فلسفهٔ آمار•

(۱)

دنیس لیندلی*
ترجمهٔ حمید پزشک

۱. مقدمه

این مقاله به جای بحث راجع به یک مسألهٔ معین، چشم‌اندازی ارائه می‌دهد که از طریق آن می‌توان غالب موضوعات آماری را بررسی کرد. در عنوان مقاله، «فلسفه» به معنی «مطالعهٔ اصول کلی شاخهٔ خاصی از دانش، تجربه، یا فعالیت» [۲۲] آمده است. کلمهٔ «فلسفه» این روزها بیشتر یادآور مفاهیم مجرد و جدا از واقعیت است. ولی قصد من در اینجا این است که از تجرید بیش از حد دوری جسته به موضوعات عملی مرتبط با آمار بپردازم. اگر آمارپیشه‌ای که این مقاله را می‌خواند آن را برای خود سودمند نیابد نوشتهٔ من به یکی از هدفهایش نرسیده است. در مقاله تلاش می‌شود روشی برای نگرش به آمار عرضه شود که ما آماردانان به کمک آن بهتر بتوانیم راه‌حل صحیح را برای مسائلی که به آنها برمی‌خوریم بیابیم. از بحثهای فنی اغلب پرهیز کرده‌ام نه به دلیل آنکه این بحثها مهم نیستند بلکه چون خواسته‌ام بیشترِ توجهمان معطوف به روشن شدن این موضوع باشد که چگونه یک مسألهٔ آماری را می‌توان بررسی کرد. در برخی جاها جزئیات را نیاورده‌ام تا ایده‌های اساسی مشخص‌تر شوند. برای مثال؛ چگالیهای احتمال بدون اشارهٔ صریحی به اندازهٔ کنترل‌کنندهٔ آنها مورد استفاده قرار گرفته‌اند.

مقاله با پذیرش این نظر آغاز می‌شود که مقولات آماری با عدم حتمیت سروکار دارند، و در ادامه نتیجه‌گیری می‌کند که عدم حتمیت را فقط می‌توان با احتمال اندازه‌گیری کرد. با این نتیجه‌گیری شرحِ نظام‌مندِ «استنباط» بر اساس حساب احتمال میسر می‌شود، که نشان خواهیم داد متفاوت با بعضی رویکردهای متداول است. اصل درستنمایی از نقشی پایه‌ای که احتمال ایفا می‌کند نتیجه می‌شود. سپس نقش تحلیل داده‌ها در ساختن الگوهای احتمال و طبیعت الگوها مورد بحث قرار می‌گیرد. این بحث به روشی برای تصمیم‌گیری می‌انجامد و طبیعت مخاطره بررسی می‌شود. روش علمی و کاربرد آن در چند مقولهٔ حقوقی در چارچوب احتمالاتی توضیح داده خواهد شد. نتیجه این است که به این ترتیب، مجموعه‌ای از روشهای عمومی و رضایت‌بخش آماری داریم که اجرای آنها باید موجبات پیشرفت کار آمار را فراهم کند.

دیدگاه ارائه‌شده در این مقاله بیشتر به ساخت الگو توجه دارد تا به استنباط صوری، و از این لحاظ با جدیدترین آرا توافق دارد. یک دلیل برای این تغییرکانون توجه، این است که استنباط صوری یک روال یا کارِ نظام‌مند در قالب حساب احتمال است. ولی ساخت الگو نمی‌تواند خیلی نظام‌مند باشد.

این مقاله حاصل تجربیات من از ششمین کنفرانس آمار بیزی است که در شهر والنسیا در ژوئن سال ۱۹۹۸ برگزار شد [۵]. اگر چه من تحت‌تأثیر کیفیت کلی مقالات عرضه‌شده و پیشرفتهای به‌دست آمده قرار گرفتم ولی به نظرم رسید که بسیاری از شرکت‌کنندگان فلسفهٔ بیزی را آنچنان که شایسته است قدر نمی‌نهند. این مقاله حاصل تلاشی است برای توضیح دیدگاه من دربارهٔ این فلسفه، و بازتاب پنجاه سال تجربهٔ آماری و تحول دیدگاه شخصی من است که از فراوانی‌گرایی به مکتب عینی‌گرای بیز و سپس به نگرش ذهنی‌گرا که در این مقاله دربارهٔ آن بحث می‌شود، روی آورده‌ام. هیچ تلاشی نکرده‌ام که فلسفه‌های دیگر را به تفصیل تحلیل کنم، فقط نشان داده‌ام که نتایج آنها در چه مواردی با نتایج ما مغایر هستند و روشهای عملی حاصل را با هم مقایسه کرده‌ام.

۲. آمار

برای بحث راجع به فلسفهٔ آمار لازم است تا حدی روشن کنیم که منظور از آمار چیست؛ البته نه آنکه تعریف دقیقی بیاوریم، بلکه در حدی که بتوانیم تصوری کلی از موضوع داشته باشیم. نظر این مقاله این است که آمار مطالعهٔ عدم حتمیت است [۲۵b] و آماردانان متخصص رسیدگی به مسألهٔ عدم حتمیت‌اند. آنان ابزارهایی نظیر خطاهای استاندارد و سطوح معنی‌دار بودن را برای اندازه‌گیری عدم حتمیت ارائه داده‌اند. برای تحقیق در این مطلب که این تعریف چقدر با کاری که ما عملاً انجام می‌دهیم مطابقت دارد،

می‌توان به چهار سری مجلهٔ منتشر شده توسط انجمن سلطنتی آمار[1] در سال ۱۹۹۷ مراجعه کرد. در این انتشارات می‌توان مقوله‌هایی گوناگون حتی راجع به حسابداری اجتماعی و قوانین حقوقی ثابت یافت. بجز چند استثنا نظیر بخش الگوریتمها (که بعداً حذف شده است) و مقاله‌ای دربارهٔ آموزش، تمام مقالات یا مستقیماً با عدم حتمیت سروکار دارند یا دربارهٔ مشخصه‌های مطرح‌شده در مسائلی هستند که به‌نحوی با عدم حتمیت مرتبط‌اند. شاهدی بر دیدگاه ما، این است که آمار نقش بسیار مهمتری در مقوله‌های شامل تغییرات، که به عدم حتمیت منجر می‌شوند، دارد تا در مقوله‌هایی که با موضوعات دقیق سروکار دارند. برای مثال، کشاورزی ارتباط نزدیکی با آمار دارد، در صورتی‌که فیزیک دارای چنین ارتباطی نیست. توجه کنید که فقط بررسی عدم حتمیت است که برای ما اهمیت دارد نه خودِ موضوعی که غیرحتمی است. مثلاً ما سازوکار بارش باران را مطالعه نمی‌کنیم بلکه علاقه‌مندیم بدانیم باران می‌بارد یا خیر. این امر، باعث شده که علم آمار در موقعیتی عجیب قرار بگیرد که ما آماردانان را به دیگران وابسته می‌کند. پیش‌بینی بارش باران هم به هواشناسان و هم به آماردانان مربوط می‌شود. ما تنها در مقام نظریه‌پرداز است که می‌توانیم بدون وابستگی به دیگران عمل کنیم. حتی در آن حالت هم چنانچه خیلی از علوم فاصله بگیریم آسیب خواهیم دید. در این مقاله لفظ «مشتری» به فردی مانند دانشمند یا حقوقدان اطلاق می‌شود که در رشتهٔ کاری خودش با عدم حتمیت مواجه است و برای حل مسائلش به آمار مراجعه می‌کند.

پس موضع فلسفی‌ای که در اینجا می‌پذیریم این است که آمار اساساً مطالعهٔ عدم حتمیت است و نقش آماردان کمک کردن به مشتریهایی از حوزه‌های دیگر است که در حوزهٔ کار خودشان با عدم حتمیت مواجه می‌شوند. در عمل این محدودیت وجود دارد که آمار غالباً با داده‌ها سروکار دارد و رابطهٔ بین عدم حتمیت، یا تغییرپذیری داده‌ها و عدم حتمیت در خود مقوله، آن چیزی است که آماردان به بررسی‌اش می‌پردازد. برخی از نویسندگان حتی داده‌ها را به داده‌های دارای فراوانی که قابل تکرار به‌صورت تقریباً یکسان باشند محدود می‌کنند. عدم حتمیت در خارج از داده‌ها، جاذبهٔ زیادی برای آماردان ندارد. مطالعهٔ عدم حتمیت منحصر به آماردانان نیست. احتمال‌دانان در این باره بحث می‌کنند که تصادفی بودنِ بخشی از یک سیستم چگونه بر دیگر بخشهای آن تأثیر می‌گذارد. بنابراین الگویی [مدلی] از یک فرایند تصادفی پیش‌بینی‌هایی در مورد داده‌هایی که آن فرایند تولید می‌کند به‌دست می‌دهد. گذر از فرایند به داده‌ها روشن است؛ مشکل زمانی پیش می‌آید که می‌خواهیم از داده‌ها به سوی فرایند برویم. این مقاله بیشتر به این مرحله، که آن را معمولاً استنباط می‌نامند، و عملی که تولید می‌کند اختصاص دارد.

