غیرمترقبه‌ها در آمار و احتمال

حمید پزشک
دانشگاه تهران

مقدمه

اگر به لغت‌نامهٔ «آکسفورد» نگاه کنیم، کلمهٔ «آمار» را به‌عنوان شیوه یا علم جمع‌آوری و تجزیه و تحلیل داده‌های عددی در مقیاس بزرگ تعریف کرده است که عمدتاً برای استنباط در مورد نسبت‌ها در **کل** بر اساس نسبت‌های در **جزء** یا نمونه به‌کار گرفته می‌شود. این تعریف درواقع بیان دیدگاهی است که به آن دیدگاه **«فراوانی‌گرا»** از آمار» می‌گویند. در این دیدگاه که دیدگاه جزء به‌کل نیز نامیده می‌شود، برای شناسایی یک جامعه به بخشی از آن توجه می‌کنند. به این بخش که طبق قاعده‌ای خاص انتخاب می‌شود، «نمونه» می‌گویند. در مقابل این دیدگاه از آمار دیدگاه دیگری وجود دارد که به آن دیدگاه **«بیزی»** می‌گویند. این نام‌گذاری به افتخار **توماس بیز** (۱۷۶۱ـ ۱۷۰۱)، دانشمند و فیلسوف انگلیسی انجام شده است. قضیهٔ بیز که بی‌تردید یکی از زیباترین قضیه‌های عالم ریاضی است، اساس دیدگاه بیزی را شکل می‌دهد. در این دیدگاه هم به نحوی از جزء برای کل استنباط می‌شود، فقط با این تفاوت که از اطلاعات و دانش قبل از نمونه‌گیری استفاده می‌شود. به عبارت دیگر، در دیدگاه فراوانی‌گرا، اساس تصمیم‌گیری دربارهٔ کل (جامعه) فقط اطلاعات به دست آمده از جزء (نمونه) است، اما در دیدگاه بیزی فرض بر این است که شما قبل از نمونه‌گیری، اطلاعاتی (حتی مبهم) از جامعه دارید و سپس نمونه می‌گیرید. حال با ترکیب این دو اطلاعات استنباط را در مورد جامعه انجام می‌دهید.

هدف این نوشتهٔ کوتاه که با الهام از یکی از یادداشت‌های **والت هیکی**، دبیر ارشد نشریهٔ داده‌ها، نوشته شده است، تجزیه و تحلیل یک مسئلهٔ آماری خاص نیست. حتی قصد ندارد به لحاظ تاریخی به واقعه‌ای مرتبط به علم آمار و احتمال بپردازد. هدف این است که در شرحی کوتاه بحث کنیم، در دنیای پر از پدیده‌های غیرقطعی اطراف ما، استفاده از علم آمار و احتمال چقدر در توجیه پدیده‌ها توانمند است. جالب است که بدانید، حتی نگاه محدود و غیرفنی به «آمار و احتمال» می‌تواند به شما ایده‌های جدیدی در مورد تفسیر داده‌ها بدهد.

در این نوشته چند موضوع کلاسیک را که به نحو غیرمتعارف حل شده‌اند، انتخاب کرده‌ایم که کمک می‌کنند دریابیم، چقدر داده‌ها می‌توانند گمراه‌کننده باشند و چگونه وقایع می‌توانند ما را به مسیرهای درست یا اشتباه بفرستند!

۱. اطلاع از نمونه‌های قبلی

الف) در جعبه‌ای سه مهرهٔ سیاه و چهار مهرهٔ سفید وجود دارد. همهٔ مهره‌ها هم‌شکل‌اند و فقط رنگشان فرق می‌کند.

حالت اول: شما یک مهره به تصادف از جعبه بیرون می‌آورید. چقدر احتمال دارد مهرهٔ انتخاب شده سفید باشد؟

پاسخ به نظر سرراست می‌آید: $\frac{۴}{۷}$ کلاً هفت مهره وجود دارد که چهارتای آن‌ها سفیدند. بنابراین احتمال $\frac{۴}{۷}$ درست است.

