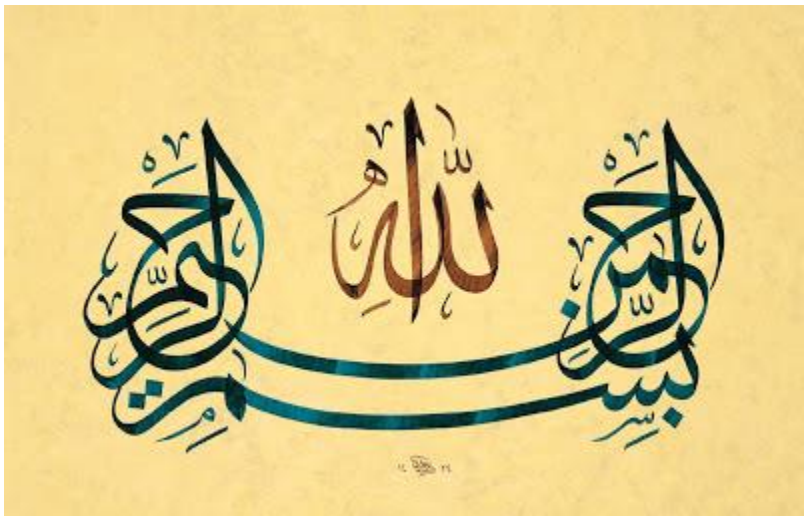
آمار بیزین

نویسندگان : آنتونی اوگان، بریان آرلوس
مترجم : دکتر فرشید علاءالدینی

نویسندگان: آنتونی اوگان، بریان آرلوس
آمار بیزین
مترجم: دکتر فرشید علاء الدینی


بسم الله الرحمن الرحيم

آمار بیزین

نویسندگان : آنتونی اوگان، بریان آرئوس

مترجم : فرشید علاءالدینی

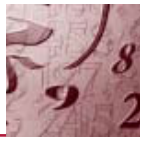
ناشر : مولف

تیراژ : 1000 جلد

نوبت چاپ : اول

قیمت : 30000 تومان

سال 1397



فهرست

پیش‌گفتار و تاریخچه‌ای مختصر ۱
مرور کلی ۷
بخش ۱- استنباط ۹
بخش ۲- روش بیزین ۱۴
بخش ۳- اطلاعات پیشین ۱۸
بخش ۴- مشخص‌سازی پیشین ۲۲
بخش ۵- محاسبات ۲۵
بخش ۶- طراحی و تحلیل کارآزمایی‌ها ۲۷
بخش ۷- مدل‌های اقتصادی ۲۹
ضمیمه ۳۲

پیش‌گفتار و تاریخچه‌ای مختصر

اجازه دهید سخنم را با گفتن این مطلب آغاز کنم که در دهه‌ی ۱۹۷۰ با این‌که دانش تحلیل به شیوه‌ی بیزین را فراگرفته بودم اما به علت آن‌که محاسبات ما عملی به نظر نمی‌آمدند به تدریج این حیطه را به فراموشی سپردم. دو دهه‌ی بعد، در دهه‌ی ۱۹۹۰، دریافتم که تحلیلگران بیزین به واسطه‌ی روش‌های محاسباتی زنجیره‌ی مارکوف مونت کارلویی (MCMC) به پیشرفت‌های قابل ملاحظه‌ای دست یافته‌اند و نهایتاً نرم‌افزار محاسباتی این شیوه نیز ابداع شد. از آن زمان به بعد، مجدداً به بکارگیری ابزارهای بیزین متمایل شدم و شوق پیوستن به جامعه‌ی جهانی پرشور و رو به فزونی تحلیلگران بیزین که در حل مشکلات زندگی واقعی پیشتاز شده بودند در من زنده شد.

برای برخی از خوانندگان مخصوصاً آن‌هایی که اهل آمار کلاسیک یا فراوانی‌گرا هستند، شیوه‌ی گفتار این درسنامه‌ی مقدماتی ممکن است کمی مشکل به نظر برسد. هنگامی که طرفداران آمار بیزین دوباره فرصت ارائه‌ی دیدگاه‌هایشان را پیدا کنند بحثی بسیار قدیمی مجدداً مطرح خواهد شد. تحلیلگران بیزین مدت‌ها بر این عقیده بودند که ریاضیِ احتمالات و استنباط کاملاً به نفع آن‌ها است و صرفاً آمار رایج است که این مسئله را نفی می‌کند. طبعاً این واقعیت برای آن‌ها ناخوشایند بود. با این‌حال، زمانه در حال تغییر است و امروزه شاهد آغاز یک همگرایی هستیم، به نحوی که تحلیلگران فراوانی‌گرا به ارزش اهداف و شیوه‌های تحلیلگران بیزین پی‌برده‌اند و حامیان شیوه‌ی بیزین به آن دسته از تکنیک‌های محاسباتی اشراف پیدا کرده‌اند که بین روش‌های مورد استفاده‌ی آن‌ها و نیازهای علوم کاربردی پلی ایجاد می‌کنند. گفتگو در مورد دیدگاه بیزین ممکن است بی‌سرانجام باشد زیرا به باور ما شیوه‌های فعلی به لحاظ منطقی ایراد دارند و در عین حال بسیاری از افراد آن‌ها را مسلّم فرض می‌کنند. در واقع، برخی از فراوانی‌گراها نیز که معتقدند تحلیلگران بیزین در تلاش برای مجاز دانستن تغییر عقاید در حیطه‌ی علم هستند به همین اندازه احساس عجز می‌کنند. عجیب این‌که با وجود شدت گرفتن بحث، اختلاف‌نظری بر سر صحت ریاضیات وجود ندارد. اختلاف‌نظر اساسی در مورد تعریف واحدی است که همه چیز از آن منشاء می‌گیرد.

چرا این بحث قدیمی چنین هیجانی را برانگیخته است؟ در سال ۱۹۲۵، ایگون پی‌یرسون در ژورنال جدیدی به نام بیومتریکا در مقاله‌اش چنین نوشت:

> هم حامیان و هم مخالفان آن‌چه قضیه‌ی بیز نامیده می‌شود تقریباً به طور کامل به منطق استدلالشان متکی بوده‌اند؛ این امر از زمانی که پرایس، که یادداشت‌های [توماس] بیز را در محضر انجمن سلطنتی در سال ۱۷۶۳ بیان می‌کرد، و برای اولین بار دست روی قاعده‌ی مسلّمی گذاشت که فردی که تازه پا به این جهان می‌گذارد باید انتظارش از طلوع‌های بعدی آفتاب را بر آن مبنا قرار دهد، تا این اواخر که کینز طی رساله‌ای در باب احتمال (۱۹۲۱) متذکر شده است که اتکا به چنین قضیه‌ی احمقانه‌ای شرم‌آور است مصداق داشته است.

شایان ذکر است که پی‌یرسون که بعدها عمدتاً با مکتب فراوانی‌گرا و به طور خاص با لم نیمن پی‌یرسون شناخته شد، از صحت روش بیزین در این مقاله حمایت می‌کند. مقاله‌ای مروری و قابل دسترس پیرامون فلسفه و روش‌های بیزین که مقاله‌ای کلاسیک قلمداد می‌شود توسط ادواردز، لیدمن، و ساوج در سال ۱۹۶۳ نوشته شده است. نقل کردن شرح سابقه‌ی آن‌ها ارزش دارد:

> قضیه‌ی بیز واقعیتی ساده و بنیادین راجع به احتمالات است که به نظر می‌آید که برای توماس بیز هنگامی که مقاله‌ی مشهورش را می‌نوشته واضح بوده است ...، در عین حال که وی این قضیه را به وضوح بیان نکرده است. به این ترتیب، از آمار بیزین به سبب این دلیل نسبتاً ناکافی نام برده می‌شود که موارد فراوان‌تری برای اعمال قضیه‌ی بیز در مقایسه با [اعمال] آمار کلاسیک وجود دارد. بنابراین، از یک دیدگاه کلی، آمار بیزین حداقل از سال ۱۷۶۳ قدمت دارد.
>
> به بیان دقیق‌تر می‌توان گفت که آمار بیزین از سال ۱۹۵۹ با انتشار احتمالات و آمار برای اتخاذ تصمیم‌های اقتصادی توسط رابرت شلیفر پا به عرصه گذاشت. این درس‌نامه‌ی مقدماتی برای اولین بار پیاده‌سازی عملی ایده‌های مهم آمار بیزین را شرح می‌دهد: این‌که احتمال مجموعه‌ای از دیدگاه‌های سازمان‌یافته است، و این‌که استنباط با تکیه بر داده‌ها چیزی بیش از بازبینیِ چنین نظری (احتمال قبلی) به سبب اطلاعات مربوطه‌ی جدید نیست.

این بخش دو عقیده‌ی مهم را در بر دارد. اولی در رابطه با تعریف «احتمال» است، و دوم اینکه، گرچه اندیشه‌های در پسِ آمار بیزین در پایه‌ی علم آمار جای دارند، اما آمار بیزین آنچنان پیشرفت کرده که تصمیم‌گیری را ساده کرده است.

احتمال ریاضیاتی است که برای توصیف عدم قطعیت استفاده می‌شود. دیدگاه غالب امروزی در علم آمار، که در این درس‌نامه دیدگاه «فراوانی‌گرا» نامیده شده است، احتمال یک رویداد را حد فراوانی نسبی تعریف می‌کند که به آن میزان در مجموعه‌ای از مشاهدات مرتبط و قابل انتظار که احتمال وقوع آن رویداد در آن‌ها وجود دارد به وقوع می‌پیوندد؛ شایان ذکر است که احتمال دارد این مجموعه کاملاً فرضی باشد. در نظر یک فراوانی‌گرا، جایگاه عدم قطعیت در رویدادها است. به بیان دقیق‌تر، فراوانی‌گرا صرفاً در تلاش برای تعیین «احتمال یک رویداد» است که مشخصه‌ی مجموعه‌ای از رویدادهای مشابه محسوب می‌شود که لااقل در کل نسخه‌هایی تکراری هستند. یک تحلیلگر بیزین هر رویداد را منحصر به فرد می‌داند، به طوری که این رویداد ممکن است رخ دهد یا ندهد؛ وی معتقد است که احتمال یک رویداد عددی برای اشاره به نظر مشاهده‌گر مربوطه درباره‌ی این مسئله است که آیا در یک مشاهده‌ی خاص آن رویداد رخ می‌دهد یا نمی‌دهد. در نظر تحلیلگر بیزین، عدم قطعیتی که با احتمال توصیف می‌شود به مشاهده‌گر ربط پیدا می‌کند. به این ترتیب یک تحلیلگر بیزین به صحبت پیرامون احتمال وقوع رویدادی منحصر به فرد تمایل دارد. خوانندگان علاقه‌مند می‌توانند طرز عمل ریاضی و فلسفیِ

مفاهیم مختلف احتمال را در مقاله‌ی کایبورگ و اسموکلر (۱۹۶۴) بیابند.

مایه‌ی تاسف است که این دو تعریف با معانی مازاد برچسب خورده‌اند. فراوانی‌گراها احتمالاتشان را «عینی - objective» می‌دانند؛ در حالی که احتمالات تحلیلگران بیزین «ذهنی - subjective» نامیده می‌شوند. به دلیل مفاهیم مازاد جای گرفته در این برچسب‌ها این تعاریف قطب متضاد یکدیگر قلمداد می‌شوند. ذهنیت برای یک فرآیند علمی ویژگی نامطلوبی محسوب می‌شود و به دلبخواهی بودن و سوگیری دلالت دارد. گفته می‌شود که روش‌های فراوانی‌گرا عینی هستند، بنابراین، دلبخواهی بودن در آن‌ها جایگاهی ندارد، و به همین سبب برای تحقیقات علمی و تحقیقات به شکلی که آزمونگر از نزدیک با آزمودنی در تعامل نیست مناسب‌تر هستند.

هیچکدام از این دو کرانه، این دو دیدگاه را به طور کامل وصف نمی‌کنند. متاسفانه، نزدیک به یک قرن است که سردرگمی ناشی از این برچسب‌ها هیجانات بی‌جهتی را در هر دو جبهه برانگیخته است.

از دیدگاه بیزین، برای یک رویداد ممکن است به تعداد مشاهده‌گران احتمالات مختلفی وجود داشته باشد. در واقع دلیل مسابقات اسب دوانی همین است. این تعدّد برای فرد فراوانی‌گرا تشویش‌برانگیز است؛ جهان‌بینی او با (به طور کلی) نمونه‌گیری مکرر بلندمدت حکم می‌کند که برای هر رویداد احتمال یگانه‌ای وجود داشته باشد. اما داشتن دیدگاه ذهنی نسبت به احتمال به مفهوم به صورت دلبخواه بودن احتمال نیست. ادواردز و همکاران از صفت سازمان‌یافته برای توصیف «دیدگاه‌ها» استفاده کرده‌اند. احتمال ذهنیِ بیزین subjective probability of the Bayesian در معنای خاص باید سازمان‌یافته باشد به گونه‌ای که از تمام قوانین ریاضیات مربوط به محاسبه‌ی احتمال تبعیت کند، و مخصوصاً باید به شیوه‌ی خاصی که قضیه‌ی بیز حکم می‌کند و با در نظر گرفتن داده‌های جدید بازبینی شود. این قضیه - که دو دیدگاه احتمال را به هم پیوند می‌زند - بیان می‌کند که در شرایطی که مجموعه‌ی طویلی از مشاهدات مربوط به رخدادها و عدم رخدادهای یک رویداد داریم، مهم نیست که نظرات مشاهده‌گران بیزین در ابتدای این مجموعه تا چه حد از یکدیگر متفاوت بوده است، زیرا آن‌ها نظراتشان را در حین اضافه شدن هر مشاهده‌ی جدیدی تغییر می‌دهند. پس از مشاهدات بسیار، نظرات آن‌ها به مقدار عددی تقریباً یکسانی برای احتمال نزدیک خواهد شد. به علاوه، از آن‌جایی که می‌توان برای این رویداد مشاهدات بلند مدتی را ارائه کرد، لم lemmaقضیه می‌گوید که آن مقدار عددی که مشاهده‌گران روی آن به توافق رسیده‌اند، در حد، دقیقاً برابر با فراوانی نسبی بلند مدت long-run relative frequency است.

از این‌رو، هنگامی که مشاهدات فراوانی وجود دارد، تحلیلگران بیزین و فراوانی‌گراها در احتمالاتی که به رویدادها نسبت می‌دهند به همگرایی تمایل دارند. پس مشکل چیست؟

دو مشکل وجود دارد. نخست، رویدادهایی وجود دارند - شاید هم کسی بگوید اکثر رویدادهای موردنظر در زمان تصمیم‌گیری‌های دنیای واقعی - که برای آن‌ها صرفاً در یک آزمون به داده‌های فراوان مرتبطی دست پیدا نمی‌کنیم. در این موارد، هم بیزین‌ها و هم فراوانی‌گراها باید در مورد این‌که کدام

داده‌ها را ادغام کرده و کدام داده‌ها را ادغام نکنند به قضاوت‌های ذهنی اتکا کنند. تحلیلگران بیزین تمایل به جامع نگری دارند، اما وزنی که در تحلیل ادغام شده pooled analysis به داده‌ها می‌دهند طبق درکی است که از ارتباط داده‌ها با برآورد در دسترس دارند. تحلیلگران بیزین ممکن است نهایتاً برآوردهای متفاوتی از احتمال ارائه دهند زیرا در ابتدا نظرات کاملاً متفاوتی داشته‌اند و در ضمن این‌که داده‌ها مهمتر از توزیع‌های پیشین priors نیستند، و یا ممکن است به علت این‌که این تحلیلگران قضاوت متفاوتی نسبت به ارتباط داشته‌اند به داده‌های ادغام شده به طرز متفاوتی وزن داده باشند. فراوانی‌گراها به طور ذهنی تصمیم می‌گیرند (زیرا معیاری کاملاً عینی برای «ارتباط» وجود ندارد) که کدام داده‌ها مرتبط و کدام‌ها غیرمرتبط هستند و آن داده‌های به نظر مرتبط را با حداکثر وزن با توجه به داده‌های موجود ادغام می‌کنند. فراوانی‌گراهایی که درباره‌ی ارتباط مجموعه داده‌های مختلفِ از پیش موجود اختلاف نظر دارند، در مورد برآوردهایشان از احتمالات نهایی رویدادها نیز اختلاف نظر دارند. مثال مشخص در این مورد در سال ۲۰۰۲ در بحث پرسروصدایی راجع به این پرسش روی داد که آیا ماموگرافی برای غربالگری موجب کاهش مرگ‌ومیر ناشی از سرطان پستان می‌شود یا خیر. این بحث هنوز به نتیجه نرسیده است.

دومین مشکل این است که بیزین‌ها و فراوانی‌گراها بر سر این مسئله که کدام رویدادها برای تعیین احتمالات مناسب و معنی‌دار هستند اختلاف نظر دارند. بیزین‌ها احتمال یک فرضیه‌ی خاص را با توجه به داده‌های مشاهده شده محاسبه می کنند. ادواردز و همکاران شروع دوران بیزین را از زمان انتشار کتابی درباره‌ی استفاده از آمار برای اتخاذ تصمیم‌های تجاری به حساب آورده‌اند، دلیل این امر این است که احتمالی که یک رویداد خاص کسب می‌کند (یا احتمال اینکه یک فرضیه صحیح است)، با در نظر گرفتن داده‌ها، دقیقاً آن چیزی است که برای تصمیم‌گیری‌ها لازم است و به آن رویداد (یا فرضیه) بستگی دارد. متاسفانه، در بحث ریاضیات، نمی‌توان این احتمال خاص را بدون ارجاع به برخی از احتمالات پیشین آن رویدادِ، پیش از جمع‌آوری داده‌ها محاسبه کرد. و، در نظر گرفتن احتمال پیشین موضوع ذهنیت در احتمال را مطرح می‌کند.

برای اجتناب از این تنگنا، فراوانی‌گراها - مخصوصاً آر.ای فیشر، جِی نیمن و ای. پی‌یرسون - تلاش کردند تا قدرت شواهد را مستقل از احتمالات پیشین فرضیه‌ها بیان کنند. فیشر مقدار احتمال (P) را ابداع کرد، و نیمن و پی‌یرسون آزمون فرضیه‌ی صفر را با استفاده از مقدار P اختراع کردند.

چندین سال پیش (۱۹۹۹)، گودمن به زیبایی تاریخچه و نتایج این روند را در مقاله‌ای که شیوه‌ی بیان فوق العاده‌ای دارد خلاصه کرده است. یک متخصص آمار با بهره گرفتن از روش نیمن و پی‌یرسون و مقادیر احتمال برای رد فرضیه‌های صفر در سطح ۵٪، به طور متوسط در طولانی‌مدت (فرض کنیم در تمام طول مدت حرفه‌ی آن متخصص آمار)، فقط در حدود ۵٪ از مواقع به اشتباه یک فرضیه‌ی صفرِ صحیح را رد می‌کند. با این‌حال، محاسبات درباره‌ی نمونه‌ای خاص با مجموعه‌ی خاصی از داده‌ها و فرضیه‌ی صفری خاص که رویدادی منحصر به فرد و نه قابل تکرار است اطلاعاتی به دست نمی‌دهند.

صرفاً با تکیه بر داده‌ها نمی‌توان گفت که چقدر احتمال دارد که فرضیه‌ی صفر در نمونه‌ای خاص صحیح باشد. در بهترین حالت داده‌ها به شما می‌گویند که چقدر باید از احتمال پیشین خود دور شوید تا فرضیه صحیح باشد. یک تحلیلگر بیزین می‌تواند این احتمال را محاسبه کند، زیرا از نظر یک بیزین توضیح احتمال پیشین یک رویداد بی‌همتا منطقی است.

در واقع، همان‌طور که گودمن شرح بیشتری بر آن داد، نیمن و پی‌یرسون زیرک بودند و همانند سایر آماردانان هوشمند دریافتند که آزمون فرضیه‌ها مشکل آن‌ها را برطرف نمی‌کند. یک پاسخ که در جامعه‌ی فراوانی‌گراها مطرح شد جایگزین کردن آزمون فرضیه‌ها با برآورد فاصله‌ای interval estimation بود؛ به نظر می‌رسد که برآورد به اصطلاح فواصل اطمینان در بردارنده‌ی مقدار پارامتری باشد که فرضیه بر آن استوار است. متاسفانه برآورد فاصله هم این مسئله را حل نکرد، بلکه آن را به سمت ریاضیات پیچیده سوق داد و، اگر هم برای آن راه حلی ارائه داده باشد، صرفاً آن را غامض کرده است.

پس همه‌ی این‌ها برای کسی که آمار فراوانی‌گرا را فراگرفته است یا برای کسی که می‌خواهد از شیوه‌های بیزین سر در بیاورد چه مفهومی دارد؟ بگذارید این شخص را «شما» تصور کنیم.

همه‌ی این مسائل حداقل به این مفهوم است که «شما» برای محاسبه‌ی فواصل به روش جدیدی دست پیدا می‌کنید که بسیار شبیه روش‌هایی است که در محاسبات مرسوم از فواصل اطمینان از آن‌ها استفاده می‌شود. تنها امتیاز شما در دانستن این مسئله است که یک فاصله‌ی مشخص به احتمال معینی پارامتر را در بر می‌گیرد، که این امر در مورد فاصله‌ی تقریباً یکسان محاسبه شده به شیوه‌ی مرسوم صحت ندارد. گرچه که این یافته‌ی چندان بزرگی هم نیست.

همه‌ی این‌ها این مفهوم را نیز می‌رساند که باید در مورد مسئله‌ی آماری که در حال حل آن هستید متفاوت بیاندیشید و این یعنی زحمت بیشتر. مخصوصاً ممکن است که برای تعیین احتمال پیشینی که قابل اثبات به دیگران باشد مجبور به تلاش بسیار باشید. گرچه ممکن است معذب شوید، اما تحلیلگران بیزین در تلاش برای یافتن روش‌هایی برای کمک به «شما» در مورد فرآیند درک و تعیین احتمالات پیشین هستند و همچنین در تلاش‌اند تا استدلال‌هایی را برای اثبات این احتمالات به «شما» پیشنهاد دهند.

توضیح می‌دهیم که در عوض چه عایدتان می‌شود. نخست، در هر تحلیل معینی برای مجموعه داده‌ای معین و فرضیه‌ای معین (نه فقط فرضیه‌ی صفر) قادر به محاسبه‌ی احتمال صحیح بودن فرضیه خواهید بود. یا، مورد سودمندتر این‌که، قادر به تعیین احتمال این خواهید بود که مقدار واقعی پارامتر درون هر فاصله‌ی مفروضی قرار دارد. این همان چیزی است که برای تصمیم‌گیری های کمّی و برای سنجش هزینه‌ها و فواید تصمیم‌ها بسته به این برآوردها لازم است.

