بنام خدا

خودآموز آمار و تحلیل کمّی در آزمون‌های

CFA و FRM

همراه با حل تشریحی بیش از ۳۰۰ مسأله

(قابل استفاده برای تمامی رشته‌های مدیریت، حسابداری و اقتصاد)

جلد نخست

تألیف: دکتر زهره طغرایی

انتشارات پندار پارس

سرشناسه	: طغرایی، زهره، ۱۳۵۳-
عنوان و نام پدیدآور	: خودآموز آمار و تحلیل کمی در آزمون‌های FRM و CFA همراه با حل تشریحی بیش از ۳۰۰ مساله.../ تألیف زهره طغرایی .
مشخصات نشر	: تهران : پندار پارس ، ۱۴۰۴ -
مشخصات ظاهری	: ج.
شابک	: ج.۱ : 0-47-7785-622-۹۷۸
وضعیت فهرست نویسی	: فیپا
موضوع	: سرمایه‌گذاری -- تجزیه و تحلیل -- آزمون‌ها-Investment analysis -- Examinations
	مدیریت سبد سرمایه‌گذاری -- آزمون‌ها-Portfolio management-- Examinations
رده بندی کنگره	: HG۴۵۲۹
رده بندی دیویی	: ۶/۳۳۲
شماره کتابشناسی ملی	: ۱۰۱۲۰۶۸۵
اطلاعات رکورد کتابشناسی	: فیپا

انتشارات پندارپارس

دفتر فروش: انقلاب، ابتدای کارگرجنوبی، کوی رشتچی، شماره ۱۴، واحد ۱۶
www.pendarepars.com

تلفن: ۶۶۵۷۲۳۳۵ – ۶۶۹۲۶۵۷۸ همراه: ۰۹۱۲۲۴۵۲۳۴۸

نام کتاب	**: خودآموز آمار و تحلیل کمّی در آزمون‌های FRM و CFA همراه با حل تشریحی بیش از ۳۰۰ مسأله (قابل استفاده برای تمامی رشته‌های مدیریت، حسابداری و اقتصاد) – جلد ۱**
ناشر	**: انتشارات پندار پارس**
تالیف	**: دکتر زهره طغرایی**
چاپ نخست	**: تیر ماه ۱۴۰۴**
شمارگان	**: ۲۰۰ نسخه**
چاپ، صحافی	**: وحید**
قیمت	**: ۶۰۰.۰۰۰ تومان شابک : ۰-۴۷-۷۷۸۵-۶۲۲-۹۷۸**

تقدیم به پدر و مادر عزیزم که
آموختن را از آنان آموختم

پیش‌گفتار

در دنیای امروز، دانش مالی به یکی از ارکان اساسی تصمیم‌گیری‌های حرفه‌ای تبدیل شده است. آزمون‌های معتبر بین‌المللی همچون FRM[1] و CFA[2] به متخصصان مالی این امکان را می‌دهند که مهارت‌های خود را به سطحی بالاتر برده و جایگاهی برجسته در دنیای مالی کسب نمایند. یکی از کلیدی‌ترین مباحث این آزمون‌ها، آمار و تحلیل کمّی است؛ دانشی که نه تنها در امتحانات، بلکه در مسیر حرفه‌ای هر تحلیلگر مالی نقشی حیاتی ایفا می‌کند.

این کتاب به عنوان یک راهنمای خودآموز جامع طراحی شده است تا شرکت‌کنندگان برای شرکت در دو آزمون مذکور بتوانند با زبانی ساده و مثال‌های کاربردی، مفاهیم پیچیده آمار و تحلیل کمّی را درک کنند. هدف نویسنده این بوده که پیچیدگی‌های ریاضی را با توضیحات روشن و قابل فهم کاهش داده و ابزارهای تحلیل را به شیوه‌ای عملی ارائه نماید.

در این کتاب با تحلیل و حل تشریحی مسائل و تست‌های دو مجموعه‌ی SchweserNotes و Quantitative Methods CFA (**سطح ۱**) تلاش شده است تا مفاهیم به صورت روشن و قابل فهم ارائه شوند. هدف اصلی این اثر، ارتقاء درک خوانندگان و تسهیل مسیر یادگیری آنهاست.

با این وجود، همه دانشجویان رشته‌های مدیریت، حسابداری و اقتصاد و علاقمندان به یادگیری آمار و مدل‌سازی نیز می‌توانند از محتوای این کتاب بهره ببرند.

در جلد نخست، به مباحث پایه‌ای و کلیدی که لازمه درک مدل‌های احتمالی مانند رگرسیون است پرداخته می‌شود. جلد دوم به مباحث پیشرفته‌ای چون روش‌های شبیه‌سازی مونت کارلو و بوت استرپ، سری زمانی، روش‌های یادگیری ماشین و ملاک‌های مهم ارزیابی این روش‌ها و ... اختصاص خواهد داشت.

در این کتاب، سفری را آغاز می‌کنیم که مقصد آن آموختنِ روش‌های مختلف تحلیل داده و در هسته مرکزی آن، شناسایی روابط بین پدیده‌ها و وقایع است. از منظر علم داده، می‌توان این روش‌ها را به دو گروه روش‌های احتمالی و غیر احتمالی تقسیم نمود که تمرکز اصلی این کتاب بر روش‌های احتمالی است؛ هرچند در جلد دوم کتاب به برخی از روش‌های یادگیری ماشین که غیر احتمالی هستند نیز پرداخته خواهد شد.

به منظور رسیدن به مقصد مذکور، لازم است گام به گام، ابزارها و وسایل مورد نیاز این سفر را تهیه و جمع‌آوری نماییم. این ابزارها، شامل مفاهیم پایه‌ای و مهم تحلیل داده شامل احتمال، متغیرتصادفی و انواع آن، توزیع احتمال متغیر تصادفی و نهایتاً استفاده از این توزیع‌های احتمال به منظور پاسخ به سؤالات احتمالی و شناسایی روابط بین متغیرهای تصادفی است. در هر فصل، به هر یک از این ابزارها پرداخته می‌شود. این سفر را مانند قصّه‌ای در زمان آغاز می‌کنیم و در هر گام که پیش می‌رویم مشکلات مربوط به ابزارهای معرفی شده و راه حل آنها در هر زمان را

[1] Financial Risk Management
[2] Chartered Financial Analyst

مطرح می‌کنیم. خواننده توجه داشته باشد در هر گام که پیش می‌رویم، مقصد اصلی را به خود یادآوری نماید تا بتواند همزمان و به درستی در مسیر گام بردارد.

نویسنده با تکیه بر سال‌ها تجربه در تدریس و تحقیق در حوزه تحلیل داده در رشته‌های مختلف علمی، سعی بر آن نموده است که به شیوه‌ای ساده و کاربردی مباحث پایه‌ای در آمار را تشریح نماید. امیدواریم این کتاب بتواند راهگشای شما در مسیر موفقیت باشد و به شما کمک کند که با "اعتمادبه‌نفس و دانش کافی" آزمون‌های پیش رو را پشت سر بگذارید. موفقیت شما، افتخار ماست.