حالت دوم: از همان جعبه (با هفت مهره) چهار مهره بیرون می‌آوریم و بدون نگاه کردن به رنگشان، کناری می‌گذاریم. حال مهرهٔ پنجم را بیرون می‌آوریم. چقدر احتمال دارد مهرهٔ انتخاب شده سفید باشد؟

پاسخ به سرراستی پاسخ قبلی به نظر نمی‌رسد! ولی باز هم: $\frac{۴}{۷}$. چرا؟ قبل از اینکه جواب را به نحوی غیرمتعارف توجیه کنیم، به این سؤال هم توجه کنید:

حالت سوم: از همان جعبه (با هفت مهره) شش مهره بیرون می‌آوریم و بدون نگاه کردن به رنگشان، کناری می‌گذاریم. حال

مهـــرۀ هفتم را بیرون می‌آوریم. چقـــدر احتمال دارد مهرۀ انتخاب شده سفید باشد؟ باز هم: $\frac{۴}{۷}$.

با اینکه با استفاده از روابط نه خیلی پیچیده نظریۀ احتمال می‌توان نشان داد پاسخ به سؤال‌های مطرح شده در حالت‌های دوم و ســـوم هم $\frac{۴}{۷}$ اســـت، ولی این توجیه زیبا و غیرفنی نیز کمک‌کننده است: در حالت‌های دوم و سوم، **با توجه با اینکه هیچ اطلاعی از رنگ مهره‌هایی که قبلاً بیرون آورده شده، نداریم**، مســـئله مانند آن است که از ابتدا می‌خواهیم از همان جعبه (با هفت مهره) یک مهرۀ سفید بیرون بیاوریم. احتمال این رویداد $\frac{۴}{۷}$ است!

ب) شما دانش‌آموز یک کلاس ۲۰ نفری هستید. روزی معلم کیســـه‌ای را به شما نشـــان می‌دهد که در آن ۱۹ مهرۀ سیاه و یک مهرۀ ســـفید وجود دارد. باز هم همۀ مهره‌ها هم‌شکل‌اند و فقط رنگشان فرق می‌کند. معلم می‌گوید صف ببندید و یکی‌یکی بیایید جلوی میز من و هر کدام به تصادف یک مهره از کیســـه بیرون بیاورید و بدون برگرداندن مهره به کیســـه، نوبت را به نفر بعدی در صف بدهید تا او هم مهره‌ای بیرون بیاورد. هر کس مهرۀ سفید را بیرون بیاورد، جایزه‌ای نفیس می‌گیرد. حال به شما اجازه داده‌اند در مکان اول یا هفتم یا بیستم صف بایستید. کدام مکان را انتخاب می‌کنید؟ چرا؟

پاسخ: هیچ ارجحیتی بین مکان‌ها وجود ندارد و همۀ مکان‌ها هم‌شـــانس هستند. احتمال برنده شـــدن برای هر سه مکان $\frac{۱}{۲۰}$ اســـت. برای مکان اول احتمال برنده شدن $\frac{۱}{۲۰}$ است و برای مکان‌های هفتم و بیستم، به دلیل آنکه **هیچ اطلاعی از رنگ مهره‌هایی که قبلاً بیرون آورده شده‌اند، نداریم**، باز هم احتمال برنده شدن $\frac{۱}{۲۰}$ است! البته یادآوری می‌کنیم که با استفاده از روابط سرراست نظریۀ احتمال می‌توان این احتمال را برای مکان‌های هفتم و بیستم به دست آورد و نشان داد $\frac{۱}{۲۰}$ است.

۲. مسئلۀ مونتی هال

فرض کنید شما در یک مسابقۀ تلویزیونی برای بردن جایزۀ اصلی شرکت می‌کنید. مجری برنامه به شما سه در بسته نشان می‌دهد که پشـــت دوتای آن‌ها بز است و پشت یکی خودرویی گران‌قیمت. مجری به شـــما می‌گوید: اگر شما دری را انتخاب کنید، او یک در دیگر را که پشـــتش یک بز اســـت، باز می‌کند. بعد به شما دو انتخاب می‌دهد: یا همچنان روی دری که انتخاب کرده‌اید، پافشاری کنید، یا انتخابتان را تغییر دهید و در دیگری را که مجری باز نکرده است، انتخاب کنید. آیا روی انتخابی که کرده‌اید پافشاری می‌کنید؟ به عبارت دیگر، آیا فرقی می‌کند که انتخابتان را عوض کنید؟