توجه کنید که عدم حتمیت همه جا هست و منحصر به علوم یا حتی داده‌ها نیست. از جمله، بعضی از جنبه‌های الهیات از آن نشأت می‌گیرد [۲]. بنابراین شناسایی آمار به‌عنوان مطالعهٔ عدم حتمیت نقش گسترده‌ای را به آماردانان محول می‌کند. اگر بتوان مکتبی فلسفی پدید آورد که همهٔ عدم حتمیت‌ها را در بر بگیرد، پیشرفت مهمی در شناخت ما از جهان به‌دست خواهد آمد. ولی در حال حاضر دستیابی به چنین هدف بزرگی حتمی به نظر نمی‌رسد.

1. Royal Statistical Society

۳. عدم حتمیت

با پذیرش این نظر که آمار مطالعهٔ عدم حتمیت است، تحقیق دربارهٔ این پدیده لازم می‌آید. رهیافت علمی به عدم حتمیت به معنی اندازه‌گیری آن است زیرا همچنان‌که کلوین گفته است، فقط با انتساب عدد به مفهوم علمی است که آن مفهوم را می‌توان به درستی درک کرد. دلیل اندازه‌گیری فقط این نیست که می‌خواهیم برداشت دقیق‌تری مثلاً از این موضوع داشته باشیم که حتمیت ما در پیش‌بینی وضع بازار بورس کمتر از پیش‌بینی طلوع خورشید در روز بعد است. بلکه تلفیق عدم حتمیت‌ها هم هست. فقط در حالات استثنایی ممکن است یک مورد عدم حتمیت در یک مسأله وجود داشته باشد و در حالت کلی موارد متعددی از آن وجود دارد. هنگام جمع‌آوری داده‌ها، در واحد نمونه‌گیری عدم حتمیت وجود دارد و در عدد گزارش‌شده در نمونه‌گیری نیز عدم حتمیت هست. در یک مسألهٔ آماری نوعی هم در داده‌ها و هم در پارامتر عدم حتمیت وجود دارد. بنابراین مسألهٔ اساسی، ترکیب عدم حتمیت‌هاست. قوانین ترکیب اعداد بسیار آسان است. اعداد از دو طریق جمع و ضرب ترکیب می‌شوند و این ترکیب به غنای مفاهیم ریاضی منجر می‌شود. ما می‌خواهیم عدم حتمیت‌ها را اندازه بگیریم تا بتوانیم آنها را ترکیب کنیم. سیاستمداری گفته است که قیود را به اعداد ترجیح می‌دهد. ولی متأسفانه ترکیب قیود مشکل است.

این اندازه‌گیری چگونه میسر می‌شود؟ همهٔ اندازه‌گیریها بر مبنای مقایسه با یک استاندارد صورت می‌پذیرد. ما برای اندازه‌گیری طول از خط قرمز-نارنجی مربوط به ایزوتوپ کریپتون ۸۶ به‌عنوان استاندارد استفاده می‌کنیم. نقش اساسی مقایسه دال بر این است که هیچ مطلقی در دنیای اندازه‌گیری وجود ندارد. این نکته‌ای است که در بخش ۱۱ به آن برمی‌گردیم. بنابراین لازم است استانداردی برای عدم حتمیت بیابیم. استانداردهای گوناگونی پیشنهاد شده است اما ساده‌ترین آنها یعنی بازیهای شانسی از نظر تاریخی اولین پیشنهاد است. این بازیها متضمن نخستین مواردی از عدم حتمیت هستند که به‌طور نظام‌مند بررسی شده‌اند. بنابراین بگذارید یک بازی ساده را به‌عنوان استاندارد به‌کار ببریم.

فرض کنید کیسه‌ای شامل N توپ کوچک است که آنقدر که مهندسیِ مدرن توانسته، آنها را شبیه به هم ساخته است. تصور کنید یک توپ به‌طور تصادفی از کیسه بیرون آورده می‌شود. برای اینکه مفهوم این جمله روشن باشد باید معنی «تصادفی» مشخص شود. فرض کنید که توپها از ۱ تا N شماره‌گذاری شده باشند و همچنین فرض کنید، بدون آنکه هزینه‌ای برای شما منظور شود، چنانچه توپ شمارهٔ ۵۷ را بیرون بیاورید جایزه‌ای به شما تعلق گیرد. اکنون این حالت را در نظر بگیرید که همان جایزه به شما تعلق می‌گیرد اگر شمارهٔ ۱۲ را بیرون بیاورید. اگر این دو حالت یا به‌طور کلی‌تر هر شمارهٔ بین ۱ تا N برای شما تفاوت نداشته باشد، آنگاه توپ به‌طور تصادفی بیرون آورده شده است. توجه کنید که تعریف تصادفی بودن در اینجا ذهنی و وابسته به شماست. آنچه برای یک نفر تصادفی است ممکن است برای شخص دیگری تصادفی نباشد. ما به این مقوله در بخش ۸ برخواهیم گشت.

حال که مفهوم بیرون کشیدن تصادفی توپ از کیسه مشخص شد، اعداد روی توپها را فراموش کنید و فرض کنید R تا از این N توپ قرمز و بقیه

سفیدند و رنگ توپها تأثیری در انتخاب شما ندارد. این پیشامد غیرحتمی را در نظر بگیرید که توپی که به تصادف بیرون کشیده می‌شود قرمز باشد. عقیده بر این است که به این ترتیب استانداردی برای عدم حتمیت به‌دست می‌آید و اندازهٔ آن R/N یعنی نسبت تعداد توپهای قرمز به کل توپهای داخل کیسه است. در اینجا هیچ نکتهٔ عمیقی وجود ندارد و این فقط شکلی از یک فرض است که بازیهای شانسی مبتنی بر آنها هستند. حال این را به پیشامد یا گزاره‌ای که می‌تواند رخ بدهد یا ندهد و یا درست باشد یا نباشد، تعمیم می‌دهیم. پیشنهاد می‌شود که عدم حتمیت وقوع یک پیشامد را از طریق مقایسه با استاندارد اندازه‌گیری کنید. مثلاً اگر فکر می‌کنید عدم حتمیت یک پیشامد معادل است با درآوردنِ تصادفیِ یک توپ قرمز از کیسه‌ای شامل N توپ که Rتای آن قرمزند، آنگاه پیشامد مورد نظر دارای عدم قطعیتی برابر با R/N برای شماست. انتخاب R و N به شما بستگی دارد. به‌ازای N داده شده، به آسانی می‌توان نشان داد که بیش از یک چنین Rای وجود ندارد. حال شاخصی از عدم حتمیت برای هر پیشامد در دست است. قبل از آنکه جلوتر برویم بگذارید فرایند اندازه‌گیری را به دقت بررسی کنیم.

در اینجا یک فرض جدی اتخاذ شده مبنی بر اینکه مفهوم عدم حتمیت را می‌توان از جنبه‌های دیگر جدا کرد. در بحث راجع به تصادفی بودن، مفید بود که یک جایزه را تحت شرایط مختلف، مثلاً بیرون آوردن توپ ۵۷ یا توپ ۱۲، مقایسه کنیم. اما لزوماً میزان جوایز نمی‌تواند در مقایسهٔ یک پیشامد با یک استاندارد مؤثر باشد. به‌عنوان مثال فرض کنید پیشامدی که عدم قطعیتش مورد مطالعه است این باشد که سال بعد قرار است یک بمب اتمی در فاصلهٔ ۵۰ مایلی شما منفجر شود. در این‌صورت جایزهٔ ۱۰۰۰۰ دلاری برای درآوردن یک توپ قرمز و برای گریز از تشعشعات اتمی دارای ارزش کاملاً متفاوتی خواهد بود. فرایند اندازه‌گیری که توضیح داده شد مبتنی بر این فرض است که شما می‌توانید عدم حتمیت انفجار هسته‌ای را از پیامدهای ناخوشایند آن جدا کنید. رمزی [۲۴] که کار او در بخش ۴ مورد بحث قرار خواهد گرفت، مفهومی به نام «پیشامدِ اخلاقاً خنثی» را معرفی کرد که در مورد آن، مقایسه با مسألهٔ کیسهٔ توپها مشکلات کمتری ایجاد می‌کند. بمب اتمی از نظر اخلاقی خنثی نیست.

حال به فرضی که اتخاذ نشده است توجه کنید. برای هیچ پیشامدی، از جمله انفجار بمب اتمی، فرض نشده است که شما توانایی اندازه‌گیری و تعیین R را دارید (و همچنین تعیین N را که دقت شما را در اندازه‌گیریِ عدم حتمیت منعکس می‌کند). بلکه فرض می‌کنیم که دوست دارید این کار را انجام دهید اگر بدانید چگونه انجام می‌شود. چیزی که در مورد هر فرایند اندازه‌گیری فرض می‌شود این است که معقول باشد و نه آنکه به‌سادگی انجام شود. اگر نمی‌دانیم چطور فاصلهٔ خود را تا ماه اندازه بگیریم به این معنی نیست که باور نداریم فاصله‌ای تا آن وجود دارد. دانشمندان دقت فراوان صرف اندازه‌گیری دقیق طول کرده‌اند زیرا مفهوم فاصله را با توجه به نور کریپتون باور دارند. به‌طور مشابه، معقول به نظر می‌رسد که سعی کنیم عدم حتمیت را اندازه بگیریم.