خواننده گرامی، دریافت نظرها، پیشنهادها، و انتقادهای شما در بهبود کیفیت این کتاب مهم و مؤثر است. با نشانی ایمیل زیر می‌توانید با نویسنده مکاتبه فرمایید:

zohreh_toghrayee@yahoo.com

https://www.linkedin.com/in/zohreh-toghrayee-3339-a668/

زهره طغرایی

بهار ۱۴۰۴

فهرست

فصل نخست؛ مقدمه‌ای بر احتمال ۱۱

مقدمه ۱۲

مفاهیم مرتبط با احتمال ۱۶

احتمالات شرطی و استقلال ۲۰

پیشامدهای مستقل شرطی ۳۱

دو پیشامد ناسازگار ۳۲

مسائل حل شده ۳۲

فصل دوم؛ متغیرهای تصادفی ۴۱

متغیرهای تصادفی ۴۲

متغیرهای تصادفی کمّی ۴۶

متغیرهای کمّی گسسته ۴۶

متغیرهای کمّی پیوسته ۴۷

متغیرهای تصادفی کیفی ۴۸

توزیع های احتمال ۵۰

امید ریاضی ۶۱

گشتاورهای توزیع‌های آماری ۶۴

ترکیب خطی از یک متغیر تصادفی ۷۱

مسائل حل شده ۷۳

فصل سوم؛ توزیع‌های آماری متغیرهای تصادفی گسسته ۷۵

مقدمه ۷۶

توزیع دوجمله‌ای ۷۸

میانگین و واریانس توزیع دو جمله‌ای ۸۳

توزیع پواسن ۸۵

کاربرد توزیع پواسن به‌جای توزیع دوجمله‌ای ۸۷

میانگین و واریانس توزیع پواسن ۸۷

توزیع یکنواخت گسسته ۸۸

مسائل حل شده ۸۹

فصل چهارم؛ توزیع‌های آماری متغیرهای تصادفی پیوسته ۹۱

توزیع یکنواخت پیوسته ۹۲

توزیع نرمال ۹۳

مسائل حل شده ۱۰۷

توزیع لُگ نرمال ۱۱۸
توزیع تی-استودنت ۱۲۰
توزیع کای اسکوئر ۱۲۱
توزیع F ۱۲۲
توزیع نمایی ۱۲۴
توزیع بتا ۱۲۵
توزیع‌های آمیخته ۱۲۶
مسائل حل شده ۱۲۹
فصل پنجم؛ تحلیل‌های توصیفی دو متغیره ۱۳۳
مقدمه ۱۳۴
توزیع احتمال توأم دو متغیره ۱۳۵
تابعی از دو متغیر تصادفی ۱۳۸
کواریانس و ضریب همبستگی بین دو متغیر تصادفی ۱۴۰
ضریب همبستگی ۱۴۲
ضریب هم‌چولگی و هم‌کشیدگی ۱۴۸
ترکیب خطی ساده از یک متغیر تصادفی و ویژگی‌های آماری آن ۱۵۱
واریانس جمع موزون از دو متغیر تصادفی ۱۵۲
مقدار مورد انتظار شرطی ۱۵۵
متغیرهای مستقل با توزیع احتمال یکسان ۱۵۷
مسائل حل شده ۱۵۹
فصل ششم؛ گشتاورهای نمونه ۱۷۱
مقدمه ۱۷۲
برآورد نقطه‌ای ۱۷۳
برآورد میانگین ۱۷۷
برآورد واریانس و انحراف معیار ۱۸۲
بهترین برآوردگرهای خطی نا اریب ۱۸۸
قانون اعداد بزرگ ۱۹۰
قضیه حدّ مرکزی ۱۹۰
برآورد سایر ویژگی‌های توزیع احتمال جامعه ۱۹۱
برآورد ویژگی‌های توزیع‌های چند متغیره ۱۹۵
نقاط پرت ۱۹۷
مسائل حل شده ۲۰۲

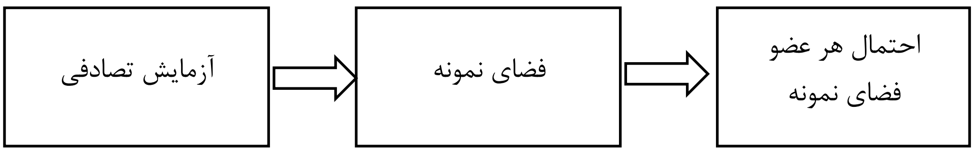
فصل هفتم؛ آزمون فرض ۲۲۷
مقدمه ۲۲۸
چرایی آزمون فرض ۲۲۸
تصمیم‌گیری درباره آزمون فرض دوطرفه ۲۳۲
تصمیم‌گیری درباره آزمون فرض یک‌طرفه رو به بالا ۲۳۶
تصمیم‌گیری درباره آزمون فرض یک‌طرفه رو به پایین ۲۳۷
خطای نوع اول و دوم ۲۴۲
مقدار احتمال یا P-value در یک آزمون فرض ۲۴۵
فاصله اطمینان و رابطه آن با آزمون فرض ۲۴۷
آزمون فرض درباره مقایسه میانگین دو جامعه ۲۵۱
آزمون فرض‌های چندگانه ۲۵۲
آزمون فرض درباره واریانس ۲۵۳
مسائل حل شده ۲۵۶
فصل هشتم؛ رگرسیون ۲۷۷
مقدمه ۲۷۸
سفر از یک متغیر تصادفی به دو متغیر تصادفی ۲۷۹
برآوردِ پارامترهای مدل رگرسیون خطی ساده ۲۸۵
ویژگی برآوردگرهای پارامترهای رگرسیون خطی ساده ۲۹۰
فرض‌های مدل رگرسیون خطی ساده و نحوه ارزیابی آنها ۲۹۱
آزمون فرض ۲۹۸
ضریب تعیین ۳۰۶
پیش‌بینی‌های به‌دست آمده برای متغیر وابسته ۳۰۷
مدل رگرسیون خطی چندگانه ۳۱۲
ضریب تعیین تعدیل شده ۳۱۸
فرض همگنی واریانس ۳۱۸
سایر اشکال مدل رگرسیون خطی ۳۱۹
مدل Lin-log ۳۲۱
مدل Log-log ۳۲۲
متغیرهای مستقل نشانگر (موهومی) ۳۲۵
مسائل حل شده ۳۲۶
پیوست ۱؛ جدول‌های آماری ۳۷۷
پیوست ۲؛ واژه نامه ۳۸۳

فصل نخست

مقدمه‌ای بر احتمال

سرآغاز: پژوهشگری درباره یک پدیده مشخص، ادّعایی مطرح نموده است. این ادّعا، یا صرفاً درباره همان پدیده است یا در مورد ارتباطش با سایر پدیده‌ها می‌باشد.

گام اول:



مقدمه

موضوع احتمال در ریاضیات به موضوع اندازه‌ها و آزمایش‌های تصادفی باز می‌گردد و اگر تمایل داشته باشیم این مفهوم را به شکل ساده‌تر و عمیق‌تر بیاموزیم، بهتر است کمی به دوران پیش‌تر برگردیم و چگونگی شکل‌گیری این دو مفهوم را بررسی نماییم.

به همه ما از کودکی مفاهیم مختلفی همچون رنگ‌ها، شکل و اسم حیوانات و ... آموزش داده شده است. یکی از این مفاهیم مهم، اندازه است که با استفاده از ابزارهایی مانند متر و ترازو یاد گرفتیم برخی از اشیاء از برخی دیگر بلندتر و یا برخی سنگین‌تر هستند.

امروزه، تقریباً تمام آنچه در پیرامون خود مشاهده می‌کنیم با ابزارهای مختلفی قابل اندازه‌گیری است؛ با این تفاوت که برای برخی از این موارد به طور مستقیم ابزار اندازه‌گیری وجود نداشته و با استفاده از تعاریف و طبقه‌بندی‌های مشخصی اندازه‌گیری آنها انجام می‌شود.

در علم ریاضی به منظور ارائه تعریف دقیق و مشخص از اندازه‌گیری، مجموعه‌ای شامل سه مؤلفه به صورت زیر ارائه شد:

(ویژگی، اندازه‌ها، مجموعه مورد نظر)

این مجموعه‌ی سه مؤلفه‌ای فوق نشان می‌دهد با استفاده از حداقل یک *ویژگی* می‌توان *اندازه‌ای* به *مجموعه مورد نظر* اختصاص داد. برای مثال، در مجموعه (محیط، اندازه ضلع، مجموعه مربع‌ها) می‌توان با استفاده از اندازه ضلع (ویژگی) هر مربع (مجموعه مورد نظر)، محیط(اندازه) آن را به صورت زیر به‌دست آورد:

$$\text{محیط} = 4 \times \text{اندازه ضلع}$$

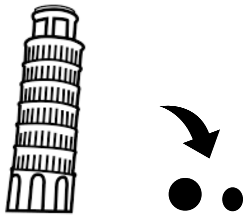
به همین ترتیب، برای تمام اشکال هندسی می‌توان اندازه‌های محیط، مساحت، حجم و ...را تعریف نمود. بنابراین، اگر مسأله‌ای شامل یک مجموعه داشته باشیم، می‌توانیم بر اساس ماهیت آن مجموعه با استفاده از حداقل یک ویژگی، اندازه‌ای به آن مجموعه اختصاص دهیم.