پاسخ: به نظر می‌رسد این مسئله هم مثل مسئلۀ قبل است. ابتدا که هر ســـه در بسته‌اند، شانس برد خودرو $\frac{۱}{۳}$ است و حال که مجری یک در را باز کرده، احتمال برد $\frac{۱}{۲}$ شده است و فرقی نمی‌کند شما انتخابتان را عوض کنید یا خیر. اما با کمی استدلال ساده درمی‌یابید: **به نفع شماست انتخابتان را عوض کنید، زیرا احتمال برنده شدن بعد از آنکه مجری یک در را باز کند $\frac{۱}{۲}$ نیست بلکه $\frac{۲}{۳}$ است!**

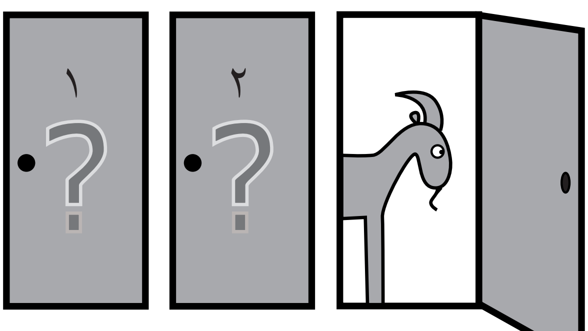
فرض کنید شـــما ابتدا در شمارۀ ۱ را انتخاب کردید. به حالت‌های جدول ۱ توجه کنید تا متوجه شـــوید، بعد از آنکه مجری یک در را که پشتش بز بود، باز کرد، به نفع شماست که انتخابتان را عوض کنید.

جدول ۱. حالت‌های متفاوت در مسئلۀ هال

در شمارۀ ۱	در شمارۀ ۲	در شمارۀ ۳	نتیجه، اگـــر انتخاب اولتان را عوض نکنید	نتیجه، اگـــر انتخاب اولتان را عوض کنید
خودرو	بز	بز	خودرو	بز
بز	خودرو	بز	بز	**خودرو**
بز	بز	خودرو	بز	**خودرو**

همان‌طـــور که در جدول بالا می‌بینیـــد، **برای این حالت خاص که شما ابتدا در شمارۀ ۱** را انتخاب کرده بودید، با عوض کردن انتخابتان بعد از آنکه مجری در دیگری را باز می‌کند، در $\frac{۲}{۳}$ اوقات خودرو را می‌برید و فقط در $\frac{۱}{۳}$ اوقات بازنده می‌شوید!

تفاوت مسئلۀ مونتی هال با مسئلۀ قبلی در چیست؟



تصویر ۱. پشت کدام در خودرو وجود دارد؟

پاسخ: در مسئلۀ قبلی نکتۀ کلیدی در این بود که اطلاعی از مرحله‌های قبلی نمونه‌گیری نداشتیم و همه‌چیز بر حسب تصادف رخ می‌داد. ولی در مسئلۀ مونتی هال، **مجری مطلعی** وجود دارد که مطمئن است، پشت دری را که باز می‌کند، بز هست.

۳. پارادکس روز تولد

فرض کنید شما رئیس اداره‌ای با ۲۳ کارمند هستید. چقدر

احتمال دارد حداقل دو نفر از کارمندانتان دارای روز تولد مشترک باشند (یا در یک روز از سال به دنیا آمده باشند)؟ فرض کنید سال ۳۶۵ روز دارد.

پاسخ: با توجه به اینکه سال ۳۶۵ روز دارد، به نظر می‌رسد که خیلی محتمل نباشد که دو نفر یا بیشتر از دو نفر روز تولد مشترک داشته باشند ولی این‌طور نیست. این احتمال بیش از ۵۰ درصد است!

تصویر ۲. پارادکس روز تولد

واضح است اگر ۳۶۶ نفر در اداره کار می‌کردند، با احتمال ۱ حداقل دو نفر روز تولد مشترک داشتند.