در ثانی، همان‌طور که داده‌هایی جدید در رابطه با مسئله‌تان جمع‌آوری می‌کنید، روش آسانی را برای ارزیابی برآورد خود به شیوه‌ای سازمان‌یافته و قابل دفاع در اختیار خواهید داشت.

این دو فایده مورد سومی را هم حاصل می‌کنند: با این طرز تفکر و محاسبه دیگر نباید نگران لو رفتن داده‌هایتان قبل از پایان از پیش تعیین‌شده‌ی آزمون باشید. در واقع، مجموعه‌ی جدیدی از ابزارها را در اختیار خواهید داشت و با بهره‌گیری از آن‌ها و با قواعد توقف اختیاری optional stopping rules قادر به بهینه‌سازی اندازه‌ی آزمون‌ها خواهید بود. این موضوع بسیار پیشرفته‌ای در روش‌های بیزین محسوب می‌شود که بسیار فراتر از سطح این درس‌نامه‌ی مقدماتی است، اما منابع مطالعاتی مربوط به آن مدام بیشتر می‌شوند.

فایده‌ی دیگر این است که افراد دیگری که به نتایج انتشار یافته‌ی شما به عنوان نسبت هزینه - اثربخشی برای انجام محاسبات مشابه متکی هستند قادر خواهند بود تا عدم قطعیت را به شیوه‌ای معنادار برای تعیین دقت نتایجشان مستقیماً لحاظ کنند. گرچه این واقعیت دستاورد غیرمستقیمی برای شما خواهد بود اما به ارزش تحلیل‌هایتان خواهد افزود.

نفع پنجم - که ناشی از پیشرفت در روش‌های محاسباتی ایجاد شده به سبب نیازهای بیزین‌ها است - این است که شما، به منظور نشان دادن فرآیندهای ایجاد داده‌ها، به سهولت قادر به برآورد توزیع‌ها برای توابع پارامترهایی هستید که به نوبت در مدل‌های آماری بسیار پیچیده برآورد شده‌اند. «شما» از اتکا به صورت‌بندی‌های formulation ساده‌انگارانه‌ی درست‌نمایی داده‌ها data likelihood صرفاً به منظور امکان استفاده از آزمون‌های استاندارد خلاص خواهید شد. این امر به شیوه‌های مختلف با پیشرفت‌های عظیمی قابل قیاس است که در برآورد مدل‌های رگرسیون بسیار پیچیده در مقایسه با مدل‌های خطی ساده در سال‌های گذشته داشته‌ایم.

در نهایت، این را بدانید که با این دانش به سرعت سازگار خواهید شد. تغییر عظیمی در علم آمار در حال وقوع است، و توانایی درک، به کار گرفتن و انتقاد از تحلیل‌های بیزین برای محققان و پزشکان در آینده‌ای نزدیک حائز اهمیت خواهد بود. امیدوارم همه‌ی این منافعی که برشمردیم روزی برایتان مفید واقع شوند، که البته، لازمه‌ی آن این است که برای یادگیری برخی از اصول که ممکن است از همان لحظه‌ی اول چندان کارساز به نظر نیایند وقت صرف کنید. اگر در این مسیر ثابت قدم بمانید احتمال موفقیتتان را بالا می‌دانم.

مرور کلی

این درس‌نامه‌ی مقدماتی برای اقتصاددانان سلامت، متخصصین پژوهش و آماردانانی است که به درک مبانی آمار بیزین و این‌که چگونه می‌توان از روش‌های بیزین در ارزیابی اقتصادیِ فن‌آوری‌های مراقبت‌های سلامت بهره گرفت علاقمند هستند. برای مطالعه‌ی این درس‌نامه به دانش قبلی پیرامون آمار بیزین نیازی نیست. تنها انتظاری که از خواننده می‌رود این است که از تکنیک‌های مرسوم غیر بیزین مانند برآورد نااریب، فواصل اطمینان و آزمون‌های معنادار بودن درک پایه‌ای داشته باشد؛ آن رویکرد مرسوم در آمار «فراوانی‌گرا» نامیده می‌شود.

این درس‌نامه‌ی مقدماتی در پاسخ به علاقه‌ی روزافزون و همچنین پذیرش روش‌های بیزین در حوزه‌ی اقتصاد سلامت نوشته شده است. برای مثال، در انگلستان، National Institute for Clinical Excellence در راهنمای خود به طور خاص به پذیرش شیوه‌های بیزین اشاره کرده است تا حامیان مالی‌اش در این مورد اعلام نظر کنند. در ایالات متحده، سازمان غذا و دارو (FDA) نیز مخصوصاً در حوزه‌ی تجهیزات پزشکی رویکرد مثبتی نسبت به عقاید بیزین دارد. این بروز علاقه به رویکرد بیزین به هیچ وجه منحصر به این حوزه نیست، با این‌حال، در آغاز قرن بیست و یکم شاهد افزایش چشمگیر کاربرد روش‌های بیزین در علم، فن‌آوری، علوم اجتماعی، مدیریت و تجارت هستیم. یافتن علت‌های مربوطه دشوار نیست، به طوری‌که این علل در تمام حوزه‌های کاربردی یکسان هستند. این علت‌ها مبنی بر این مزیت‌های مهم رویکرد بیزین هستند که در ذیل به آن‌ها اشاره شده است:

۱. روش‌های بیزین در مقایسه با روش‌های فراوانی‌گرا استنباط‌های طبیعی‌تر و سودمندتری ارائه می‌دهند.

۲. روش‌های بیزین قادرند از اطلاعات موجود بیشتر بهره ببرند و از همین رو معمولاً نتایج قوی‌تری را در مقایسه با روش‌های فراوانی‌گرا حاصل می‌آورند.

۳. روش‌های بیزین در مقایسه با روش‌های فراوانی‌گرا قادر به حل مسائل پیچیده‌تری هستند.

۴. روش‌های بیزین در مورد مسائل مربوط به تصمیم‌گیری‌ها ایده‌آل هستند، در صورتی‌که روش‌های فراوانی‌گرا به آن دسته از تحلیل‌های آماری محدود می‌شوند که به صورت غیرمستقیم روی تصمیم‌ها تاثیر می‌گذارند.

۵. روش‌های بیزین در مقایسه با روش‌های فراوانی‌گرا ابهام کمتری در مورد قضاوت‌های لازم برای استنباط دارند.

در ادامه‌ی این درس‌نامه خواهیم دید که این مزیت‌ها چگونه پدید می‌آیند و پیامدهایشان برای اقتصاد سلامت و پژوهش چیست. با این‌حال، حتی نظری اجمالی به این مزایا خواننده را به فکر فرو می‌برد که چرا روش‌های فراوانی‌گرا همچنان استفاده می‌شوند. پاسخ این است که ایرادهای بسیاری نیز به رویکرد بیزین وارد است:

۱. روش‌های بیزین حاوی عنصر ذهنیت هستند که علناً در روش‌های فراوانی‌گرا وجود ندارد.

۲. در عمل دشوار است که بتوان اطلاعات افزوده‌ی مورد استفاده در روش‌های بیزین را با اطمینان مشخص کرد.

۳. روش‌های بیزین در مقایسه با روش‌های فراوانی‌گرا پیچیده‌تر هستند، و نرم‌افزار پیاده‌سازی آن‌ها اندک یا ناموجود است.

نویسندگان این درس‌نامه به رویکرد بیزین کاملاً متعهد هستند و باور دارند که می‌توان بر این ایرادها فائق آمد. درباره‌ی این باور خود توضیح خواهیم داد اما سعی خواهیم کرد که در گفتارمان درباره‌ی استدلال‌های رقیب و آخرین دستاوردها صادق باشیم.

این درس‌نامه با بحث کلی پیرامون مزایا و ایرادهای روش بیزین در برابر رویکرد فراوانی‌گرا آغاز می‌شود و مشتمل بر توضیح مفاهیم و ابزارهای اولیه‌ی آمار بیزین است. این قسمت شامل پنج بخش با عناوین استنباط، روش بیزین، اطلاعات پیشین، مشخص‌سازی پیشین - Prior Specification و محاسبات است که تمامی حقایق و استدلال‌های مهم مرتبط با استفاده از آمار بیزین را به شیوه‌ای ساده و غیرتکنیکی ارائه می‌دهد.

امیدواریم که سطح جزئیات مندرج در این بخش‌ها نیازهای عموم خوانندگان را برطرف کند، اما درک عمیق‌تر و توجیه ادعاهای موجود در متن اصلی در ضمیمه نیز آمده است. تاکید می‌کنیم که ضمیمه برای خواننده‌ی عام نگارش یافته و قصد ما این بوده که تکنیکی نباشد.

در دو بخش آخر که تحت عناوین طراحی و تحلیل کارآزمایی‌ها و مدل‌های اقتصادی نگارش شده‌اند، توضیح داده شده است که آمار بیزین چگونه در اقتصاد سلامت و پژوهش مشارکت دارد. باید تاکید کنیم که این حوزه‌ی پژوهشی به سرعت در حال پیشرفت است، و ممکن است این بخش‌ها به سرعت از دور خارج شوند. امیدواریم که خوانندگان برای ایفای نقششان در این پیشرفت‌های هیجان‌انگیز تشویق شوند، به نحوی که تکنیک‌های جدیدی ابداع کنند یا تکنیک‌های از پیش موجود را در برنامه‌های خودشان به کار ببندند.

و در نهایت، بخش نتیجه‌گیری خلاصه‌ای از استدلال‌های به کار رفته در این درس‌نامه را ارائه می‌دهد. فهرست مطالعات بیشتر نیز راجع به روش‌های بیزین و کاربرد آن‌ها در اقتصاد سلامت مجموعه‌ای از پیشنهادات کلی را فراهم آورده است.

بخش ۱
استنباط

به منظور حاصل آمدن درک واضحی از مزایا و ایرادهای رویکرد بیزین، ابتدا باید تفاوت‌های اصلی بین استنباط بیزین و فراوانی‌گرا را متوجه شویم. این بخش به ماهیت احتمال، پارامترها و استنباط های ذیلِ دیدگاه‌های بیزین و فراوانی‌گرا می‌پردازد.

روش‌های فراوانی‌گرا و بیزین بر پایه‌ی مفاهیم متفاوتی از احتمال بنیان نهاده شده‌اند. بر طبق تئوری فراوانی‌گرا، صرفاً رویدادهای تکرارپذیر هستند که احتمال برایشان مطرح می‌شود. در چارچوب بیزین، احتمال صرفاً عدم قطعیت را توصیف می‌کند. اصطلاح «عدم قطعیت - uncertainty» را باید در مفهوم عام آن تفسیر کرد. یک رویداد به سبب ذاتاً غیرقابل پیش‌بینی بودن ممکن است غیرقطعی باشد، زیرا در معرض تغییرپذیری تصادفی random variability است، برای مثال می‌توان به پاسخ یک بیمار تصادفاً انتخاب شده به یک دارو اشاره کرد. یک رویداد به سبب این‌که دانش ناقصی از آن داریم هم ممکن است غیرقطعی باشد. در این مورد می‌توان به میانگین پاسخ به یک دارو در بین همه‌ی بیماران یک جمعیت مورد بررسی اشاره کرد. فقط اولین نوع عدم قطعیت در آمار فراوانی‌گرا به رسمیت شناخته شده است، در حالی‌که رویکرد بیزین هر دو نوع عدم قطعیت را به یک میزان شامل می‌شود.

مثال:

فرض کنید که ماری سکه‌ای را پرت کرده است و نتیجه را، این‌که شیر است یا خط، می‌داند، اما آن‌را به جمال نشان نداده است. جمال چه احتمالی باید بدهد که شیر باشد؟ وقتی که این پرسش مطرح می‌شود، اغلب مردم می‌گویند که شانس ۵۰ - ۵۰ است، یعنی، احتمال یک دوم است. این امر موافق با رویکرد بیزین در مورد احتمال است، به این ترتیب که نتیجه‌ی پرتاب سکه برای جمال غیرقطعی است و بنابراین او قاعدتاً می‌تواند عدم قطعیت را با یک احتمال بیان کند. با این‌حال، از دیدگاه فراوانی‌گرا، سکه یا شیر است یا خط و رویداد تصادفی در این بین وجود ندارد. از نظر فراوانی‌گرا معنایی ندارد که جمال - در مقایسه با ماری که نتیجه را می‌داند و نامطمئن نیست - احتمالی برای رویداد منظور کند. رویکرد بیزین به وضوح بین دانش ماری و جمال تمایز قائل می‌شود.

شیوه‌های آماری معمولاً به نحوی فرموله می‌شوند که راجع به پارامترهای مجهول استنباط هایی ارائه دهند. پارامترها نماینده‌ی چیزهای مجهولی هستند، و معمولاً می‌توان آن‌ها را ویژگی‌های

جمعیتی تلقی کرد که داده‌ها از آن حاصل می‌شوند. هر پرسشی را می‌توان به صورت سوالی راجع به مقادیر مجهول این پارامترها بیان کرد. دلیل مهم بودن تفاوت رویکردهای فراوانی‌گرا و بیزین نسبت به مفهوم احتمال این است که این تفاوت تاثیر مهمی در چگونگی فکر کردن ما درباره‌ی پارامترها می‌گذارد. پارامترها به مسئله‌ی خاصی مربوط می‌شوند و معمولاً در معرض تغییرپذیری تصادفی نیستند. بنابراین، آمار فراوانی‌گرا پارامترها را تصادفی نمی‌داند و به همین ترتیب اظهار احتمال راجع به آن‌ها را معنادار نمی‌داند. در مقابل، از دیدگاه بیزین کاملاً منطقی است که احتمالات را راجع به پارامترها مطرح کنیم، چرا که پارامترها مجهول هستند.

دقت کنید که در آمار بیزین، به خاطر سهولت بیان اصطلاح‌ها، هر کمیّت غیرقطعی را متغیّر تصادفی می‌نامیم، حتی زمانی که عدم قطعیت نه به علت تصادفی بودن بلکه به علت دانش ناقصمان باشد.

مثال:

> این گزاره را در نظر بگیرید که برای یک ارائه‌دهنده‌ی مراقبت سلامت درمانِ شماره‌ی ۲ مقرون‌به‌صرفه‌تر از درمان شماره‌ی ۱ است. پارامترهای مجهولی در این گزاره مستتر هستند، برای مثال، هزینه‌ی متوسط هر یک از درمان‌ها و میانگین اثربخشی در بین تمام بیماران موجود در جمعیتی که این ارائه‌دهنده در قبال‌شان مسئول است. از دیدگاه بیزین، از آن جایی که نمی‌دانیم این گزاره درست باشد یا نه، این عدم قطعیت به عنوان احتمال مطرح می‌شود. در واقع، نتیجه‌ی تحلیل بیزین در مورد این مسئله ممکن است محاسبه‌ی این احتمال باشد که درمان ۲ برای ارائه‌دهنده مقرون‌به‌صرفه‌تر از درمان ۱ است. با این حال، از دیدگاه فراوانی‌گرا، این‌که آیا درمان ۲ مقرون‌به‌صرفه‌تر است یا نه، یک گزاره‌ای one-off proposition است که به دو درمان معین در زمینه‌ای معین اشاره می‌کند، و چون تکرارپذیر نیست نمی‌توانیم راجع به احتمال آن صحبت کنیم.

در این مثال آخر، تحلیلگر فراوانی‌گرا می‌تواند آزمون معنادار بودن فرضیه‌ی صفر را اجرا کند، مبنی بر این‌که درمان ۲ مقرون‌به‌صرفه‌تر نیست، و به این ترتیب مقدار P را به دست بیاورد. در این مرحله، خواننده باید با دقت ادعاهای موجود در جدول «تفسیر مقدار P» را که در زیر آمده است بسنجد و تصمیم بگیرد که کدام‌ها صحیح هستند.

تفسیر مقدار P

این فرضیه‌ی صفر که درمان ۲ مقرون‌به‌صرفه‌تر از درمان ۱ نیست در سطح ۵٪ رد شده است (یعنی، P = ۰٫۰۵.) این چه مفهومی دارد؟

۱. صرفاً ۵٪ از بیماران باید به شیوه‌ی مقرون‌به‌صرفه‌تری با درمان ۱ درمان شوند.

۲. اگر بنا باشد که تحلیل را بارها تکرار کنیم، و هر بار از داده‌های جدیدی استفاده کنیم، و اگر فرضیه‌ی صفر واقعاً صحیح باشد، آن‌وقت فقط در ۵٪ آن مواقع (به اشتباه) آن‌را رد کرده‌ایم.

۳. شانس این‌که فرضیه‌ی صفر صحیح باشد فقط ۵٪ است.

ادعای ۳ این است که مقدار P معمولاً چگونه تفسیر می‌شود؛ با این‌حال چنین تفسیری صحیح نیست زیرا فرضیه احتمالی را مطرح می‌کند که در حوزه‌ی مفاهیم بیزین، و نه فراوانی‌گرا، قرار می‌گیرد. تفسیر صحیح مقدار P بسیار پیچیده‌تر است و در ادعای ۲ ارائه شده است. (ادعای ۱ یکی دیگر از سوء تفسیرهای نسبتاً رایج است. از آن‌جایی که فرضیه راجع به هزینه‌ی میانگین و اثربخشی‌های میانگین است، اطلاعاتی راجع به هر یک از بیماران ارائه نمی‌دهد).

دلیل اصلی این‌که چرا نمی‌توانیم مقدار P را به این شیوه تفسیر کنیم این است که این موضوع را در نظر نمی‌گیرد که پیشین بودن فرضیه‌ی صفر تا چه اندازه‌ای منطقی است.

مثال:

آزمایشی انجام می‌شود تا امکان انتقال فکر از یک آزمودنی به آزمودنی دیگر مشخص شود. آزمودنی «الف» یک دست کارت بُر خورده دارد و تلاش می‌کند تا با آزمودنی «ب» ارتباط ذهنی برقرار کند، به نحوی که آزمودنی «ب» فقط با فکر تشخیص دهد که کارت در دستِ آزمودنی «الف» قرمز است یا سیاه. در آزمایش انجام گرفته آزمودنی «ب» رنگ ۳۳ عدد کارت را درست تشخیص می‌دهد. فرضیه‌ی صفر این است که هیچ انتقال فکری صورت نگرفته و آزمودنی «ب» بر حسب تصادف حدس می‌زند. مشاهده‌ی صحیح بودن رنگ ۳۳ عدد کارت با مقدار P (یک طرفه) برابر با ۳٫۵٪ معنادار است. آیا اکنون باید قبول کنیم که ۹۶٫۵٪ قطعیت دارد که آزمودنی «الف» قادر به انتقال افکارش به آزمودنی «ب» بوده باشد؟

اکثر دانشمندان انتقال فکر را کاملاً غیرمنطقی می‌دانند و به هیچ وجه صرفاً با یک آزمایش کوچک از این دست متقاعد نخواهند شد. پس از مشاهده‌ی نتیجه‌ی این آزمایش، اکثر افراد همچنان به فرضیه‌ی صفر اعتقاد دارند و نتیجه را ناشی از شانس می‌دانند.

در عمل، آمارشناسان فراوانی‌گرا تشخیص دادند که برای رد یک فرضیه‌ی صفر کاملاً منطقی، همانند آن‌چه در مثال بالا دیده شد، در مقایسه با رد فرضیه‌ی صفری که با ابهام بیشتری همراه است

به شواهد بسیار قوی‌تری نیاز است. این امر مشخص می‌کند که مقدار P در تمام مواقع نمی‌تواند به یک معنا باشد و تفسیر مقدار P به شکل احتمال فرضیه‌ی صفر نه تنها اشتباه است، بلکه هنگامی‌که پیشین بودن این فرضیه بسیار منطقی (یا بسیار غیرمنطقی) باشد از این هم اشتباه‌تر می‌شود.

برای بسیاری از کاربران آمار و حتی بسیاری از آمارشناسان عجیب است که کسی نتواند مقدار P را احتمال صحیح بودن فرضیه‌ی صفر تفسیر کند. به همین ترتیب، عجیب است که کسی نتواند فاصله‌ی اطمینان ۹۵٪ در مورد تفاوت درمان‌ها را به این صورت تفسیر کند که تفاوت واقعی به احتمال ۹۵٪ در این بازه قرار می‌گیرد.

با این وجود، این تفاسیر اشتباه هستند ... و ممکن است واقعاً گمراه‌کننده باشند. تفسیرهای صحیح به‌مراتب غیرمستقیم‌تر و غیر شهودی‌تر هستند. (برای مثال‌های بیشتر ضمیمه را ببینید.)

استنباط‌های بیزین دقیقاً تفسیرهای مورد نظر را دارند. تحلیل به شیوه‌ی بیزین در مورد یک فرضیه دقیقاً به احتمال صحیح بودن آن می‌انجامد. علاوه بر این، فاصله‌ی بیزین ۹۵٪ برای یک پارامتر دقیقاً به این معنی است که به احتمال ۹۵٪ آن پارامتر در آن فاصله قرار دارد. این امر ماهیت مزیت شماره‌ی ۱ است؛ «استنباط‌های طبیعی‌تر و تفسیر پذیرتر»ی که توسط شیوه‌های بیزین ارائه می‌شوند.

جدول ۱

خلاصه‌ای از تفاوت‌های مهم بین رویکردهای فراوانی‌گرا و بیزین

فراوانی‌گرا	بیزین
ماهیت احتمال	
احتمال همان فراوانی حدّی و بلندمدت است.	احتمال میزان باور شخصی را می‌سنجد.
احتمال صرفاً در مورد رویدادهایی مطرح می‌شود (که حداقل در اصل) تکرارپذیر باشند.	احتمال در مورد هر رویداد یا گزاره‌ای که نسبت به آن اطمینان نداشته باشیم مطرح می‌شود.
ماهیت پارامترها	
پارامترها تکرارپذیر یا تصادفی نیستند.	پارامترها مجهول هستند.
از همین‌رو، پارامترها متغیرهایی تصادفی نیستند، بلکه کمیّت‌هایی ثابت (مجهول) هستند.	از همین‌رو، پارامترها متغیرهایی تصادفی هستند.
ماهیت استنباط	
ادعایی راجع به پارامترها مطرح نمی‌کند (گرچه خلاف آن به نظر می‌رسد).	ادعاهای مستقیمی راجع به احتمال پارامترها مطرح می‌کند.
برحسب تکرار بلندمدت تفسیر می‌شود.	بر حسب شواهدی برگرفته از داده‌های مشاهده شده تفسیر می‌شود.
مثال	
این فرضیه را در سطح معناداری ۵٪ رد می‌کنیم.	احتمال این‌که این فرضیه صحیح باشد ۰٫۰۵ است.
در ۵٪ از نمونه‌ها که فرضیه صحیح است، فرضیه رد خواهد شد (اما توضیحی درباره‌ی این نمونه داده نشده است)	ادعا بر اساس این نمونه مطرح می‌شود (به عنوان میزانی از باور).