۴. عدم حتمیت و احتمال

گفتیم که یک دلیل اصلی اندازه‌گیری عدم حتمیت‌ها، این است که بتوانیم آنها را با هم ترکیب کنیم. بنابراین بگذارید ببینیم که روش پیشنهادشده چگونه ما را به این هدف می‌رساند. فرض کنید از N توپ داخل کیسه، R تا قرمز، B تا آبی و بقیه سفید هستند. در این‌صورت عدم حتمیتِ مربوط به استخراج یک توپ رنگی از کیسه عبارت است از $(R+B)/N = R/N + B/N$ یعنی برابر با مجموع عدم حتمیت‌های مربوط به قرمز و مربوط به آبی است. همین نتیجه برای هر دو پیشامد ناسازگار که عدم حتمیت‌هایشان به ترتیب R/N و B/N باشد به‌دست می‌آید و ما قانون جمع را برای عدم حتمیت‌های پیشامدهای ناسازگار داریم.

حال فرض کنید R توپ قرمز و $N-R$ توپ سفید درکیسه است و Sتا از N توپ خالدار و $N-S$تا ساده‌اند. کیسه حالا دارای چهار نوع توپ است که تعدادی از آنها، مثلاً Tتا، قرمز و خالدارند. بنابراین عدم حتمیت مربوط به استخراج توپ قرمز خالدار T/N می‌باشد که برابر است با $R/N \times T/R$ یعنی حاصلضرب عدم حتمیت یک توپ قرمز در عدم حتمیت یک توپ خالدار در میان قرمزها. باز همین نتیجه برای هر دو پیشامد که با توپهای رنگی و توپهای خالدار مقایسه شوند، به‌دست می‌آید و ما قانون ضرب را برای عدم حتمیت‌ها داریم.

این قوانین ضرب و جمع به همراه قانون تحدب که می‌گوید اندازهٔ R/N همیشه داخل بازهٔ (محدب) به طول واحد قرار می‌گیرد، قوانینی هستند که احتمال را، لااقل برای تعداد متناهی از پیشامدها، تعریف می‌کنند (بخش ۵ را ببینید). بنابراین نتیجه می‌گیریم که اندازه‌های عدم حتمیت را می‌توان با استفاده از حساب احتمال بیان کرد. در جهت عکس، قوانین احتمال به سادگی به قوانین مربوط به نسبتها تحویل می‌یابد. با توجه به این مطلب می‌توان دریافت که چرا روشهای فراوانی‌گرا غالباً بسیار مفید هستند: تلفیق عدم حتمیت‌ها را می‌توان بر اساس نسبتها یا فراوانیها در یک گروه، در اینجا گروه توپها، مطالعه کرد. پایهٔ ریاضی احتمال بسیار ساده است و شاید تعجب‌آور باشد که این پایهٔ ساده منتج به نتایج مفید و پیچیده می‌شود.

نتیجه می‌گیریم که آماردانان با عدم حتمیت سروکار دارند و عدم حتمیت را می‌توان به‌وسیلهٔ احتمال اندازه‌گیری کرد. ممکن است تصور شود که توضیح مجملی که در اینجا داده شد برای چنین نتیجه‌گیری مهمی کفایت نمی‌کند، لذا بیایید به دیدگاههای دیگر نیز نظری بیندازیم.

از لحاظ تاریخی، عدم حتمیت مرتبط با بازیهای شانسی و قمار بوده است. بنابراین یک راه اندازه‌گیری عدم حتمیت از طریق قمارهایی است که به عدم حتمیت وابسته‌اند. شرط‌بندی به مبلغ s بر روی وقوع پیشامدی خاص برای بردی به مبلغ w در واقع شاخصی از عدم حتمیت وقوع آن پیشامد را برحسب بخت [نسبت] w/s به ۱ به‌دست می‌دهد. حال به جای ترکیبی از پیشامدها خانواده‌ای از قمارها داریم. برای مشخص‌تر شدن این مفهوم، مسابقات اسبدوانی را در نظر بگیرید. در مورد دو اسب که در یک مسابقه شرکت دارند، هم شرط‌بندی جداگانه روی آنها و هم شرط‌بندی روی این پیشامد که حداقل یکی از آنها مسابقه را ببرد، قابل بررسی است؛ و اگر دو اسب در مسابقه‌های مختلف شرکت کنند می‌توان روی این پیشامد شرط‌بندی کرد که هر دو اسب مسابقه‌هایشان را ببرند. با استفاده از این‌گونه ترکیبات می‌توان چیزی را که در انگلستان «دفترچهٔ هلندی»[1] خوانده می‌شود

1. Dutch book

به‌کار گرفت. می‌گویند دنباله‌ای از شرط‌بندیها با بختها [نسبتها]ی مشخص تشکیل یک دفترچهٔ هلندی می‌دهد هرگاه بتوان مجموعه‌ای از سرمایه‌گذاریها را چنان انجام داد که در کل، سرمایهٔ گذاشته شده افزایش یابد. مسؤول شرط‌بندی هرگز بختهایی را که بر اساس آنها می‌توان دفترچهٔ هلندی ساخت فاش نمی‌کند. به آسانی می‌شود نشان داد [۱۶] که اجتناب از دفترچهٔ هلندی معادل است با احتمالات مستخرج از بختها، که در قوانین تحدب، جمع، و ضرب احتمالها که در بالا گفته شد تبعیت می‌کنند. برای بخت ۰، احتمال متناظر برابر است با $p = (۱ + ۰)^{-۱}$. به‌طور خلاصه، استفاده از بخت به همراه اجتناب از دفترچهٔ هلندی ما را مانند قبل به احتمال رهنمون می‌سازد.

دوفینتی روش دیگری را نیز معرفی کرده است. فرض کنید از شما خواسته شده است که میزان عددی عدم حتمیت یک پیشامد را از نظر خودتان مشخص کنید و به شما گفته شده که امتیازی به اندازهٔ $S(x, E)$ می‌گیرید هرگاه این میزان را x بگیرید؛ $E = ۱$ و $E = ۰$ به ترتیب متناظر با وقوع و عدم وقوع پیشامدند. برای پیشامدهای متفاوت امتیازها جمع می‌شوند. جریمه‌ای که دوفینتی به‌کار گرفت عبارت است از $S(x, E) = (x - E)^۲$، ولی بجز چند تابع خاص که منجر به نتایج مضحکی می‌شوند هر تابع دیگری را نیز می‌توان تابع جریمه در نظر گرفت. بنابراین اگر شما x را به‌عنوان تابعی از احتمال به‌کار بگیرید یعنی تابعی از چیزی که از سه قانون احتمال تبعیت می‌کند، مطمئن خواهید بود که از شخص دیگری که روش دیگری به‌کار می‌گیرد بهتر عمل می‌کنید به این معنی که جریمهٔ کمتری می‌گیرید. روش دوفینتی روش جذابی است زیرا عملاً می‌توان آن را به‌کار برد و برای تعلیم کسانی که وضع هوا را پیش‌بینی می‌کنند مورد استفاده قرار گرفته است. این روش معمولاً محکی تجربی برای کیفیت برآوردهای احتمالاتی شما، و در نتیجه آزمونی از توانایی شما به‌عنوان آماردان، به‌دست می‌دهد (بخش ۹ را ببینید). نکتهٔ مهم این است که این روش هم به احتمال ختم می‌شود. بعداً گروه سومی از روشها را مطالعه می‌کنیم که در آنها هم استفاده از احتمال برای توصیف عدم حتمیت الزامی است.

انجمن سلطنتی آمار انگلستان و بسیاری دیگر از گروههای آماری در آغاز به‌منظور جمع‌آوری و انتشار داده‌ها تأسیس شدند. این هدف با بحث ما در مورد عدم حتمیت توافق دارد زیرا یکی از انگیزه‌های جمع‌آوری داده‌ها، کاهش عدم حتمیت است. امروز هم بخش مهمی از فعالیتهای آماری به این منظور انجام می‌شود. اغلب دولتها دارای اداراتی هستند که وظیفهٔ اصلی آنها تهیه و ارائهٔ آمار است. خیلی طول نکشید که آماردانان به این فکر افتادند که چگونه از داده‌ها می‌توان به بهترین وجه استفاده کرد، و استنباط آماری نوین پا به عرصهٔ وجود نهاد. آنها شاهد بودند که استنباط آماری به‌عنوان مبنای تصمیمها و عملها به‌کار می‌رود و همراه دیگران شروع به بررسی مسألهٔ تصمیم‌گیری کردند. حال توضیح می‌دهیم که چگونه این موضوع هم به احتمال منتهی می‌شود.

به نظر می‌آید کسی که اولین قدم را به سمت تصمیم‌گیری برداشت رمزی [۲۴] بود. وی سؤال ساده‌ای را مطرح کرد: «وقتی با عدم حتمیت مواجه‌ایم چگونه تصمیم بگیریم؟» و چند فرض منطقی اتخاذ کرد و از آنها قضایایی را استنتاج نمود. قضیه‌ای که الان منظور ماست آن است که می‌گوید شاخصهای عدم حتمیت باید از قوانین حساب احتمال تبعیت کنند. بنابراین باز هم به احتمال برمی‌گردیم. مقالهٔ رمزی چندان مورد توجه قرار نگرفت تا آنکه ساویج [۲۵] همان سؤال را مطرح کرد و با فرضهایی متفاوت با فرضهای رمزی ولی مرتبط با آنها به همان نتیجه یعنی به احتمال رسید. ساویج همچنین ارتباطی بین این موضوع و ایده‌های دوفینتی برقرار کرد. از آن زمان تا کنون کسان دیگری نیز در این حوزه کندوکاو کرده و نتایج مشابهی به‌دست آورده‌اند. متن محبوب من در این زمینه، در میان متنهایی که از دقت کامل برخوردارند، فصل ششم [۱۱] است. اثر برناردو و اسمیت [۷] هم شرح عالی جدیدی از این موضوع است.