اجازه دهید پیشامد A را تعریف کنیم: پیشامد آنکه حداقل دو نفر در یک روز به دنیا آمده باشند و A^c را پیشامد متمم آن در نظر بگیریم، با استفاده از رابطهٔ زیر می‌توانیم احتمال پیشامد A را بیابیم:

$$P(A) = ۱ - P(A^c)$$

محاسبهٔ $P(A^c)$ آسان‌تر است. شما دو نفر را در اداره انتخاب کنید. احتمال آنکه نفر دوم روز تولدی متفاوت از نفر اول داشته باشد، $\frac{۳۶۴}{۳۶۵}$ است. احتمال آنکه نفر سوم روز تولدی متفاوت از نفر اول و دوم داشته باشد، $\frac{۳۶۳}{۳۶۵}$ است. اگر همین‌طور در اداره جلو برویم، می‌توانیم $P(A^c)$ را برای ۲۳ نفر به‌صورت زیر بیابیم:

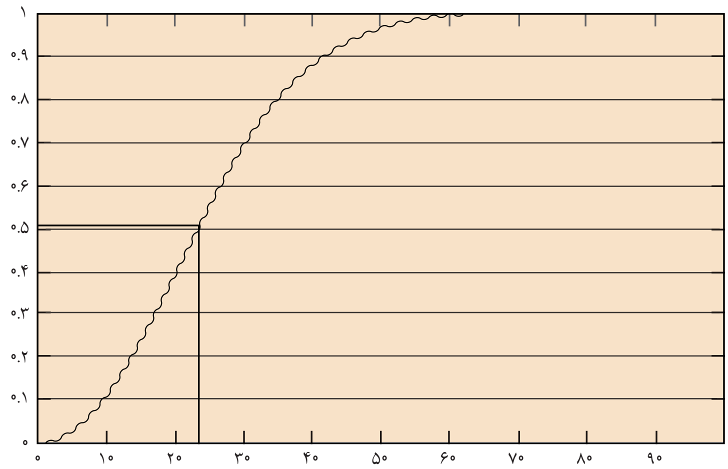
$$P(A^c) = \frac{۳۶۵}{۳۶۵} \times \frac{۳۶۴}{۳۶۵} \times \frac{۳۶۳}{۳۶۵} \times \frac{۳۶۲}{۳۶۵} \times \ldots \times \frac{۳۴۳}{۳۶۵} = ۰/۴۹۲۷$$

لذا در اداره‌ای با ۲۳ کارمند، احتمال آنکه هیچ دونفری روز تولد مشترک نداشته باشند، ۰/۴۹۲۷ است. پس احتمال آنکه حداقل دو نفر در یک روز از سال به دنیا آمده باشند، برابر است با: ۰/۵۰۷۳=۰/۴۹۲۷-۱.

جالب است بدانید، اگر ۵۷ نفر در اداره کار کنند، احتمال آنکه حداقل دو نفرشان روز تولد مشترک داشته باشند (یا در یک روز از سال به دنیا آمده باشند)، بیش از ۰/۹۹ است! نمودار ۱ P(A) را به‌عنوان تابعی از تعداد کارمندان نشان می‌دهد.

۴. توصیهٔ آبراهام والد

در طول جنگ جهانی دوم به آبراهام والد مأموریت داده



نمودار ۱. احتمال آنکه حداقل دو نفر در یک روز به دنیا آمده باشند، به‌عنوان تابعی از تعداد افراد داخل یک گروه.

شد در مورد هواپیماهای صدمه‌دیده‌ای که از پرواز بر فراز آلمان برگشته بودند، تحقیق کند. او باید آسیب‌های هواپیماها را بررسی می‌کرد تا مشاوره دهد از کدام قسمت‌ها باید بیشتر محافظت شود.