بخش ۲
روش بیزین

مبانی آمار بیزین بسیار ساده هستند. پارادایم بیزین پارادایم یادگیری از داده‌ها است.

داده‌ها به دانش ما می‌افزایند و به سبب آن‌ها حرف‌های جدیدتری درباره‌ی پارامترها و فرضیه‌های مرتبط خواهیم داشت. به این ترتیب، هرگاه که بخواهیم مجموعه‌ی جدیدی از داده‌ها را یاد بگیریم، باید بدانیم که پیش از مشاهده‌ی آن داده‌ها چه چیزی معلوم بوده است. به این‌ها اطلاعات پیشین prior information می‌گویند. از طریق ادغام اطلاعات پیشین است که رویکرد بیزین از اطلاعات بیشتری در مقایسه با رویکرد فراوانی‌گرا استفاده می‌کند. بحث در مورد این‌که اطلاعات پیشین دقیقاً بیانگر چه چیزی هستند و از کجا سرچشمه می‌گیرند در بخش بعدی با عنوان اطلاعات پیشین مطرح شده است. به منظور توضیح چگونگی عملکرد پارادایم بیزین، فرض می‌کنیم که اطلاعات پیشین مشخص شده‌اند و به شکل توزیع پیشین برای پارامترهای نامشخص مدل آماری بیان شده‌اند. توزیع پیشین آن‌چه را بیان می‌کند که قبل از دیدن داده‌های جدید معلوم است (یا به صحت آن باور داریم). به این ترتیب، این اطلاعات با اطلاعات موجود در داده‌ها تلفیق می‌شوند تا توزیع پسین ایجاد شود؛ توزیع پسین آن‌چیزی را بیان می‌کند که در حال حاضر و پس از دیدن داده‌ها در مورد پارامترها می‌دانیم (معمولاً به این توزیع‌ها با عناوین «پیشین» و «پسین» اشاره می‌کنیم).

مکانیسم ریاضیِ این تلفیق قضیه‌ی بیز نامیده می‌شود، و به همین دلیل است که این رویکرد آماری «بیزین» خوانده می‌شود. از منظر تاریخی، این نام از نام کشیش توماس بیز، وزیر قرن هجدهمی، برگرفته شده است که برای اولین بار استفاده از این قضیه را به این شیوه نشان داد و موجب پیدایش آمار بیزین شد.

بیان ساده‌ی این فرآیند در جعبه‌ی «مثالی از قضیه‌ی بیز» ارائه شده است.

تصویر۱ یک نمودار سه تایی و روشی برای مشاهده‌ی این است که روش‌های بیزین چگونه دو منبع اطلاعاتی را ادغام می‌کنند. قدرت هر منبع اطلاعاتی با میزان باریکی narrowness منحنی‌اش مشخص می‌شود؛ به طوری‌که منحنی باریک‌تر مقدار پارامترهای بیشتری را منتفی می‌داند و به همین سبب اطلاعات قوی‌تری را ارائه می‌دهد. در تصویر۱، مشاهده می‌کنیم که داده‌های جدید (منحنی قرمز) در مقایسه با توزیع پیشین (منحنی خاکستری) کمی بیشتر اطلاعات به دست می‌دهند. از آن‌جایی که قضیه‌ی بیز قادر به تشخیص قدرت هر یک از منابع است، در تصویر۱، توزیع پسین (منحنی نقطه‌چین سیاه) از داده‌ها کمی بیشتر تاثیر گرفته است تا توزیع پیشین. برای مثال، اوج منحنی توزیع پسین در ۱٫۳۳ است و مقداری بیشتر به اوج منحنی داده‌ها نزدیک است تا

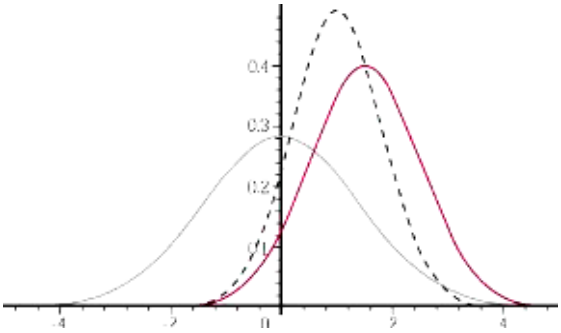
به اوج منحنی توزیع پیشین. دقت کنید که منحنی توزیع پسین در مقایسه با منحنی توزیع پیشین و منحنی داده‌ها باریک‌تر است که این امر نشان‌دهنده‌ی این است که توزیع پسین قدرت هر دو منبع اطلاعاتی را به سمت خود کشیده است.

منحنی داده‌ها به زبانِ فنی درست‌نمایی likelihood نامیده می‌شود و در استنباط فراوانی‌گرایانه نیز اهمیت دارد. در هر دو پارادایم استنباط، نقش این منحنی توصیف کردن قدرت پشتیبانی از داده‌ها برای مقادیر محتمل پارامتر است. آشکارترین تفاوت بین روش‌های فراوانی‌گرا و بیزین این است که آمار فراوانی‌گرا صرفاً از درست‌نمایی likelihood استفاده می‌کند، در حالی‌که آمار بیزین هم از درست‌نمایی و هم از اطلاعات پیشین استفاده می‌کند.

در تصویر ۱، تحلیل بیزین به دلیل آن‌که هم از اطلاعات پیشین و هم از داده‌ها استفاده می‌کند استنباط‌های متفاوتی را نسبت به رویکرد فراوانی‌گرا ایجاد می‌کند. برآورد فراوانی گرا، با صِرف استفاده از داده‌ها، تقریباً برابر با ۱٫۵ است. تحلیل بیزین از این واقعیت نفع می‌برد که بر مبنای اطلاعات پیشین بعید است که مقدار صحیح پارامتر برابر با ۲ یا بیشتر از آن باشد.

در نتیجه، برآورد بیزین در حدود ۱ است. تحلیل بیزین اطلاعات پیشین و اطلاعات داده‌ها را به شیوه‌ای مشابه با یک فراتحلیل ادغام می‌کند؛ در فراتحلیل‌ها اطلاعات حاصل از چندین آزمایش گزارش شده با هم ادغام می‌شوند. برآورد پسین حد واسط بین برآورد پیشین و برآورد داده‌ها است و در مقایسه با برآورد مجزای هر یک از این دو منبع اطلاعاتی برآورد دقیق‌تری محسوب می‌شود (همان‌طور که در چگالی پسین به صورت منحنی باریک‌تر دیده می‌شود). این دومین مزیت مهم استفاده از رویکرد بیزین است؛ «توانایی بهره‌گیری از اطلاعات بیشتر و کسب نتایج قوی‌تر».

مثالی از قضیه ی بیز



تصویر ۱. توزیع پیشین (خاکستری) و اطلاعات حاصل از داده‌های جدید (قرمز) به منظور ایجاد توزیع پسین (نقطه‌چین سیاه).

در این مثال، اطلاعات پیشین (منحنی خاکستری) تقریباً با قطعیت می‌گویند که پارامتر بین ۴- و ۴+ قرار دارد، و به احتمال زیاد بین ۲- و ۲+ است، و بهترین برآورد ما از این منحنی برابر با ۰ است.

بر طبق داده ها (منحنی قرمز)، مقدار پارامتر بین ۰ تا ۳ است و این مقدار به هیچ وجه کمتر از ۲- نیست.

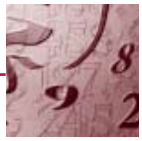
توزیع پسین (منحنی نقطه‌چین سیاه) این دو منبع اطلاعاتی را در کنار هم قرار می‌دهد. از همین‌رو، چگالی پسین برای مقادیر کمتر از ۲- اندک است زیرا با توجه به داده‌ها این مقادیر کاملاً غیرمنطقی هستند. مقادیر بالاتر از ۴+ با توجه به توزیع پیشین منتفی هستند؛ و مجدداً توزیع پسین موافق با این مسئله است. منحنی داده‌ها به مقادیر در حدود ۱٫۵ گرایش دارد، در حالی‌که توزیع پیشین مقادیر نزدیک به ۰ را ترجیح می‌دهد. توزیع پسین به هر دو گوش می کند و تلفیق حاصل حد واسط آن دو خواهد بود. پس از دیدن داده‌ها، حالا تصور می‌کنیم که پارامتر به احتمال زیاد نزدیک به ۱ است.

طبق پارادایم بیزین، همه‌ی استنباط‌ها از توزیع پسین حاصل می‌شود. یک برآورد از پارامتر ممکن است مد - mode این توزیع باشد (یعنی، نقطه‌ای که در آن به حداکثر خود می‌رسد). انتظار پسین posterior expectation یکی دیگر از رایج‌ترین برآوردها است. اگر فرضیه‌ای داشته باشیم، آن‌وقت احتمال صحیح بودن آن فرضیه نیز از توزیع پسین حاصل می‌شود. برای مثال، در تصویر ۱ احتمال این‌که پارامتر مثبت باشد برابر با مساحت زیر منحنی نطقه‌چین سیاه به سمت راست مبدأ و مساوی با ۰٫۸۹ است.

بر خلاف استنباط فراوانی‌گرایانه - که باید تمام سوالات را به صورت آزمون‌های معناداری، فواصل اطمینان و برآوردگرهای نااریب بیان کند - استنباط بیزین می‌تواند از توزیع پسین استفاده کرده و برای هر نوع سوالی پاسخ‌های مرتبط و مستقیم ارائه دهد. مثالی در این مورد ارتباط ذاتی بین آمار بیزین و نظریه‌ی تصمیم‌گیری decision theory است. با ادغام توزیع پسین و تابع مطلوبیت utility func tion (که پیامد تصمیمات مختلف را اندازه می‌گیرد) قادر به تعیین تصمیم بهینه optimal decision خواهیم بود. تصمیم بهینه، مطلوبیت مورد انتظار را به حداکثر می‌رساند. در ارزیابی‌های اقتصادی، این امر بسته به تابع مطلوبیت ممکن است هزینه‌ی مورد انتظار را به حداقل یا میزان اثر مورد انتظار را به حداکثر برساند. با این‌حال، از منظر مقرون‌به‌صرفه بودن، سود خالص مناسب‌ترین معیار برای مطلوبیت است (به صورت متوسطِ میزان اثر ضربدر تمایل به پرداخت، منهای هزینه‌ی مورد انتظار تعریف می‌شود).

برای مثال، یک ارائه‌دهنده‌ی مراقبت سلامت را در نظر بگیرید که باید از دو روش یکی را برای بازپرداخت انتخاب کند. تصمیم بهینه انتخاب آن روشی است که سود خالص مورد انتظار بیشتری دارد. تحلیل بیزین بی‌درنگ این پاسخ را برمی‌گزیند، اما تحلیل فراوانی‌گرای مشابهی در این مورد وجود ندارد. صرفاً آزمودن این فرضیه که یکی از سودهای خالص از دیگری بیشتر است پاسخ درستی به سوال نمی‌دهد (همان‌طور که محاسبه‌ی این احتمال که سود خالص روش ۲ بیشتر از سود خالص روش ۱ است پاسخ بیزین مناسبی محسوب نمی‌شود). جزئیات بیشتر در مورد این مثال و قضیه‌ی بیز را می‌توانید در ضمیمه بیابید.

این امر بیانگر یکی دیگر از مزایای مهم آمار بیزین یعنی مزیت چهارم آن است : «روش‌های بیزین برای تصمیم‌گیری‌ها ایده‌آل هستند».



بخش ۳
اطلاعات پیشین

اطلاعات پیشین هم قدرت و هم ضعف بالقوه‌ی رویکرد بیزین محسوب می‌شوند. دیده‌ایم که اطلاعات پیشین چگونه امکان دسترسی به اطلاعات بیشتر و به این ترتیب امکان ایجاد استنباط‌های قوی‌تر را برای شیوه‌های بیزین میسّر کرده‌اند. [بهره‌گیری از] این اطلاعات یکی از مزایای مهم رویکرد بیزین است. از سوی دیگر، بیشتر انتقادهای وارد به تحلیل بیزین معطوف به اطلاعات پیشین هستند.

انتقاد اصلی این است که اطلاعات پیشین ذهنی هستند: اطلاعات پیشین شما نسبت به اطلاعات پیشین من متفاوت است، و به همین سبب توزیع پیشین شما با توزیع پیشین من تفاوت دارد. همین امر باعث می‌شود که توزیع پسین، و تمام استنباط‌های حاصله از آن، ذهنی باشند. به این معنا ادعا می‌شود که کل رویکرد بیزین ذهنی است. در حقیقت، روش‌های بیزین بر مبنای تفسیر ذهنی از احتمال بنا نهاده شده‌اند. این مسئله در جدول ۱ با عنوان «میزان باور شخصی» توضیح داده شده است. اگر بنا باشد که احتمالاتی را به پارامترها و فرضیه‌ها نسبت دهیم این صورت‌بندی (ضمیمه را برای جزئیات بیشتر ببینید) لازم است، زیرا تفسیر فراوانی‌گرا راجع به احتمال بسیار بسته narrow است. با این‌حال، برای بسیاری از دانشمندانی که طبق آموزه‌های قبلی‌شان باید تا جای ممکن عنصر ذهنیت را نفی کنند، پذیرش مزایای رویکرد بیزین بسیار سخت خواهد بود. از دیدگاه منتقدان رویکرد بیزین، «ذهنیت» ایراد اصلی آن است.

به باور ما این مخالفت در کلیات و در عمل ناموجه است. در کلیات ناموجه است زیرا ممکن نیست که علم کاملاً عینی باشد. در عمل ناموجه است زیرا روش بیزین ماهیت واقعی یک روش علمی را در موارد زیر به وضوح نشان می‌دهد:

- با بنا نهادن اطلاعات پیشین روی شواهد و استدلال‌های قابل دفاع، [عنصر] ذهنیت در توزیع پیشین به حداقل می‌رسد.
- تفاوت در موقعیت‌های پیشین prior positions به سبب تجمع داده‌ها برطرف شده و اجماع حاصل می‌آید.

با در نظر گرفتن مورد دوم باید بگوییم که قضیه‌ی بیز، به منظور استخراج توزیع پسین، با توجه به قدرت نسبی اطلاعات پیشین و داده‌ها به آن‌ها وزن می‌دهد و به این ترتیب توزیع پسین را حاصل می‌آورد. اگر اطلاعات پیشین مبهم و ناکافی باشند وزن قابل‌اغماضی در تلفیق داده‌ها می‌گیرند، و توزیع پسین به طور کامل بر اساس اطلاعات داده‌ها (همان‌طور که در تابع درست‌نمایی توضیح داده شده است) خواهد بود. به همین ترتیب، همان‌طور که داده‌های بیشتر و بیشتری کسب می کنیم،

وزنی که قضیه‌ی بیز در مقایسه با توزیع پیشین به داده‌های تازه کسب شده می‌دهد افزایش پیدا می‌کند. مجدداً، توزیع پسین عملاً به طور کامل به اطلاعات موجود در داده‌ها متکی است. این ویژگی قضیه‌ی بیز بازتابی از فرآیند علم است؛ به طوری‌که انباشتگی شواهد عینی فرآیند اولیه‌ای است که به موجب آن اختلاف نظرها برطرف می‌شوند. وقتی که داده‌ها شواهد قطعی را فراهم آورند دیگر جایی برای نظرات شخصی باقی نمی‌ماند.

به مورد اول برمی‌گردیم، این جمله چنین توضیح داده است که در صورت وجود اطلاعات پیشین درست و بنیادی، این اطلاعات باید بر مبنای شواهد و استدلال‌های قابل دفاع باشند. این مسئله به خصوص هنگامی اهمیت دارد که داده‌های جدید آن‌قدر جامع نباشند که بر اطلاعات پیشین چیره شوند، در این موقع است که قضیه‌ی بیز به توزیع پیشین وزن غیر قابل‌اغماضی در تلفیق با داده‌ها می‌دهد. چنین اطلاعات پیشینی معمولاً در کاربری‌های پزشکی، و مخصوصاً در ارزیابی اقتصادی از فن‌آوری‌های رقابتی وجود دارند.

دو مثال در ضمیمه ارائه شده است. یکی در مورد تحلیل تفاوت زیرگروه‌ها است، که در آن شک‌گرایی پیشین راجع به وجود چنین تاثیرهایی بدون یک مکانیسم بیولوژیکی منطقی به طور طبیعی در تحلیل بیزین نهاده شده است.

مثال دیگر به موردی مربوط می‌شود که تصمیم راجع به مقرون‌به‌صرفه بودن یک داروی جدید در مقابل درمان استاندارد بستگی زیادی به شواهد موجود درباره‌ی بستری‌شدن در بیمارستان دارد. یک کارآزمایی کوچک کاهش نسبتاً قابل توجهی (و، در اصطلاح فراوانی‌گرا، معنادار) را در متوسط روزهای بستری بودن نشان داد. با این‌حال، کارآزمایی بزرگتری که قبلاً انجام شده بود از متوسط روزهای بستری‌شدن برای دارویی مشابه برآورد چندان رضایت‌بخشی ارائه نکرد. برای کارآزمایی اولیه در نتیجه‌ی تحلیل فراوانی‌گرا دو پاسخ محتمل وجود دارد:

۱.در نظر گرفتن این‌که دلیلی وجود ندارد که نرخ بستری‌شدن به علت استفاده از داروی قدیمی همانند داروی جدید باشد؛ در این صورت از کارآزمایی اولیه صرف‌نظر می‌شود زیرا اطلاعات جدیدی در مورد داروی جدید به دست نمی‌دهد.

۲.در نظر گرفتن این‌که دو دارو ضرورتاً باید نرخ بستری‌شدن‌های یکسانی داشته باشند؛ به این ترتیب داده‌های هر دو کارآزمایی را ادغام می‌کنیم.

گزینه‌ی دوم منجر به داده‌های جدیدی می‌شود که مملو از [داده‌های] کارآزمایی اولیه‌ی بسیار بزرگتر هستند، که غیر منطقی به نظر می‌رسد، اما گزینه‌ی اول مستلزم حذف اطلاعات بالقوه سودمند است. در عمل، یک فراوانی‌گرا احتمالاً گزینه‌ی اول را برمی‌گزیند، اما با توجه به نکته‌ای که کارآزمایی اول نشان می‌دهد احتمالاً این تصمیم نرخ واقعی را کمتر از حد برآورد خواهد کرد.

معمولاً واقع‌بینانه‌تر است که این دیدگاه را برگزینیم که دو نرخ بستری‌شدن متفاوت اما همانند هستند. در ضمیمه توضیح داده می‌شود که تحلیل بیزین چگونه می‌تواند از کارآزمایی اولیه به عنوان

اطلاعات پیشین استفاده کند، گرچه که لازمه‌ی آن قضاوت در مورد تشابه داروها است. قبل از اجرای کارآزمایی جدید فکر می‌کردیم که تفاوت نرخ بستری‌شدن‌های آن دو دارو چقدر متفاوت باشد؟

تحلیل بیزین به جای این استنباط مبهم که: «کارآزمایی اولیه روی دارویی مشابه میانگین روزهای بستری‌شدن بیشتری را نشان داد، و به همین خاطر در مورد کاهش مشاهده شده در این کارآزمایی شک دارم»، یک تلفیق قطعی و کمّی از دو منبع اطلاعاتی فراهم می‌آورد. این تلفیق ناشی از تفسیری روشن، منطقی و واضح از اطلاعات پیشین است. این نکته بخشی از مزیت شماره‌ی پنجم رویکرد بیزین است: «قضاوت‌هایی واضح‌تر». بدون تحلیل بیزین طبیعی است که ادعاهای کارآزمایی جدید را تعدیل کنیم. احتمالاً میزان این تعدیل همچنان بسته به قضاوت شخصی افراد باشد، اما، طبق قضیه‌ی بیز، مبنای این قضاوت چندان مشخص نیست و نتیجه نیز به وضوح از این قضاوت ناشی نمی‌شود.

۳. این امر به شیوه‌ی مهم دیگری منجر می‌شود که روش‌های بیزین در آن حضور دارند. هنگامی‌که توزیع پیشین و درست‌نمایی مشخص شوند (و علناً قابل استفاده باشند)، محاسبه‌ی توزیع پسین و استخراج استنباط‌های پسین یا تصمیم‌ها انجام می شود. در مقابل، هنگامی که درست‌نمایی در تحلیل فراوانی‌گرا تعیین می‌شود، هنوز آزادی انتخاب داریم که از بین بسیاری از قوانین استنباط کدامیک را برگزینیم. برای مثال، گرچه تعیین برآوردگرهای بهینه در مسائل ساده امکان‌پذیر است، اما در کل ممکن است برآوردگرهای ناریب بسیاری وجود داشته باشند که هیچکدام از آن‌ها به لحاظ داشتن واریانس یکنواخت کوچکتر uniformly smaller variance بر دیگری برتری ندارد. به این ترتیب، پزشک در استفاده از هر یک از این قوانینِ استنباط یا فکر به سایر آن‌ها بر حسب ضرورت آزاد است. این ویژگی رویکرد فراوانی‌گرا از آن‌جایی که انتخاب‌های مربوطه اساساً دلبخواهی هستند به عدم شفافیت منجر می‌شود.

در مورد انتقاد اول چطور؟؛ این‌که شیوه‌های بیزین ذاتاً ذهنی هستند. واقعیت این است که می‌توان با در اختیار داشتن توزیع پیشین و بر مبنای حدس، پیش‌داروی، یا تفکر آرزومندانه به تحلیل بیزین دست زد. در حقیقت قضیه‌ی بیز تمام این اقدامات بدسرانجام را می‌پذیرد، اما آمار بیزین به هیچ وجه آن‌ها را قبول نمی‌کند. همچنین، به خاطر بیاورید که، در یک تحلیل بیزین صحیح، اطلاعات پیشین نه تنها واضح‌اند بلکه بر اساس شواهد قابل دفاع و استدلال بنا نهاده شده‌اند، و اگر تحلیل بیزین بر اساس چنین شواهدی انجام بگیرد، آن‌وقت ماهیت تمام اقدامات اشاره شده در بالا آشکار شده و در نتیجه رد خواهند شد.

بیان مختصری راجع به این‌که اطلاعات پیشین باید شامل چه چیزهایی باشند در جعبه‌ی «شواهد» ارائه شده است.