۵. احتمال

نتیجهٔ بحث بالا این است که شاخصهای عدم حتمیت باید از قوانین حساب احتمال تبعیت کند. قواعد دیگر نظیر قوانین منطق فازی یا نظریهٔ امکان[1] که به مینیمم و ماکسیمم وابسته‌اند نه به مجموع و حاصلضرب، و همچنین چند قانون دیگر که آماردانان از آنها استفاده می‌کنند، کنار گذاشته می‌شوند. بخش ۶ را ببینید. تمام این استنتاجات، چه بر پایهٔ توپهای داخل کیسه‌ها، چه بر پایهٔ قمار و چه بر پایهٔ قوانین امتیازبندی و یا تصمیم‌گیری، مبتنی بر فرضهاست. چون این فرضیات منجر به نتیجه‌گیریهای بسیار مهم می‌شوند، لازم است به دقت بررسی شوند. متأسفانه اکثریت بزرگ آماردانان این کار را نمی‌کنند. بعضی از آنها نقش محوری احتمال را نفی می‌کنند بدون آنکه دلایلی برای آن ارائه کنند. هدف این مقاله این نیست که بررسی عمیقی راجع به این مطلب انجام دهد ولی اجازه دهید به یک فرض به خاطر نتایج بعدی‌اش نگاهی بیندازیم. همانند همهٔ فرضهای دیگر تصور می‌شود که این فرض هم خودبه‌خود بدیهی است و خلاف آن را احمقانه می‌دانیم. این فرض مبتنی است بر یک مفهوم اولیه یعنی اینکه یک پیشامد از پیشامد دیگر محتمل‌تر است، کلمهٔ («محتمل» در اینجا به مفهوم فنی آن مورد نظر نیست و فقط به معنای معمولی آن در زبان طبیعی به‌کار رفته است). می‌نویسیم «$A > B$» هرگاه «A محتمل‌تر از B» باشد. فرض مورد نظر این است که اگر $A_۱$، $A_۲$ با هم و $B_۱$، $B_۲$ با هم ناسازگار باشند آنگاه $A_i > B_i$ ، $i = ۱, ۲$، نتیجه می‌دهد $A_۱ \cup A_۲ > B_۱ \cup B_۲$ (در اینجا $A_۱ \cup A_۲$ به معنی وقوع پیشامدی است که رخ می‌دهد اگر و تنها اگر $A_۱$ یا $A_۲$ رخ بدهد). اگر بدانیم فرد بعدی که از در وارد می‌شود محتمل‌تر است زنی غیر سفیدپوست باشد ($A_۱$) تا مردی غیر سفیدپوست ($B_۱$) یا زنی سفیدپوست باشد ($A_۲$) تا مردی سفیدپوست ($B_۲$)، آنگاه ممکن است احساس ناخرسندی کنیم که زن بودن واردشونده ($A_۱ \cup A_۲$) کمتر از مرد بودن او ($B_۱ \cup B_۲$) محتمل باشد. مباحثی که در بالا به آنها اشاره شد با فرضها یا اصول موضوعی از این نوع شروع می‌شوند و نکتهٔ مهم این است که همهٔ آنها به این واقعیت می‌انجامند که احتمال تنها وسیلهٔ قانع‌کننده برای بیان عدم حتمیت است.

جملهٔ اخیر کاملاً دقیق نیست. برخی از نویسندگان با تعمق در اصول موضوع، مخالفتهایی با آن ابراز کرده‌اند. والی در نقد زیبایی [۲۷] سعی کرده با یک جفت عدد به نام احتمالهای بالایی و پایینی به جای یک عدد به نام احتمال، سیستمی را بنا کند. نتیجهٔ کار، سیستمی پیچیده‌تر است، که به نظر من این پیچیدگی ضرورتی ندارد. با این حال می‌پذیرم که گاهی اوقات

1. possibility theory

روش احتمالاتی کاملاً دقیق نیست و این عدم دقت می‌تواند با به‌کارگیری احتمالهای بالایی و پایینی ترمیم شود. هدف این است که این جفت عدد، گزاره‌های احتمالاتی را دقیق‌تر کنند، ولی احتمال به تنهایی شامل شاخصی از دقت خودش هست. عقیدهٔ من این است که ساده بر پیچیده ترجیح دارد، به شرط آنکه کارساز باشد و من اساساً با تیغ اُکام[1] موافقم.

با این نتیجه‌گیری که عدم حتمیت تنها توسط احتمال به طرز قانع‌کننده‌ای بیان می‌شود، سه قانون یا سه اصل موضوع حساب احتمال را به‌طور رسمی بیان می‌کنیم. احتمال به دو عنصر وابسته است: پیشامد غیرحتمی و شرایطی که تحت آنها به آن پیشامد توجه می‌کنید. مثلاً در بیرون کشیدن گوی از کیسه، احتمالی که برای قرمز بودن گوی قائل می‌شوید به این شرط وابسته است که گوی به تصادف از کیسه بیرون کشیده شود. برای احتمال وقوع A وقتی که می‌دانیم یا فرض می‌کنیم که B رخ داده است می‌نویسیم $p(A|B)$ و می‌خوانیم احتمال A به شرط (یا به فرض) B، این قوانین عبارت‌اند از:

(الف) قانون ۱ (تحدب): برای همهٔ A و Bها، $0 \leqslant p(A|B) \leqslant 1$ و $p(A|A) = 1$

(ب) قانون ۲ (جمع): اگر A و B ناسازگار باشند و C مفروض باشد،

$$p(A \cup B|C) = p(A|C) + p(B|C)$$

(ج) قانون ۳ (ضرب): برای همهٔ A و B و C ها

$$p(AB|C) = p(A|BC)p(B|C)$$

(اینجا AB به معنی پیشامدی است که رخ می‌دهد اگر A و B هر دو رخ دهند). قانون تحدب گاهی اوقات توسط رابطهٔ زیر تقویت می‌شود: $p(A|B) = 1$ تنها اگر A نتیجهٔ منطقی B باشد. قانون جمع به قانون کراموف معروف است.

قانون جمع ممکن است فقط یک نوع ظریف‌کاری ریاضی به نظر برسد ولی در حقیقت نتایج عملی مهمی دارد که در بخش ۸ نشان داده خواهد شد. دربارهٔ این قانون نکته‌ای شایان ذکر است. با سه قانونی که در بالا ذکر شد، قانون جمع دو پیشامد را می‌توان به هر تعداد متناهی از پیشامدها تعمیم داد. هیچ‌یک از رویکردهای بسیاری که قبلاً دربارهٔ آنها بحث شد به برقراری این قانون برای بینهایت پیشامد منجر نمی‌شود. ولی معمول است که فرض شود قانون جمع برای تعداد نامتناهی از پیشامدها برقرار است زیرا خلاف آن منتهی به نتایج ناخوشایند می‌شود. برای صراحت بخشیدن به این فرض می‌توان قانون جمع را اصلاح کرد و یا قانون چهارمی اضافه کرد که همراه سه قانون بالا تعمیم قانون جمع را به بینهایت پیشامدِ دوبه‌دو ناسازگار به‌دست دهد. من شخصاً ترجیح می‌دهم قانون چهارمی به شکل زیر اضافه کنیم.

(د) قانون ۴ (قابلیت اختلاط): اگر $\{B_n\}$ افرازی، متناهی یا نامتناهی، از C باشد و به‌ازای همهٔ مقادیر n، و k $p(A|B_nC) = k$ آنگاه $p(A|C) = k$.

(به آسانی می‌توان نشان داد که قانون ۴ وقتی که افراز متناهی است از قوانین ۱ تا ۳ به‌دست می‌آید. این تعریف از آنِ دوفینتی است). قابلیت اختلاط از نوع دسته‌ای از قوانین است که به نام «چیزهای مطمئن» معروف‌اند. به تسامح می‌توان گفت، چنانچه هر اتفاقی بیفتد (هر چه باشد B_n) باور شما در مورد یک پدیده برابر k باشد آنگاه این باور به‌طور غیرشرطی k است. برخی از آماردانان به نظر می‌رسد که از قابلیت اختلاط فقط زمانی که به نفع آنهاست استفاده می‌کنند: ما جنبهٔ کاربردی آن را در بخش ۸ بررسی می‌کنیم. توجه کنید که قوانین احتمالی که اینجا به آنها اشاره کردیم مانند اصول موضوعی که درکتابهای درسی احتمال آورده می‌شود نیستند. اینها، بجز قانون ۴، نتایجی از فرضهای ساده‌تر دیگری هستند.

۶. معنی‌دار بودن و اطمینان

عکس‌العمل بسیاری از آماردانان نسبت به این حکم که باید از احتمال استفاده کنند این است که آنها این کار را می‌کنند و دیدگاههایی که در اینجا توضیح داده شده فقط مهر تأییدی است بر کاری که از قبل می‌کرده‌اند. مجلات پر است از احتمالات: توزیعهای نرمال و دوجمله‌ای به‌وفور دیده می‌شود؛ خانوادهٔ نمایی همه جا هست. حتی ممکن است ادعا شود از هیچ شاخص دیگری برای اندازه‌گیری عدم حتمیت استفاده نمی‌شود: عدهٔ کمی از آماردانان ممکن است به منطق فازی اشاره کنند. ولی این عکس‌العمل درست نیست؛ آماردانان شاخصهایی از عدم حتمیت را به‌کار می‌برند که با استفاده از قوانین حساب احتمال با یکدیگر ترکیب نمی‌شوند.

