تفکر بیزی داشته باش

آمار بیزی ساده شد!

نویسنده: آلن. ب. داونی

مترجم: امیرعباس اسدی

فهرست مطالب

فصل ۱

قضیه‌ی بیز

۱.۱ احتمال شرطی

ایده‌ی بنیادی پشت تمام آمار بیزی قضیه‌ی بیز[1] است، قضیه ای که به طرز شگفت انگیزی بدست آوردنش آسان بوده و همچنین امکان درک احتمال شرطی را فراهم می کند.بحث را از احتمال آغاز کرده، سپس احتمال شرطی، قضیه‌ی بیز و در آخر آمار بیزی را بررسی می کنیم.

احتمال عددی بین 0 و 1 (شامل این دو هم می شود) و نشانگر درجه‌ی باور ما به یک حقیقت یا پیشبینی است. مقدار 1 نشان دهنده اطمینان از درستی یک حقیقت و یا اطمینان از به وقوع پیوستن یک پیشبینی است. مقدار 0 هم نشانگر نادرستی یک حقیقت است.

مقادیر میانی نشان دهنده‌ی درجه‌ی اطمینان و قطعیت هستند. مقدار 0.5 که اغلب به صورت 50% نوشته می شود، به این معناست که وقوع یک پیشبینی به اندازه‌ی عدم وقوع آن محتمل است. مثلا احتمال اینکه یک سکه‌ی پرتاب شده شیر باشد بسیار نزدیک 50% است.

یک احتمال شرطی[2] احتمالی بر اساس برخی از اطلاعات قبلی است. برای مثال من می خواهم بدانم احتمال اینکه در سال آینده دچار یک حمله‌ی قلبی شوم چقدر است. بر طبق CDC، سالانه حدود $785,000$ آمریکایی اولین حمله‌ی قلبی خود را تجربه می کنند.[3] جمعیت ایالات متحده حدودا 311 میلیون نفر است، بنابراین احتمال اینکه یک آمریکایی که به طور تصادفی انتخاب شده در سال آینده دچار یک حمله‌ی قلبی شود، حدودا 0.3% است. اما من که یک آمریکایی که به صورت تصادفی انتخاب شده باشد نیستم.همه گیر شناسان پی برده اند که عوامل زیادی بر میزان خطر وقوع حمله‌ی قلبی تاثیر می گذارند؛ بر اساس آن عوامل، احتمال اتفاق افتادن یک حمله‌ی قلبی برای من می تواند بالاتر یا پایین تر از میانگین باشد. من یک مرد هستم، 45 سال سن و کلسترول تقریبا بالایی دارم.این عوامل شانس ابتلاء را افزایش می دهند. در حالیکه، فشار خون من پایین است، سیگار نمی کشم و این عوامل شانس را کاهش می دهند. من با وارد کردن اطلاعاتم در یک محاسبه کننده‌ی آنلاین به آدرس `http://cvdrisk.nhlbi.nih.gov/calculator.asp` متوجه شدم احتمال اینکه سال آینده یک حمله‌ی قلبی را تجربه کنم چیزی حدود 0.2% و کمتر از میانگین است.این مقدار یک احتمال شرطی است چون بر اساس عواملی است که شرایط من محسوب می شوند.

معمولا احتمال شرطی را به صورت $p(A|B)$ نمایش می دهند که معنای آن احتمال درستی A با فرض درستی B است.در این مثال، A نشان

[1] Bayes's Theorem
[2] Conditional Probability
[3] `http://www.cdc.gov/heartdisease/facts.htm`

دهنده‌ی ابتلای من به حمله‌ی قلبی در سال آینده و B نشان دهنده‌ی شرایطی که نام بردم است.

۲.۱ احتمال توأم

احتمال توأم [۴] روشی برای بیان احتمال درستی دو چیز است. منظورمان از $p(A\,and\,B)$، احتمال درستی هر دو رخداد A و B است. اگر شما در مورد احتمال پرتاب سکه و تاس بدانید احتمالا فرمول زیر را آموخته اید.