شواهد

اطلاعات پیشین باید بر اساس شواهد درست و قضاوت‌های منطقی استوار باشند. یک راه برای درک این مطلب یادآوری این نقل‌قول آشنا است: توزیع پیشین باید «[برگرفته] از شواهد باشد، کل شواهد باشد و چیزی غیر از شواهد نباشد».

- «شواهد» ؛ اطلاعاتی خالص که به درستی تفسیر شوند
- «کل شواهد» : اطلاعات مربوطه حذف نشوند (ترجیحاً اجماعی باشد که دانش متخصصان بسیاری را در خود جای داده باشد)
- «چیزی غیر از شواهد نباشد» : عاری از سوگیری و پیش‌داوری باشد.

بخش ۴
مشخص سازی پیشین

امیدواریم بخش‌های قبلی خواننده را متقاعد کرده باشند که اطلاعات پیشین وجود دارند و می‌بایست به منطقی‌ترین، عینی‌ترین و واضح‌ترین شیوه‌ی ممکن از آن‌ها بهره گرفت. در این‌جا به این مسئله می‌پردازیم که چگونه توزیع احتمال پیشین - یا همان منحنی خاکستری در تصویر اول - را مشخص کنیم.

به مثال ذکر شده در بخش قبل رجوع کنید که در آن اطلاعات پیشین حاوی اطلاعاتی راجع به بستری‌شدن در پی تجویز دارویی مشابه در یک کارآزمایی بود. این مسئله در ضمیمه به شکل توزیع پیشینی با میانگین ۰٫۲۱ (متوسط روزهای حضور در بیمارستان به ازای هر بیمار) و انحراف استاندارد ۰٫۰۸ مشخص شده است. این امر با رجوع به کارآزمایی مطرح شده در پرسش توجیه می‌شود، به این ترتیب که میانگین روزهای حضور در بیمارستان به سبب استفاده از دارویی متفاوت اما مشابه به میزان ۰٫۲۱ برآورد می‌شود در حالی‌که خطای استاندارد نیز ۰٫۰۳ است. اما توزیع پیشین تعریف شده چطور از اطلاعات پیشین داده شده حاصل شده است؟

در فرآیند مشخص کردن توزیع پیشین ناگزیر قضاوت روی می‌دهد. همانند مورد بالا، قضاوت معمولاً به علت نیاز به تفسیر اطلاعات پیشین و ارتباطش با داده‌های جدید به وقوع می‌پیوندد. نرخ بستری‌شدن‌های این دو دارو چه تفاوتی دارند؟ کارشناسان متفاوت ممکن است اطلاعات پیشین را به شیوه‌های متفاوتی تفسیر کنند. همان‌طور که ممکن است یک کارشناس اطلاعات را در زمان‌های متفاوت به شکل‌هایی متفاوت تفسیر کند، همانند مثال مربوط به تصمیم‌گیری در مورد این‌که انحراف استاندارد پیشین ۰٫۷۵ باشد یا ۰٫۸؟

حتی اگر اطلاعات پیشین ما مبتنی بر شواهد خالص بوده و ارتباط واضحی با داده‌های جدید داشته باشند، نمی‌توانیم آن‌ها را با دقت و اطمینان کامل به شکل توزیع پیشین تبدیل کنیم. این ایراد شماره‌ی ۲ است: «مشخص‌سازی پیشین غیر قابل‌اطمینان است».

با این حال، در عمل صرفاً باید با قابلیت اطمینان و دقت بسنده‌ای توزیع پیشین را مشخص کنیم. می‌توانیم به بررسی گستره‌ای از مشخص‌سازی‌های پیشین منطقی بر اساس تفسیرهای معقول از شواهد موجود بپردازیم و در زمان لزوم قضاوت کردن، نادقیق بودن را جایز بشماریم. اگر استنباط‌ها یا تصمیم‌های پیشین اساساً حساسیتی به آن تغییرات نداشته باشند، آن وقت است که دیگر غیر قابل‌اطمینان بودن ذاتی موجود در فرآیند مشخص سازی پیشین اهمیتی ندارد. این عملِ تحلیل حساسیت با در نظر گرفتن مشخص‌سازی پیشین یکی از ویژگی‌های مهم روش بیزین است که در تمام

برنامه‌های تحلیل تصمیم وجود دارد.

دقت لازم در مشخص‌سازی پیشین به منظور دستیابی به استنباط‌ها و تصمیم‌های قوی به قدرت داده‌های جدید بستگی دارد. همان‌طور که دیده‌ایم، با در اختیار داشتن داده‌های قوی و کافی، اهمیت اطلاعات پیشین بسیار کم می‌شود و تفاوت‌های موجود در قضاوت‌ها در مرحله‌ی تفسیر داده‌ها بی‌اهمیت خواهند شد. هنگامی که داده‌های جدید چندان قوی نبوده و اطلاعات پیشین محسوس باشند، آن وقت است که تحلیل حساسیت ضرورت می‌یابد. همچنین لازم به ذکر است که گاهی عقیده‌ی کارشناس، علی‌رغم ایرادهای آشکاری که بر آن وارد است، جزء بسیار سودمندی در اطلاعات پیشین محسوب می‌شود. روش‌های برون‌کشی - elicit قضاوت‌های کارشناسان موضوع پژوهشیِ پویایی است که مورد علاقه‌ی آماردانان و روانشناسان است.

بررسی توزیع‌های پیشین آگاهی‌بخش تا به امروز ادامه یافته است. روش‌های دیگر مشخص کردن توزیع پیشین در تحلیل به شیوه‌ی بیزین در جعبه‌ی «انواع و تعریف‌های توزیع پیشین» ارائه شده‌اند.

به دلیل دشوار بودن برون‌کشیِ دقیق و قابل اعتماد توزیع‌های پیشین، برخی از تحلیلگران راه‌حل‌های مرسوم را پیشنهاد داده‌اند که باورهای پیشین یا موضع پیشین توام با شک skeptical prior position در آن‌ها جایگاهی ندارند.

استدلالی که موافق ارائه ندادن اطلاعات پیشین است از هر انتقادی راجع به ذهنیت اجتناب می‌کند. تلاش زیادی برای یافتن فرمولی شده است که ناآگاهی پیشین را نشان دهد که نتیجه‌ای در بر نداشته است و قطعاً جست‌وجویی بی‌نتیجه است. با این وجود، بازنمودهای استخراج شده‌ی مختلف حداقل به منظور ارائه‌ی اطلاعات پیشینِ نسبتاً ضعیف سودمند خواهند بود.

در صورت قوی بودن اطلاعات جدید (مرتبط با اطلاعات پیشین) انتظار نمی‌رود که اطلاعات پیشین سهم قابل ملاحظه‌ای در [اطلاعات] پسین داشته باشند. در این وضعیت، بی‌فایده (و نه مقرون‌به‌صرفه) خواهد بود که با دقت به برون‌کشی داده‌های پیشین موجود بپردازیم. در عوض، در چنین وضعیتی معمول است که برخی از پیشین‌های «غیرآگاهی‌بخش»، «ناشی از قصور»، «عطفی»، «نادرست»، «مبهم»، «ضعیف» یا «ناشی از جهل» را به کار بگیریم (گرچه که آخرین مورد حقیقتاً اسم بی‌مسمّایی است). این اصطلاحات در آمار بیزین کم و بیش به جای هم به کار می‌روند تا به این نکته اشاره کنند که توزیع پیشین تقریباً از اطلاعات پیشین مستقل است. اصطلاح «نادرست - improper» به این منظور استفاده شده است که، به بیان فنی، بسیاری از این توزیع‌ها عملاً وجود ندارند، زیرا که یک توزیع نرمال با واریانس بی‌نهایت وجود ندارد.

انواع و تعاریف توزیع پیشین

توزیع‌های پیشین آگاهی‌بخش (یا اصل): اطلاعات پیشین اصل و بهترین قضاوت در مورد قدرت این اطلاعات و ارتباط آن‌ها با داده‌های جدید را ارائه می‌دهند.
توزیع‌های پیشین غیر آگاهی‌بخش (یا ناشی از قصور، رجوعی، نادرست، ضعیف، ناشی از جهل): فاقد اطلاعات پیشین قابل قبول هستند.
توزیع‌های پیشین شکاک: وضعیتی را نشان می‌دهند که یک فرضیه‌ی صفر احتمالاً صحیح خواهد بود.
توزیع‌های پیشین ساختاری (یا سلسله مراتبی): اطلاعات پیشین اصل در مورد ارتباطات بین پارامترها را ادغام می‌کنند.

هدف استفاده از توزیع‌های پیشین «شکاک» این است که اگر داده‌ها بتوانند شخص شک‌اندیشی را متقاعد کنند، آن‌وقت است که هر کسی از موضع پیشینِ توام با شکّی کمتر از این فرد نیز متقاعد خواهد شد. بنابراین، اگر کسی در موردِ برخی از فرضیه‌ها در موضع پیشینِ توام با شک قرار داشته باشد و با این‌حال توسط داده‌ها متقاعد شود، آن‌وقت است که احتمال پسین او برای یک فرضیه‌ی معین بالا می‌رود، و فرد دیگری در موضع پیشینِ توام با شکّی کمتر در مقایسه با این فرد نهایتاً برای این فرضیه احتمال پسین بالاتری در مقایسه با فرد اول لحاظ می‌کند. در چنین موردی داده‌ها آن‌قدر قوی هستند که به نتیجه‌گیری قاطعانه‌ای منتهی شوند. از سوی دیگر، اگر در هنگام استفاده از توزیع پیشینِ شکاک داده‌ها آن‌قدر قوی نباشند که احتمال پسین بالایی را برای آن فرضیه‌ی خاص حاصل کنند، در این‌صورت نباید راجع به آن فرضیه یک استنباط قطعی ارائه دهیم. گرچه این فکر هم وسوسه‌انگیز است، اما در مورد این‌که توزیع پیشین شکاک باید به چه شکل باشد توافق نظری وجود ندارد.

عقاید نسبتاً پیچیده‌تر در مورد توزیع‌های پیشین ساختاری یا سلسله مراتبی (آخرین دسته در جعبه‌ی «انواع و تعریف‌های توزیع پیشین») در ضمیمه شرح داده شده‌اند.

بخش ۵
محاسبات

استفاده از نرم‌افزار در ساده‌ترین روش‌های آماری ضروری است، و روش‌های بیزین نیز از این قاعده مستثنی نیستند. در آمار بیزین، پیاده‌سازی قضیه‌ی بیز و سپس استخراج استنباط‌ها یا تصمیم‌های مربوطه از توزیع پیشین از جمله اعمال اصلی محسوب می‌شوند. در مسائل بسیار ساده می‌توان چنین وظایفی را با استفاده از جبر انجام داد، اما در مورد مسائلی که قدری پیچیده هستند نمی‌توان به این شیوه عمل کرد.

تا دهه‌ی ۱۹۹۰ از روش‌های بیزین استقبال خوبی می‌شد اما از آن‌جایی که نرم‌افزار و ابزارهای ضروری محاسباتی آن‌ها هنوز توسعه نیافته بودند کاربری عملی این روش‌ها بسیار محدود بود. هر کس که می‌خواست به تحلیل‌های آماری مهمی بپردازد چاره‌ای نداشت جز آن‌که به روش‌های فراوانی‌گرا روی بیاورد. موقعیت این دو دیدگاه آماری در کمی بیش از یک دهه تغییری اساسی کرد. ابزارهای محاسباتی به‌طور خاص برای تحلیل بیزین توسعه یافتند، به گونه‌ای که بسیار قدرتمندتر از تمام ابزارهای در اختیار روش‌های فراوانی‌گرا بودند، و در نتیجه‌ی این امر تحلیلگران بیزین امروزه قادر به حل مسائل بسیار پیچیده‌ای هستند که فراوانی‌گراها حتی از پرداختن به آن‌ها نیز عاجزاند. این امر همچنان مصداق دارد که روش‌های بیزین بسیار پیچیده‌تر هستند، و، گرچه که تکنیک‌های محاسباتی این روش‌ها در جوامع علمی کاملاً پذیرفته شده‌اند، اما هنوز هم نرم‌افزار کاربرپسندی در اختیار پزشکان عمومی قرار نگرفته است.

این تحول هنوز هم جریان دارد و توسعه‌ی نرم‌افزارهای محاسباتی در حال جابه‌جا کردن تعادل بین ایراد شماره‌ی ۳ « پیچیدگی و نبود نرم‌افزار» و مزیت شماره‌ی ۳ «توانایی حل مسائل بسیار پیچیده‌تر» است. ابزار اصلی یک روش شبیه‌سازی به نام زنجیره‌های مارکوف مونت کارلویی (MCMC) است. هدف MCMC به عبارتی نادیده گرفتن عملیات ریاضی به جای اجرا کردن آن‌ها است. استنباط بیزین با بیرون کشیدن تصادفی نمونه‌ی شبیه‌سازی شده‌ی بسیار بزرگی از توزیع پسین حل می‌شود. نکته این‌جا است که اگر از هر توزیعی نمونه‌ای در اختیار داشته باشیم که به اندازه‌ی کافی بزرگ باشد آن موقع است که کل توزیع را در مقابلمان خواهیم داشت. می‌توانیم هر آن چیزی را که می‌خواهیم راجع به توزیع بدانیم از روی نمونه محاسبه کنیم. برای مثال، اگر بخواهیم میانگین پسین را بشناسیم لازم است که فقط میانگین این «نمونه‌ی استنباطی» را محاسبه کنیم. اگر نمونه به قدر کافی بزرگ باشد در آن صورت است که میانگین نمونه یک تقریب بسیار دقیق از میانگین واقعی توزیع خواهد بود، به نحوی که می‌توانیم تفاوت بین این دو را نادیده بگیریم.

دسترسی به روش‌های محاسباتی از جمله MCMC حتی در پیچیده‌ترین مدل‌ها نیز استنباط‌های دقیق بیزین را امکان‌پذیر می‌کند. برای مثال، مدل‌های خطی تعمیم‌یافته generalized linear models را کاملاً می‌توان با روش‌های بیزین تحلیل کرد، در صورتی‌که روش‌های فراوانی‌گرا به تقریب‌ها approximation وابسته هستند. در واقع، در مسائل بسیار پیچیده، مدل‌سازی بیزین اجزای مختلف روش‌های مدل‌سازی را بدون هیچ محدودیتی با اطلاعات پیشین ساختاری ادغام می‌کند، و مهم نیست که ترکیب این مدل‌ها تا پیش از آن بررسی و تحلیل شده باشد یا خیر. آماردان برای مدل‌سازی در استفاده از داده‌ها و سایر اطلاعات موجود به نحوی که بتواند واقع‌بینانه‌ترین شکل ممکن را حاصل بیاورد، کاملاً مختار است. مهم نیست که مدل حاصل چقدر به‌هم‌ریخته باشد زیرا با استفاده از MCMC (حداقل در کل) می‌توان استنباط‌های پسین را محاسبه کرد.

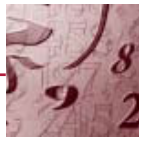
روش‌های بیزین تنها ابزارهای قابل اجرا در بسیاری از زمینه‌ها از جمله آنالیز تصویر، اپیدمیولوژی فضایی و آنالیز شجره‌نامه‌های ژنتیکی هستند.

گرچه که نرم‌افزارهای همیار در تحلیل‌های بیزین در حال بیشتر شدن هستند اما بیشتر آن‌ها بسیار تخصصی‌اند و برای تحلیلگران معمولی مناسب نیستند. متاسفانه در حال حاضر نرم‌افزاری وجود ندارد که در عین قدرتمند بودن کاربرپسند هم باشد و انتظار اکثر مردم از پکیج‌های آماری را برآورده کند. پکیج‌های آماری First Bayes و WinBUGS رایگان هستند و در دسترس عموم پژوهشگران قرار دارند.

پکیج آماری First Bayes برنامه‌ی بسیار ساده‌ای با هدف آموزش و کمک به درک افراد مبتدی پیرامون نحوه‌ی عملکرد روش‌های بیزین است. این برنامه برای تحلیل‌های پیچیده استفاده نمی‌شود و مدعی آموزش آمار بیزین هم نیست، اما در بسیاری از دانشگاه‌های جهان از آن در دوره‌های آموزشی آمار بیزین استفاده می‌شود. این برنامه در کنار یک کتاب درسی - همانند آن‌هایی که در بخش مطالعات بیشتر این درس‌نامه‌ی مقدماتی معرفی شده‌اند - بسیار سودمند خواهد بود و می‌توان به رایگان آن‌را از آدرس http://www.shef.ac.uk/~st۱ao/ دانلود کرد.

برنامه‌ی WinBUGS برنامه‌ی قدرتمندی برای انجام محاسبات MCMC است که در تحلیل‌های پیچیده‌ی بیزین استفاده‌ی فراوانی دارد. این برنامه عامل موثری در رشد نرم‌افزارهای کاربردی بیزین بوده است و از آدرس http://www.mrc-bsu.cam.ac.uk/bugs به صورت رایگان قابل دانلود است. لطفاً دقت کنید که برنامه‌ی WinBUGS در حال حاضر چندان کاربرپسند نیست و گاهی اوقات با پیغام‌های خطایی سردرنیاوردنی از ادامه‌ی کار باز می‌ماند. با توجه به محبوبیت روزافزون روش‌های بیزین احتمال می‌رود که نرم‌افزارهای تجاری کاربرپسندتر و قدرتمندتری در سال‌های آتی به بازار عرضه شوند.

بخش ضمیمه جزئیات بیشتری راجع به دو سمت سکه‌ی محاسباتی بیزین ارائه می‌دهد: ایراد شماره‌ی ۳ «پیچیدگی و نبود نرم‌افزار» و مزیت شماره‌ی ۳ - «توانایی حل مسائل پیچیده‌تر».



بخش ۶
طراحی و تحلیل کارآزمایی ها

روش‌های بیزین برای طراحی کارآزمایی‌های بالینی ذاتاً مفید هستند، به این علت که این کارآزمایی‌ها معمولاً مرحله‌ای هستند، و، طراحی هر کارآزمایی به میزان زیادی بر اساس شواهد موجود از کارآزمایی‌های پیشین صورت می‌گیرد. منابع مهمی که برای طراحی کارآزمایی‌های بالینی با بهره‌گیری از روش‌های بیزین در دسترس هستند قطعاً در طراحی کارآزمایی‌های هزینه اثربخشی کاربرد دارند.

کارآزمایی‌های هزینه اثربخشی همیشه حاوی اطلاعات بالینی پیشین و احتمالاً شکلی از اطلاعات اقتصادی هستند، که در رویکرد فراوانی‌گرا از این اطلاعات برای تعیین الزامات قدرت کارآزمایی و به این ترتیب برای شناسایی اندازه‌ی نمونه استفاده می‌شود. از آن‌جایی که اطلاعات پیشین صراحتاً در روش‌های طراحی بیزین بیان می‌شوند، وابستگی طراحی انتخاب شده به اطلاعات پیشین کاملاً مشهود است. در تحلیل بیزین برای مشخص کردن میزان بزرگی اثر حاصل به تعیین دانش پیشین می‌پردازیم. برای مثال، در طراحی مرحله‌ی سوم کارآزمایی بالینی، اطلاعات حاصل از بررسی‌های مرحله‌ی دوم مبنای توزیع پیشین اثر قرار می‌گیرند. این امر امکان رویکرد آگاهی‌بخش در تنظیم اندازه‌ی نمونه را فراهم می‌آورد.

برای اندازه‌ی مشخصی از یک نمونه، در محاسبه‌ی بیزین احتمال این‌که کارآزمایی با موفقیت به نتیجه‌ی مثبتی ختم شود تعیین می‌شود. این امر را می‌توان مستقیماً به این تصمیم ربط داد که آیا یک کارآزمایی با اندازه‌ای مشخص (و از همین رو هزینه‌ای مشخص)، که این میزان اطمینان از موفقیت آن (و بازگشت مالی متعاقب آن) وجود دارد ارزش اجرا دارد یا خیر. این امر با محاسبات قدرت [رویکرد] فراوانی‌گرا در تضاد است، این محاسبات صرفاً احتمال اثبات یک اثر مشروط به اثر مجهول واقعی را که مقدار معینی برای آن لحاظ می‌شود ارائه می‌دهند.

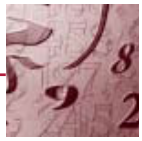
ویژگی تسهیل‌کنندگی مهم طراحی بیزین این است که امکان انجام تحلیل‌های موقتی interim analyses را بدون آن‌که تاثیری روی نتایج نهایی داشته باشند میسّر می‌کند و لازم نیست که چنین تحلیل‌هایی از قبل برنامه‌ریزی شده باشند. دلیل این امر آن است که تحلیل بیزین از تضاد موجود در تحلیل موقتی رویکرد فراوانی‌گرا متاثر نمی‌شود؛ مبنی بر این‌که، دو حامی مالی که کارآزمایی‌های یکسانی را انجام می‌دهند و نتایج یکسانی کسب می‌کنند در صورتی‌که یکی از آن‌ها یک تحلیل موقتی انجام دهد (اما به این علت کارآزمایی را متوقف نکند) و دیگری چنین تحلیلی انجام ندهد احتمالاً به نتیجه‌گیری‌های متفاوتی خواهند رسید. یک کارآزمایی مبتنی بر رویکرد بیزین را

می‌توان پیش از موعد متوقف کرد یا آن‌را به هر علت قابل قبولی طولانی‌تر کرد، بدون آن‌که به علت انجام چنین کارهایی در تحلیل‌های بعدی نیاز به جبران باشد.

رویکرد بیزین به غیر از طراحی کارآزمایی‌ها در تحلیل نتایج کارآزمایی‌ها نیز سودمند است. امروزه شاهد گرایش روزافزونی به بحث ارزیابی اقتصادی هستیم که موجب در نظر گرفتن هزینه اثربخشی به عنوان هدف ثانویه‌ی کارآزمایی‌های بالینیِ مرسوم شده است. این امر به بیان ساده به مفهوم جمع‌آوری داده‌های مربوط به استفاده از منابع در کنار کارآزمایی‌های متداول اثربخشی است، اما ممکن است داده‌های اقتصادی جامع‌تر، ثبت‌نام‌های واقع‌بینانه‌تر، و معیارها و یا ابزارهای مرتبط‌تر با نتایج را نیز دربر بگیرد. روش‌های تحلیلِ آماریِ چنین کارآزمایی‌هایی در حال ارتقاء هستند. مقاله‌ی مروری مفیدی در رابطه با اقدامات بیزین در این حوزه توسط اوهگان و استیونس (۲۰۰۲) نوشته شده است.