گیریم فرض H این باشد که یک دارو بی‌تأثیر است و یا یک عامل اجتماعی تأثیری بر میزان جنایت ندارد. پزشک یا جامعه‌شناس دربارهٔ H نظر حتمی ندارد و داده‌هایی جمع‌آوری می‌شوند به این امید که این عدم حتمیت برطرف شود و یا دست‌کم کاهش یابد. برای راهنمایی در مورد عدم حتمیت از آماردانی کمک خواسته می‌شود و این آماردان ممکن است پیشنهاد کند که مشتری شاخصی از عدم حتمیت مثل سطح زیر منحنی یا سطح معنی‌دار بودن را در مورد H به‌عنوان فرض صفر به‌کار گیرد. یعنی فرض کند H درست است و احتمال داده‌های مشاهده شده را تحت این فرض محاسبه کند. در واقع این شاخصِ اعتقادی است که می‌توان به H داشت و هر چه این احتمال کوچکتر باشد اعتقاد به H کمتر است.

این روش در واقع مغایر با بحث بالاست زیرا در بالا تأکید بر این بود که عدم حتمیت در مورد H باید با احتمال H اندازه‌گیری شود. سطح معنی‌دار بودن چنین احتمالی را به‌دست نمی‌دهد. این تفاوت را می‌توان به‌صورت زیر مطرح کرد.

سطح معنی‌دار بودن ــ احتمال مشاهدهٔ جنبهٔ خاصی از داده‌ها، با فرض اینکه H درست است؛

احتمال ــ احتمالی که شما برای H قائل‌اید، با فرض مشاهدهٔ داده‌ها.

«مغلطهٔ دادستان» که در محافل حقوقی معروف است، در واقع در اثر خلط کردن $p(A|B)$ با $p(B|A)$، دو کمیتی که به‌ندرت با هم برابرند، رخ می‌دهد. تفاوت بین سطح معنی‌دار بودن و احتمال تقریباً مانند مغلطهٔ دادستان است: «تقریباً» به این دلیل که گرچه در ساختار مثال دادستان، B را می‌توان برابر با H در نظر گرفت ولی داده‌ها به نحو دیگری بررسی می‌شوند. A به‌عنوان داده‌ها در احتمال مورد استفاده قرار می‌گیرد. طرفداران سطح معنی‌دار بودن خیلی زود دریافتند که کافی نیست فقط از داده‌ها استفاده کنند بلکه لازم

1. Occam's razor

اشاره‌ای است به اصلی منسوب به فیلسوف انگلیسی ویلیام اُکام (در اصل، آکم) مبنی بر اینکه در توضیح یک پدیده باید حداقلِ مفروضات آورده شود.ـم.

است تعدادی دادهٔ «فرین‌تر» را به شکل دُم یک توزیع وارد کنند. همان‌طور که جفریز [۱۹] اشاره می‌کند، این سطح شامل داده‌هایی است که ممکن بود مشاهده شوند ولی نشدند.

حال از آزمون فرض به برآورد نقطه‌ای می‌رویم. پارامتر θ ممکن است اثر یک دارو یا تأثیر یک عامل اجتماعی بر میزان جنایت باشد. باز θ غیرحتمی است، لذا ممکن است داده‌هایی جمع‌آوری شده از آماردانی مشورت خواسته شود (امید این است که ترتیب غیر از این باشد). آماردان معمولاً یک بازهٔ اطمینان را پیشنهاد می‌کند. در این امر که مبتنی بر عدم حتمیتِ اندازه‌گیری شده است، از چگالی احتمال θ و شاید بازه‌ای از آن چگالی استفاده می‌شود. دوباره ما با تضادی شبیه به آنچه در مغلطهٔ دادستان بود مواجه‌ایم:

اطمینان ــ احتمال آنکه بازه شامل θ باشد؛

احتمال ــ احتمال آنکه θ در این بازه قرار گیرد.

عبارت اول گزاره‌ای احتمالاتی در مورد یک بازه، مشروط به مشاهدهٔ θ است؛ حال آنکه عبارت دوم گزاره‌ای در مورد θ مشروط به داده‌هاست. غالباً این دو را با هم اشتباه می‌گیرند. مهمتر از این اشتباه این است که نه سطوح معنی‌دار بودن و نه بازه‌های اطمینان هیچ‌یک بر اساس قوانین حساب احتمال ترکیب نمی‌شوند.

آیا این اشتباه حائز اهمیت است؟ از لحاظ نظری مسلماً چنین است زیرا استفاده از شاخصی که از قوانین حساب احتمال تبعیت نکند نهایتاً ناقض بعضی از فرضهای پایه‌ای خواهد بود، فرضهایی که بدیهی در نظر گرفته شده‌اند و نقض آنها ایجاد اشکال می‌کند. ولی از لحاظ عملی این موضوع آنقدر واضح نیست و لازم است دربارهٔ پیامدهای عملی آنها توضیح داده شود. آماردانان تمایل دارند هر مسأله را جداگانه مطالعه کنند و نیازی نمی‌بینند که نتایج را با هم ترکیب کنند؛ این ترکیب نتایج است که مشکل به‌وجود می‌آورد، و این را در مثال رنگ‌ـجنسیت در بخش ۵ دیدیم. به‌عنوان مثال به‌ندرت امکان دارد که بتوان دفترچهٔ هلندی را با استفاده از گزاره‌های سطوح معنی‌داری ساخت. معلوم است که برخی از برآوردگرهای رایج غیر قابل‌قبول‌اند. روشن‌ترین مثال از یک اِشکال مهم، ناشی از رابطهٔ بین سطح معنی‌دار بودن و اندازهٔ (n) نمونه‌ای است که این سطح بر آن بنا شده است. تعبیر «معنی‌دار در سطح ۵٪» بستگی به n دارد حال آنکه احتمال ۵٪ همیشه ثابت است. آماردانان به ارتباط بین گزاره‌ای که بیان می‌کنند و اندازهٔ نمونه‌ای که گزاره مبتنی بر آن است توجه چندانی نکرده‌اند. دلایلی نظری وجود دارد که فکر کنیم به‌دست آوردن سطح معنی‌دار ۵٪ خیلی ساده است [۳]. و اگر این مطلب درست باشد بسیاری از آزمایشها امیدی واهی به تأثیری سودمند، که در واقع وجود ندارد، ایجاد می‌کنند.

گزاره‌های منفرد آماری معمولاً اشکالی ندارند. مشکل زمانی به‌وجود می‌آید که بخواهیم آنها را با هم تلفیق کنیم. به‌عنوان مثال، بازه‌های اطمینان برای یک پارامتر غالباً قابل‌قبول هستند ولی برای چندین پارامتر نه. حتی میانگین یک نمونه از توزیع نرمال نیز در ابعاد بالا مناسب به نظر نمی‌رسد. در آزمایشی با تیمارهای گوناگون، آزمونهای منفرد خوب‌اند ولی مقایسه‌های چندگانه دچار مشکل می‌شوند. اما حقیقت علمی با تلفیق نتایج آزمایشهای گوناگون محقق می‌شود: هنوز مبحث فراتحلیل[1] حوزه‌ای دشوار از آمار است.

1. meta-analysis

چگونه می‌توان چند مجموعهٔ داده دربارهٔ یک فرض را که هر یک سطح معنی‌دار بودنی برای خود دارد با هم ترکیب کرد؟ نتایج دو آزمون tای استودنت برای میانگینهای μ_1 و μ_2 با آزمون Tی دو متغیره برای (μ_1, μ_2) سازگار نیست [۱۸].

پس این نتیجه‌گیری که احتمال تنها شاخصِ عدم حتمیت است صرفاً یک تعریف و تمجید خشک و خالی نیست بلکه به بسیاری از فعالیتهای پایه‌ای آماری مربوط می‌شود. ساویج ایده‌های خودش را در تلاشی برای توجیه عمل آماری ارائه داد. او با تشکیک در جنبه‌هایی از آن نتایجی به‌دست آورد که برای خودش شگفت‌انگیز بود. حال از اختلاف نظرها بگذریم و به سمت ایده‌های سازنده‌ای برویم که از شناسایی نقش اساسی احتمال در آمار نتیجه شده‌اند.

۷. استنباط

فرمولبندی‌ای که در سراسر این قرن خدمات شایسته‌ای به آمار کرده بر اساس داده‌هایی است که به‌ازای هر مقدار از پارامتر مورد نظر، یک توزیع احتمال دارند. این فرمولبندی با این نظر که عدم حتمیت داده‌ها باید به زبان احتمالاتی بیان شود، سازگار است. مطلوب آن است که از داده‌ها چیزی دربارهٔ پارامتر استنباط کنیم. معمولاً همهٔ جنبه‌های پارامتر، مورد نظر نیست، بنابراین می‌نویسیم (θ, a) به این معنی که می‌خواهیم چیزی در مورد θ استنباط کنیم و a در اصطلاح فنی پارازیت یا پارامتر مزاحم است. اگر داده‌ها را با x نمایش دهیم آنگاه $p(x|\theta, a)$ احتمال x است وقتی که θ و a داده شده باشند. یک مثال ساده می‌تواند توزیعی نرمال با میانگین θ و واریانس a باشد ولی این فرمولبندی بسیاری از حالتهای پیچیده را نیز در بر دارد.

