline{x}_k$) مربوط است.

برآوردکننده‌ای بدیهی عبارت است از

$$\hat{\boldsymbol{\mu}}(\overline{\mathbf{x}}) = \overline{\mathbf{x}} \qquad (۷.۸)$$

که هر μ_i را با $\overline{x}_i$ به صورتی که در (۱.۳) آمده، برآورد می‌کند. امید ریاضی مربع زیان خطای این برآورد برای هر بردار پارامتر μ برابر است با:

$$E||\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}||^۲ = k \qquad (۸.۸)$$



شکل ۴ برآورد استاین $\tilde{\mu}$ با انقباض برآورد بدیهی $\hat{\mu} = \bar{\mathbf{x}}$ به سمت O به دست آمده است. ضریب انقباض، هر چه $||\bar{x}||$ نزدیکتر به O باشد قوی‌تر است. استاین و جیمز نشان دادند که برای هر μ، $E||\tilde{\mu} - \mu||^2 < E||\hat{\mu} - \mu||^2$. می‌توانیم هر مبدأ دیگر O′ را انتخاب کنیم و برآورد استاین دیگری به دست آوریم که بر $\hat{\mu}$ نیز ارجح باشد.

آنچه استاین نشان داد آن است که اگر k، تعداد میانگینهایی که باید برآورد شوند، بزرگتر از ۳ یا برابر با آن باشد، آنگاه برآوردکنندهٔ

$$\tilde{\mu}(\bar{\mathbf{x}}) = [1 - \frac{k-2}{||\bar{\mathbf{x}}||^2}]\bar{\mathbf{x}} \qquad (۹.۸)$$

برای هر μ دارای

$$E||\tilde{\mu} - \mu||^2 < k \qquad (۱۰.۸)$$

است! (این صورت خاص $\tilde{\mu}$ با همکاری جیمز[1] در ۱۹۶۰ به دست آمد.) از دیدگاه فراوانی‌گرا، $\tilde{\mu}$ بردار μ را به طور یکنواخت بهتر از $\hat{\mu}$ برآورد می‌کند. این برآوردکننده از دیدگاه بیزگرا هم بهتر است: با فرض هر توزیع پیشین برای μ، برآوردکردن این بردار با $\tilde{\mu}$ به جای $\hat{\mu}$، باعث می‌شود امید مربع خطای برآورد کلی کمتر باشد (در این حال، متوسط‌گیری به شرط تصادفی‌بودن μ و تصادفی‌بودن $\mathbf{x}$ انجام می‌شود).

برآوردکنندهٔ استاین مبتنی بر (۳.۸) است. چون $||\hat{\mu}|| = ||\bar{\mathbf{x}}||$ با احتمال زیاد گرایش به بزرگتربودن از $||\mu||$ دارد. یک ضریب انقباض $[1 - (k-2)/||\bar{\mathbf{x}}||^2]$ برای به دست‌آوردن برآوردکننده‌ای که به μ نزدیکتر است به کار می‌رود. ضریب انقباض، وقتی $||\bar{\mathbf{x}}||^2$ کوچک است مؤثرتر است. با $k = ۱۰$، $||\bar{\mathbf{x}}||^2 = ۱۲$، داریم $\tilde{\mu} = [۰٫۳۳۳]\bar{\mathbf{x}}$. اگر به جای آن $||\bar{\mathbf{x}}||^2 = ۸۰۰$، آنگاه $\tilde{\mu} = [۰٫۹۹]\bar{\mathbf{x}}$. شکل ۴ شمایی از این موضوع را نشان می‌دهد.

توجه کنید که مبدأ O نقشی خاص در ساختن $\tilde{\mu}$ دارد، گرچه O در بیان مسألهٔ برآورد نقشی ندارد. درواقع، می‌توانیم مبدأ را به هر نقطهٔ دیگر فضای k بعدی، مثلاً O′ ببریم، و برآورد استاین دیگری به صورت

$$\tilde{\mu}' = \mathrm{O}' + [1 - \frac{k-2}{||\bar{\mathbf{x}} - \mathrm{O}'||^2}](\bar{\mathbf{x}} - \mathrm{O}') \qquad (۱۱.۸)$$

به دست آوریم که باز هم به طور یکنواخت بهتر از $\hat{\mu}$ است.