کارهای آماری اولیه روی حصول استنباط برای نسبت هزینه - اثربخشی افزایشی تمرکز داشتند، اما ویژگی‌های غیرعادی نسبت‌ها به دلایل مختلف منجر به راه‌حل‌هایی زیر حد مطلوب شدند. اخیراً به استنباط برای سود خالص (افزایشی) که به لحاظ آماری ساده‌تر است بیشتر پرداخته می‌شود. تحلیل‌های بیزین منحصراً رویکرد سود خالص را اتخاذ کرده‌اند. در واقع، هنگام استفاده از مزایای خالص، منحنی قابلیت پذیرش هزینه - اثربخشی ساده‌ترین بیان برای هزینه - اثربخشی نسبیِ دو درمان است؛ ماهیت ذاتاً بیزین این معیار در ضمیمه بحث شده است.

هزینه‌ها در کارآزمایی‌ها، مانند هر جای دیگری، همیشه تحریف‌شده هستند. روش‌های بیزین به سادگی با این ویژگی سازگار می‌شوند. اوهگان و استیونس مثال خوبی ارائه کردند که در آن نتیجه‌ی اثربخشی مضاعف بود. آن‌ها هزینه‌ها را به شکل توزیع لگاریتمی نرمال مدل دادند (که به طور آشکار مطابق با چولگی است) و توزیع‌های لگاریتمی نرمال را برای بیمارانی با نتایج اثربخشی مثبت و منفی در هر یک از گروه‌های درمانی حساب کردند. همچنین آن‌ها نشان دادند که در مجموعه داده‌ای که دو بیمار با هزینه‌های بسیار بالا در یک گروه قرار می‌گیرند چگونه حتی اطلاعات پیشین ساختاریِ ساده به ارائه‌ی استنباط‌های پسین واقع‌بینانه‌تر کمک خواهند کرد. چنین مدلی به سادگی توسط MCMC تحلیل می‌شود. استیونس و همکاران (۲۰۰۳) جزئیات کاملی از کد WinBUGS را به منظور محاسبه‌ی استنباط‌های پسین ارائه کردند. مثال خوب دیگری در این مورد مقاله‌ی ساتن و همکاران (۲۰۰۳) است.



بخش ۷
مدل های اقتصادی

ارزیابی اقتصادی با ساخت مدل‌های اقتصادی به شکل گسترده‌ای به کار بسته می‌شود. حتی هنگامی که داده‌های مربوط به هزینه‌ها همراه با اثربخشی در یک کارآزمایی بالینی جمع‌آوری می‌شوند باز هم برای ارزیابی مناسب اقتصادی کافی نیستند. دلیل این امر تا آن‌جا که مشخص شده است این است که الگوهای درمانی در کارآزمایی‌های بالینی تفاوت بسیاری با الگوهای درمانی در جامعه‌ی پزشکی دارند (در دومی بحث علاقه‌ی عملی practical interest مطرح است). در واقع چنین داده‌هایی برخی از ورودی‌ها را (برای مثال، داده‌های اثربخشی بالینی) به مدل هزینه - اثربخشی اعلام می‌کنند، در حالی‌که سایر مقادیر ورودی (برای مثال، استفاده از منابع، قیمت‌ها) از دیگر منابع حاصل می‌شوند.

ورودی‌های مدل‌های اقتصادی در بهترین حالت صرفاً برآوردی از مقادیر مجهول این پارامترها هستند، و این واقعیتی است که در عمل اجرای تحلیل حساسیت مشخص شده است. عموماً، این امر شامل یک تحلیل حساسیت یک‌طرفه‌ی فرمالیته می‌شود که در آن یک ورودی در آنِ واحد به ارزش‌های جایگزین موقتی ad hoc alternative value تغییر می‌کند و مدل مجدداً اجرا می‌شود تا مشخص شود که آیا نتیجه‌ی هزینه - اثربخشی تغییر کرده است یا خیر. حتی زمانی که هیچ‌یک از این تغییرات در مقدار پارامترهای واحد برای تغییر نتیجه کافی نباشد و در نتیجه مشخص نشود که کدام درمان هزینه - اثربخش تر است، این تحلیل مقدار [بازه‌ی] اطمینان را نشان نمی‌دهد که بتوانیم با پیوست آن به این استنباط آن‌را استنباطی صحیح بدانیم. مقادیر واقعی ورودی‌های افراد ممکن است خارج از محدوده‌ی بررسی شده باشند. علاوه بر این، اگر دو یا بیش از دو ورودی با هم در آن محدوده ها تغییر کنند، در آن صورت ممکن است نتیجه را تغییر دهند.

روش آماری صحیح برای ارزیابی عدم قطعیت در خروجی مدلی که از عدم قطعیت در ورودی آن حاصل می‌شود آنالیز حساسیت احتمالاتی probabilistic sensitivity analysis (PSA) نام دارد. این روش توسط NICE، دیگر سازمان‌های قانونی و بسیاری از متون دانشگاهی توصیه شده است. در این روش توزیع‌های احتمال به ورودی‌ها تخصیص داده می‌شوند، تا به این ترتیب عدم قطعیت موجود در مقادیر حقیقی آن‌ها نمایش داده شود، و سپس این عدم قطعیت از طریق مدل انتشار داده می‌شود. آگاهی در مورد مزیت‌های PSA در حال افزایش است، مقاله‌های بریگز و همکاران (۲۰۰۲) و پارمیگیانی (۲۰۰۲) دو مثال در این رابطه هستند.

مهم است بدانیم که در PSA، توزیع‌های احتمال را به پارامترهای مجهول نسبت می‌دهیم و

واضح است که این یک تحلیل بیزین خواهد بود. در واقع، روش‌های بیزین برای سالیان متمادی در اقتصاد سلامت به کار گرفته شده‌اند. تشخیص ماهیت بیزین این توزیع‌های احتمال نتایج مهمی را در پی دارد. توزیع‌ها باید با استفاده از نظرات مورد بحث در بخش مشخص‌سازی پیشین، تعیین شوند. به‌خصوص، شواهدی که در صدد پر کردن مدل‌های اقتصادی هستند، ندرتاً به طور مستقیم به پارامترهایی که آن مدل در نرم‌افزارها واقعاً به آن‌ها نیاز دارد ارتباط پیدا می‌کنند. داده‌های کارآزمایی از جمعیتی متفاوت (احتمالاً در کشوری دیگر) و نحوه مصرف متفاوتی به دست می‌آیند، داده‌های ریجستری ذاتاً سوگیری دارند و غیره. همان‌طور که درباره‌ی استفاده از داده‌های پیشین در تحلیل‌های معمول بیزین مطرح شد، ارتباط بین داده‌های استفاده شده برای پر کردن مدل و پارامترهایی که استفاده‌ی مطلوب ما از مدل را تعریف می‌کنند هنوز مشخص نشده است. معمولاً این تفاوت ها نادیده گرفته می‌شوند. با این حال، استفاده از برآوردها و خطاهای استاندارد گزارش شده در منابع مطالعاتی برای تعریف توزیع‌های ورودی موجب دست‌کم گرفتن عدم قطعیت واقعی خواهد شد.

تکنیک معمول برای PSA[روش] شبیه‌سازی مونته کارلو است، که در آن مجموعه‌های تصادفی از مقادیر ورودی کشیده می‌شوند و مدل برای هر مجموعه اجرا می‌شود. به این ترتیب نمونه‌ای از توزیع خروجی حاصل می‌آید (که بسیار شبیه نمونه‌برداری MCMC از توزیع پسین است). این کار زمانی امکان‌پذیر است که مدل به قدری ساده باشد که تقریباً بلافاصله روی کامپیوتر اجرا شود، در صورتی‌که برای مدل‌های پیچیده‌تر احتمالاً به دست آوردن نمونه‌ی به اندازه‌ی کافی بزرگی از اجراها ممکن نیست. برای چنین مواقعی، استفنسون و همکاران (۲۰۰۲) روش جایگزینی را بر اساس آمار بیزین برای محاسبه‌ی توزیع خروجی با استفاده از تعداد اجراهای بسیار کمتر ارائه کرده‌اند.

هنگامی که عدم قطعیت در خروجی مدل در PSA از طریق توزیع احتمال آن تعیین می‌شود، راه ساده برای بیان عدم قطعیت در مورد هزینه - اثربخشی مجدداً به‌وسیله‌ی منحنی قابلیت پذیرش هزینه - اثربخشی خواهد بود. همان‌طور که قبلاً اشاره شد، این یکی دیگر از ساختارهای ذاتاً بیزین است.

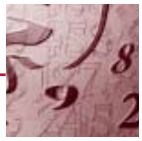
پاسخی ساده به عدم قطعیت درباره‌ی هزینه - اثربخشی این است که از خود بپرسیم آیا کسب داده‌های بیشتر عدم قطعیت را کاهش می‌دهد یا خیر. برای مثال، در انگلستان، یکی از تصمیماتی که NICE در هنگام پرسشی مبنی بر تصمیم‌گیری در مورد هزینه - اثربخشی یک دارو می‌گیرد این است که بگوید در حال حاضر شواهد کافی نیستند، و تایید دارو به منظور بازپرداخت توسط سازمان ملی خدمات سلامت را به هنگامی موکول کند که داده‌های بیشتری حاصل شده باشند. نظریه‌ی تصمیم‌گیری بیز روش ساده‌ای را به لحاظ مفهومی برای آگاهی دادن از چنین تصمیمی فراهم می‌آورد، که در واقع از طریق محاسبه‌ی مقدار مورد انتظار expected value از اطلاعات نمونه‌ها است.

مقدار مورد انتظار از محاسبات اطلاعات توسط فلی و هازن (۱۹۹۸)، کلاکستون و پوزنت (۱۹۹۶)،

برنان و همکاران (۲۰۰۳) و یک محاسبه‌ی بیزین برای مدل‌های پیچیده که توسط اوکلی (۲۰۰۲) توسعه یافته تایید شده است. این پیوند مستحکمی بین چنین تحلیل‌هایی و طراحی کارآزمایی‌ها است، از آن‌جایی که تعدیل مقدار مورد انتظار از اطلاعات نمونه‌ها در برابر هزینه‌های نمونه‌گیری در واقع یک روش استاندارد بیزین برای تشخیص اندازه‌ی بهینه‌ی نمونه‌ها محسوب می‌شود.

ارتباط مهم دیگری بین تحلیل مدل‌های اقتصادی و تحلیل کارآزمایی‌های هزینه اثربخشی وجود دارد. زمانی که شواهد برای پارامترهای منفرد در یک مدل اقتصادی از یک کارآزمایی یا سایر داده‌های آماری حاصل می‌شوند، توزیع طبیعی به منظور اختصاص به آن پارامترها همان توزیع پسین آن‌ها است که از تحلیل بیزین داده‌های خام ایجاد شده است. به این ترتیب فرض گرفته می‌شود که داده‌ها ارتباط مستقیمی با پارامتر مورد نیاز در مدل دارند، نه آن که ضرورتاً به پارامتری مشابه اما متفاوت مربوط می‌شوند. در مورد دوم، پیوند دادن توزیع پسینِ حاصل از تحلیل داده‌ها به پارامترهای مورد نیاز برای مدل، با استفاده از اطلاعات پیشین ساختاری، ساده است.

این پیوند تحلیل آماریِ داده‌های کارآزمایی به ورودی‌های مدل اقتصادی شکلی از تلفیق شواهد evidence synthesis است و ماهیت کل‌گرای رویکرد بیزین را نشان می‌دهد. مثال‌هایی در این مورد توسط آدس و لو (۲۰۰۲) و کوپر و همکاران (۲۰۰۲) ارائه شده است. آدس و همکارانش شواهد را از منابع همپوشان داده‌ها درون یک تحلیل بیزین تلفیق می‌کنند. تلفیق شواهد دقیقاً همان کاری است که قضیه‌ی بیز انجام می‌دهد.



ضمیمه

استنباط - جزئیات

زیربخش‌های بعدی جزئیاتی از استدلال‌های مطرح شده در بخش «استنباط» و جزئیاتی از تفاوت‌های اصلی بین آمار بیزین و فراوانی‌گرا را که در جدول ۱ مشخص شده است ارائه می‌دهند.

ماهیت پارامترها

اساسی‌ترین تمایز بین رویکردهای بیزین و فراوانی‌گرا این است که در رویکرد بیزین، پارامترهای مجهول موجود در مدل آماری متغیرهای تصادفی هستند. بر همین اساس، در یک تحلیل بیزین، پارامترها توزیع احتمال دارند، در صورتی که در تحلیل فراوانی‌گرا پارامترها ثابت هستند اما مقادیرشان مجهول است و مجاز به بیان احتمال راجع به آن‌ها نیستیم. این مسئله ممکن است فرد غیر حرفه‌ای را سردرگم کند، زیرا یکی از بیانیه‌های فراوانی‌گراها مثل فاصله‌ی اطمینان مشخصاً به بیان احتمال درباره‌ی پارامترهای مجهول می‌پردازد. اگر این ادعا را مشاهده کنیم که [۳٫۵، ۱۱٫۶] مساوی با فاصله‌ی اطمینان ۹۵٪ برای پارامتر μ است، صراحتاً به این معنی است که ۹۵٪ احتمال دارد که μ بین ۳٫۵ و ۱۱٫۶ قرار گرفته باشد. نه، این ادعا نمی‌تواند چنین معنایی داشته باشد زیرا μ در استنباط فراوانی‌گرا یک کمیّت تصادفی محسوب نمی‌شود. هنگام بحث ماهیت استنباط‌ها خواهیم دید که این ادعا واقعاً چه معنایی دارد.

در آمار بیزین، پارامترها توزیع احتمال دارند، و اگر یک تحلیل بیزین فاصله‌ی ۹۵٪ ایجاد کند دقیقاً چنین تفسیری خواهد داشت که معمولاً روی فاصله ی اطمینان قرار می‌گیرد.

ماهیت احتمال

اساس تمایز بین این دو رویکرد آماری در مورد ماهیت پارامترها تفاوتی است که در چگونگی تفسیر آن‌ها از احتمال وجود دارد. در آمار فراوانی‌گرا، احتمال صرفاً به صورت فراوانی بلندمدت، حدّی، و نسبی تعریف می‌شود. این تعریف آشنایی است که در دوره‌های آموزشی مقدماتی استفاده می‌شود، که غالباً با مثال‌هایی همچون پرتاب سکه به دفعات بسیار زیاد و دقت به فراوانی بلندمدت و حدّی «شیرها» بیان می‌شود. علت این امر آن است که احتمال فراوانی‌گرا بر پایه‌ی این تعریف فراوانی از احتمال بنا نهاده شده است که ما آن روش‌های مرسوم را «فراوانی‌گرا» می‌نامیم. در مقابل،

آمار بیزین، به چنین تفسیری از احتمال تکیه دارد که احتمال به میزانی از باور افراد اطلاق می‌شود. گرچه که این بیان برای برخی از افراد «سردرگم‌کننده» و «غیرعلمی» به نظر می‌رسد، اما باید بپذیریم که آمار بیزین بر اساس همین تفسیر به تحلیل‌های بسیار پیشرفته و موفقی دست یافته است. همان‌طور که در اطلاعات پیشین شرح داده شد، این امر به ذهنیت و اقدامات غیرعلمی خارج از کنترل منتهی نمی‌شود.

لزوم چنین تفسیری هنگامی بیشتر مشخص می‌شود که بدانیم چه تعداد از رویدادهایی که می‌خواهیم وجود احتمال را برایشان متصور شویم هرگز ممکن نیست که احتمال فراوانی‌گرا داشته باشند. احتمال یک فرضیه مثالی واضح است، زیرا یک فرضیه‌ی معین یا صحیح است و یا صحیح نیست، و نمی‌توانیم برای آن همانند مثال پرتاب سکه تکرارهای تجربی را فرض کنیم. احتمال بارش باران در روز بعد یا احتمال این‌که در طول زندگی‌تان دچار سکته‌ی قلبی (MI) شوید از جمله مثال‌های دیگر هستند. فردا روز منحصر به فردی است، و وضعیت هواشناسی آن از قبل دقیقاً مشخص نیست، و شما هم فرد منحصر به فردی هستید که آرایش ژنتیکی و سبک زندگی‌تان شما را در برهه‌ای از زندگی کمتر یا بیشتر در معرض سکته‌ی قلبی قرار می‌دهد و این الگو دقیقاً با الگوی فردی دیگر همخوان نیست. با وجود یگانه بودن این رویدادها، معمولاً به تصور در مورد احتمال آن‌ها می‌پردازیم. اکثر رویدادها و متغیرهای غیرقطعی مورد علاقه‌ی دانشمندان و پزشکان آن‌هایی هستند که یگانه‌اند، و تفسیر فراوانی از احتمال قادر به بیان احتمال آن‌ها نیست.

ماهیت استنباط‌ها

در رویکرد فراوانی‌گرا احتمالات باید بر مبنای تکرار باشند. این ادعا که [۳٫۵ ، ۱۱٫۶] مساوی با فاصله‌ی اطمینان ۹۵٪ برای پارامتر μ است به این معنا است که اگر این تجربه را بارها تکرار کنیم، و اگر هر بار فاصله را با استفاده از قاعده‌ای محاسبه کنیم که این‌بار برای به دست آوردن فاصله‌ی [۳٫۵ ، ۱۱٫۶] استفاده کردیم، در آن صورت ۹۵٪ از آن فاصله‌ها حاوی μ هستند. احتمال ۹۵٪ یکی از ویژگی‌های قاعده‌ی مورد استفاده در ایجاد فاصله، و نه خود فاصله، است. ما مجاز نیستیم، و همینطور صحیح هم نیست، که این احتمال را به فاصله‌ی حقیقی [۳٫۵ ، ۱۱٫۶] نسبت دهیم. این استدلال بسیار غیرشهودی است. ادعاهای رویکرد فراوانی‌گرا همانند این موارد حتی توسط آماردانان حرفه‌ای نیز به اشتباه تفسیر شده‌اند. این تفسیر غلط از فاصله‌ی اطمینان، مبنی بر این‌که پارامتر μ به احتمال ۹۵٪ در فاصله‌ی معینی [۳٫۵ ، ۱۱٫۶] قرار می‌گیرد، هم توسط آماردانان و هم افراد غیرحرفه‌ای بیان شده است. با این وجود، این تفسیر اشتباه است. اما یک فاصله‌ی بیزین دارای چنین تعبیری است.

رویکرد بیزین از اصطلاحات متفاوتی نسبت به اصطلاحات مشابه در رویکرد فراوانی‌گرا استفاده می‌کند. فاصله‌های بیزین معمولاً فاصله‌های قابل باور credible intervals نامیده می‌شوند تا تفاوت

آن‌ها با فواصل اطمینان مشخص شود. بیشترین فاصله‌ی چگالی density interval فاصله‌ی قابل باوری است که کوتاه‌تر از تمام فاصله‌های چگالی موجود باشد و احتمال وجود مقدار واقعی پارامتر نیز در مورد آن مطرح شود.

استدلال کاملاً مشابهی را می‌توان در مورد آزمون‌های معناداری رویکرد فراوانی‌گرا مطرح کرد. اگر یک فرضیه با مقدار P برابر با ۱٪ رد شود، این عبارت به این معنا نیست که تنها به احتمال ۱٪ آن فرضیه درست است. در تحلیل فراوانی‌گرا، برای فرضیه‌ها بیشتر از پارامترها احتمال وجود ندارد. مجدداً مقدار P باید مبنی بر تکرار قاعده باشد. در این مورد، این قاعده است که می‌گوید هنگامی‌که داده‌ها برخی شرایط را برآورده می‌کنند ما رسماً فرضیه را رد می‌کنیم. تفسیر صحیح مقدار P این است که، اگر آزمایش را بارها تکرار کنیم، و اگر فرضیه هر بار واقعاً صحیح باشد، آن‌وقت فقط در ۱٪ از این آزمایش‌ها است که قاعده (به اشتباه) ما را به رد آن فرضیه می‌رساند. این تصور گمراه‌کننده و پیچیده است. جای تعجب نیست که پزشکان معمولاً مقدار P را به اشتباه احتمال درست بودن فرضیه تفسیر می‌کنند.

تفسیر مقدار P به این شیوه نه تنها اشتباه است بلکه ممکن است بسیار گمراه‌کننده باشد. خطر از آن‌جا ناشی می‌شود که این تفسیر این مسئله را نادیده می‌گیرد که فرضیه در وهله‌ی اول چقدر منطقی بوده است. سه مثال را مشاهده می‌کنید:

مثال‌ها:

۱) **غربالگری:** آزمون غربالگری برای بیماری نادری را در نظر بگیرید. آزمون بسیار دقیق است، و نرخ مثبت کاذب و منفی کاذب آن ۰٫۱٪ است (یعنی، تنها به یک نفر از هزار نفری که به بیماری مبتلا نیستند نتیجه‌ی مثبت داده می‌شود، و تنها یک نفر از هزار نفر مبتلا به بیماری نتیجه‌ی منفی دریافت می‌کند). شما آزمون را انجام می‌دهید و نتیجه‌تان مثبت می‌شود. حالا باید چه فکری بکنید؟ از آن‌جایی که این غربالگری فقط یک مورد از بین هزار مورد اشتباه می‌کند، آیا نتیجه‌ای که گرفته‌اید به این معنا نیست که با قطعیت ۹۹٫۹٪ به بیماری مبتلا هستید؟ در شرایط آزمودن فرضیه، نتیجه‌ی مثبت این امکان را به شما می‌دهد که فرضیه‌ی صفر مبنی بر این‌که به بیماری مبتلا نیستید را در سطح ۰٫۱٪ معناداری رد کنید، نتیجه‌ی کاملاً معناداری که با آن تشخیص ۹۹٫۹٪ موافق است. اما این بیماری نادر است، و می‌دانیم که در عمل اکثر نتیجه مثبت‌هایی که برای آزمایش‌های بیشتر گزارش می‌شوند در دسته‌ی مثبت‌های کاذب قرار خواهند گرفت. اگر فقط یک نفر از بین ۵۰۰۰۰ نفر به این بیماری مبتلا باشد، احتمال شما مبنی بر مبتلا بودن پس از آزمون غربالگری مثبت کمتر از ۱ در ۵۰ خواهد بود.