$$p(A\,and\,B) = p(A)p(B) \quad \text{هشدار: این فرمول همیشه درست نیست}$$

برای مثال اگر دو سکه را پرتاب کنیم و معنی A شیر بودن سکه اول و معنی B شیر بودن سکه دوم باشد، خواهیم داشت $p(A) = p(B) = 0.5$ و بنابراین $p(A\,and\,B) = p(A)p(B) = 0.25$. اما این فرمول صرفا به این دلیل درست است که A و B مستقل هستند. به عبارت دیگر دانستن نتیجه‌ی رخداد اول، احتمال رخداد دوم را تغییر نمی دهد، مطابق تعریف $p(B|A) = p(B)$.
اکنون مثالی متفاوت که در آن A و B مستقل نیستند را بررسی می کنیم. فرض کنید A یعنی امروز باران می آید و B یعنی اینکه فردا باران می آید. اگر بدانیم که امروز باران آمده، امکان اینکه فردا هم باران ببارد بیشتر است بنابراین $p(B|A) > p(B)$.
به صورت کلی احتمال توأم برابر است با

$$p(A\,and\,B) = p(A)p(B|A)$$

برای هر دو رخداد A و B. با این اوصاف اگر احتمال باران آمدن در هر روز 0.5 باشد، احتمال باران آمدن در دو روز پشت سر هم 0.25 نبوده و احتمالا کمی بیشتر است.

۳.۱ مسئله‌ی کوکی

به زودی به سراغ قضیه‌ی بیز خواهیم رفت اما قبل از آن می خواهم اهمیت این قضیه را با یک مثال به نام مسئله‌ی کوکی روشن کنم[5]. فرض کنید دو کاسه پر از کوکی داریم. کاسه‌ی اول حاوی 30 کوکی وانیلی و 10 کوکی شکلاتی و کاسه‌ی دوم حاوی 20 عدد از هر کدام است.اکنون فرض کنید که یکی از کاسه ها را به صورت تصادفی انتخاب و بدون نگاه کردن در آن، یک از کوکی هایش را برمی داریم. اگر کوکی وانیلی باشد، چقدر احتمال دارد که از کاسه‌ی اول انتخاب شده باشد؟ این یک احتمال شرطی است؛ میخواهیم p(وانیلی|کاسه‌ی اول) را محاسبه کنیم که روش محاسبه اش چندان واضح به نظر نمی رسد. اما اگر به جای آن قرار بود احتمال وانیلی بودن کوکی را با فرض اینکه از کاسه‌ی اول انتخاب شده باشد محاسبه کنیم، سوال راحتی بود:

$$p(\text{کاسه‌ی اول|وانیلی}) = \frac{3}{4}$$

متاسفانه، $p(A|B)$ با $p(B|A)$ برابر نیست. اما راهی وجود دارد که با استفاده از یکی، دیگری را بدست آوریم: قضیه‌ی بیز.

۴.۱ قضیه‌ی بیز

اکنون هرآنچه که برای بدست آوردن قضیه‌ی بیز نیاز داریم را در اختیار داریم. بحث را از این حقیقت که ترکیب عطفی جابجایی پذیر است آغاز می کنیم؛ یعنی:

$$p(A\,and\,B) = p(B\,and\,A)$$

[4] Conjoint Probability

[5] بر اساس مثالی در http://en.wikipedia.org/wiki/Bayes'_theorem ، که دیگر در این صفحه وجود ندارد

برای هر دو رخداد A و B.
سپس مقدار احتمال توأم را جایگذاری می کنیم:

$$p(A\,and\,B) = p(A)p(B|A)$$

از آنجایی که معنای خاصی برای A و B در نظر نگرفته ایم، می توانیم آن ها را جابجا کنیم. با جابجایی آن ها خواهیم داشت:

$$p(B\,and\,A) = p(B)p(A|B)$$

این تمام چیزی بود که نیاز داشیم. اکنون اگر این ها را کنار هم بگذاریم، بدست می آید:

$$p(B)p(A|B) = p(A)p(B|A)$$

این یعنی دو راه برای محاسبه‌ی احتمال توأم داریم. اگر مقدار $p(A)$ را داشته باشید، می توانید آن را در احتمال شرطی $p(B|A)$ ضرب کنید. و یا اینکه می توانید از ضرب $p(B)$ در $p(A|B)$ برای محاسبه استفاده کنید. هر کدام را که انتخاب کنید، نتیجه یکسان است. در آخر طرفین را بر $p(B)$ تقسیم می کنیم:

$$p(A|B) = \frac{p(A)p(B|A)}{p(B)}$$

و این قضیه‌ی بیز است! شاید در نگاه اول به نظر نرسد، اما به طرز شگفت انگیزی قدرتمند است.
برای مثال می توانیم با استفاده از آن مسئله‌ی کوکی را حل کنیم. از B_1 برای نشان دادن رخداد بیرون آمدن کوکی از کاسه‌ی اول و از V برای نشان دادن وانیلی بودن کوکی استفاده می کنیم. با استفاده از قضیه‌ی بیز خواهیم داشت:

$$p(B_1|V) = \frac{p(B_1)p(V|B_1)}{p(V)}$$

عبارت سمت چپ همان چیزی است که به دنبالش بودیم: احتمال اینکه کوکی از کاسه‌ی اول انتخاب شده باشد، اگر بدانیم طعم آن وانیلی است. سمت راست عبارت است از:

- $p(B_1)$: احتمال اینکه بدون در نظر گرفتن طعم کوکی، کاسه‌ی اول را انتخاب کنیم. از آنجایی که طبق مسئله کاسه به صورت تصادفی انتخاب می شود، می توانیم فرض کنیم $p(B_1) = \frac{1}{2}$.