از سویی دیگر، تا پیش از زمان گالیله، هر نظریه بر مبنای در کنار هم قرار گرفتن دو یا چند گزاره درست ارائه می‌شد. برای مثال، دو گزاره زیر را در نظر بگیرید:

- ارسطو یک انسان است.
- هر انسانی می‌میرد.

از دو گزاره صحیح فوق می‌توان نتیجه گرفت ارسطو خواهد مرد.

پس از آزمایش معروف گالیله که دو گلوله ناهم وزن را از بالای برج کج پیزا به صورت همزمان رها و مشاهده نمود این دو گلوله بر خلاف نظریه‌های قبلی همزمان به زمین اصابت کردند، تمام آنچه زیربنای نظریه‌پردازی تا آن زمان بود فرو ریخت و مقرر شد ارائه هر نظریه‌ای مستلزم انجام آزمایش‌های مرتبط و مناسب موضوع مورد بررسی باشد. چرا که تا پیش از آن تصور می‌شد اشیاء سنگین‌تر زودتر به زمین اصابت می‌کنند و این آزمایش این نظریه را رد نموده و مبنای قانون جدید در فیزیک شد.



بنابراین، دانشمندان در حوزه‌های مختلف علم برای بررسی هر فرضیه‌ای ابتدا آزمایش مربوطه را انجام داده و نتایج را مشاهده نموده و سپس در صورت کافی بودن شواهد، فرضیه را به نظریه ارتقاء دادند. در راستای انجام آزمایش‌های مختلف، پژوهشگران دریافتند این آزمایش‌ها را می‌توانند به دو گروه آزمایش قطعی[1] و تصادفی[2] تقسیم‌بندی نمایند.

آزمایش‌های قطعی، آزمایش‌هایی هستند که پس از آنکه یکبار انجام و نتیجه آن مشاهده می‌شود در صورت تکرار آن آزمایش، نتیجه آن از پیش مشخص است. برای مثال، در آزمایش‌های مختلف نشان داده شده آب در دمای ۱۰۰ درجه فارنهایت به جوش می‌رسد. بنابراین، اگر چند ظرف مختلف آب را روی حرارت قرار دهیم، انتظار داریم همه آنها در دمای ۱۰۰ درجه بجوشند و البته این اتفاق هم رخ می‌دهد. از این‌رو، نتایج آزمایش‌های قطعی پیش از انجام آنها مشخص است.

[1] Deterministic Trials
[2] Random Trials

از سویی دیگر، آزمایش‌های تصادفی آزمایش‌هایی هستند که اگر بارها و بارها انجام شوند، نتیجه[1] هر بار، پیش از انجام آن مشخص نیست. برای مثال، یک سکه را بارها و بارها پرتاب می‌کنیم و پیش از هر بار پرتاب آن نمی‌دانیم نتیجه پرتاب شیر است یا خط. ممکن است این سؤال پیش آید که وقتی نتایج یک آزمایش تصادفی مشخص نیست، فایده انجام این آزمایش‌ها چیست؟ در پاسخ به این سؤال باید گفت درست است که نتایج را قبل از انجام آزمایش نمی‌دانیم، ولی می‌دانیم این نتایج چه چیزهایی می‌توانند باشند. برای مثال، در پرتاب یک سکه می‌دانیم نتیجه یا شیر است یا خط. در پرتاب یک تاس، نتیجه یکی از اعداد صحیح ۱ تا ۶ است.

به منظور مدوّن نمودن آزمایش‌های تصادفی، مجموعه نتایج ممکن هر آزمایش تصادفی را "فضای نمونه"[2] می‌نامند و هر یک از این اعضاء یا ترکیباتی از آنها نیز با عنوان پیشامد[3] شناخته می‌شود. کاربرد این‌گونه آزمایش‌ها در حوزه‌های مختلف علم بسیار وسیع است. جدول زیر مثال‌های متنوعی از آزمایش‌های تصادفی و فضای نمونه آنها را ارائه می‌نماید.

همان‌گونه که مشاهده می‌گردد در انجام یک آزمایش تصادفی، نتایج بر اساس یک ویژگی بررسی می‌شود؛ برای مثال، در انتخاب تصادفی یکی از سهام‌های بازار بورس، وضعیت تغییر قیمت آن سهام مورد نظر ماست. هرچند، می‌توان ویژگی‌های دیگری همچون حجم خرید و فروش و ...را نیز در نظر گرفت. بنابراین، پژوهشگر از ابتدا آگاه است چه ویژگی از این آزمایش تصادفی برای وی اهمیت دارد و بر این اساس منتظر مشاهده چه نتایجی است.

---

[1] Outcome
[2] Sample Space or Event Space
[3] Event

ردیف	آزمایش تصادفی	فضای نمونه
۱	پرتاب یک سکه سالم	{خط ،شیر}
۲	پرتاب دو سکه سالم به صورت همزمان	{(خط ، خط)، (شیر ، خط)، (خط ، شیر)، (شیر ،شیر)}
۳	پرتاب یک تاس شش وجهی سالم[1]	{1،2،3،4،5،6}
۴	انتخاب تصادفی یک پرونده تسهیلاتی در یک بانک به منظور بررسی وضعیت پرونده	{امهالی ، مشکوک الوصول ،معوق ،گذشته سررسید ،جاری}
۵	انتخاب تصادفی یکی از سهام‌های بورس و بررسی وضعیت تغییر قیمت آن نسبت به دوره قبل	{بدون تغییر قیمت ،کاهش ، افزایش}
۶	انتخاب تصادفی و همزمان دو سهام از یک سبد ۱۵ سهامی و بررسی وضعیت تغییر قیمت آنها نسبت به دوره قبل	{(افزایش قیمت دومی ، افزایش قیمت اولی) ، (کاهش قیمت دومی ،افزایش قیمت اولی) ، (افزایش قیمت دومی ، کاهش قیمت اولی) ، (کاهش قیمت دومی ، کاهش قیمت اولی)}

بر این اساس، با یک مجموعه مواجهیم که بر مبنای حداقل یک ویژگی ساخته می‌شود. آشنا به نظر می‌رسد!! یک مجموعه و حداقل یک ویژگی. پس، می‌توان برای این مجموعه و بر اساس این ویژگی(ها) یک اندازه تعریف نمود و آن اندازه چیزی نیست جز *احتمال*. به عبارتی، درست است که پیش از انجام یک آزمایش تصادفی نمی‌دانیم نتیجه آزمایش چیست، ولی می‌دانیم چه چیزهایی می‌تواند باشد و برای هر یک از موارد ممکن یا همان اعضاء فضای نمونه می‌توان اندازه‌ای بنام "احتمال" به‌دست آورد. احتمال در تعریف مردم جامعه، معنی همان شانس را می‌دهد. به عبارتی، هر یک از این نتایج ممکن چقدر شانس دارند که اتفاق بیفتند. بنابراین، مجموعه سه مؤلفه‌ای که قبلاً به آن اشاره کردیم به صورت زیر تعریف می‌شود:

(احتمال ، ویژگی ، فضای نمونه)

برای مثال، پیش از پرتاب یک سکه نمی‌دانیم نتیجه چیست، ولی می‌دانیم نتیجه شیر یا خط است و اگر سکه سالم باشد شانس مشاهده شیر و یا خط ۰.۵ می‌باشد. گفتنی است در صورت در نظر گرفتن تنها یک ویژگی با مفاهیم ساده‌تری از احتمال سرو کار داریم و محسوب نمودن بیش از یک ویژگی مستلزم این است که ارتباطات این