تصویر ۳. از کدام قسمت از هواپیما باید بیشتر محافظت شود؟

آبراهام درمی‌یابد، بدنهٔ هواپیما و سیستم سوخت از قسمت‌هایی هستند که احتمال بیشتری دارد توسط گلوله‌ها یا ترکش‌ها آسیب ببینند تا موتورهایشان. حال او باید به مقامات بالاترش چه پیشنهادی بدهد؟

پاسخ: محافظت‌های بیشتر را به قسمت‌های آسیب‌ندیده معطوف کنید. بله درست متوجه شدید، آبراهام والد که در آن زمان عضو گروه پژوهش‌های آماری بود، پیشنهادی غیرمتعارف کرد که حاصلش نجات جان تعداد قابل توجهی از خلبان‌ها و هواپیماها شد.

بر اساس مشاهدات و آسیب‌های هواپیماهای بازگشته به

پایگاه، محافظ‌ها را برای قسمت‌هایی از هواپیما که بیشترین آسیب را دیده‌اند، قرار ندهید! در حقیقت این قسمت‌ها آن‌قدر مقاوم بوده‌اند که آسیب را تحمل کنند. اگر یک قسمت اساسی یک هواپیمای برگشته به پایگاه ـ مثلاً موتورش ـ بدون آسیب مانده، بدان معنی است که به احتمال زیاد هواپیماهایی که موتورشان آسیب دیده‌اند، اصلاً نتوانسته‌اند برگردند.

توصیهٔ والد برای این موضوع نه تنها یک مدرک تاریخی آماری قابل توجه به شمار می‌رود، بلکه مقدمه‌ای بر استفادهٔ بیشتر از آمار در طول جنگ جهانی می‌شود که جهان را به‌سوی پیدایش حوزهٔ علمی «تحقیق در عملیات» پیش می‌برد.

۵. برآورد

فرض کنید سکه‌ای در دست دارید که روی یک وجهش عدد ۰ و روی وجه دیگرش عدد ۱ حک شده است. قرار است در پرتاب‌های متوالی از این سکه، هرگاه وجه ۱ مشاهده شد، ۱۰۰ تومان بگیرید و اگر عدد ۰ مشاهده شد، ۱۰۰ تومان بپردازید. قبل از بازی به شما اجازه می‌دهند که شانس مشاهدهٔ وجه‌ها را برای خودتان برآورد کنید تا تصمیم بگیرید، آیا در بازی شرکت می‌کنید یا نه؟ چه ایده‌ای برای برآورد شانس مشاهدهٔ ۱ دارید؟

پاسخ: سکه را به تعداد دفعات نسبتاً زیاد ـ مثلاً ۴۰ بار ـ پرتاب می‌کنیم و تعداد دفعاتی که عدد ۱ را مشاهده کردیم به تعداد کل پرتاب‌ها (۴۰ بار) تقسیم می‌کنیم. پاسخ قانع‌کننده‌ای است. این کار برآورد مناسبی از شانس مشاهدهٔ عدد ۱ می‌دهد. جالب است بدانید، نظریهٔ ریاضی قوی از این ایده پشتیبانی می‌کند. این نحوه برآورد که با استفاده از آزمایش‌های مستقل از هم به دست می‌آید، درواقع اساس «روش آماری فراوانی‌گرا» برای برآورد یک پارامتر ناشناخته است که به آن نام «برآورد ماکزیمم» درست‌نمایی می‌گویند. مثلاً اگر در این ۴۰ پرتاب ۱۸ بار عدد یک مشاهده شده باشد، برآورد مورد نظر برابراست: $\frac{۱۸}{۴۰} = ۰/۴۵$. جالب است بدانید در این نظریهٔ فراوانی‌گرا، نمونهٔ نسبتاً زیاد غالباً به نمونه‌های بزرگ‌تر یا مساوی ۳۰ گفته می‌شود.

حال فرض کنید دوست شما قبل از پرتاب‌های شما این سکه را ۵۰ بار پرتاب کرده و ۲۸ بار عدد ۱ را مشاهده کرده است. او این موضوع را به شما می‌گوید. آیا اطلاعی که او می‌دهد، روی برآورد شما، یعنی $\frac{۱۸}{۴۰}$ تأثیر خواهد گذاشت؟ پاسخ این سؤال مثبت است. به کمک این مثال ساده به روش دیگر برآورد به نام «برآورد بیزی»، اشاره می‌کنیم. در آمار بیزی، هم به نتایج به دست آمده از نمونه ($\frac{۱۸}{۴۰}$)، و هم به اطلاعاتی که خارج از نمونه در دسترس است، توجه می‌کنیم. در مثال ما، یک برآورد بیزی برای شانس مشاهدهٔ عدد ۱ عبارت است از $۰/۵۱ = \frac{(۱۸+۲۸)}{(۵۰+۴۰)}$.