- $p(V|B_1)$: یعنی احتمال انتخاب یک کوکی وانیلی از کاسه‌ی اول که برابر است با $\frac{3}{4}$.

- $p(V)$: یعنی احتمال انتخاب کوکی وانیلی از هر دو کاسه. از آن جایی که شانس انتخاب کاسه ها یکسان است و کاسه ها حاوی تعداد یکسانی کوکی هستند، شانس انتخاب هر کوکی یکسان است. در دو کاسه مجموعا 50 کوکی وانیلی و 30 کوکی شکلاتی است، پس $p(V) = \frac{5}{8}$.

با جایگذاری در قضیه‌ی بیز بدست می آید:

$$p(B_1|V) = \frac{\frac{1}{2} \times \frac{3}{4}}{\frac{5}{8}}$$

که با $\frac{3}{5}$ برابر است. پس انتخاب کوکی وانیلی برهانی [6] برای کاسه‌ی اول است، چون یرون آمدن کوکی های وانیلی از کاسه‌ی اول محتمل تر است.
این مثال یکی از استفاده های قضیه‌ی بیز را نشان می دهد: فراهم کردن روشی برای بدست آوردن $p(B|A)$ از $p(A|B)$. این استراتژی در مواردی مانند مسئله‌ی کوکی، که محاسبه‌ی سمت راست قضیه‌ی بیز آسان تر از سمت چپ آن است، به کار می آید.

[6] evidence

۵.۱ تفسیر در طول زمان

می توان به قضیه‌ی بیز از زاویه‌ی دیگری هم نگاه کرد: این قضیه راهی برای تصحیح و بروزرسانی احتمال یک فرض مثل H را در صورت مشاهده‌ی داده هایی جدید مانند D فراهم می کند. این شیوه‌ی فکر کردن به قضیه‌ی بیز **تفسیر در طول زمان**[۷] نامیده می شود.احتمال فرضیات در طول زمان مشاهده‌ی داده های جدید تغییر می کند. اگر قضیه‌ی بیز را با H و D بازنویسی کنیم، خواهیم داشت:

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)}$$

در این تفسیر هر کدام از عبارات بالا نامی دارند:

- $p(H)$ احتمال فرض پیش از مشاهده‌ی داده ها است که احتمال پیشین یا به صورت مختصر پیشین[۸] نامیده می شود.
- $p(H|D)$ چیزی است که می خواهیم محاسبه کنیم، احتمال فرضیات پس از مشاهده‌ی داده ها است که پسین[۹] نامیده می شود.
- $p(D|H)$ احتمال تولید ها داده ها تحت فرض است که درست نمایی[۱۰] نامیده می شود.
- $p(D)$ احتمال تولید داده ها تحت هر فرض ممکن است که ثابت نرمال سازی [۱۱] نامیده می شود.

گاهی اوقات می توانیم احتمال پیشین را بر اساس اطلاعات پیشفرض خود محاسبه کنیم. مثلا در مسئله‌ی کوکی مشخص شده است که احتمال انتخاب کاسه ها یکسان است. در دیگر موارد ممکن است پیشین به صورت ذهنی انتخاب شود و افراد به دلیل اطلاعات پیشفرض متفاوت و یا تفسیر متفاوت از اطلاعات یکسان، بر سر شیوه‌ی محاسبه‌ی آن دچار اختلاف شوند. محاسبه‌ی درست نمایی معمولا آسان ترین قسمت کار است.در مسئله‌ی کوکی، اگر بدانیم کوکی از کدام کاسه انتخاب شده، می توانیم احتمال وانیلی بودن کوکی را محاسبه کنیم. محاسبه‌ی ثابت نرمال سازی گاهی نیاز به حقه دارد.این مقدار به عنوان احتمال مشاهده‌ی داده تحت هر فرض ممکن تعریف می شود اما در موارد کلی تعیین اینکه دقیقا چه معنایی دارد، دشوار است. در اغلب موارد با متمایز کردن مجموعه هایی از فرضیات که ویژگی های خاصی دارند، ساده سازی انجام می دهیم.این دو ویژگی عبارتند از:
دو به دو ناسازگار[۱۲] : حداکثر یکی از فرض های مجموعه می تواند درست باشد و
کاملا در بر گیرنده[۱۳] : حداقل یکی از فرضیات باید درست باشد.
در ادامه برای مجموعه هایی از فرضیات که دارای این ویژگی ها هستند کلمه‌ی مناسب را استفاده می کنیم. در مسئله‌ی کوکی تنها دو فرض داشتیم – کوکی از کاسه اول یا دوم بیرون آمده است – که این دو فرض دو به دو ناسازگار و کاملا در بر گیرنده هستند.