با این روش، عدم حتمیت درون داده‌ها مطابق میل هر شخص کنترل می‌شود. پارامتر نیز غیرحتمی است. در حقیقت این عدم حتمیت است که دغدغهٔ اصلی آماردان است و طبق این روش، آن نیز باید با یک احتمال $p(\theta, a)$ توصیف شود. با این کار ما از روش معمول عدول کرده‌ایم. غالباً می‌گویند که پارامترها بنا به فرض کمیتهایی تصادفی هستند. ولی چنین نیست. این اصول موضوع هستند که فرض می‌شوند و خاصیت تصادفی بودن از آنها نتیجه می‌شود. با در دست بودن هر دو احتمال و استفاده از حساب احتمال می‌توان در میزان عدم حتمیت با توجه به داده‌ها تجدیدنظر کرد:

$$p(\theta, a|x) \propto p(x|\theta, a)p(\theta, a) \qquad (۱)$$

که ثابت تناسب فقط به x بستگی دارد و نه به پارامترها. از آنجا که a مورد علاقهٔ ما نیست می‌توانیم با استفاده از حساب احتمال آن را حذف کنیم:

$$p(\theta|x) = \int p(\theta, a|x)\,\mathrm{d}a\,. \qquad (۲)$$

رابطهٔ (۱) قانون ضرب و رابطهٔ (۲) قانون جمع است. مسألهٔ استنباط به کمک این دو رابطه حل می‌شود و یا به عبارت بهتر چارچوبی برای حل آن فراهم می‌آید. رابطهٔ (۱) قضیه بیز است که نام آن بنا به تصادفی تاریخی به کل این رویکرد داده شده است و به آن روش بیزی می‌گویند. این نام‌گذاری تأسف‌انگیز با چند نام‌گذاری بدتر همراه است. غالباً $p(\theta)$ را توزیع پیشین و $p(\theta|x)$ را توزیع پسین می‌نامند. اینها نامناسب‌اند زیرا پیشین و پسین اصطلاحاتی

نسبی هستند که به داده‌ها مربوط می‌شوند. پسین امروز، پیشین فرداست. ولی این دو اصطلاح چنان رواج یافته‌اند که اجتناب کامل از آنها ناممکن است.

حال مواضعی را که به آنها دست یافته‌ایم خلاصه‌وار ذکر می‌کنیم.

(الف) آمار مطالعهٔ عدم حتمیت است.

(ب) عدم حتمیت باید با احتمال اندازه‌گیری شود.

(ج) عدم حتمیت مربوط به داده‌ها به این طریق، مشروط به پارامترها، اندازه‌گیری می‌شود.

(د) عدم حتمیت پارامتر نیز با احتمال اندازه‌گیری می‌شود.

(ه) استنباط در قالب حساب احتمال، عمدتاً به کمک روابط (۱) و (۲)، انجام می‌شود.

دربارهٔ موارد (الف) و (ب) قبلاً بحث کردیم و (ج) را نیز عموماً می‌پذیرند. مورد (ه) از موارد (الف) تا (د) نتیجه می‌شود. اعتراض عمده به روش بیزی، به مورد (د) مربوط می‌شود. بنابراین اجازه دهید به آن بپردازیم.

۸. ذهنیت

در سطح پایه‌ای که بحث را از آنجا شروع کردیم، روشن است که عدم حتمیت‌ها در نظر اشخاص مختلف معمولاً متفاوت است. ممکن است دربارهٔ بازیهای شانسی که خوب اجرا شوند توافقی وجود داشته باشد به‌طوری که بتوان آنها را به‌عنوان استاندارد به‌کار برد. ولی در بسیاری از موضوعات، حتی در علوم، ممکن است توافق وجود نداشته باشد. در زمان نوشتن این سطور دانشمندان دربارهٔ برخی از مسائل مربوط به GMF (مواد غذایی‌ای که به‌طور ژنتیکی دستکاری شده‌اند) اختلاف‌نظر دارند. بنابراین لازم به نظر می‌رسد که ذهنیت را به‌نحوی مناسب در نمادهایمان منعکس کنیم. راه مرجح این است که مفهوم آگاهی شخص را، در زمانی که قضاوتِ احتمالاتی می‌کند، وارد کار کنیم. فرض کنید K نمایانگر آن باشد، لذا یک نماد بهتر از $p(\theta)$، نماد $p(\theta|K)$، احتمال θ به شرط K، است که بر اساس داده‌ها به $p(\theta|x,K)$ تغییر می‌یابد. این نماد ارزشمند است زیرا بر شرطی بودن احتمال تأکید می‌کند (بخش ۵ را ببینید). این نماد به دو مؤلفه وابسته است: چیزی که عدم حتمیتِ آن توصیف می‌شود و آگاهی‌ای که این عدم حتمیت بر آن بنا شده است. حذف مؤلفهٔ شرطی‌کننده غالباً به سردرگمی می‌انجامد. تمایز بین پیشین و پسین با تکیه بر شرایط متفاوتی که احتمال پارامتر مورد نظر در آن شرایط تعیین می‌شود، بهتر توصیف می‌شود.

این نظر مطرح شده است که اگر دو نفر آگاهی یکسانی راجع به یک موضوع داشته باشند، عدم حتمیت و در نتیجه احتمال از نظر آنها باید یکسان باشد. این را دیدگاه لاژم می‌نامند. این دیدگاه دو اشکال دارد. اولاً توضیح اینکه منظور از داشتن آگاهی یکسان چیست و همچنین تشخیص آن در عمل، دشوار است. نکتهٔ دوم اینکه اگر احتمال لزوماً به‌دست می‌آید پس باید بتوانیم آن را بدون ارجاع به شخص محاسبه کنیم. تاکنون هیچ‌کس نتوانسته این محاسبه را به‌صورت کاملاً قانع‌کننده‌ای انجام دهد. یک راه، استفاده از مفهوم جهل است. اگر سطح آگاهی I مشخص شود، آنگاه فقدان آگاهی در مورد θ مشخص می‌شود و $p(\theta|I)$ تعریف می‌گردد و $p(\theta|K)$ را به‌ازای هر K توسط فرمول بیز یا رابطهٔ (۱) با تبدیل I به K می‌توان محاسبه کرد. متأسفانه این کار معمولاً منجر به تناقضی با خاصیت اختلاط می‌شود. برای مثال، فرض کنید می‌دانید که θ فقط مقادیر صحیح مثبت را می‌تواند اختیار کند و اطلاع دیگری دربارهٔ مقادیر آن ندارید. آنگاه از مفهوم معمولی جهل چنین برمی‌آید که همهٔ مقادیر همشانس هستند: $p(\theta=i|I)=c$ به‌ازای همهٔ iها. قانون جمع برای n پیشامد دوبه‌دو ناسازگار $\{\theta=i\}$ با ضابطهٔ $nc>1$ نتیجه می‌دهد $c=0$ زیرا بنابر اصل تحدب، هیچ احتمالی نباید از یک بیشتر شود. حال اعداد صحیح مثبت را به مجموعه‌هایی سه عضوی افراز کنید که در هر یک از آنها دو عدد فرد و یک عدد زوج وجود داشته باشد مثلاً به‌صورت $A_n=(4n-3,4n-1,2n)$. اگر E پیشامد آن باشد که θ زوج است آنگاه $p(E|A_n)=1/3$ و با توجه به خاصیت اختلاط، $p(E)=1/3$. افراز دیگری را مرکب از مجموعه‌هایی در نظر بگیرید که هر یک شامل دو عدد زوج و یک عدد فرد است مثلاً به‌صورت $B_n=(4n-2,4n,2n-1)$. در این‌صورت $p(E|B_n)=2/3$ و بنابراین $p(E)=2/3$ که با مقدار قبلی متفاوت است. در واقع با انتخاب مناسب افزار، $p(E)$ می‌تواند هر مقداری را در بازهٔ یکه اختیار کند. چنین توزیعی را ناسره می‌نامند.

متأسفانه بیشتر تلاشها در جهت تولید $p(\theta|I)$ با استفاده از دیدگاه «لازم» به یک ناسرگی منجر می‌شود که نه تنها با خاصیت اختلاط سازگار نیست بلکه پیامدهای غیر قابل‌قبول دیگری هم دارد. در این زمینه، مثلاً [۱۰] را ببینید. دیدگاه لازم را اولین بار جفریز [۱۹] مورد بررسی قرار داد. برناردو [۶] و دیگران پیشرفتهای واقعاً چشمگیری را موجب شده‌اند ولی مسأله هنوز حل نشده است. در این مقاله این نظر اتخاذ می‌شود که احتمال گزاره‌ای است که یک شخص با آگاهی مشخص در مورد یک کمیت غیرحتمی می‌سازد. از این شخص با عنوان «شما» یاد می‌کنیم. $p(A|B)$ میزان باور شما در مورد A است وقتی B را می‌دانید. با این دید، درست نیست از احتمال به‌صورتی یاد کنیم که گویی منحصر به‌فرد است؛ فقط «احتمال از نظر شما» معنی دارد. بیان شرطها به همان اندازه اهمیت دارد که پیشامد غیرحتمی مورد نظر.