1. W. James

نتیجهٔ استاین برای فراوانی‌گراها و بیزیها مشکلات زیادی پدید آورده است که در اینجا نمی‌توانیم به آنها بپردازیم. استلزامهای این نتیجه برای بیزگرایان عینی و فیدوسیالیها بسیار نگران‌کننده بوده است. توزیع پیشین ظاهراً همواری که به (۴.۸) منجر می‌شود ابداً هموار نیست: این توزیع، بردار پارامتر را وامی‌دارد که نسبتاً دور از هر مبدأ از قبل انتخاب‌شدهٔ O′ باشد. اگر نظریهٔ رضایت‌بخشی از استنباط بیزگرای عینی وجود داشته باشد، برآوردکنندهٔ استاین نشان می‌دهد که باید بسیار پیچیده‌تر از آنچه باشد که قبلاً انتظار می‌رفت.

دردسر مسألهٔ برآورد چند پارامتری این نیست که سخت‌تر از برآورد کردن یک پارامتر است. بلکه آسانتر هم هست، به این معنا که همزمان، با مسائل زیادی سر و کار دارد که ممکن است اطلاعاتی اضافی به دست دهند که در غیر این حالت در دسترس نیستند. دردسر، مربوط به یافتن و به‌کار بستن این اطلاعات اضافی است. مدل بیزی (۱.۴) را در نظر بگیرید. اگر تنها یک μ برای برآوردکردن داشته باشیم، پارامترهای این مدل را باید صرفاً بر اساس اعتقاد یا تجربهٔ شخصی اختیار کرد. اما، اگر چندین میانگین μ_1، μ_2، ...، μ_k را برای برآوردکردن داشته باشیم که هر یک مستقلاً از یک جامعهٔ $\mathcal{N}(m, s^2)$ استخراج شده باشد، داده‌های $\bar{x}_1$، $\bar{x}_2$، ...، $\bar{x}_k$ به ما امکان می‌دهند که m و s^2 را برآورد کنیم، به جای اینکه مقادیری برای آنها مفروض بگیریم. با قراردادن برآوردها در (۶.۴)، یک «قاعده بیز تجربی» که خیلی شبیه قاعدهٔ استاین (۱۱.۸) است به دست می‌آید. نظریهٔ بیزی تجربی که ابتدا هربرت رابینز[1] در اوایل دههٔ ۱۹۵۰ آن را مطرح کرد، نویدبخش مصالحه‌ای نسبی بین فراوانی‌گراها و بیزگراهاست.

۹. چند نکتهٔ پایانی

علی‌رغم مجادله بر سر مبانی آمار و تا حدی به سبب آن، رشد و شکوفایی آمار همچنان ادامه دارد. بدون مبالغه میلیونها تحلیل آماری در ۵۰ سال

1. Herbert Robbins

گذشته اجرا شده‌اند که این تعداد تحلیل محققاً بر قابلیت اعتماد روشهای معمول آماری، وقتی با احتیاط به کار گرفته شوند، صحه می‌گذارد. من در کار مشاوره‌ای خود همواره توان روشهای استاندارد را در تحلیل و توضیح مجموعه‌های داده‌ای بزرگ از رشته‌های علمی مختلف می‌بینم. از لحاظی، مهمترین باور، صرف‌نظر از مجادله‌های فراوانی‌گرایان و بیزیها آن است که: کم و بیش تکنیکهایی کلی برای استخراج اطلاعات از داده‌های نوفه‌ای وجود دارند که برای تقریباً هر زمینه از نیازها مناسب‌اند. به عبارت دیگر، آماردانان بر این باورند که آمار اصالتاً به‌عنوان یک نظام وجود دارد حتی اگر نتوانند در ماهیت دقیق آن هم‌نظر باشند.