۶.۱ مسئله‌ی M & M

M&M ها آب نبات های کوچکی هستند که در رنگ های مختلف عرضه می شوند. شرکت مارس، که تولید کننده این آب نبات ها است، هر چند وقت یکبار ترکیب رنگی آن ها را عوض می کند. در سال ۱۹۹۵ آب نبات های آبی معرفی شدند. پیش از آن ترکیب رنگی آب نبات های داخل یک بسته عبارت بود از 30% قهوه ای، 20% زرد، 20% قرمز، 10% سبز، 10% نارنجی و 10% برنزی. بعد از آن به 24% آبی، 20% سبز، 16% نارنجی، 14% زرد، 13% قرمز و 13% قهوه ای تغییر کرد.

[7] The Diachronic Interpretation
[8] prior
[9] posterior
[10] likelihood
[11] normalizing constant
[12] Mutually exclusive
[13] Collectively exhaustive

فرض کنید دوست من دو کیف از این آب نبات ها دارد و به من می‌گویید که یکی از سال ۱۹۹۴ و یکی از سال ۱۹۹۶ است.او به من نمی‌گوید که کدام، کدام است ولی از هر کدام یک آب نبات به من می دهد.یکی زرد است و دیگری سبز. احتمال اینکه آب نبات زرد از کیف سال ۱۹۹۴ باشد، چقدر است؟ این مسئله، مشابه مسئله‌ی کوکی است با این تفاوت که از هر کاسه یک نمونه برداشته شده است.این مسئله بهانه ای شد تا روش جدولی را معرفی کنم که برای حل اینگونه مسائل با استفاده از قلم و کاغذ مفید است.در فصل بعدی این مسائل را با استفاده از روش های الگوریتمی حل خواهیم کرد. قدم اول بررسی فرض است. کیفی که آب نبات زرد از آن بیرون آمده را کیف ۱ و دیگری را کیف ۲ می گوییم. با این اوصاف فرضیات عبارتند از:

- A : کیف ۱ از سال ۱۹۹۴ و کیف ۲ از سال ۱۹۹۶ است.
- B : کیف ۱ از سال ۱۹۹۶ و کیف ۲ از سال ۱۹۹۴ است.

اکنون جدولی تشکیل می دهیم که هر سطر آن متناظر با یک فرض و هر ستون آن متناظر با قسمتی از قضیه‌ی بیز است:

فرض H	پیشین $p(H)$	درست نمایی $p(D\|H)$	$p(H)p(D\|H)$	پسین $p(H\|D)$
A	$\frac{1}{2}$	$(20)(20)$	200	$\frac{20}{27}$
B	$\frac{1}{2}$	$(14)(10)$	70	$\frac{7}{27}$

ستون اول شامل احتمالات پیشین است. بر اساس داده های مسئله منطقی است که فرض کنیم $p(A) = p(B) = \frac{1}{2}$.ستون دوم درست نمایی را نشان می دهد که با استفاده از اطلاعات مسئله محاسبه شده است.مثلا اگر A درست باشد، آب نبات زرد از کیف سال ۱۹۹۴ با احتمال 20% و آب نبات سبز از کیف سال ۱۹۹۶ با احتمال 20% آمده است.اگر B درست باشد، آب نبات زرد با احتمال 14% از کیف سال ۱۹۹۶ و آب نبات سبز از کیف سال ۱۹۹۴ با احتمال 10% آمده است. چون انتخاب ها مستقل از یکدیگر انتخاب شده اند، احتمال توأم را با ضرب آن ها بدست می آوریم. ستون سوم صرفا حاصلضرب دو ستون قبلی است. مجموع این ستون یعنی 270 همان ثابت نرمال سازی است. برای محاسبه‌ی آخرین ستون، ستون سوم را بر ثابت نرمال سازی تقسیم می کنیم. تموم شد. به همین راحتی!

۷.۱ مسئله‌ی مونتی هال

۸.۱ بحث