همان‌طور که در ابتدای این نوشته اشاره شد، صورت ساده‌ای که می‌توان از روش آمار بیزی بیان کرد، این‌گونه است: «شما در مورد پدیده‌ای یک باور دارید که غالباً به آن **باور قبلی** می‌گویند. حال اطلاعات جدیدی به دستتان می‌رسد (یا نمونه می‌گیرید). واضح است با نوری که اطلاعات جدید می‌تاباند، باور قبلی تغییر می‌کند و باور جدید یا باور بعدی حاصل می‌شود». زیبایی آنالیزی که توماس بیز معرفی کرد، در این است که چگونه اطلاعات قبلی را با نمونه‌ای که خودمان می‌گیریم، ترکیب کنیم تا اطلاعات به‌روز شده را به دست آوریم. در مثال پرتاب سکه، ما این ترکیب را با جمع زدن تعداد یک‌های مشاهده شده در پرتاب‌های دوستتان و پرتاب‌های خودتان و تقسیمش بر کل پرتاب‌ها انجام دادیم. اما بحث در موارد کلی‌تر پیچیده‌تر از مثال مطرح شده است.

۶. سخن آخر

سؤال بااهمیتی که به ذهن می‌رسد این است که: «چرا علم آمار این‌قدر با احتمال عجین شده است؟» پاسخ به این سؤال دشوار نیست. به عبارت دیگر، واضح است زمانی که می‌خواهیم با استفاده از یکی از روش‌های فراوانی‌گرا یا بیزی دربارهٔ جامعه استنباط کنیم و دیدی از جامعه به دست آوریم که این دید برای قدم‌های بعدی، یعنی تصمیم‌گیری و برنامه‌ریزی لازم است، با پدیدهٔ «عدم قطعیت» مواجهیم. عدم قطعیت همه‌جا هست؛ از داشتن باور قبلی که مطمئناً با تردید همراه و بسیار شخصی است تا اخذ نمونه که غالباً تصادفی است و همچنین انجام استنباط که بر مبنای اطلاعات نمونه انجام می‌شود. ابزار مواجهه با عدم قطعیت، احتمال است. به گفتهٔ آماردان بزرگ **دنیس لیندلی** (۲۰۱۳-۱۹۲۳) احتمال درواقع «علم حساب دیفرانسیل عدم قطعیت» است. اگر بخواهیم با عدم قطعیت برخورد کنیم و یا به عبارتی آن را تحت کنترل درآوریم، باید از حساب احتمالات استفاده کنیم.

در مثال‌های کوتاه بالا به نقش احتمال در تجزیه و تحلیل حوادث پرداختیم. ادبیات غنی موجود در زمینهٔ آمار و احتمال و فعالیت‌های پژوهشی محققان، شامل مقاله‌ها، کتاب‌ها و کارگاه‌ها در این زمینه موجب ظهور و توسعهٔ حوزه‌های متفاوت علمی بین‌رشته‌ای، مانند آمار حیاتی، نظریهٔ بازی‌ها، هوش مصنوعی، یادگیری ماشین، داده‌کاوی، ریاضیات مالی، بیوانفورماتیک، یادگیری عمیق، پردازش تصویر و ... شده است.

منابع

۱. ایزد دوستدار، نوروز و پزشک، حمید (۱۳۹۶). احتمال و استنباط آماری (ج ۱). انتشارات دانشگاه تهران. چاپ ششم.

۲. حمید پزشک و نوروز ایزد دوستدار (۱۳۹۸). احتمال و استنباط آماری (ج ۲). انتشارات دانشگاه تهران. چاپ چهارم.

3. https://www.businessinsider.com