چشم‌انداز آینده چیست؟ در یک همایش اخیر، دنیس لیندلی[1] از دانشگاه لندن، سخنرانیی تحت عنوان «آیندهٔ آمار ــ قرن بیست و یکم قرن بیزگرایی» ایراد کرد. احتمال ذهنی شخصی من برای وقوع این پیشامد ۱۵ر۰ است. مزیت عمدهٔ بیزگرایی ذهنی، که پروفسور لیندلی به آن اشاره کرده است، سازگاری منطقی آن است. معمولاً فیلسوفانی که مبانی استنباط علمی را بررسی می‌کنند، سرانجام از فراوانی‌گرایی روی‌گردان می‌شوند. و به استدلال بیزی روی می‌آورند.

امّا، سازگاری کافی نیست. بیزگرایی ذهنی باید با چالش عینیت علمی مواجه شود. این سنگرِ نهایی دیدگاه فراوانی‌گراست. اگر سدهٔ بیست و یکم، سدهٔ بیزی باشد، گمان من آن است که ترکیبی از بیزگرایی ذهنی، عینی و تجربی است، که پیچیدگی و تناقض آن چندان کمتر از وضع موجود نخواهد بود. پیچیدگی مسائلی که بررسی آنها را از آماردانان می‌خواهند با سرعتی ترسناک رو به افزایش است. این روزها سر و کار داشتن با مجموعه‌های داده‌ای میلیونی، و مدلهایی با هزاران پارامتر، غیرمعمول نیست. همان‌طور که از بخش ۸ برمی‌آید، این روند، احتمالاً مشکلات ایجاد یک نظریهٔ آمارِ منطقاً سازگار را بیشتر می‌کند.

مراجع

1. V. Barnett. Comparative Statistical Inference, Wiley, New York, 1973.

[شامل بحث روشنی دربارهٔ دیدگاه فراوانی‌گرا در مقایسه با روشهای بیزی.]

2. A. Birnbaum, On the foundations of statistical inference (with discussion). J. Amer. Statist. Assoc. 57 (1962) 269-326.

[بحث عمیقتری دربارهٔ مجادلات بر سر مبانی. این بحث فی‌نفسه عالی است.]

3. B. DeFinetti, Foresight: Its logical laws, its subjective sources. Studies in Subjective Probability, ed. by M. Kyburg and H. Smokler, 93-158, Wiley, New York, 1964.

[این اثر مهم را افراطی‌ترین، و (همراه با ساوج) پرنفوذترین، آماردان ذهنی‌گرای عصر ما در ۱۹۳۵ نوشته است. این جلد شامل مقاله‌هایی نیز از ون، بورال، رمزی، کوپمن، و ساوج است.]

4. B. Efron, Biased versus unbiased estimation. Advances in Math., No. 3, 16 (1975) 259-277.

[برآوردکنندهٔ استاین در نظریه و عمل.]

5. R. A. Fisher, Statistical Methods and Scientific Inference, Oliver and Boyd, London, 1956.

[آخرین رسالهٔ عمدهٔ فیشر. بحث و استدلالهای فیدوسیالی و شرطی در حد قانع‌کننده‌ای پیشرفته‌اند. ولی این مطالب را باید با تردید و احتیاط مطالعه کرد!]

6. H. Jeffreys, Theory of Probability, 3rd Edition, Clarendon Press, Oxford, 1967.

[مهمترین رسالهٔ نوین دربارهٔ بیزگرایی عینی.]

7. D. V. Lindley, Bayesian Statistics – A Review. SIAM Monographs in Applied Mathematics, SIAM, Philadelphia, (1971).

[مرجع خوبی برای دیدگاه بیزی، هم عینی و هم ذهنی.]

8 ________, The future of statistics – a Bayesian 21st century. Proccedings of the Conference on Directions for Mathematical Statistics, (1974). Special supplement to Advances in Applied Probability, September 1975.

[مقاله‌های هوبر و رابینز هم به آیندهٔ آمار مربوط‌اند.]

9. L. J. Savage, The Foundations of Statistics. Wiley, New York, 1954.

[این کتاب توجه به دیدگاه بیزی ذهنی را دوباره برانگیخت.]

******

• Bradley Efron, "Controversies in the foundations of statistics", *Amer. Math. Monthly*, (4) **85** (1978) 231-246.

* بردلی افرون، بخش آمار دانشگاه استنفرد، آمریکا

1. Dennis Lindley