---

[1] The role of a fair six-sided die

ویژگی‌ها مورد بررسی قرار گیرد و در این صورت ابعاد محاسبه احتمالات را با پیچیدگی بیشتری مواجه می‌سازد. این موضوع، در سه حوزه مختلف مورد بررسی قرار می‌گیرد:

- حوزه احتمال: قسمت احتمالات توأم و شرطی
- حوزه متغیرهای تصادفی: قسمت کواریانس و ضریب همبستگی
- حوزه مدل‌سازی و شناسایی الگو : قسمت مدل‌های مختلف آماری و یادگیری ماشین

بنابراین، تا اینجا با فلسفه وجودی مفهوم احتمال آشنا شدیم. از این پس، قدم به قدم با مفاهیم بیشتر در مورد فضای نمونه و احتمال آشنا شده و در مسیری گام می‌نهیم که منتهی به شناسایی الگوها در مجموعه‌ای از مشاهدات به‌دست آمده از انجام آزمایش‌های تصادفی می‌گردد تا بتوان از آن برای استنباط و استنتاج و نیز پیش‌بینی بهره برد.

مفاهیم مرتبط با احتمال

۱. *آزمایش تصادفی*: فرآیندی است که بر مبنای شانس بوده و نتایج آن را نمی‌توان با قطعیت بیان کرد. پرتاب یک تاس یا سکه را می‌توان به عنوان آزمایش‌های تصادفی نام برد. نتایج یک آزمایش تصادفی را "رخداد" یا "اتفاق" می‌نامند.

۲. *فضای نمونه*: مجموعه تمام نتایج ممکن یک آزمایش تصادفی است. فضای نمونه را به صورت یک مجموعه نشان می‌دهند. در مثال پرتاب یک سکه، فضای نمونه عبارتست از :

$\{$خط ،شیر$\}$

۳. *پیشامد*: زیرمجموعه‌ای از فضای نمونه است که می‌تواند یکی از نتایج ممکن یا ترکیبی از نتایج ممکن آزمایش تصادفی باشد. برای مثال در پرتاب یک تاس، پیشامد می‌تواند ظاهر شدن یک عدد فرد باشد که به صورت زیر نشان می‌دهیم:

$$\{1و3و5\}$$

توجه: به هریک از اعضای فضای نمونه، *نتیجه* گفته می‌شود، در حالیکه به هر یک از زیرمجموعه‌های ممکن یک فضای نمونه، *پیشامد* می‌گویند.

مثال ۱-۱: فرض کنید آزمایش تصادفی عبارتست از تعیین جنسیت فرزندان یک خانواده ۳ فرزندی. اگر B را برای نشان دادن پسر و G را برای نشان دادن دختر استفاده کنیم، فضای نمونه آزمایش را تعیین کنید.

توجه داشته باشید آزمایش مورد نظر یک آزمایش تصادفی است، یعنی جنسیت فرزندان، قبل از انتخاب خانواده مورد نظر مشخص نیست و به معنی نداشتن قطعیت درباره نتیجه آزمایش است. نتایج ممکن عبارتست از:

حالت	فرزند اول	فرزند دوم	فرزند سوم
۱	B	B	B
۲	B	B	G
۳	B	G	B
۴	B	G	G
۵	G	G	G
۶	G	G	B
۷	G	B	G
۸	G	B	B

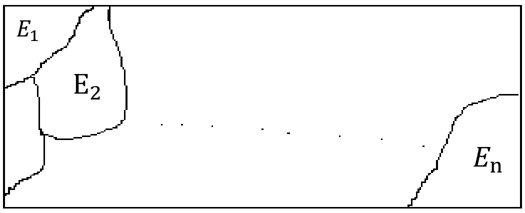
بنابراین فضای نمونه را به صورت زیر نشان می‌دهیم:

$$\{BBB, BBG, BGB, BGG, GGG, GGB, GBG, GBB\}$$

اکنون، براساس تعریف پیشامد و فضای نمونه، به موضوع محاسبه احتمال می‌پردازیم.

در تعریف، می‌توان این‌گونه بیان نمود که احتمال، اندازه‌ای است که نشان دهنده شانس وقوع یک پیشامد خاص است و با P[1] نشان داده می‌شود. در قسمت‌های پیش ملاحظه نمودیم یکی از راه‌های تقسیم‌بندی یک مجموعه این بود که زیرمجموعه‌ها هیچ نقطه مشترکی نداشته باشند. بنابراین، فضای نمونه را نیز می‌توان به زیرمجموعه‌هایی از پیشامدها تقسیم کرد که کاملاً از هم مجزا و بدون نقطه مشترک باشند. می‌توان این زیرمجموعه‌ها را با $\{E_1, E_2, \ldots\}$ نشان داد؛ طوری که هر زیرمجموعه نیز احتمال مربوط به خود را داشته باشد:

$$\{P(E_1), P(E_2), \ldots\}$$



[1] Probability

در این صورت احتمال یا مقدار P هر یک از پیشامدها دو ویژگی مهم زیر را داراست:

۱) $0 \leq P(E_i) \leq 1$

۲) $\mathrm{P}(\mathrm{E}_1)+\mathrm{P}(\mathrm{E}_2)+\cdots+\mathrm{P}(\mathrm{E}_n)=1$

ویژگی شماره (۱) بیان می‌کند اندازه احتمال هر پیشامدی در حالت کلی یک عدد مثبت و کوچکتر از یک است. احتمال، در هر مسئله و با هر روشی محاسبه شود این ویژگی را دارد. اما ویژگی شماره (۲) بازگو کننده این واقعیت است که اگر نحوه تقسیم‌بندی فضای نمونه به گونه‌ای باشد که هیچ یک از نتایج ممکن نقطه مشترکی نداشته باشند، جمع احتمال تمام آنها برابر یک است. به عبارتی، احتمال فضای نمونه یک است. می‌توان گفت وقتی یک آزمایش تصادفی انجام می‌شود، فضای نمونه به عنوان مجموعه مرجع و بزرگترین مجموعه در ارتباط با آزمایش انجام شده، احتمال وقوع یک دارد؛ به این دلیل که به هرحال یکی از اعضای فضای نمونه رخ می‌دهد.

بنابراین در محاسبه احتمال، ابتدا آزمایش تصادفی و سپس فضای نمونه مورد نظر باید تعیین شوند. محاسبه احتمالات به ما کمک می‌نماید درباره اینکه هر پیشامدی چقدر شانس وقوع دارد اطلاعاتی به‌دست آوریم، چرا که گاهی به دنبال پیشامدی با بیشترین شانس وقوع و گاهی به دنبال پیشامدی با کمترین شانس رخ دادن هستیم. از این‌رو، لازم است به دنبال روش یا روش‌هایی برای محاسبه احتمال باشیم.

به طور کلی سه رویکرد برای محاسبه احتمال وجود دارد:

۱. روش کلاسیک: در این روش، فضای نمونه آزمایش به عنوان مجموعه‌ای در نظر گرفته می‌شود که در دسترس بوده و تعداد اعضایش قابل شمارش است. همچنین، فرض می‌شود تمام اعضای فضای نمونه شانس یکسان برای رخ دادن دارند. بنابراین، احتمال رخ دادن پیشامدی مانند E به صورت زیر محاسبه می‌شود:

$$P(E)=\frac{n(E)}{n(S)}=\frac{\text{تعداد نتایج موجود در } E}{\text{تعداد اعضاء فضای نمونه}}$$

مثال ۱-۲: فرض کنید آزمایش مورد نظر، پرتاب یک تاس باشد. احتمال اینکه عدد ظاهر شده یک عدد فرد باشد چقدر است؟

$$\text{فضای نمونه} = \{1,2,3,4,5,6\}$$
$$E = \text{فردبودن عدد ظاهر شده در پرتاب تاس} = \{1,3,5\}$$
$$n(E)=3$$
$$n(S)=6$$
$$P(E)=\frac{n(E)}{n(S)}=\frac{3}{6}=\frac{1}{2}$$

۲. روش تجربی: در این روش، به فضای نمونه اصلی دسترسی نداریم و به همین دلیل، محاسبه احتمال بر مبنای تجربه‌های موجود انجام می‌شود. مثلاً، احتمال اینکه بچه بعدی که در خانواده متولد می‌شود دختر باشد و یا احتمال اینکه یک فرد ۸۰ ساله یک سال دیگر زنده بماند. احتمالات ذکر شده از روش کلاسیک قابل محاسبه نیستند، زیرا نتایج مختلف آزمایش مورد نظر، شانس یکسان ندارند و البته در بسیاری از مواقع به تمام اعضای فضای نمونه دسترسی نداریم.