گرچه که ممکن است این مثال مشخصاً درباره‌ی آزمون فرضیه‌ها نباشد اما عملاً شباهتی مستقیم با آن دارد. می‌توانیم از این مشاهده‌ی نتیجه‌ی غربالگری مثبت به عنوان داده‌ی استفاده کنیم که با آن این فرضیه‌ی صفر را بیازماییم که به موجب آن شما به بیماری مبتلا نیستید. اگر فرضیه‌ی صفر صحیح باشد، آن‌وقت این مشاهده کاملاً بعید خواهد بود، و آن‌وقت رسماً می توانیم فرضیه‌ی صفر را با مقدار P برابر با ۰٫۰۰۱ رد کنیم. اما، احتمال واقعی فرضیه‌ی صفر بیش از ۰٫۹۸ است. این مثال جالبی

است که در آن احتمال فرضیه (< ۰٫۹۸) کاملاً متفاوت از مقدار P (۰٫۰۰۱) است. این متفاوت بودن از آن‌جا ناشی می‌شود که فرضیه‌ی صفر مبنی بر این‌که شما مبتلا به بیماری هستید با احتمال بسیار اندکی مطرح می‌شود. کسی که با ماهیت آزمون‌های غربالگری آشنا است بعید است که به اشتباه نرخ مثبت کاذب را احتمال ابتلا به بیماری تفسیر کند (اما لازم است که این تمایز برای بیماران مشخص شود!). به همین ترتیب، تفسیر مقدار P به صورت احتمال فرضیه‌ی صفر اشتباه است، زیرا چنین تفسیری احتمال پسین را دقیقاً به همین روش درنظر نمی‌گیرد.

۲) تحلیل زیرگروه‌ها: متخصصین آمار پیوسته راجع به جستجو در داده‌ها برای یافتن تاثیرات معنادار در زیرگروه‌ها هشدار داده‌اند. در کارآزمایی‌های بالینی، معمولاً تحلیل زیرگروه‌ها صرفاً در صورتی مجاز است که آن‌ها از پیش تعیین شده و دارای مکانیسم بیولوژیکی منطقی باشند. در این مورد به وضوح مشخص می‌شود که تفسیر معناداری بستگی به این دارد که فرضیه در وهله‌ی اول چقدر منطقی بوده باشد.

۳) توسعه‌ی داروها: شرکت‌های داروسازی ترکیبات فراوانی را سنتز می‌کنند و در آن‌ها به دنبال تاثیرات مطلوب بالینی می‌گردند. همانند مثال مربوط به غربالگری، حتی پس از آن‌که یک کارآزمایی تاثیر بسیار معناداری را ایجاد می‌کند، احتمال واقعی بودن این تاثیر ممکن است زیاد نباشد؛ از آن‌جایی که انتظار مثبت‌های کاذب را داریم. علی‌رغم این موضوع، و علی‌رغم تجربیات حاصل از بررسی داروهای فراوانی که به مرحله‌ی سوم و پرهزینه‌ی کارآزمایی‌های بالینی راه پیدا می‌کنند تا بی‌اثر بودنشان مشخص شود، شرکت‌های داروسازی همچنان به اشتباه و با خوش‌بینی زیاد مقادیر معنادار P را تفسیر می کنند.

تحلیل بیزین بر خلاف رویکرد فراوانی‌گرا می‌تواند احتمالی مبنی بر درست یا غلط بودن فرضیه ارائه دهد. آزمون فرضیه‌ی بیزین هم همین کار را انجام می‌دهد؛ صحیح یا غلط بودن احتمال فرضیه‌ی موجود در مسئله را بیان می‌کند. یک تحلیل بیزین صحیح باید قادر به تشخیص صحیح احتمالات در تمامی مثال‌های ذکر شده در بالا باشد.

برآوردهای نقطه‌ایِ point estimates رویکرد فراوانی‌گرا نیز بر پایه‌ی تکرار هستند، و گرچه که در روش‌های بیزین ندرتاً به اشتباه تفسیر شده‌اند، اما همچنان تفاوت های مهمی در این مورد بین دو دیدگاه وجود دارد. تصور کنید که یک تحلیل فراوانی‌گرا برآورد نااریب μ را برابر با ۹٫۱ گزارش می‌کند. نااریبی یکی از ویژگی‌های قاعده‌ی برآورد کردن، و نه خود برآورد، است، و در این مورد به این معنا است که اگر آزمایش را بارها تکرار کنیم و هر بار قاعده‌ی یکسانی را برای به دست آوردن برآورد μ به کار بگیریم، در آن صورت مقدار میانگین این برآوردها μ خواهد بود. به طور متوسط، برآوردها نه خیلی بالا و نه خیلی پایین خواهند بود، اما نمی‌توان راجع به این‌که آیا بنا بر انتظار مقدار ۹٫۱ بسیار بالا یا بسیار پایین است اظهار نظر کرد. تحلیل بیزین احتمالاً اینطور گزارش خواهد کرد که ۹٫۱ مقدار مورد انتظار از μ است، و بنابر تفسیر آن انتظار نمی‌رود که ۹٫۱، به عنوان برآورد μ، بسیار بالا یا بسیار پایین باشد.

مثال:

مقدار عوارض جانبی یک دارو را در نظر بگیرید. در یک کارآزمایی که مشتمل بر ۵۰ بیمار بود، داروی مورد نظر موجب پدید آمدن عوارض جانبی نشد. برآوردگر نااریب استانداردِ نرخ عوارض جانبی به ازای هر بیمار صفر (۰/۵۰) است. بر چه اساس قبول کنیم که این مقدار «میانگین است و بسیار بالا یا بسیار پایین نیست»؟ مشخص است که این مقدار بسیار بالا نیست و تقریباً با قاطعیت معلوم است که بسیار پایین است. واقعیت این است که قاعده‌ی برآورد (که یعنی درنظر گرفتن تعداد بیمارانی با عوارض جانبی و تقسیم آن عدد بر ۵۰) برآوردهایی را حاصل می‌آورد که به طور متوسط نه بسیار بالا و نه بسیار پایین هستند، البته در صورتی که با مجموعه داده‌های جدیدی این قاعده را تکرار کنیم. همچنین مشخص است که نمی‌توانیم این تفسیر را به برآوردهای فردی تعمیم دهیم. انجام این کار همانند تفسیر مقدار P به شکل احتمال صحیح بودن فرضیه‌ی صفر است که مشخصاً اشتباه است. در تحلیل بیزین، با توجه به این‌که موردی از عوارض جانبی در بین ۵۰ بیمار مشاهده نشده است، نرخ مورد انتظار از عوارض جانبی مثبت خواهد بود. به علاوه، از تفسیر انتظار پسین هم چنین برمی‌آید که این برآورد نه بسیار بالا و نه بسیار پایین است.

استنباط‌های طبیعی‌تر و سودمندتر

در استنباط‌های فراوانی‌گرایانه، واژگان «اطمینان»، «معنادارای» و «نااریبی» اصطلاحاتی فنی هستند، و باید آن‌ها را بر اساس تعاریفشان تفسیر کرد.

این واقعیت که پزشکان همیشه، اما به اشتباه، مایل به تفسیر فاصله‌ی اطمینان به شکل بیانی از احتمال پارامتر هستند، شاهدی بر این موضوع است که رویکرد بیزین شهودی‌تر و طبیعی‌تر است و پاسخ‌های واضح‌تری به سوالات مراجعان می‌دهد. تفسیر مقدار P به صورت احتمال آن‌که فرضیه‌ای صحیح باشد نیز وسوسه‌انگیز است، زیرا این دقیقاً همان چیزی است که پزشک می‌خواهد بشنود. مجدداً بیان می‌کنیم که استنباط بیزین به صورت طبیعی و مستقیم به نیازهای پزشک می‌پردازد.

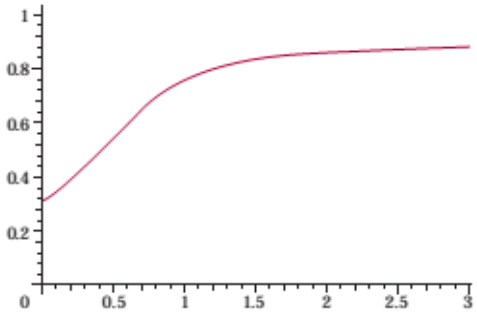
این مزیت مهم شماره‌ی ۱ رویکرد بیزین در بخش مرور کلی است «استنباط‌هایی طبیعی‌تر و سودمندتر» و به یک مزیت واقعی برای اقتصاددانان سلامت تبدیل شده است. برای مثال، یکی از متداول‌ترین روش‌های ارائه‌ی تحلیل اثربخشی هزینه‌ها به واسطه‌ی منحنی قابلیت پذیرش هزینه - اثربخشی (CEAC) است که توسط وان هوت و همکاران (۱۹۹۴) مطرح شد. مثالی در این مورد در تصویر الف۱ ارائه شده است. برای هر مقدار از آستانه‌ی تمایل به پرداخت CEAC، (۸) احتمالی را نشان می‌دهد که یک درمان از دیگری هزینه - اثربخش‌تر است.

تاکنون باید برای خواننده روشن شده باشد که این احتمال صرفاً در چارچوب بیزین معنادار است. این احتمال به احتمال یک رویداد یگانه اشاره می‌کند (هزینه - اثربخشی نسبیِ این دو درمان خاص یگانه است و تکرارپذیر نیست)، و آن رویداد به صورت پارامترهای مجهولِ مدل آماریِ استفاده شده در

تحلیل شواهدِ موجود بیان شده است. خاطرنشان می‌کنیم که ایجاد یک جایگزین فراوانی‌گرایانه برای CEAC امکان‌پذیر است، که به صورت مقادیر P تعریف می‌شود (اوهگان و همکاران، ۲۰۰۰؛ لوتگرن و زتریاس، ۲۰۰۰)، و هر مقدار از ۸ را روی نمودار نشان می‌دهد، با این احتمال که داده‌ها در مجموعه‌ای محدود به داده‌های مشاهده شده قرار می‌گیرند، تصور می‌شود که واقعیت فرضیه این است که دو درمان به یک میزان هزینه - اثربخش هستند. با این‌حال، هنگامی که تحلیل بیزین به CEAC مستقیم‌تر و سودمندتری منتهی می‌شود که احتمال هزینه -اثربخش تر بودن درمان ۲ را در مقایسه با درمان ۱ نشان می‌دهد، دیگر نیازی به اتخاذ رویکرد فراوانی‌گرا نیست.

تصویر الف ۱

نمودار CAEC



احتمال این‌که درمان جدید مقرون‌به‌صرفه‌تر از دارونما باشد، در مقابل تمایل به پرداخت رسم شده است [در واحدهای ۱۰۰۰۰ دلار به ازای هر سال زندگی تعدیل شده بر حسب کیفیت (QALY)].

روش بیزین - جزئیات

زیربخش‌های بعدی جزئیات بیشتری درباره‌ی روش بیزین ارائه می‌دهند.

قضیه‌ی بیز

ساده‌ترین راه برای بیان قضیه‌ی بیز بدون استفاده از نمادهای ریاضی از این قرار است:

[توزیع] پسین متناسب است با [توزیع] پیشین ضربدر درست‌نمایی

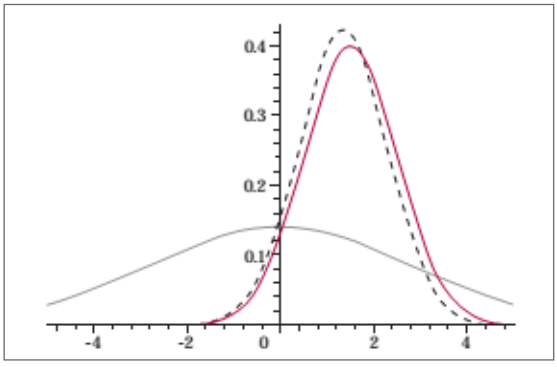
در این جمله باید چندین اصطلاح را تعریف کرد و توضیح داد. بهتر است که به «تصویر ۱» در

جعبه‌ی «مثالی از قضیه‌ی بیز» رجوع کنید. «پسین» به معنای توزیع پسین پارامتر(های) مجهول است. به بیان دقیق‌تر، این تابع چگالی احتمال پسین است که به شکل منحنی نقطه‌چین سیاه در تصویر۱ نشان داده شده است. به همین ترتیب، «پیشین» به معنای توزیع پیشینِ پارامتر(های) ناشناخته، و نیز در شکل تابع چگالی احتمال پیشین است و با خط خاکستری در تصویر۱ نشان داده شده است. «درست‌نمایی» تابع مشترکی در هر دو نظریه‌ی فراوانی‌گرا و بیزین است، و بیان‌گر اطلاعات موجود در داده‌ها است و به شکل منحنی قرمز در تصویر۱ نشان داده شده است. برای هر مقدار معینی از پارامتر مجهول، درست‌نمایی معمولاً احتمال مشاهده‌ی اطلاعاتی را نشان می‌دهد که در واقع مشاهده شده‌اند. قضیه‌ی بیز توضیح می‌دهد که باید دو منحنی را ضرب کنیم. از آن‌جایی که مساحت زیر هر یک از منحنی‌های چگالی باید برابر با ۱ باشند، محصول را به منظور ایجاد چنین وضعیتی میزان scale می‌کنیم، که قضیه‌ی بیز آن را این چنین بیان می‌کند: توزیع پسین «متناسب با» محصول است. از این‌رو، منحنی نقطه‌چین سیاه در تصویر۱ حاصل ضرب منحنی‌های خاکستری و قرمز و میزان کردن نتیجه به نحوی است که مساحت زیر آن برابر با ۱ باشد.

به علاوه، این مکانیسم ضرب کردن دو منحنی مشخص می‌کند که قضیه‌ی بیز به هر یک از منابع اطلاعاتی بر اساس قدرتش وزن می‌دهد. شرایطی را در نظر بگیرید که اطلاعات پیشین خیلی ضعیف هستند. این وضعیت را می‌توان با منحنی خاکستری کاملاً مسطحی نشان داد، و به این ترتیب احتمال پیشینِ کم و بیش یکسانی به طیف وسیعی از مقادیر پارامترها داده می‌شود. هنگامی که قضیه‌ی بیز را اعمال می‌کنیم، توزیع پسین تقریباً برابر با درست‌نمایی خواهد شد، و از آن‌جایی که برای کامل شدن تا ۱ باید میزان شود، توزیع پسین در واقع همانند منحنی قرمز است. این وضعیت در تصویر۲ که در آن اطلاعات پیشین را نسبت به تصویر۱ تضعیف کرده‌ایم نشان داده شده است.

تصویر الف۲

نمودار سه تایی با اطلاعات پیشین ضعیف‌تر



هنگامی که اطلاعات پیشین نسبت به اطلاعات داده‌ها بسیار ضعیف باشند، توزیع پیشین اهمیت بسیار کمی در قضیه‌ی بیز پیدا می‌کند به نحوی که توزیع پسین عملاً همان درست‌نمایی می‌شود. در چنین وضعیتی انتظار داریم، و در مسائل ساده غالباً در می‌یابیم که، روش‌های بیزین به استنباط‌های مشابهی با استنباط‌های حاصل از روش‌های فراوانی‌گرای مرسوم منتهی می‌شوند. روش‌های بیزین قادر به بهره‌گیری از اطلاعات بیشتری در مقایسه با روش‌های فراوانی‌گرا هستند، اما اگر به هر منبع اطلاعاتی وزن درخور آن را بدهیم، در آن صورت است که وزن اطلاعات ضعیف طبیعتاً کاهش می‌یابد.
یکی دیگر از ویژگی‌های سودمند پارادایم بیزین در این عبارت ساده به زیبایی بیان شده است:

پسینِ امروز پیشینِ فردا است.

این پارادایم درباره‌ی یاد گرفتن است، و همیشه می‌توانیم بیشتر یاد بگیریم. هنگامی که داده‌های بیشتری کسب می‌کنیم، قضیه‌ی بیز به ما می‌گوید که چگونه دانشمان را به منظور ادغام داده‌های جدید به‌روز رسانی کنیم. توزیع پسین قدیمی حاوی تمام چیزهایی است که پیش از دیدن داده‌های جدید می‌دانستیم، و به همین خاطر به توزیع پیشین جدید تبدیل می‌شود. قضیه‌ی بیز این توزیع را با داده‌های جدید تلفیق می‌کند تا توزیع پسین جدیدی ارائه دهد. و این روال به همین ترتیب ادامه می‌یابد ... روش های بیزین برای کارآزمایی‌های پی‌درپی ایده‌آل هستند!
قضیه‌ی بیز همچنین بیشتر مشخص می‌کند که چرا تفسیرهای معمول از استنباط‌های اشتباه فراوانی‌گرا غلط هستند. درست‌نمایی احتمال به دست آوردن داده‌های واقعی را - با درنظر گرفتن هر مقدار مشخصی از پارامتر - بیان می‌کند. به بیان ساده،

درست‌نمایی = احتمال (داده‌ها | پارامترها)
Likelihood = P (data | parameters)

این توزیع مبنای استنباط فراوانی‌گرا محسوب می شود. از سوی دیگر، توزیع پسین نیز مبنای استنباط بیزین است، که توزیع احتمالِ پارامترها با توجه به داده‌های واقعی است،

توزیع پسین = احتمال (پارامترها | داده‌ها)
Posterior = P (parameters | data)

در واقع جابه‌جایی غیر موجه بین پارامترها و داده‌ها است که منجر به تفسیرهای اشتباه می‌شود.

برای مثال، مقدار P برابر است با (فرضیه | داده‌ها) P، در حالی‌که آن‌چه یک تصمیم‌گیرنده می‌خواهد، و آن‌چه از استنباط بیزین حاصل می‌شود، این است که (داده‌ها | فرضیه) P. کاملاً واضح است که این دو کاملاً متفاوت هستند، و قضیه‌ی بیز ارتباط بین آن‌ها را نشان می‌دهد: تنها با ترکیب مقدار P و توزیع پیشین است که می‌توانیم احتمال پسین مطلوبمان را استخراج کنیم.

استنباط بیزین

در رویکرد بیزین، تمام استنباط‌ها از توزیع پسین حاصل می‌شوند. هنگامی‌که تحلیل بیزین فاصله‌ی احتمالی (فاصله‌ی قابل باور) را برای یک پارامتر گزارش می‌دهد، این همان فاصله‌ی پسین است که از توزیع پسین پارامتر - و نه تنها بر اساس داده‌ها بلکه هرگونه اطلاعات یا دانش دیگری که پژوهشگر در اختیار دارد - حاصل شده است. این احتمال که یک فرضیه صحیح است احتمال پسین محسوب می‌شود و در مورد برآورد یک پارامتر می‌توان میانگین پسین (مقدار «مورد انتظار») را مثال زد.

این‌ها نظایر بیزین - Bayesian analogues از سه نوع استنباط موجود در چارچوب [رویکرد] فراوانی‌گرا هستند. با این حال، استنباط بیزین بسیار منعطف‌تر است.

معمولاً پرسش حقیقی مورد نظر با این حالت‌های استنباط فراوانی‌گرایانه جور درنمی‌آید. برای مثال، پژوهشگر معمولاً می‌خواهد بداند «حالا که داده‌ها را دیده‌ایم درباره‌ی این پارامتر چه می‌دانیم؟». رویکرد فراوانی‌گرا حتی پاسخ ساده‌ای هم برای این سوال عادی ندارد. پاسخ بیزین همان سادگی است؛ چگالی پسین را رسم می‌کنیم. از این‌رو، منحنی نقطه‌چین سیاه در تصویر ۱ آن‌چه را که پس از تلفیق تمام شواهد موجود در مورد پارامتر می‌دانیم به طور کامل بیان می‌کند.

نظریه‌ی تصمیم‌گیری مثال خوب دیگری از انعطاف‌پذیری استنباط بیزین ارائه می‌دهد. در این تئوری، مجموعه‌ای از تصمیمات احتمالی و تابع مطلوبیت وجود دارد، و تابع مطلوبیت مشخص می‌کند اگر پارامترها دارای مقادیر خاصی باشند چطور می‌توان تصمیم خاصی گرفت. برای مثال، برای یک آزمون فرضیه می‌توانیم تابع مطلوبیتی را تعریف کنیم که می‌گوید اگر صحیح بودن فرضیه مشخص شود پذیرفتن آن درست است (مطلوبیت بالا)، یا در صورت غلط بودن آن باید آن را رد کرد (مطلوبیت پایین).

اگر پارامترها را می‌شناختیم تصمیم‌گیری ساده بود؛ برای آن مقدار پارامترها باید آن تصمیمی را انتخاب کنیم که بالاترین مطلوبیت را دارد. با این حال، پارامترها عمدتاً مجهول هستند. بر طبق نظریه‌ی تصمیم‌گیری، باید تصمیمی را برگزینیم که بالاترین مطلوبیت مورد انتظار (پسین) را دارد. این انتظار برابر با میانگین مقادیر مطلوبیت است، که میانگین آن با توجه به توزیع پسین پارامترها محاسبه شده است. این یک بیان فنی است، اما نکته این‌جا است که روش‌های فراوانی‌گرا ما را به آن تصمیم بهینه نمی‌رسانند. این مورد ضرورتاً یک ساختار بیزین است و در عین حال روش دیگری است که توزیع پسین ما را در پاسخ دادن به سوال مورد نظرمان یاری می‌رساند.

برای یک ارائه‌دهنده‌ی مراقبت سلامت که در حال انتخاب بین دو داروی جایگزین است، سود خالص net benefit تابع مطلوبیت مناسبی محسوب می‌شود. سود خالصِ (به بیان دقیق، سود پولی خالص) یک داروی معین به این ترتیب به دست می‌آید: با میانگین گرفتن از اثربخشی دارو، ضرب آن در قیمتی که این ارائه‌دهنده تمایل به پرداخت آن دارد، به نحوی که یک واحد افزایش در اثربخشی دیده شود، و سپس منها کردن میانگین هزینه‌ای که ارائه‌دهنده برای استفاده از این دارو می‌پردازد. به این ترتیب، قاعده‌ی به حداکثر رساندنِ مطلوبیتِ مورد انتظار می‌گوید که ارائه‌دهنده باید دارویی با سود خالص مورد انتظار بیشتر را برگزیند. یعنی، اگر سود خالص افزایشی مورد انتظار expected incremental net benefit داروی ۲ نسبت به داروی ۱ مثبت باشد، ارائه‌دهنده باید داروی ۲ را برگزیند (کلاکستون، ۱۹۹۹).

اطلاعات پیشین - جزئیات

زیربخش‌های بعدی راجع به توضیحات بخش اطلاعات پیشین در متن اصلی جزئیات بیشتری ارائه می‌دهند.

ذهنیت

در زیربخش ماهیت احتمال در این ضمیمه به این واقعیت اشاره شده است که روش‌های بیزین بر اساس تفسیر ذهنی از احتمال هستند. در آن‌جا توضیح داده‌ایم که چنین نحوه‌ی بیانی در صورتی که بخواهیم احتمالات را به پارامترها و فرضیه‌ها نسبت دهیم ضرورت پیدا می‌کند، زیرا تفسیر رویکرد فراوانی‌گرا از احتمال بسیار بسته است. با این‌حال، [وجود عامل ذهنیت] روش‌های بیزین را در معرض اتهام [وجود عامل] ذهنیت قرار می‌دهد که از نظر بسیاری غیر قابل قبول و غیر علمی است.