در مثال بالا که θ فقط مقادیر صحیح مثبت را اختیار می‌کند به محض آنکه مفهوم چیزی را که این حرف یونانی به آن اشاره دارد فهمیدید، دیگر جاهل نیستید. در آن‌صورت $p(\theta=i)=a_i$ با شرط $\Sigma a_i=1$. برخی از تحقیقات آماری به جهت سردرگمی بین حرف یونانی و واقعیتی که حرف نمایندهٔ آن است مخدوش می‌شوند. من همه را به مبارزه می‌طلبم که کمیتی واقعی ارائه کنند که واقعاً نسبت به آن جاهل باشند. نکتهٔ دیگر این است که رایانه نمی‌تواند مثالی از θ تولید کند. از آنجا که $p(\theta\leqslant n|I)=nc=0$ و $p(\theta>n)=1$، بنابراین θ باید قطعاً بزرگتر از هر عددی باشد که شما به فکرتان می‌رسد و یا رایانه می‌تواند محاسبه کند.

توجه به مطلب دیگری نیز ضروری است. معمولاً هنگامی که بخشی از پیشامد شرطی‌کننده مجهول است اما فرض می‌شود که رخ داده است، باز از همین مفهوم و نمادها استفاده می‌شود. به‌عنوان مثال، همهٔ آماردانان نماد $p(x|\theta,a)$ یا به عبارت دقیق‌تر $p(x|\theta,a,K)$ را به‌کار می‌برند. توجه کنید که آنها در اینجا مقدار پارامتر را نمی‌دانند. چیزی که با این نماد بیان می‌شود عدم حتمیت دربارهٔ x است اگر پارامترها مقادیر θ و a را اخذ کنند

(و آگاهی ما راجع به آنها به اندازهٔ K باشد). لازم نیست که در این چارچوب بین فرض و واقعیت فرق گذاشته شود و نماد $p(A|B)$ کفایت می‌کند.

موضع فلسفی من این است که عدم حتمیت شخصی شما از طریق احتمال شخصی شما در مورد یک کمیت غیرحتمی بر اساس سطح آگاهی شما، واقعی یا فرضی، بیان می‌شود. این را نگرش شخصی یا ذهنی به احتمال می‌نامیم.

بسیاری از افراد، مخصوصاً وقتی با مفاهیم علمی سروکار دارند، فکر می‌کنند گزاره‌هایشان عینی‌اند و از طریق یک احتمالِ یکتا بیا می‌شوند، و از مزاحمت ذهنیات هراس دارند. این هراس را می‌توان با توجه کردن به واقعیت و نحوهٔ بازتاب واقعیت در حساب احتمال کاهش داد. ما در چارچوب علوم بحث می‌کنیم ولی این رویکرد به‌طور کلی درست است. علم حقوق مثال دیگری به‌دست می‌دهد. فرض کنید θ کمیت علمیِ مورد نظر باشد. (در حقوق جنایی می‌توان θ را برحسب اینکه متهم مرتکب جنایت شده یا نه، برابر ۱ یا ۰ گرفت). در ابتدا محقق در مورد θ کم می‌داند چون اطلاع پایه‌ای K اندک است، و دو محقق عقاید مختلفی خواهند داشت که با احتمالهای $p(\theta|K)$ بیان می‌شود. سپس آزمایشهایی ترتیب داده می‌شود، داده‌های xای به‌دست می‌آیند، و این احتمالات به‌صورتی که قبلاً گفته شد به $p(\theta|x,K)$ تبدیل می‌شوند. می‌توان نشان داد [۱۴] که تحت شرایطی که معمولاً حاصل می‌شود، چنانچه تعداد داده‌ها افزایش یابد، عقاید مختلف راجع به θ همگرا می‌شود و معمولاً به معلوم شدن θ و یا حداقل، تعیین آن با دقت زیاد، می‌انجامد. چیزی که ما در عمل مشاهده می‌کنیم این است: با اینکه دانشمندان ممکن است در ابتدا نظراتی متفاوت و در برخی موارد مغایر هم داشته باشند ولی نهایتاً به توافق می‌رسند. همان‌طور که شخصی گفته است: عینیت در حقیقت همان توافق است. بنابراین توافق خوبی بین تجربهٔ علمی و پارادایم بیزی وجود دارد. مواردی هست که تقریباً همه دربارهٔ یک احتمال توافق دارند. این را در مبحث تعویض‌پذیری در بخش ۱۴ بررسی خواهیم کرد.

حال می‌توان فهمید که چرا نوعی اکراه در پذیرش نکتهٔ (د) که همان استفاده از توزیع احتمال برای توصیف عدم حتمیت پارامتر است، وجود دارد. علت این است که جنبهٔ اساسی ذهنی از نظر دور مانده است. وقتی تعداد داده‌ها کم است مقدار $p(\theta,a)$ در اذهان مختلف متفاوت است. وقتی داده‌ها افزایش می‌یابند، توافق حاصل می‌شود. توجه کنید که $p(x|\theta,a)$ نیز ذهنی است. وقتی دو آماردان در تحلیل یک مجموعهٔ داده‌ها دو الگوی مختلف به‌کار می‌برند، این موضوع آشکارا تصدیق می‌شود. ما دوباره به این مطلب در بخشهای ۹ و ۱۲ برمی‌گردیم تا نقش الگوها را بررسی کنیم.

۹. الگوها

موضوع الگوها [مدلها] را دراپر [۱۲] به دقت از دیدگاه بیزی تحلیل کرده است و برای شرحی مفصل‌تر از آنچه در اینجا می‌آید باید به مقالهٔ او مراجعه کرد. موضع فلسفی ارائه‌شده در این مقاله این است که عدم حتمیت فقط باید برحسب احتمال مورد نظر شما توصیف شود. اجرای این نظر نیاز به ساختن توزیعهای احتمال برای تمام عناصر غیرحتمی در پدیدهٔ واقعی مورد مطالعه دارد. تعیین کامل احتمال، الگو یا مدل (احتمال) نامیده می‌شود، هر چند معنای آن، چنانکه توضیح داده خواهد شد، با معنای معمول این اصطلاح در آمار فرق دارد و با مدلهای رایج در علوم هم متفاوت است. کار آماردان این است که برای دنیای غیرحتمیِ تحت مطالعه الگو بسازد. هنگامی که این کار انجام شد، عدم حتمیتِ جنبه‌های خاص مورد نظر را می‌توان به‌وسیلهٔ حساب احتمال و بر اساس اطلاعات موجود محاسبه کرد. بنابراین مطالعهٔ ما دو جنبه دارد: ساختن الگو و تحلیل آن. دومی اساساً خودبه‌خود انجام می‌شود؛ علی‌الاصول می‌توان آن را توسط ماشین انجام داد. ولی اولی نیازمند به تماس نزدیک با واقعیت است. با دستکاری در جمله‌ای از دوفینتی و قدری غلّو می‌توان گفت: «وقتی که الگو می‌سازی، فکر کن؛ و وقتی الگو را داری فکر نکن و آن را به رایانه بسپار». بار دیگر این مطلب را تکرار می‌کنیم که شخصی که احتمالات مورد نظر او بررسی می‌شود، یا به زبان این مقاله، «شما»، آماردان نیست بلکه یک مشتری است؛ غالباً محققی است که خواستار اظهارنظر آماری شده است. وظیفهٔ آماردان ترجمهٔ عدم‌حتمیت‌های محقق به زبان احتمال و محاسبهٔ آنها با استفاده از اعدادی است که به‌دست می‌آورد. هر الگو صرفاً انعکاس واقعیت در فکر شماست و مانند احتمال، نه شما را و نه دنیای شما را توصیف نمی‌کند، بلکه فقط رابطهٔ بین شما و آن دنیا را وصف می‌کند. لذا نام بردن از آن به‌عنوان الگوی واقعی درست نیست. زمانی می‌توان این کلمه را به‌کار برد که غالب مردم با آن الگو موافق باشند. مثلاً توزیع نرمال دو متغیری را ممکن است بتوان الگوی واقعی قد پدران و پسران نامید.

در یک سناریوی معمولی چه نوع عدم حتمیت‌هایی وجود دارد؟ مسألهٔ اساسی استنباط و استقرا، پیش‌بینی داده‌های آتی از روی داده‌های گذشته است. رصدهای مکرر حرکت اجرام آسمانی محاسبهٔ مواضع آیندهٔ آنها را امکان‌پذیر می‌کند. مطالعات کلینیکی روی یک دارو به پزشک امکان می‌دهد که وضع آیندهٔ بیماری را که آن دارو برایش تجویز شده، پیش‌بینی کند. گاهی اوقات داده‌های غیرحتمی در گذشته‌اند، نه در آینده. هرگاه گزارشی از یک واقعهٔ تاریخی به جا نمانده باشد، مورخ از تمام شواهدی که در دست دارد استفاده می‌کند تا شاید مشخص کند چه اتفاقی به‌وقوع پیوسته است. دادگاه جنایی می‌کوشد بر اساس شواهدی که پس از جنایت به‌دست آمده، معلوم کند چه اتفاقی افتاده است. ولی ما در اینجا فرض می‌کنیم که داده‌های گذشته، x، برای استنباط راجع به داده‌های آینده، y، به‌کار می‌روند (همان‌طور که در الفبای انگلیسی x قبل از y است). با این ترتیب، مسألهٔ ما محاسبهٔ $p(y|x,K)$ است. برای سادگی، اطلاعات قبلی را که در طول کار ثابت است در نمادگذاری منظور نمی‌کنیم و می‌نویسیم $p(y|x)$.