برای محاسبه احتمال چنین پیشامدهایی، باید آزمایش مربوطه بارها و بارها تحت شرایط یکسان تکرار شود تا فراوانی نسبی آن پیشامد به‌دست آید و به این ترتیب به تقریبی از فضای نمونه اصلی دست می‌یابیم. اگر این آزمایش n بار انجام شده و پیشامد A در این آزمایش f بار مشاهده شود، احتمال وقوع پیشامد A برابر است با:

$$P(A) = \frac{f}{n}$$

که احتمال فوق به نوعی، مفهوم فراوانی نسبی را می‌رساند.

در واقع در این روش، به تعداد کافی آزمایش تصادفی انجام می‌شود تا بتوان تصویر نسبتاً دقیقی از فضای نمونه به‌دست آورد.

مثال ۱-۳: از بین ۵۰۰ خودروی تولید شده در یک کارخانه ۱۰ خودرو که به تصادف انتخاب شده‌اند رنگ لیمویی دارند. چقدر احتمال دارد خودروی بعدی را که انتخاب می‌کنیم رنگ لیمویی داشته باشد؟

همانطور که مشاهده می‌شود این احتمال بر مبنای تجربه قبلی محاسبه می‌شود. در این آزمایش:

$$n = 500$$
$$f = 10$$
$$P\left(\text{لیمویی بودن ماشین بعدی}\right) = \frac{10}{500}$$

به بیانی دیگر، احتمال واقعی پیشامد مورد نظر نیز بر اساس مفهوم فراوانی نسبی به‌دست می‌آید، طوری که فرض می‌شود آزمایش به تعداد بسیار زیاد انجام شده و فراوانی نسبی آن پیشامد در تعداد بسیار زیاد آزمایش، همان احتمال واقعی مربوط به آن است.

۳. روش قضاوتی: در این روش مقدار احتمال بر اساس حدس، تخمین، عقاید و یا اطلاعات تقریبی از افراد خبره و یا مطالعات پیشین به‌دست می‌آید. برای مثال، احتمال اینکه علی در امتحان درس کنترل کیفیت نمره الف بگیرد یا احتمال اینکه تیم بسکتبال الف در مقابل تیم ب برنده شود بر اساس این است که اطلاعاتی درباره نحوه و میزان درس خواندن علی داشته باشیم و یا اینکه بازی‌های قبلی دو تیم بسکتبال الف و ب را تماشا

کرده و اطلاعاتی درباره بازیکنان و چگونگی بازی تیمی آنها به‌دست آورده باشیم. چرا که علی بارها و بارها در این امتحان شرکت نکرده که بتوانیم تعداد دفعاتی که نمره الف گرفته را بشماریم.

احتمالات شرطی و استقلال

همانطور که ملاحظه شد پس از انجام یک آزمایش، با توجه به فضای نمونه کل آزمایش، احتمالات پیشامدهای مختلف محاسبه می‌شوند. در بسیاری از مواقع در انجام یک آزمایش تصادفی، بیش از یک پیشامد مورد نظر پژوهشگر است که در این صورت سه نوع توزیع احتمال روی فضای نمونه این آزمایش تعریف می‌شود که عبارتند از توزیع احتمال توأم[۱]، توزیع احتمال شرطی[۲] و توزیع احتمال حاشیه‌ای[۳] یا غیر شرطی[۴].

در حالت کلی، روی کل فضای نمونه، توزیع احتمال توأم تعریف می‌شود و در صورتی که برای هر یک از پیشامدها به صورت جداگانه احتمال را به‌دست آوریم، با احتمالات حاشیه‌ای مواجه هستیم. اما، گاهی می‌توان فضای نمونه را کوچکتر نمود و این زمانی اتفاق می‌افتد که اطلاعات بیشتری از آزمایش تصادفی وجود داشته باشد.

بنابراین، وقتی پیشامد خاصی مورد نظر بوده و هیچ گونه اطلاعات دیگری درباره سایر پیشامدها نداشته باشیم، واژه "احتمالات حاشیه‌ای" به‌کار می‌رود. اما وقتی پیشامد دیگری به صورت همزمان مورد نظر باشد و هدفمان محاسبه احتمال یک پیشامد با داشتن اطلاعاتی درباره پیشامد دیگر باشد، موضوع احتمالات شرطی مطرح می‌شود. به عبارتی، احتمالات شرطی همان احتمالات معمولی به شرط داشتن اطلاعات اضافه هستند.

فرض کنید روی فضای نمونه یک آزمایش تصادفی، دو پیشامد A و B تعریف شود، به گونه‌ای که پیشامد مورد نظر A بوده و نیز می‌دانیم که پیشامد B رخ داده است. دراین صورت، احتمال A به شرط B با $P(A|B)$ که همان احتمال شرطی است نشان داده شده و به شکل زیر محاسبه می‌شود:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

که در آن $P(A \cap B)$ احتمال این است که هر دو پیشامد به طور همزمان با یکدیگر رخ دهند که همان توزیع احتمال توأم است. از طرفی، $P(B)$ احتمال حاشیه‌ای مربوط به رخ دادن پیشامد B می‌باشد.

[1] Joint Probability Distribution
[2] Conditional Probability Distribution
[3] Marginal Probability Distribution
[4] Unconditional Probability

توجه داشته باشیم در احتمال شرطی فوق، پیشامد B، فضای نمونه جدید است که از فضای نمونه اصلی آزمایش تصادفی کوچکتر می‌باشد. در واقع، اطلاعات اضافی برای ما یک فضای نمونه جدید خلق می‌کنند و در محاسبه احتمالات، مبنای تعداد اعضای فضای نمونه، تعداد اعضای پیشامد B است.

حالت فوق را تعمیم داده و فرض کنید تعداد n پیشامد $A_1, A_2, \dots, A_n$ در فضای نمونه وجود دارد. در این صورت، احتمالات توأم، شرطی و غیر شرطی (حاشیه‌ای) به ترتیب، به صورت زیر نشان داده می‌شود:

- $P(A_1, A_2, \dots, A_n)$
- $P(A_i|A_j) \quad ; i,j = 1,2,\dots,n, \;\; i \neq j$
- $P(A_i) \quad ; i = 1,2,\dots,n$

مثال ۱-٤: فرض کنید از ۵۰۰ نفر از افراد جامعه این سؤال پرسش می‌شود که آیا موافق کاهش ساعات اداری می‌باشند یا خیر؟ پاسخ آنها به صورت موافق یا مخالف است. یک اطلاع اضافی هم درباره جنسیت این افراد به صورت جدول زیر داده شده است. فرض کنید یک نفر به تصادف انتخاب می‌شود. احتمال موافق بودن این فرد به شرط اینکه فرد انتخاب شده زن باشد چقدر است؟

مرد	زن	جمع
۱۱۰	۱۷۰	موافق
۱۹۰	۳۰	مخالف
۳۰۰	۲۰۰	جمع

توجه: در حل مسائل احتمالاتی این نکته مهم را در نظر بگیرید که همیشه اطلاعات ارائه شده در مسأله و نیز صورت سؤال اصلی را باید به زبان احتمالاتی نوشت که به این منظور، لازم است هر یک از پیشامدها به درستی تشخیص داده شود.