با این‌حال، نمی‌شود که علم فقط بر اساس عینیّت باشد. در اکثر رشته‌ها، مکاتب فکری و اختلاف‌نظرها فراوان هستند. علم تلاش می‌کند با استفاده از داده‌های عینی و منطق [عامل] ذهنیت را به حداقل برساند، اما جایی که داده‌ها قطعی نباشند باید از قضاوت و مهارتمان بهره ببریم.

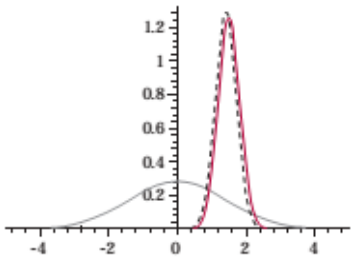
روش‌های بیزین معمولاً با این رویکرد تطبیق می‌یابند. تصویر ۱ نشان می‌دهد که قضیه‌ی بیز چگونه به طور طبیعی به هر یک از منابع اطلاعاتی بر اساس قدرتش وزن می‌دهد. در آن مثال، داده‌ها فقط قدری از توزیع پیشین آگاهی‌بخش‌تر بودند، و از همین رو توزیع پسین هم از توزیع پیشین و هم از درست‌نمایی به شدت تاثیر می‌پذیرد.

همچنین در تصویر ۲ دیدیم که اگر اطلاعات پیشین تضعیف شوند، آن‌وقت قضیه‌ی بیز چگونه تمام وزن موثر را به درست‌نمایی اختصاص می‌دهد. معمولاً به لحاظ داشتن داده‌های قوی موقعیت بهتری داریم. آن وقت است که شرایط بیشتر شبیه نمودار سه تایی موجود در تصویر الف۳ خواهد بود.

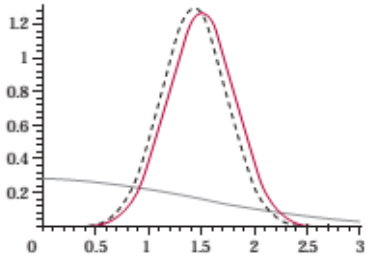
تصویر الف۳

نمودار سه تایی با داده‌هایی آگاهی‌بخش‌تر

(۱) در مقیاس مشابهی با تصویر الف۲



(۲) با مقیاس تغییریافته به منظور نشان‌دادن تفاوت بین درست‌نمایی و توزیع پسین



در این مورد، داده‌ها بر پایه‌ی ده برابر اطلاعات موجود در تصویر۱ واقع شده‌اند. منحنی درست‌نمایی قرمز در مقایسه با چگالی پیشین خاکستری بسیار باریک‌تر است. توزیع پیشین اطلاعات بسیار اندکی را در تلفیق وارد می‌کند، و چگالی توزیع پسین (منحنی نقطه‌چین سیاه) تقریباً مانند درست‌نمایی است. هنگامی که داده‌ها به قدر کافی قوی باشند، نسبت به همه‌ی اطلاعات پیشین ذهنی وزن بیشتری خواهند گرفت. گرچه ممکن است کارشناسان مختلف این زمینه دانش و عقاید پیشین متفاوتی را ارائه دهند، اما با این‌حال داده‌ها از توزیع‌های پیشین ارائه شده توسط این افراد وزن بیشتری می‌گیرند، و توزیع‌های پسین تقریباً یکسانی پدید می‌آیند. این مدلی عالی برای علم است.

اگر داده‌ها در این مورد قطعی نباشند چه؟ آن‌وقت کارشناسان مختلف این حوزه به توزیع‌های پسین بسیار متفاوتی خواهند رسید. در این مورد اتفاق نظری وجود ندارد، اما به طور کلی، تفاوت در نتیجه‌ها توسط داده‌ها کاهش خواهد یافت. در واقع این یکی از نقاط قوت رویکرد بیزین است که با درنظر گرفتن نتایجِ استفاده از توزیع‌های پیشین متفاوت مشخص می‌شود که آیا داده‌ها برای چربیدن بر آن تفاوت‌ها کافی هستند یا خیر. تحلیل بر اساس داده‌هایی که قدرت کافی را ندارند گمراه‌کننده می‌شود.

[توزیع] پیشینِ چه کسی؟

درنظر بگیرید که حامی مالیِ شماری از فن‌آوری‌های پزشکی (مثلاً، یک شرکت داروسازی یا

سازنده‌ی دستگاه) مایل به تحلیل بیزین در تایید موردی مبنی بر هزینه اثربخشی آن فن‌آوری است. چه چیزهایی به عنوان اطلاعات پیشین قابل قبول هستند؟ تنها یک راه برای پرداختن به این سوال وجود دارد و آن این است که بپرسیم از [توزیع] پیشین چه کسی باید استفاده شود؟

همان طور که قبلاً توضیح داده شد ممکن است این مسئله‌ای بی‌اهمیت باشد. اگر داده‌ها به قدر کافی قوی باشند به نحوی که بتوان تفاوت‌های موجود در عقاید و دانش پیشین تمام گروه‌های ذینفع را درنظر نگرفت، در آن‌صورت می‌توان به اتفاق نظر رسید. با این‌حال، اگر یک شخص اطلاعات یا عقاید پیشین قدرتمندی در اختیار داشته باشد این استدلال غلط از آب در می‌آید. اگر فقط یک فرد بسیار خودرای وجود داشته باشد توافقی به دست نمی‌آید. ما با چنین فردی که دیدگاه‌هایش در مورد برخی مسائل آن‌قدر مغرضانه است که به هیچ واقعیت یا استدلال مخالفی گوش نمی‌کند آشنا هستیم؛ قضیه‌ی بیز هم راجع به این افراد توضیح می‌دهد!

این مسئله برخی از ابعاد ذهنیت را آشکار می‌کند. در عین حال که باید بپذیریم که افراد مختلف ممکن است دانش پیش‌زمینه‌ی متفاوتی داشته یا اطلاعات یکسانی را به شکلی متفاوت تفسیر کنند، جایی هم برای غرض‌ورزی یا تفسیرهای اشتباه عمدی درباره‌ی حقایق موجود در اقتصاد سلامت (یا هر جای دیگر) وجود ندارد. یکی از ابعاد مهم تحلیل بیزین این است که توزیع پیشین علناً ارائه می‌شود، و اگر بر اساس استفاده‌ی منطقی از اطلاعات و تجربیات نباشد در آن‌صورت تحلیل حاصل کسی را متقاعد نخواهد کرد. این مزیت مهم شماره‌ی ۵ تحلیل بیزین است؛ «قضاوت‌های علنی‌تر». همه چیز باید آشکار باشد و پنهان‌کاری در بین نباشد.

به پرسش مطرح شده برمی‌گردیم؛ حامی مالی باید از [توزیع] پیشین چه کسی استفاده کند؟ چنین احتمال می‌رود که توزیع پیشین خود این فرد غیر قابل قبول است. در واقع، نظرات آن‌ها بر مبنای تجربیات مهم شرکت در توسعه و آزمودن محصولات قابل دفاع خواهد بود، اما خطر انتخاب گزینشی اطلاعات تا هنگام صورت نگرفتن افشای کامل وجود دارد. برای نقل به قول از جعبه‌ی «شواهد» موجود در متن اصلی، حامیِ مالی باید بتواند نشان دهد که [توزیع] پیشین او نه فقط بیانی از شواهد، بلکه بیانی از کل شواهد است.

توزیع پیشینِ بررسی‌شده و قابل دفاعی که متعلق به فردی کارشناس در این زمینه باشد بهترین گزینه برای انتخاب توزیع پیشین خواهد بود. موسسه‌ای که هزینه اثربخشی محصولی در آن بررسی می‌شود، شاید مایل به دانستن این باشد که در صورت درنظر گرفتن دیدگاه‌های سایر کارشناسان آیا استنباط‌ها تغییر می‌کنند یا خیر. به‌طور ایده آل، اجماع نظر حرفه‌ای‌ها گزینه‌ی خوبی است.

مثال‌ها:

این نظرات به شکل دو مثال زیر نشان داده شده‌اند که مختصراً در متن اصلی به آن‌ها پرداخته شده است. تحلیل زیرمجموعه‌ها مسئله‌ای بسیار پردردسر است. دوباره مطرح کردن داده‌ها به منظور

یافتن زیرگروه‌هایی از بیماران که پاسخ‌هایشان متفاوت است ریسک بالایی دارد. برای مثال، فرض بگیرید که حاصل پژوهش روی هزینه‌ها این جدول از هزینه‌های میانگین بوده است. ردیف‌های افقی نوع درمان هستند و ردیف‌های عمودی براساس اولین حرف نام خانوادگی بیماران تنظیم شده‌اند.

جدول ۲

هزینه‌های میانگین در یک کارآزمایی فرضی

	اولین حرف نام خانوادگی (A-D)	اولین حرف نام خانوادگی (E-Z)
درمان ۱	۸۰۰ دلار	۸۰۰ دلار
درمان ۲	۴۵۰ دلار	۸۵۰ دلار

به نظر می‌رسد که درمان ۲ برای بیمارانی که نامشان با حروف A تا D آغاز می‌شود ارزان‌تر است. بسیار بعید است که دلیلی منطقی برای این اثر زیرگروه وجود داشته باشد. برای جلوگیری از خطر اعلام شدن اثرهای کاذب زیرگروه‌ها، راهنماهای استاندارد کارآزمایی‌های بالینی ایجاب می‌کنند که تحلیل زیرگروه‌های احتمالی باید پیش از آغاز کارآزمایی مشخص شده باشد، و برای اثرهای پیشنهادی زیرگروه‌ها نیز باید مکانیسمی منطقی وجود داشته باشد.

از دیدگاه بیزین، نبود یک مکانیسم منطقی فقط اطلاعات پیشین را تشکیل می‌دهد؛ و اثر زیرگروه احتمال پیشین بسیار اندکی خواهد داشت. ادغام اطلاعات پیشین با داده‌ها - صرف‌نظر از این‌که داده‌ها چقدر متقاعدکننده به نظر می‌رسیدند - به احتمال پسین اندکی منجر می‌شود. اطلاعات پیشین آن‌قدر قدرت دارند که داده‌ها را تحت‌الشعاع قرار دهند. از این‌رو، راهنمای استاندارد به میزان برابری در مورد تحلیل‌های بیزین اعمال می‌شود؛ تحلیل‌های زیرگروه‌ها باید از پیش تعیین شده باشند تا بتوان اطلاعات پیشین راجع به منطقی‌بودن آن‌ها را تعیین کرد.

به این ترتیب، روش‌های بیزین داده‌ها را به طور خودکار تعدیل می‌کنند و مانع از آن می‌شوند که اثرهای غیرمنطقی برخاسته از احتمال در داده‌ها را صحیح تلقی کنیم. اطلاعات پیشین بسیار مهم هستند. روش‌های موجود در تحلیل فراوانی‌گرایانه‌ی زیرگروه‌ها اطلاعات پیشین را درنظر می‌گیرند (و از همین‌رو آن‌ها نیز ناخودآگاه بیزین هستند). با این‌حال، تحلیل فراوانی‌گرا فرضیه‌های مربوط به

زیرگروه‌ها را به آن‌هایی که منطقاً پیشین هستند و [منطقاً پیشین] نیستند تقسیم می‌کند، در حالی‌که تحلیل بیزین احتمال پیشینی را تعیین می‌کند که ممکن است هر مقداری بین ۰ تا ۱ داشته باشد و از همین رو درجه‌بندی دقیق‌تری را ممکن می‌کند.

بستری‌شدن

تصور کنید، پس از آن‌که شواهد اثربخشی یک داروی جدید مربوط به درمانی استاندارد خوب ارزیابی شدند، تصمیم در مورد هزینه اثربخش تر بودن این دارو به این بستگی داشته باشد که نرخ بستر۳ی‌شدن را کاهش دهد. در این‌جا داده‌ها حاصل از یک کارآزمایی هستند که در آن، تعداد روزهای ماندن در بیمارستان برای دو گروه صد نفره از بیماران ثبت شده بود. مجموع ۲۵ روز (میانگین ۰٫۲۵ به ازای هر بیمار) در گروهی ثبت شد که تحت درمان استاندارد قرار گرفته بودند و فقط ۵ روز بستری‌شدن (میانگین ۰٫۰۵ به ازای هر بیمار) در گروه دریافت‌کننده‌ی داروی جدید ثبت شد. در اصطلاح فراوانی‌گرا، تفاوت در سطح ۵٪ معنادار است (یک‌طرفه). (در این‌جا جزئیات فنی محاسبات را ارائه نمی‌دهیم، اما کل این تحلیل، و تحلیل بعدی، را می‌توان با استفاده از این اطلاعات افزوده بازسازی کرد: واریانس نمونه در گروه درمان استاندارد برابر با ۱٫۲ و در گروه درمان جدید برابر با ۰٫۲۴۸ است). شرکت داروسازی نیز می‌تواند، به بیان مرسوم فراوانی‌گرا، اثری را به بستری‌شدن مربوط بداند، میانگین تعداد روزهای بستری‌شدن را تحت درمان استاندارد به ازای هر بیمار برابر ۰٫۲۵ و تحت درمان با داروی جدید برابر با ۰٫۰۵ برآورد کند، و احتمالاً چنین نتیجه بگیرد که داروی جدید هزینه اثربخش‌تر از درمان استاندارد است.

با این وجود، این کارآزمایی کوچکی بوده است و داده‌ها به هیچ وجه قطعی نیستند. اگر شواهد دیگری در دست بود در نظر گرفتن آن‌ها در تحلیل معقول به نظر می‌رسید. تصور کنید که کارآزمایی بسیار بزرگتری (در شرایط مشابه) از همین دارو این نتیجه را حاصل کند که متوسط روزهای بستری‌شدن به ازای هر بیمار ۰٫۲۱ است، و این‌که خطای استاندارد این برآورد فقط در حدود ۰٫۰۳ باشد. این اطلاعات افزوده حاکی از خوشبینانه بودن نرخ مشاهده شده‌ی برابر با ۰٫۰۵ برای داروی جدید هستند، و در مورد بزرگی تفاوت واقعی بین درمان به وسیله‌ی این دارو و درمان استاندارد تردید ایجاد می‌کنند. با این‌حال، چنین تفسیری از شواهد قطعاً عجولانه است. کسی ادعا نمی‌کند که دو دارو ضرورتاً باید به نرخ‌های بستری‌شدن مشابهی منجر شوند، اما تصور این‌که این نرخ‌ها تفاوت آشکاری ندارند هم منطقی به نظر می‌رسد. از آن جایی که نمی‌توان آن‌ها را کاملاً مشابه با داده‌های کارآزمایی داروی جدید تلقی کرد، باید این کارآزمایی بزرگتر را بخشی از داده‌ها فرض کنیم و آن‌را با داده‌های جدید ادغام کنیم. روشی برای استفاده از شواهد مربوط به داروی جدید در تحلیل فراوانی‌گرا وجود ندارد. با این‌حال هر پزشک یا ارائه‌دهنده‌ی مراقبت سلامت اگر از این شواهد خارجی مطلع باشد، تمایلی به پذیرفتن شواهد کارآزمایی جدید نخواهد داشت.

تحلیل بیزین این مسئله را با این فرض حل می‌کند که کارآزمایی قبلی اطلاعات پیشین را فراهم می‌آورد، و البته عنصر ذهنیت را دخالت می‌دهد. تصور کنید تفسیرتان از اطلاعات پیشین این است که انتظار پیشین شما از متوسط روزهای بستری‌شدن برای داروی جدید باید برابر با ۰٫۲۱ اما با انحراف استانداردی برابر با ۰٫۰۸ باشد تا این واقعیت را منعکس کند که دو دارو یکسان نیستند. حالا قضیه‌ی بیز برای شما برآورد پسین ۰٫۰۹۵ را برای متوسط روزهای بستری‌شدن تحت درمان با داروی جدید حاصل می‌آورد. همچنان احتمالی قوی و منطقی (۹۰٪) وجود دارد که داروی جدید این نرخ بستری‌شدن را کاهش دهد، اما تفاوت برآورد شده ممکن است دیگر آن‌قدر بزرگ نباشد تا تضمین مشابهی مبنی بر هزینه اثربخش بودن داروی جدید در مقایسه با درمان استاندارد ارائه دهد.

در این تحلیل، ذهنیت در قضاوت در مورد این مسئله نمود می‌یابد که تفاوت نرخ‌های بستری‌شدن برای این دارو باید چقدر باشد. پزشک یا تصمیم‌گیرنده‌ی دیگری ممکن است اطلاعات پیشین را به شیوه‌ای متفاوت تفسیر کند و توزیع پیشین متفاوتی را به‌کار بگیرد. خصوصاً ممکن است آن‌ها انحراف استانداردِ پیشینِ متفاوتی داشته باشند. در واقع، این پاسخی نسبتاً قوی به تغییرات منطقی در توزیع پیشین است. بر این مبنا ممکن است این‌طور نتیجه‌گیری کنیم که تفسیر «عینی» از داده‌های ادغام شده این است که داروی جدید به احتمال زیادی (احتمالاً در حدود ۹۰٪ اما نه ۹۵٪) میانگین روزهای بستری‌شدن را کاهش می‌دهد اما این میانگین نزدیکتر به ۰٫۱ است تا به ۰٫۰۵.

این مثال نشان می‌دهد که تحلیل بیزین چگونه با بهره‌گیری از اطلاعات خالص پیشین و با درنظر گرفتن تفسیرهای منطقی از آن اطلاعات به نتیجه‌ای منطقی و علمی می‌رسد. پاسخ حاصل با تحلیل فراوانی‌گرا - که راهی برای بهره‌گیری از اطلاعات افزوده ندارد - تفاوت‌هایی اساسی دارد. علاوه بر این، پاسخ بیزین به این جهت منطقی است که با دانستن نتیجه‌ی کارآزمایی بعدی، به آن واکنش طبیعی و شهودی که هر کسی به تحلیل فراوانی‌گرا دارد رسمیت می‌دهد.

در این مثال و مثال‌های پیشین، به قدرت رویکرد بیزین در تعدیل تفسیرهای بسیار خوشبینانه نسبت به مقادیر P اشاره کردیم، اما وضعیت معکوس هم به همین میزان متداول و مهم است. مدیران شرکت‌های داروسازی و آماردانان زیستی با چنین مواردی کاملاً آشنا هستند که مرحله‌ی سوم کارآزمایی یک دارو در ایجاد اثری معنادار موفق نبوده است، در عین حال که شواهد زیادی (از داروهای مرتبط، از مرحله‌ی دوم کارآزمایی که به موارد حاد منحصر بوده است، و غیره) حاکی از اثربخشی حقیقی آن دارو هستند. یک تحلیل بیزین که به طرز مطلوبی در حال اجرا است امکان الحاق این شواهد افزوده را برای اثبات اثربخشی واقعی دارو فراهم می‌آورد. هر دو وضعیت برای توسعه‌دهندگان و کاربران فن‌آوری‌های مراقبت‌های سلامت بسیار مهم هستند؛ اولین مورد از اشتباهات پرهزینه به سبب خوشبینی بسیار زیاد جلوگیری می‌کند، و دومی این امکان را فراهم می‌آورد که محصولات سودمند وارد بازار شوند، چرا در غیر اینصورت یا باید کنار گذاشته شوند یا ورودشان به علت لزوم آزمایش‌های بیشتر به تاخیر بیفتد.

مشخص‌سازی پیشین - جزئیات

زیربخش‌های بعدی جزئیاتی راجع به سه جنبه از مشخص‌سازی پیشین ارائه می‌دهند؛ برون کشی توزیع‌های پیشین کارشناسان، توزیع‌های پیشین مزدوج conjugate priors و ایجاد توزیع‌های پیشین ساختاری.

برون‌کشی - Elicitation

روندبرون‌کشی توزیع پیشین توسط فرد کارشناس را بدون ارجاع به ماهیت واقعی اطلاعات بنیادین پیشین درنظر خواهیم گرفت. البته، در عمل، متخصص بنای تحلیلش را روی آن اطلاعات می‌گذارد، اما در این زیربخش به جزئیات اطلاعات بنیادین نمی‌پردازیم.

تصور کنید تصمیم گرفته‌ایم که توزیع پیشین را برای یک پارامتر خاص (مثل میانگین مطلوبیت حاصل از برخی درمان‌ها) مشخص کنیم، که نشان‌دهنده‌ی دانش یک کارشناس درباره‌ی آن پارامتر خواهد بود. اولین مشکلی که با آن مواجه می‌شویم این است که کارشناس به طور حتم متخصصی در حوزه‌ی احتمال یا آمار نباشد، به این معنا که برای این شخص آسان نباشد که به شکل احتمالاتیِ خاصی که توسط قضیه‌ی بیز مطالبه شده است عقایدش را بیان کند.

ممکن است کارشناس ما مایل به ارائه‌ی پارامتر باشد، اما این را چگونه تفسیر خواهیم کرد؟ آیا باید آن‌را میانگین (یا انتظار) توزیع پیشین، میانه‌ی آن توزیع، حالت آن، یا چیزدیگری فرض کنیم؟ در آمار، میانگین، میانه، و مُد ممکن است برآوردهای نقطه‌ای منطقیِ یک کمیت تلقی شوند، اما در واقع با یکدیگر متفاوت هستند. میانگین، «مقدار مورد انتظار»؛ میانه، «مقدار مرکزی»؛ و مد «محتمل‌ترین مقدار» است. ممکن است با کارشناس به جستجو در مورد این تفاوت‌های معنایی بپردازیم، اما برای کسی که آمار نیاموخته است ساده نخواهد بود که این تفاوت‌ها را درک کند و تفسیر قابل اعتمادی برای برآورد خود ارائه دهد.

می‌توانیم برخی از مشخصه‌های توزیع ارائه شده توسط کارشناس - از جمله میزان پراکنش - را از وی جویا شویم تا به میزان کلی عدم قطعیت او درباره‌ی مقدار حقیقی پارامتر پی ببریم. (به یاد بیاورید که، قدرت اطلاعات با میزان باریکیِ توزیع آن اطلاعات مشخص می‌شود).