یک راه ممکن این است که سعی کنیم $p(y|x)$ را مستقیماً محاسبه کنیم. این کار معمولاً دشوار است، هر چند می‌توان آن را نمونه‌ای از عمل در نظام استاد و شاگردی دانست. اینجا شاگرد باید پای درس معلم بنشیند و اطلاعات (داده‌های x) را جذب کند. پس از سالها کسب تجربه در این زمینه، شاگرد ما می‌تواند خودش به تنهایی استنباط کند که چه ممکن است رخ دهد. مشاهدهٔ مکرر استفاده از فلان مواد برای ساختن تایر این توانایی را به او می‌دهد که از آن مواد در ساختن تایر برای خودش استفاده کند. ولی راه بهتری وجود دارد و آن مطالعهٔ ارتباطهای بین x و y و سازوکار عمل‌کننده است. قوانین نیوتن امکان محاسبهٔ جذر و مد را می‌دهند. علم مواد به طراحی و ساخت تایر

کمک می‌کند. اغلب روشهای مدرن استنباطی را می‌توان از طریق پارامتری مانند θ بیان کرد که نشان‌دهندهٔ ارتباط بین دو مجموعه از داده‌هاست. این بحث را با استفاده از حساب احتمال بسط می‌دهیم تا θ را نیز شامل شود:

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

معمولاً فرض می‌کنند که وقتی θ معلوم است داده‌های گذشته مطرح نیست. به زبان احتمال، با فرضِ در دست داشتن θ، داده‌های x و y مستقل‌اند. با توجه به این واقعیت و با تعیین $p(\theta|x)$ به‌وسیلهٔ قضیهٔ بیز خواهیم داشت

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

حال باید عدم حتمیت مربوط به پارامتر یا $p(\theta)$ و عدم حتمیت‌های دو دسته داده، $p(x|\theta)$ و $p(y|\theta)$، را به شرطِ در دست داشتن پارامتر محاسبه کرد. در بسیاری موارد، استنباط را می‌توان در $p(\theta|x)$ متوقف کرد و $p(y|\theta)$ را به‌عهدهٔ دیگران گذاشت. در مثال بالا در مورد دارو، دکتر نیاز دارد توزیع کارایی θ را بداند تا $p(y|\theta)$ را برای مریض مورد نظرش محاسبه کند.

گرچه موضوع اصلی استنباط در بسیاری از موارد تعیین $p(\theta|x)$ است و استنباط بحق برحسب آن بیان می‌شود ولی بررسی $p(y|x)$ هم فایدهٔ مهمی دارد. این فایده از این حقیقت نشأت می‌گیرد که y در نهایت مشاهده خواهد شد؛ دکتر نهایتاً خواهد دید چه بر سر مریضش خواهد آمد. پارامتر معمولاً مشاهده نمی‌شود. عدم حتمیت θ غالباً باقی می‌ماند ولی عدم حتمیت y از میان می‌رود. این خصیصه امکان می‌دهد که کارایی استنباط را با استفاده از تعمیم قانون امتیازدهی بخش ۴ نمایش دهیم. اگر استنباط مورد نظر، $p(y|x)$ باشد و بعداً مشاهده کنیم که y برابر y_0 است، آنگاه تابع امتیاز $S\{y_0, p(\cdot|x)\}$ نشان می‌دهد که استنباط تا چه حدی خوب بوده و به این وسیله توانایی مشتری و آماردان محک زده می‌شود. این روش در علم هواشناسی مثلاً برای پیش‌بینی بارندگی فردا، مورد استفاده قرار گرفته است. چنین روشهایی به آسانی برای $p(\theta|x)$ در دسترس نیستند. یکی از انتقادهایی که علیه سطوح معنی‌دار بودن مطرح می‌شود این است که بررسی چندانی در این باره به عمل نیامده که چه تعداد از فرضهایی که در سطح ۵٪ رد شده‌اند، بعداً درست از آب در آمده‌اند. دلیلی ندارد که فکر کنیم این سطح ۵٪ است. از نظریه چنین برمی‌آید که باید خیلی بالاتر باشد و در نتیجه معنی‌دار بودن خیلی سریع حاصل می‌شود، اگر هواشناسی پیش‌بینی کند فقط ۵٪ روزها بارندگی خواهد بود و بعداً دیده شود ۲۰٪ روزها باران باریده، آن هواشناس زیاد هم قابل قدردانی نیست. بهترین راه ارزیابی کیفیت استنباطها بررسی $p(\theta|x)$ از طریق احتمالهای داده‌ها یا $p(y|x)$ای است که آنها تولید می‌کنند.

همان‌طور که قبلاً اشاره شد، معمولاً لازم است پارامتر مزاحم a، علاوه بر θ، وارد کار شود، تا ارتباط بین x و y را به قدر کفایت توصیف کند، و استقلال بین آنها را به شرط (θ, a) برقرار سازد. در مثال دارو، a ممکن است خصوصیات شخصی بیمار باشد. پارامترهای مزاحم مشکلات بسیار بزرگی را برای برخی از اشکال استنباط ایجاد می‌کنند، مثلاً استنباطهایی که فقط مبتنی بر درستنمایی‌اند. اما این مسائل را می‌توان به راحتی در چارچوب احتمالاتی و با گذار از توزیع توأم (θ, a) به چگالی حاشیه‌ای θ، رابطهٔ (۲)، برطرف کرد.

معرفی پارامترها ساخت الگو را به محاسبهٔ $p(x|\theta, a)$ و $p(\theta, a)$ تقلیل می‌دهد. $p(y|\theta, a)$ نیز مطرح می‌شود ولی تعیین آن شبیه به $p(x|\theta, a)$ است و نیازی نیست به‌طور جداگانه مورد بحث قرار گیرد. در اینجا تفاوت بین استفادهٔ ما از «الگو» و استفادهٔ معمولی از آن که فقط مبتنی بر توزیع داده‌ها به شرط در دست داشتن پارامترهاست، دیده می‌شود. تعریف ما شامل توزیع پارامترهاست زیرا آنها بخش مهمی از عدم حتمیت موجود را شکل می‌دهند. غالب متون فعلیِ مربوط به الگوها معطوف به داده‌هاست و ما در بخش بعد به این موضوع می‌پردازیم. در حال حاضر دوباره به این نکته اشاره می‌کنیم که حتی $p(x|\theta, a)$ نیز ذهنی است. معمولاً به این دلیل آن را، به غلط، عینی می‌پندارند که در بسیاری موارد، اطلاعات مردم دربارهٔ داده‌ها بیشتر است تا دربارهٔ پارامترها، و ما در بخش ۸ دیدیم که با افزایش اطلاعات، مردم به توافق نزدیک می‌شوند.

چرا سلسله مراحل تعیین $p(x|\theta, a)$ و $p(\theta, a)$ و بعد از آن $p(\theta|x)$ را انجام می‌دهیم؟ اگر $p(\theta, a)$ قابل محاسبه است چرا مستقیماً $p(\theta|x)$ را محاسبه نمی‌کنیم تا از بعضی محاسبات پیچیده پرهیز کرده باشیم؟ این سؤال را، با استفاده از اصطلاحاتی که من به آنها علاقه‌ای ندارم، می‌توان این طور نیز بیان کرد: اگر توزیع پیشین شما مستقیماً قابل تعیین است چرا توزیع پسین شما نباشد؟ بخشی از جواب در اطلاعاتی نهفته است که معمولاً دربارهٔ چگالی داده‌ها در دست است ولی دلیل اساسی، تمایل ما به سازگاری مفاهیم است. مجموعه‌ای از گزاره‌های راجع به عدم حتمیت سازگار نامیده می‌شوند هرگاه از قوانین حساب احتمال تبعیت کنند. مثلاً دو گزارهٔ $p(A|B) = ۰٫۷$ و $p(A| \sim B) = ۰٫۴$ با دو گزارهٔ $p(B|A) = ۰٫۵$ و $p(B| \sim A) = ۰٫۳$ سازگار نیستند. ($B \sim$ نمایانگر متمم مجموعهٔ B است). فرض کنید A گزاره‌ای راجع به داده‌های x است و B گزاره‌ای دربارهٔ پارامتر θ. دو گزارهٔ اول که راجع به عدم حتمیت داخل داده‌ها هستند با اولین گزاره راجع به پارامتر یعنی $p(B|A) = ۰٫۵$ سازگارند. (فرض کنید $p(B) = ۰٫۳۶ = ۱٫۱/۰٫۴$). ولی هر سه گزاره با دومین گزارهٔ راجع به پارامتر یعنی $p(B| \sim A) = ۰٫۳$ سازگار نیستند. با فرض $p(B) = ۰٫۳۶$. مقدار سازگاری ۰٫۲۲ است. روش استاندارد این اطمینان را به شما می‌دهد که برای هر مقدار داده‌ها، A یا $A \sim$، زمینه را آماده کرده‌اید، و استنباطهای نهایی راجع به B در مجموع معنی دارند.

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) 293-337.

به علت طولانی بودن مقاله، ترجمهٔ آن به دو قسمت تفکیک شده است. قسمت دوم در شمارهٔ آینده خواهد آمد.

* دنیس لیندلی، انگلستان

thombayes@aol.com