پیشامد موافق بودن را با A و پیشامد زن بودن را با B نشان می‌دهیم. بنابراین، سؤال مورد نظر عبارتست از $P(A|B)$.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

احتمال صورت و مخرج کسر فوق را با توجه به جدول به صورت زیر محاسبه می‌کنیم:

- در محاسبه احتمال زن بودن، باید تعداد کل زن‌ها را بر تعداد کل افراد تقسیم نماییم:

$$\mathrm{P(B)} = \frac{170 + 30}{500} = \frac{200}{500}$$

- برای محاسبه احتمال اینکه فرد مورد نظر هم یک زن باشد و هم با سؤال مطرح شده موافق باشد، باید تعداد چنین افرادی بر تعداد کل تقسیم شود:

$$P(A \cap B) = \frac{170}{500}$$

بنابراین، احتمال شرطی مورد نظر برابر است با:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{{}^{170}/_{500}}{200/500} = \frac{170}{200} = 0.85$$

اکنون، فرض کنید اطلاعات اضافه‌ای داده نشده باشد و بخواهیم تنها احتمال موافق بودن فرد انتخاب شده با سؤال مورد نظر را محاسبه نماییم:

$$P(A) = \frac{170 + 110}{500} = \frac{280}{500} = 0.56$$

توجه: در مسائلی که اطلاعات به صورت جدول فوق موسوم به جدول توافقی ارائه می‌شود، اعداد مندرج در هر یک از خانه‌های جدول برای محاسبه احتمال توأم یا مشترک دو پیشامد (تلاقی سطر و ستون) مورد استفاده قرار می‌گیرند. در مثال فوق، عدد ۱۷۰ در خانه مربوط به ردیف و ستون اول، تعداد افرادی را نشان می‌دهد که زن هستند و با سؤال مورد نظر موافق هستند.

مثال ۱-۵: احتمال اینکه در پرتاب یک تاس، عدد ۳ ظاهر شود چقدر است؟

در این مسأله فضای نمونه عبارتست از : $\{1و2و3و4و5و6\}$

$$P\left(\text{عدد 3 ظاهر شود}\right) = \frac{1}{6} = 0.167$$

در همین مسأله، احتمال اینکه در پرتاب تاس اعداد ۳ یا ۴ مشاهده شود چقدر است؟

$$P\left(3 \text{ یا } 4\right) = P(3) + P(4) = \frac{1}{6} + \frac{1}{6} = \frac{2}{6} = 0.33$$

اکنون، فرض کنید مسأله به صورت دیگری مطرح شود: احتمال اینکه در پرتاب یک تاس، عدد ۳ ظاهر شود به شرط اینکه می‌دانیم عدد ظاهر شده یک عدد فرد است؟

در این حالت پیشامدهای A و B را به صورت زیر تعریف می‌کنیم:

$$A = \{\text{عدد 3 ظاهر شود}\} = \{3\}$$
$$B = \{\text{عدد فرد باشد}\} = \{1و3و5\}$$

بنابراین: $A \cap B = \{3\}$ که یک عضو دارد و پیشامد B که اکنون نقش فضای نمونه جدید را ایفا می‌نماید ۳ عضو دارد. پس:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{3}$$

در حل مسأله فوق از این نکته بهره بردیم که در مسائلی که می‌توان فضای نمونه را به صورت یک مجموعه نشان داد، از فرمول‌های شمارش استفاده کرده و محاسبات فوق را انجام می‌دهیم.

مثال ۱-٦: احتمال اینکه هوا برفی بوده و اتوبوس هم دیر به ایستگاه برسد ۰.۰۲۳ است. علی از اخبار هواشناسی متوجه می‌شود که پیش‌بینی شده شانس اینکه هوا برفی باشد ۴۰٪ است. اگر برف ببارد، احتمال اینکه اتوبوس به ایستگاه دیر برسد چقدر است؟

در این مثال دو پیشامد داریم:

$$A = \text{برفی بودن هوا}$$
$$B = \text{دیر رسیدن اتوبوس به ایستگاه}$$

جمله اول در صورت مسئله، احتمال به وقوع پیوستن هر دو پیشامد است:

$$\mathrm{P(A \cap B)} = 0.023$$

در جمله دوم، احتمال برفی بودن هوا ارائه شده است:

$$\mathrm{P(A)} = 0.4$$

بنابراین، احتمال دیر رسیدن اتوبوس به شرط برفی بودن برابر است با:

$$P(\mathrm{B|A}) = \frac{P(A \cap B)}{\mathrm{P(A)}} = \frac{0.023}{0.4} = 0.057$$

توجه: در بسیاری از متون آماری در نمادگذاری به‌جای $P(A \cap B)$ از $P(AB)$ و یا $P(A,B)$ استفاده می‌شود.[1]

[1] این نمادگذاری در کتاب SchweserNotes (Quantitative Analysis) استفاده شده است.

دو نتیجه مهم از احتمالات شرطی باید مورد توجه قرار گیرد: استقلال و قاعده ضرب احتمالات که از اهمیت بسیار قابل توجهی برخوردارند.

- احتمالات شرطی به نوعی وابستگی دو پیشامد را نشان می‌دهند. در صورتی که وقوع پیشامد دوم هیچ‌گونه وابستگی به وقوع پیشامد اول نداشته باشد می‌گوییم، دو پیشامد *مستقل* هستند. به عبارتی، داشتن اطلاع از وقوع پیشامد اول هیچ گونه تغییری در محاسبه احتمال پیشامد دوم ایجاد نمی‌کند، یعنی:

$$P(A|B) = P(A)$$

در این‌صورت می‌گوییم دو پیشامد A و B از یکدیگر مستقل هستند.

مفهوم استقلال در آمار به این صورت نیز نشان داده می‌شود: $\mathrm{P}(\mathrm{A} \cap \mathrm{B}) = \mathrm{P}(\mathrm{A}) \times \mathrm{P}(\mathrm{B})$

مثال ۱-۷: جعبه‌ای شامل ۱۰۰ عدد CD تولید شده توسط دو نوع کارخانه مختلف است. تعداد CD سالم و معیوب برای هر دو برند به صورت جدول زیر نشان داده شده است. آیا دو پیشامد زیر در ارتباط با تولید این CD ها در دو کارخانه مذکور که به صورت زیر تعریف شده‌اند مستقل می‌باشند؟

A : یک CD به تصادف انتخاب شده و توسط کارخانه الف تولید شده باشد.

B : یک CD به تصادف انتخاب شده و معیوب باشد.

	معیوب	سالم	جمع
کارخانه الف	۹	۵۱	۶۰
کارخانه ب	۶	۳۴	۴۰
جمع	۱۵	۸۵	۱۰۰

$$P(B) = \frac{\text{تعداد } CD \text{ های معیوب}}{\text{تعداد کل } CD \text{ ها}} = \frac{15}{100} = 0.15$$

$$P(A) = \frac{\text{تعداد } CD \text{ های تولید شده توسط کارخانه الف}}{\text{تعداد کل } CD \text{ ها}} = \frac{60}{100} = 0.6$$

به منظور دریافتن استقلال این دو پیشامد بررسی می‌کنیم آیا احتمال‌های شرطی با احتمال‌های حاشیه‌ای برابر هستند؟ در صورت برابری، مستقل بودن دو پیشامد را می‌توان نتیجه گرفت.

$$\mathrm{P(B|A)} = \frac{P(A \cap B)}{\mathrm{P(A)}} = \frac{9}{60} = 0.15 = \mathrm{P(B)}$$

$$\mathrm{P(A|B)} = \frac{P(A \cap B)}{\mathrm{P(B)}} = \frac{9}{15} = 0.6 = \mathrm{P(A)}$$

بنابراین، دو پیشامد مورد نظر مستقل می‌باشند.