مشکل بعدی این است که هرقدر هم بتوانیم اطلاعاتی از این دست را از کارشناس حاصل بیاوریم، آن مقدار اطلاعات برای شناختن دقیق توزیع او کافی نخواهند بود. مثلاً برای شناختن منحنی خاکستری در تصویر۱ که نماینده‌ی توزیع پیشین است، باید ارتفاع آن را در تک تک نقاط مشخص کنیم. به این ترتیب، باید راجع به بینهایت واقعیت از کارشناس جویا شد. آن‌چه این مسئله را دشوارتر می‌کند این است که هر چه از کارشناس راجع به جزئیات بیشتری جویا شویم، پاسخ به این سوالات برای او سخت‌تر خواهد شد. در عمل، برون‌کشی تعدادی عبارات ساده به شکل چیزهایی مانند میانه

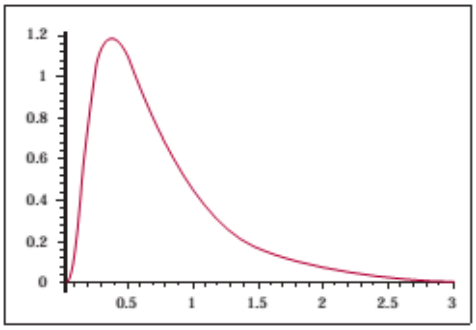
و چارک‌ها و سپس هماهنگ کردن چندین توزیع منطقی با آن اظهارات موثرترین کاری است که می‌توانیم انجام دهیم.

حتی قضاوت‌های کارشناس را نیز نمی‌توان دقیق تلقی کرد. تصور کنید که کارشناس برآوردی برابر با ۰٫۸۵ برای پارامتر دارد، و ما این مقدار را میانگین توزیع پیشین وی تفسیر می‌کنیم. حتی در صورت صحیح بودن نحوه‌ی تفسیر ما باز هم نمی‌توان این مقدار را دقیق دانست. تقریباً با قاطعیت می‌توان گفت که این کارشناس رقمی گرد ارائه کرده و نمی‌تواند ثابت کند که ۰٫۸۶ انعکاس دقیق‌تری از میانگین پیشین او نیست. خواننده هم‌اکنون باید ایراد شماره‌ی ۲ را کاملاً درک کرده باشد، این‌که «مشخص‌سازی پیشین غیر قابل اطمینان است».

با این‌حال، پژوهش‌های بسیار زیادی به چگونگی برون‌کشی دقیق و قابل اطمینانِ دانش کارشناسان معطوف شده‌اند. مشکلاتی که افراد در ارزیابی احتمالات با آن‌ها مواجه می‌شوند توسط بسیاری از پژوهشگران و به طور خاص روان‌شناسان شناختی بررسی شده‌اند، و از جمله‌ی مقالات مروری مفید در این ارتباط می‌توان به مقالات لیتنستاین و همکاران (۱۹۸۰)، می‌یر و بوکر (۱۹۸۱)، کوک (۱۹۹۱)، و کادان و ولفسن (۱۹۹۸) اشاره کرد. گرچه که روان‌شناسان روی تکالیفی تاکید داشته‌اند که افراد به طرزِ بدی به صورتِ مفهوم در می‌آورند، با این وجود اهمیت عملی این کار در این است که هم‌اکنون راجع به چگونگی اجتناب از این مشکلات بسیار می‌دانیم. این پژوهش و سایر پژوهش‌های در حال انجام در تلاش برای طرح پرسش‌هایی هستند که با احتمال بیشتری به پاسخ‌های خوب منجر می‌شوند. همچنین این پژوهش‌ها در تلاش برای جلوگیری از بروز مشکلاتی هستند که لااقل برای روان‌شناسان و آماردانان محرز شده‌اند، و علاوه بر این در جستجوی مکانیسم‌های بازخوردی برای حصول اطمینان از ارتباط خوب بین آماردانان و کارشناسان هستند.

پاسخ دوم این است که، خوشبختانه، نادقیق بودن توزیع‌های برون‌کشی شده از کارشناسان آن‌قدرها هم مهم نیست. همان‌طور که قبلاً توضیح داده شد، می‌توانیم گستره‌ای از توزیع‌های پیشین را درنظر بگیریم که با ادعاهایی که از کارشناس برون‌کشی کرده‌ایم مطابقت داشته باشند، و اگر داده‌ها به اندازه‌ی کافی قدرت داشته باشند، آن‌وقت است که تمام مشخص‌سازی‌های پیشین از توزیع پیشین الزاماً به توزیع پسین یکسانی منجر می‌شوند. جعبه ی «مثالی از برون‌کشی» این نظرات را بررسی می‌کند.

مثالی از برون‌کشی



کارشناسی پارامتر خطر نسبی (RR) را حدود ۵۰٪ برآورد می‌کند، اما در مورد مقدار واقعی آن قطعیت ندارد. او می‌گوید که این مقدار بعید است که کمتر از ۰٫۲ یا بیشتر از ۱٫۵ باشد. توزیع در جهت راست با این ادعاها مطابقت دارد، اما توزیع‌های فراوان دیگری هم ممکن است با این ادعاها مطابقت داشته باشند. سوال این است که، اگر سایر آن توزیع‌ها را در یک تحلیل بیزین با در دست داشتن داده‌هایی بیازماییم، آیا به استنباط‌های پسین بسیار متفاوتی منجر خواهند شد؟

توزیع‌های پیشین مزدوج

همان‌طور که در بحث قبلی توضیح داده شد، روش معمول برای مشخص کردن توزیع پیشین برای پارامترها عبارت از مشخص کردن (یا برون‌کشی) اندکی از مشخصه‌های توزیع مورد نظر، همچون انتظار پیشین و میزان عدم قطعیت پیشین (برای مثال، واریانس پیشین)، و سپس برگزیدن توزیعی مناسب است که با این مشخصه‌ها مطابقت داشته باشد. برای مثال، در جعبه‌ی «مثالی از برون‌کشی»، نوع خاصی از توزیع به نام توزیع گاما با دو ادعای مشخص کارشناس مطابقت یافته است.

حتی اگر مشخصه‌های واقعی محدودی تعیین شده یا از کارشناس برون‌کشی شده باشند، توزیع پیشین در عمل کاملاً مشخص خواهد بود. گرچه که انتخاب توزیع واقعی به صورت دلبخواهی انجام می‌شود، اما تمام توزیع‌های پیشینِ منطقی دیگر که با مشخصه‌ها مطابقت دارند احتمالاً بسیار مشابه هستند، و در نتیجه به استنباط‌ها یا تصمیمات مشابهی منجر می‌شوند.

بنابراین منطقی است که توزیع را بر اساس سادگی و سهولت انتخاب کنیم. از منظر ریاضیات، در بعضی از مسائل ساده‌ی آماری کلاس‌هایی از توزیع‌های پیشین با عنوان توزیع‌های پیشین مزدوج وجود دارند که بسیار مطلوب هستند. دلیل این امر در دو مشخصه خلاصه می‌شود. نخست، اگر توزیع پیشین عضوی از کلاس مزدوج مرتبط باشد، در آن صورت توزیع پسین نیز عضوی از همان کلاس

خواهد بود. دوم، توزیع‌های مزدوج آن‌قدر ساده هستند که بتوانیم استنباط‌های بسیاری را از آن‌ها و بدون توسل به روش‌های محاسباتی استخراج کنیم. هرگاه مدل آماری به نحوی باشد که یک خانواده‌ی مزدوج وجود داشته و عضوی از آن خانواده با مشخص‌سازی پیشین مطابقت داشته باشد، در آن صورت انتخاب چنین توزیعی بسیار مطلوب خواهد بود. سپس، تحلیل پسین بسیار ساده‌ای انجام می‌شود.

در واقع، در اوایل دوران آمار مدرن بیزین، در دهه‌های ۱۹۶۰ و ۱۹۷۰، تحلیل بیزین به استفاده از توزیع‌های پیشین مزدوج محدود می‌شد، زیرا ابزارهای محاسباتی برای حل مسائل پیچیده‌تر در دسترس نبودند. در حال حاضر از این توزیع‌ها بسیار کم استفاده می شود؛ آماردانان مدل‌هایی را می‌سازند که توزیع‌های پیشین مزدوجِ متناظری ندارند، و تمایل برای صورت‌بندیِ واقعی‌ترِ اطلاعات پیشین به این معنا هم هست که توزیع‌های پیشین مزدوج حتی در صورت در دسترس بودن هم دیگر انطباق‌پذیر نخواهند بود.

توزیع‌های پیشین ساختاری

توزیع‌های پیشین ساختاری معمولاً بدون آن‌که اطلاعاتی راجع به مقدار مشخص هر یک از پارامترها به‌دست دهند، اطلاعاتی درباره‌ی ارتباط بین پارامترها بیان می‌کنند. برای مثال، ممکن است توزیع پیشینی را مشخص کنیم که نشان‌دهنده‌ی ناآگاهی موثر effective ignorance درباره‌ی میانگین هزینه تحت دو درمان متفاوت باشد، اما می‌گوید که ما انتظار داریم که نسبت این میانگین‌ها بین ۰٫۲ تا ۵ باشد. میانگین هزینه تحت هر یک از درمان‌ها ممکن است هر مقداری باشد، اما هر مقداری هم که به خود بگیرد، انتظار ما این است که میانگین هزینه تحت درمان دیگر بین پنج برابر و یک پنجم این مقدار باشد.

مثال ساده‌ای در این مورد، توزیع پیشین در مثال مربوط به بستری‌شدن است که می‌توان آن‌را به دو بخش تقسیم کرد. ما اطلاعات پیشین مهمی راجع به متوسط روزهای بستری‌شدن تحت درمان با داروی مربوطه داریم، و در مورد این‌که این میانگین چه تفاوتی با میانگین متناظر بستری‌شدن به علت درمان تحت داروی جدید دارد نیز اطلاعات پیشین ساختاری داریم. در واقع، اگر داده‌های خام کارآزمایی قبلی در دسترس باشند رسماً می‌توانیم آن‌ها را به عنوان داده‌هایمان تحلیل کنیم و نتایج مربوط به کارآزمایی جدید را به‌دست بیاوریم. در این صورت، اطلاعات پیشین صرفاً ساختاری خواهند بود. چارچوب حاضر کمی شبیه به فراتحلیل‌ها است، و در واقع، فراتحلیل بیزین بر اطلاعات پیشین ساختاری استوار است.

در فراتحلیل بیزین، چندین مجموعه داده در اختیار داریم که هر کدام در شرایط اندک متفاوتی به اثربخشی درمان می‌پردازند. بنابراین، برای میانگین اثربخشی در هر کارآزمایی پارامتر جداگانه‌ای در دست داریم، اما توزیع پیشین ساختاری که نماینده‌ی انتظار پیشین است را به نحوی مشخص

می‌کنیم که این پارامترها اختلاف زیادی با هم نداشته باشند. این کار معمولاً در یک مدل سلسله مراتبی انجام می‌شود، که در آن یک پارامتر میانگین اثربخشیِ «بنیادیِ» مشترک اصل قرار داده می‌شود، و تصور می‌شود که هر یک از میانگین‌های اثربخشی کارآزمایی‌ها به طور مستقل در اطراف این پارامتر مشترک توزیع شده‌اند.

ساختار سلسله مراتبی، که یک یا چند پارامتر مشترک را ارائه می‌دهد، معمولاً برای پیوند دادن چند پارامتر مرتبط و ابراز این عقیده استفاده می‌شود که این پارامترها باید مشابه باشند و برای این ادعا به این واقعیت اشاره می‌کند که تمام این پارامترها باید مشابه با پارامتر مشترک باشند. مثال دیگر تعیین این اطلاعات پیشین ساختاری است که داده‌های هزینه که از گروه‌های مقایسه‌ی کارآزمایی ناشی می‌شوند نباید در میزان چولگی‌شان تفاوت زیادی داشته باشند. این فایده‌ی متعادل کردن اثر معدود بیمارانی را دارد که هزینه‌هایشان معمولاً بالا است.

محاسبات - جزئیات

زیربخش‌های بعدی جزئیاتی راجع به محاسبات بیزین و توانایی این محاسبات در پرداختن به مدل‌های بسیار پیچیده را ارائه می‌دهند.

پیچیدگی

منبع پیچیدگی بیش از حد در تحلیل بیزین مجدداً همان توزیع پیشین است. تصور کنید که یک مسئله‌ی ساده، کانونی، و آماری درباره‌ی برآورد میانگین توزیع نرمال (با واریانس معلوم) به درنظر گرفتن نمونه‌ای از آن توزیع مطرح شده است. در نظر فراوانی‌گرا، این مشخص‌سازی کامل مسئله است، و در واقع پاسخی یگانه برای آن وجود خواهد داشت (و در واقع میانگین نمونه به‌طور کلی به عنوان بهترین برآورد پذیرفته می‌شود). در نظر تحلیلگر بیزین، مشخص‌سازی ناقص است زیرا باید توزیع پیشین را نیز تعیین کنیم. آن‌وقت پاسخ دو منبع اطلاعاتی را تلفیق می‌کند که به توزیع پیشین بستگی خواهد داشت. با درنظر گرفتن این مسئله‌ی ساده باید بگوییم که رویکرد بیزین قادر است پاسخ‌های بسیاری را ارائه دهد.

همین مسئله موجب می‌شود که ساختن نرم‌افزاری برای تحلیل‌های بیزین بسیار دشوار باشد. خود نرم‌افزار باید پیچیده‌تر باشد. دلیل این امر آن است که نرم‌افزار باید امکان مشخص‌سازیِ توزیع پیشین را فراهم کند (که هر نوع توزیعی ممکن است باشد) و همچنین باید قادر به محاسبه‌ی استنباط‌های مورد نظر باشد؛ صرف‌نظر از این‌که توزیع پیشین ممکن است چه باشد.

از منظر ریاضی، استنباط‌های بیزین ما را قادر می‌سازند تا محصول توزیع پیشین و درست‌نمایی را یکپارچه کنیم. به این ترتیب خواهیم توانست، که برای مثال، مساحت زیر منحنی را پیدا کنیم و به

منظور رساندن مساحت زیر سطح منحنی به ۱ محصول را میزان کنیم. ادغام لازمه‌ی اعمال قضیه‌ی بیز است. آن‌وقت ادغام‌های بیشتری باید صورت بگیرد، تا برای مثال، میانگین پسین یا احتمال آن‌که یک فرضیه صحیح باشد مشخص شود. حالا حتی اگر توزیع پیشین و درست‌نمایی هر یک توابعی باشند که انتگرال‌هایشان شناخته شده است، در آن صورت هم محصول تقریباً همیشه به قدری پیچیده است که نمی‌توانیم انتگرال‌های لازم را توسط قواعد ریاضی استخراج کنیم. بنابراین انتگرال‌گیری را باید به شیوه‌ی عددی انجام داد (برای مقدمه‌ای به نظرات و روش‌های انتگرال عددی به مقاله‌ی تیستد، ۱۹۸۸، رجوع کنید). دلیل اصلی این‌که روش‌های بیزین تا دهه‌ی ۱۹۹۰ قابل پیاده‌سازی نبودند این بود که الگوریتم‌های انتگرال‌گیری موثری در دست نداشتیم. دلیل افزایش کاربردهای بیزین پس از دهه‌ی ۱۹۹۰ توسعه‌ی راه‌حل کلی و قدرتمند بود.

MCMC

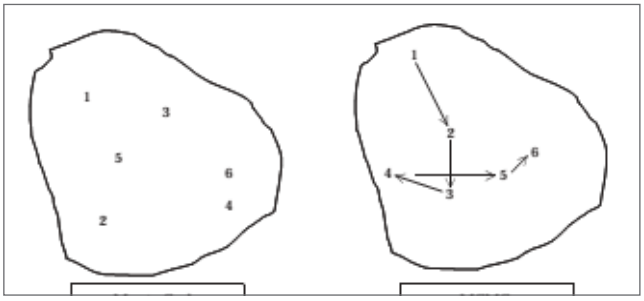
زنجیره‌ی مارکوف مونت کارلویی (MCMC) روشی است که باعث تحول محاسبات بیزین شده است. به این روش در متن اصلی پرداخته شده است: استنباط بیزین با بیرون کشیدن تصادفیِ نمونه‌ای بسیار عظیم از توزیع پسین حل می‌شود. این روش ما را قادر می‌سازد که هر استنباطی را که بخواهیم از توزیع پسین به دست بیاوریم از روی نمونه محاسبه کنیم.

در این‌جا باید تاکید کنیم که در مورد داده‌های نمونه صحبت نمی‌کنیم. (معمولاً روی میزان داده‌هایی که به‌دست می‌آوریم کنترل بسیار کمی داریم، و انتظار نداریم که نمونه‌ی ما به‌قدری عظیم باشد که بتوانیم هر آن‌چه را که می‌خواهیم به این شیوه‌ی ساده محاسبه کنیم.) در واقع موضوع صحبت ما تولید مصنوعی نمونه‌ای مشتمل بر مقدار پارامترها به‌وسیله‌ی یک روش شبیه‌سازی تصادفی است. با کمک این روش، نمونه‌ای از توزیع پسین آن پارامترها به‌دست می‌آید.

به بیان دقیق، فکر بیرون کشیدن نمونه‌ای بزرگ از توزیع پسین مونت کارلو نامیده می‌شود. شبیه‌سازی مونت کارلو در واقع شیوه‌ی محاسبه‌ی غیرمستقیم قدرتمندی از چیزهایی است که آنالیزهای مستقیم ریاضیات قادر به حل آن ها نیستند. آن‌چه موجب تمایز MCMC می‌شود روش بیرون کشیدن نمونه است. مونت کارلو را به بیان ساده می‌توان به صورت بازی پرتاب دارت تجسم کرد؛ «پرتاب» تصادفی دارت‌ها به فضایی مشتمل بر همه‌ی مقدار پارامترهای ممکن، به نحوی که هر دارت مستقل از دیگر دارت‌ها است. این رویکرد برای تحلیل بیزین عملی نیست، زیرا در مدلی که پارامترهای بسیاری را در بر دارد، بسیار مشکل است که بتوان الگوریتم کارآمدی را برای «پرتاب» تصادفی دارت‌ها با توجه به توزیع پسین مورد نظر ایجاد کرد. در روش MCMC، نقطه‌ای در اطراف فضای متشکل از مقدار پارامترهای ممکن سرگردان است. تصویر الف۴ را ببینید.

تصویر الف۴

نمونه‌گیری از توزیع پسین



به بیان فنی، یک مدل احتمالاتی به نام زنجیره‌ی مارکوف این نقطه‌ی سرگردان را تولید می‌کند. گام‌های متوالی در این زنجیره موجب تولید نمونه می‌شوند. به‌نظر می‌رسد که ساخت زنجیره‌ی مارکوف به نحوی که بتوان نمونه را از هر یک از توزیع‌های پسین مورد نظر بیرون کشید بسیار ساده باشد.

قدرت محاسبات بیزین از وجود راه‌حل‌های MCMC برای محاسبه‌ی استنباط‌های پسین از مدل‌های آماری پیچیده با پارامترهای فراوان به‌دست می‌آید. با این‌حال، در بطن این بیان ساده مسائل دیگری نیز وجود دارد.

مشخص است که این مقادیر پیاپی به هم پیوسته هستند و مانند نقاط ساده‌ی مونت کارلو از یکدیگر مستقل نیستند. اگر وابستگی زیاد باشد، در آن صورت نقاط در اطراف فضای مشتمل بر مقدار پارامترهای ممکن بسیار آهسته حرکت می‌کنند. بنابراین، نمونه‌ی بسیار بزرگی لازم است تا فضا را به نحوی پوشش دهد که نماینده‌ی شایسته‌ی توزیع پسین باشد. بنابراین، کارآمدی روش‌های MCMC وابستگی بسیاری به زنجیره‌ی مارکوفی دارد که با سرعت در اطراف فضا حرکت کند (به این ویژگی اصطلاحاً «خوب مخلوط کردن» می‌گویند). متاسفانه، گرچه که طراحی الگوریتم MCMC به نحوی که اصولاً خوب کار کند در کل آسان است، اما معمولاً مهارت و تجربه‌ی بسیار زیادی را می‌طلبد که موردی که خوب مخلوط کند ساخته شود.

مسئله‌ی دیگر این است که الگوریتم‌ها به یک دوره‌ی «گرم شدن – burn-in » نیاز دارند، تا به این ترتیب، نقطه‌ای که تصادفی حرکت می‌کند در بخشی از فضای پارامتری مستقر شود که توسط توزیع پسین پشتیبانی می‌شود. تشخیص این‌که زنجیره‌ی مارکوف چه هنگام به قدر کافی اجرا شده است اصلاً ساده نیست.

در طیف بسیار وسیعی از مدل‌های نسبتاً پیچیده (همانند آن‌هایی که می‌توان با موفقیت آن‌ها را در نرم‌افزار WinBUGS که در متن اصلی شرح داده شد پیاده‌سازی کرد) این مشکلات حداقل هستند. با این‌حال، برای مشکلات پیچیده و بزرگ همچنان باید به MCMC رجوع کرد. با این‌حال، آشنایی با این روش که استفاده‌ی زیادی در آمار بیزین دارد مداوماً بیشتر می‌شود، و منابع مطالعاتی در رابطه با الگوریتم‌های MCMC نیز مداوماً غنی‌تر می‌شوند.

حل مسائل دشوار

خط موازی فراوانی‌گرایانه‌ای نیز با مشکلات محاسباتیِ روش‌های بیزین وجود دارد، و آن این است که، به استثنای آن‌چه در مدل‌های بسیار ساده امکان‌پذیر است، حاصل آوردن آزمون‌های دقیق با رویکرد فراوانی‌گرایانه، فواصل اطمینان و برآوردگرهای نااریب بسیار دشوار است. معمولاً به این نکته توجه نمی‌شود که اکثر روش‌های فراوانی‌گرا در استفاده‌های غیر حرفه‌ای صرفاً تقریب ارائه می‌دهند، و این شامل همه‌ی روش‌های مبتنی بر مدل‌های خطی تعمیم‌یافته، آزمون‌های نسبت درست‌نمایی تعمیم‌یافته، bootstrapping و از این دست می‌شود. استنباط در مدل خطی نرمال استاندارد تنها استثنا محسوب می‌شود. حتی در این‌جا هم مقایسه‌ی مدل‌های non-nested با استفاده از روش های فراوانی‌گرا آسان نیست.

همان‌طور که در متن اصلی مطرح شد، روش‌های محاسباتی همچون MCMC، استنباط‌های دقیق بیزین را حتی در مدل‌هایی بسیار پیچیده امکان‌پذیر می‌کنند. همان‌طور که آماردانان در تلاش برای پرداختن به ساختارهای بزرگتر و پیچیده‌تر داده‌ها هستند (تحلیل داده‌های ریزآرایه، داده‌کاوی و غیره)، مزیت شماره‌ی سوم روش‌های بیزین - «توانایی حل مسائل پیچیده‌تر» - بیشتر دارای اهمیت می‌شود.