در این قسمت، استقلال پیشامدهای یک فضای نمونه را تعمیم می‌دهیم. در صورتی‌که تعداد n پیشامد مستقل از هم $A_1, A_2, \dots, A_n$ در فضای نمونه وجود داشته باشد داریم:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \times P(A_2) \times \dots \times P(A_n)$$

بنابراین، در این بخش آموختیم که شرط استقلال دو یا چند پیشامد در یک آزمایش تصادفی و نحوه نمادگذاری احتمالاتی آن چیست.

- قاعده ضرب احتمالات وقتی دو پیشامد داریم، بر اساس فرمول احتمال شرطی می‌باشد:

$$P(A|B) = \frac{P(A \cap B)}{\mathrm{P(B)}} \rightarrow P(A \cap B) = \mathrm{P(B)} \times P(A|B)$$

از طرفی:

$$P(B|A) = \frac{P(A \cap B)}{\mathrm{P(A)}} \rightarrow P(A \cap B) = \mathrm{P(A)} \times P(B|A)$$

بنابراین، احتمال اشتراک دو پیشامد یا احتمال توأم دو پیشامد را می‌توان بر حسب یک احتمال شرطی و یک احتمال غیر شرطی (حاشیه‌ای) نوشت.

اکنون، پیشامدهای $A_1, A_2, \dots, A_n$ را در نظر بگیرید. در این‌صورت، قانون ضرب احتمالات به صورت زیر است:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1|A_2 \cap \dots \cap A_n) \times P(A_2|A_3 \cap \dots \cap A_n) \times \dots \times P(A_n)$$

همانطور که مشاهده می‌شود مجموعه کل پیشامدها به n بخش مجزا تقسیم شده است. احتمال اینکه تمام پیشامدها همزمان باهم اتفاق بیفتد به صورت حاصل ضرب مجموعه‌ای از احتمالات شرطی به‌دست می‌آید.
یکی از کاربردهای مهم قاعده ضرب احتمال، محاسبه احتمال یک پیشامد (احتمال غیر شرطی) بر اساس احتمالات شرطی است. در حالت فوق فرض کنید علاوه بر n پیشامد $A_1, A_2, \dots, A_n$ ، پیشامد B نیز روی فضای نمونه تعریف شده و هدف نیز محاسبه احتمال آن است.

در این صورت، احتمال پیشامد B به صورت زیر محاسبه می‌شود:

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_n)$$
$$= P(B|A_1) \times P(A_1) + P(B|A_2) \times P(A_2) + \cdots + P(B|A_n) \times P(A_n)$$

مثال ۱-۸: فرض کنید تحصیلات دانشگاهی کارکنان یک شرکت مالی در یکی از سه گروه بدون تحصیلات دانشگاهی، لیسانس و بالاتر از لیسانس طبقه‌بندی شده‌اند؛ به طوری که ۶۰٪ کارکنان بدون تحصیلات دانشگاهی، ۳۰٪ دارای تحصیلات لیسانس و تحصیلات ۱۰٪ از آنان نیز بالاتر از لیسانس می‌باشد. از طرفی، از جنبه سیاست‌گذاری‌های هزینه‌ای برای هیئت مدیره این شرکت، افرادی که سالانه بیش از ۴۰ هزار دلار درآمد دارند از اهمیت بیشتری برخوردارند. به همین دلیل، بر اساس یک بررسی از درآمد کارکنان شرکت مذکور، هیئت مدیره متوجه شده است درآمد سالانه ۱۰٪ کارکنان بدون تحصیلات دانشگاهی بیش از ۴۰ هزار دلار است. درحالیکه، این میزان درآمد در مورد کارکنان با تحصیلات لیسانس ۷۰٪ و برای کارکنان با تحصیلات بالاتر از لیسانس ۱۰۰٪ می‌باشد. سؤال هیئت مدیره این است که با توجه به این اطلاعات، چند درصد از کارکنان این شرکت دارای درآمد سالانه بیشتر از ۴۰ هزار دلار می‌باشند؟

به منظور پاسخ به این سؤال، ابتدا باید اطلاعات داده شده در مسأله را با زبان احتمالی ترجمه نماییم. همان‌گونه که مشاهده می‌شود دو پیشامد در این بررسی در نظر گرفته شده است: تحصیلات فرد و اینکه آیا درآمد سالانه فرد از ۴۰ هزار دلار بیشتر است یا خیر. اگر پیشامد تحصیلات را با B نشان دهیم، سه گروه B_1، B_2 و B_3 برای آن می‌توان در نظر گرفت. از سویی دیگر، درآمد بیشتر از ۴۰ هزار دلار را با A نشان می‌دهیم. از این‌رو، داریم:

$$\left.\begin{array}{l} P(B_1) = 0.6 \\ P(B_2) = 0.3 \\ P(B_3) = 0.1 \end{array}\right\} \quad \text{احتمالات سه طبقه تحصیلاتی}$$

توجه داشته باشیم جمع احتمالات برابر یک است.

$$\left.\begin{array}{l} P(A|B1) = 0.1 \\ P(A|B2) = 0.7 \\ P(A|B1) = 1 \end{array}\right\} \quad \text{احتمالات درآمد بیش از ۴۰ هزار دلار در هر طبقه تحصیلاتی}$$

سؤال: $P(A) =?$

در پاسخ به این سؤال، از فرمول قاعده احتمال به صورت زیر استفاده می‌کنیم:

$$P(A) = P(A \cap B1) + P(A \cap B2) + P(A \cap B3)$$
$$= P(A|B1) \times P(B1) + P(A|B2) \times P(B2) + P(A|B3) \times P(B3)$$
$$= 0.1 \times 0.6 + 0.7 \times 0.3 + 1 \times 0.1 = 0.37$$

بنابراین، می‌توان گفت در این شرکت ۳۷٪ کارکنان، از درآمد سالانه بالای ۴۰ هزار دلار برخوردارند.

علاوه بر قاعده احتمال، احتمال شرطی از مباحث بسیار مهم در آمار و احتمالات بوده و زیربنای موضوع احتمال بیز است و اولین بار توسط توماس بیز[1](۱۷۶۱-۱۷۰۲) پیشنهاد شد و رهیافت جدیدی در محاسبه احتمالات و استنباط آماری ارائه نمود. در این رویکرد، یک احتمال شرطی بر مبنای احتمال شرطی دیگر و اطلاعات اضافه‌تر محاسبه می‌شود. نسخه ساده‌تر این روش با فرض وجود تنها دو پیشامد A و B در فضای نمونه آزمایش تصادفی به صورت زیر است:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

که $P(B) \neq 0$

در اینجا با تعمق بیشتری به این فرمول می‌پردازیم. پیشامد A به گونه‌ای است که از قبل، اطلاعاتی درباره احتمال وقوع آن داریم : $P(A)$. اکنون اطلاعاتی درباره پیشامد B به‌دست آورده‌ایم و با توجه به این اطلاعات جدید، می‌خواهیم احتمال قبلی ($P(A)$) که از این پس آن را با "*احتمال پیشین*" نامگذاری می‌نماییم را به‌روزرسانی کنیم. آنچه به دنبال آن هستیم "احتمال رخ دادن A به شرط آنکه می‌دانیم B رخ داده است" می‌باشد. به عبارتی، رویکرد بیز در محاسبه احتمالات، رهیافت به‌روزرسانی احتمال یک پیشامد *پس از* دستیابی به اطلاعات پیشامد دیگر است و این احتمال به‌روز رسانی شده را "*احتمال پسین*" می‌نامند.

از طرفی، بر اساس قاعده احتمال کل، احتمال پیشامد B را نیز به صورت زیر به‌دست می‌آوریم:

$$P(B) = P(B \cap A) + P(B \cap A') = P(B|A) \times P(A) + P(B|A')P(A')$$

توجه داشته باشید در حل مسائل مربوط به قاعده بیز، تشخیص دو پیشامد فوق و نیز احتمال پیشین از اهمیت زیادی برخوردار است.

مثال ۱-۹: فرض کنید آزمایش جدیدی برای تشخیص بیماری ایدز (HIV) ارائه شده و به شکلی است که فرد بیمار را در ۹۹ درصد مواقع شناسایی می‌کند و در ۰.۲ درصد مواقع، قادر به شناسایی فرد سالم نیست. همچنین، از اطلاعات قبلی می‌دانیم احتمال اینکه در این جامعه یک نفر به بیماری ایدز مبتلا شده باشد ۰.۰۰۰۰۱ است. فرد جدیدی در این آزمایش شرکت کرده و پاسخ آن مثبت شده است. احتمال ابتلای این فرد به بیماری مذکور چیست؟

[1] Thomas Bayes

از آنجایی که در صورت مسأله بیان شده از اطلاعات قبلی می‌دانیم احتمال ابتلا به بیماری ایدز در این جامعه ۰.۰۰۰۰۱ است یعنی اولاً ابتلا به بیماری ایدز، پیشامد اصلی است که قرار است احتمال وقوع آن را به‌روزرسانی نماییم و ثانیاً احتمال پیشین آن نیز ۰.۰۰۰۰۱ می‌باشد.

ابتدا اطلاعات داده شده در مساله را با نمادگذاری احتمالی می‌نویسیم:

$$P(HIV) = 0.00001 \Rightarrow P(No\ HIV) = 1 - 0.00001 = 0.99999$$
$$P(test + |No\ HIV) = 0.002$$
$$P(test + |HIV) = 0.999$$

در مرحله بعد، با استفاده از فرمول قاعده بیز به سؤال مسئله پاسخ می‌دهیم:

$$P(HIV|test +) = \frac{P(test + |HIV) \times P(HIV)}{P(test +)}$$

$$= \frac{P(test + |HIV) \times P(HIV)}{P(test + |HIV) \times P(HIV) + P(test + |No\ HIV) \times P(No\ HIV)}$$

$$= \frac{0.999 \times 0.00001}{0.999 \times 0.00001 + 0.002 \times 0.99999} = 0.00497$$

مثال ۱-۱۰: فرض کنید احتمال اینکه اقتصاد کشوری عملکرد بهتری داشته باشد ۰.۶ است و در این صورت، احتمال رشد یک سهام خاص ۰.۷ و احتمال کاهش ارزش آن ۰.۳ است. از طرفی، احتمال اینکه اقتصاد این کشور از عملکرد ضعیفی برخوردار شود ۰.۴ است و در این صورت، احتمال رشد همان سهام ۰.۲ و احتمال کاهش ارزش آن ۰.۸ می‌باشد. در شرایطی هستیم که مطلع شدیم قیمت سهام مذکور رشد داشته است. احتمال اینکه اقتصاد این کشور عملکرد خوبی داشته باشد چقدر است؟

همان‌گونه که مشاهده می‌شود در این مسأله دو پیشامد مطرح شده است: عملکرد اقتصاد (دو طبقه بهتر و ضعیف) و قیمت سهام (دو طبقه رشد و کاهش) که به ترتیب با A و B نشان می‌دهیم. به منظور پاسخ به این سؤال، ابتدا اطلاعات داده شده مسأله را با زبان احتمالات می‌نویسیم:

$$\left.\begin{aligned} P(B_1|A +) &= 0.7 \\ P(B_2|A +) &= 0.3 \\ P(B_1|A -) &= 0.2 \\ P(B_2|A -) &= 0.8 \end{aligned}\right\} \text{احتمالات شرطی}$$

$$\left.\begin{array}{l} P(A+) = 0.6 \\ P(A-) = 0.4 \end{array}\right\} \quad \text{احتمالات پیشین}$$

سؤال:

$$P(A+|B_1)=?$$

با استفاده از فرمول بیز به صورت زیر به سؤال فوق پاسخ می‌دهیم:

$$P(A+|B_1) = \frac{P(B_1|A+) \times P(A+)}{P(B_1)}$$

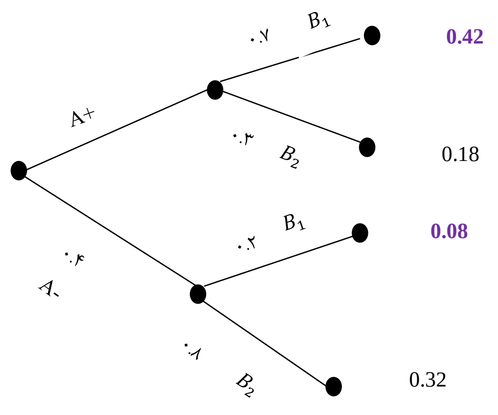
$$= \frac{P(B_1|A+) \times P(A+)}{P(B_1|A+) \times P(A+) + P(B_1|A-) \times P(A-)}$$

$$= \frac{0.7 \times 0.6}{0.7 \times 0.6 + 0.2 \times 0.4} = \frac{0.42}{0.5} = 0.84$$

به عبارتی اگر قیمت سهام رشد داشته باشد، با احتمال ۰.۸۴ اقتصاد آن کشور عملکرد بهتری داشته است.

در اینجا به ذکر نکته کاربردی می‌پردازیم. مسائل بیز را می‌توان با شیوه درختی نیز حل نمود. در این شیوه، ابتدا پیشامدی که احتمالات پیشین برای آن ارائه شده است، در ریشه درخت قرار می‌گیرد و سپس پیشامد دیگر که احتمالات آن را به شرط این ریشه در اختیار داریم در ساقه‌های درخت قرار می‌دهیم. سپس، احتمالات ارائه شده در مسأله روی ریشه‌ها و ساقه‌ها نوشته می‌شود.

در مثال فوق، عملکرد اقتصاد کشور مورد نظر در ریشه و وضعیت رشد سهام در ساقه‌های آن واقع می‌شود:



در این صورت، از آنجایی که با رشد قیمت سهام مواجه شده‌ایم (B_1)، فضای نمونه ما تنها به دو احتمال ۰.۴۲ و ۰.۰۸ محدود می‌شود که در مخرج کسر قرار می‌گیرد (0.5=0.08+0.42).

اکنون، از این فضای نمونه تنها به دنبال احتمال عملکرد بهتر هستیم (A+) و باید احتمال ۰.۴۲ را در نظر بگیریم و از این‌رو، احتمال عملکرد بهتر اقتصاد به شرط رشد قیمت سهام برابر است با:

$$\frac{0.42}{0.5} = 0.84$$

مثال ۱-۱۱: فرض کنید یک کشتی باری شامل ۱۰۰۰ دستگاه اتومبیل به سمت پارکینگ بارگیری شده است که ویژگی این اتومبیل‌ها به شرح زیر می‌باشد:

- تعداد ۶۰۰ اتومبیل با رنگ آبی وجود دارد. (B)
- در میان اتومبیل‌های آبی رنگ، تعداد ۱۵۰ اتومبیل دارای کمک راننده (DA) هستند.
- تعداد ۴۰۰ اتومبیل قرمز رنگ وجود دارد. (R)
- از میان اتومبیل‌های قرمز رنگ، تعداد ۲۰۰ اتومبیل از کمک راننده برخوردار (DA) هستند.

بر اساس اطلاعات فوق، احتمالات زیر را محاسبه نمایید:

۱. احتمالات غیر شرطی $P(B)\,, P(R)$
۲. احتمالات شرطی $P(DA|R)\,, P(DA|B)$
۳. احتمالات توأم $P(DA,B), P(DA,R)$
۴. احتمال کل $P(DA)$
۵. احتمال بیز $P(B|DA)$

پاسخ:

۱. احتمالات غیر شرطی:

$$P(B) = \frac{600}{1000} = 0.6$$

$$P(R) = \frac{400}{1000} = 0.4$$

توجه داریم از آنجایی که اتومبیل‌ها یا قرمز رنگ هستند یا آبی رنگ، جمع این دو احتمال غیر شرطی باید ۱ باشد.

۲. احتمالات شرطی:

$$P(DA|B) = \frac{150}{600} = 0.25$$