



۸۱۰۱۰۴۲۱۴



ارزیابی کارایی سیستم‌های کامپیوتری

گروه اول

مسئله روز تولد: K نفر در یک اتاق هستند، احتمال آنکه لااقل یک زوج (۲ نفر) دارای تولد یکسان باشند.

اشتباه لایب نیتز: اگر ۲ تاس را با هم بریزیم، کدام محتمل تر است، مجموع ۱۱ یا مجموع ۱۲.

مسئله نیوتن - پیپس: کدامیک محتمل تر است:

(الف) ۶ تاس را می‌ریزیم و لااقل یک ۶ رخ دهد.

(ب) ۱۲ تاس را می‌ریزیم و لااقل ۲ تا ۶ رخ دهد.

(پ) ۱۸ تاس را می‌ریزیم و لااقل ۳ تا ۶ رخ دهد.

فرض کنید X و Y اعدادی باشند که به طور تصادفی از یک جعبه که حاوی ۴ کارت به شماره‌های ۱، ۲، ۳، ۴ باشد، انتخاب می‌شوند.

جدول توزیع توأم X و Y را بیان کنید (نمایش دهید):

(الف) برای نمونه‌گیری با جایگذاری

(ب) برای نمونه‌گیری بدون جایگذاری

(ج) در هر دو حالت $P(X \leq Y)$ را حساب کنید.

(الف) هر زوج i, j متساوی احتمال است

$$P(X = i, Y = j) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16} \quad \text{for } i, j \in \{1, 2, 3, 4\}$$

$$P(X \leq Y) = \frac{1}{16}(1 + 2 + 3 + 4) = \frac{10}{16}$$

4	3	2	1	$X \setminus Y$
$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	1
$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	2
$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	3
$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	4

(ب)

$$P(X = i, Y = i) = 0 \quad i \in \{1, 2, 3, 4\}$$

$$P(X = i, Y = j) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12} \quad \text{for } i \neq j \in \{1, 2, 3, 4\}$$

4	3	2	1	$X \setminus Y$
$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	0	1
$\frac{1}{12}$	$\frac{1}{12}$	0	$\frac{1}{12}$	2
$\frac{1}{12}$	0	$\frac{1}{12}$	$\frac{1}{12}$	3
0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	4

(ج)

$$P(X \leq Y) = P(X < Y) = \frac{1}{2}$$

$$P(X < Y) = P(X > Y)$$

$$P(X = Y) = 0$$

$$P(X < Y) + P(X > Y) + P(X = Y) = 1$$

X_1 و X_2 بیانگر عدد تاس است وقتی ۲ تاس را می‌ریزیم.

$$X_1 = \text{عدد تاس اول}$$

$$X_2 = \text{عدد تاس دوم}$$

$$Y_1 = \max\{X_1, X_2\}$$

$$Y_2 = \min\{X_1, X_2\}$$

توزیع توأم برای $(X_1, X_2) = ?$

توزیع توأم برای $(Y_1, Y_2) = ?$

چون X_1, X_2 مستقل هستند:

for any pair (i, j) , $1 \leq i, j \leq 6$

$$P(X_1 = i, X_2 = j) = P(X_1 = i)P(X_2 = j) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

6	5	4	3	2	1	$X_1 \setminus X_2$
$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	1
$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	2
$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	3
$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	4
$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	5
$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	6

می‌دانیم کمینه X_1, X_2 نمی‌تواند بزرگتر از بیشینه X_1, X_2 باشد. پس:

$$P(\underbrace{Y_1}_{\max} = i, \underbrace{Y_2}_{\min} = j) = 0 \quad \text{for } j > i$$

$$P(Y_1 = i, Y_2 = i) = \frac{1}{36}$$

$$P(Y_1 = i, Y_2 = j) = \frac{2}{36} \quad \text{for } i > j$$

چون پیشامد $E = \{(1, 2), (2, 1)\}$ در دو حالت رخ می‌دهد (که در هر دو حالت $\min = 1$ و $\max = 2$ است)، احتمال آن برابر با مجموع احتمالات این دو حالت است ($\frac{1}{36} + \frac{1}{36} = \frac{2}{36}$). به همین ترتیب برای بقیه ارقام زیر قطر اصلی این قاعده برقرار است.

6	5	4	3	2	1	$Y_1 \setminus Y_2$
0	0	0	0	0	$\frac{1}{36}$	1
0	0	0	0	$\frac{1}{36}$	$\frac{2}{36}$	2
0	0	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	3
0	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	4
0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	5
$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	6

فرض کنید از یک بسته کارت، ۷ کارت به تصادف بر می‌داریم (از ۵۲ کارت، ۱۳ تا پیک، ۱۳ تا دل و... است).

X = متغیر تصادفی که تعداد پیک در ۷ کارت را بیان می‌کند. امید X را بدست آورید.

حل ۱: توزیع Hypergeometric

حل ۲: به کمک متغیر نشانگر: فرض کنید $I_j =$ متغیر نشانگر بیانگر پیشامد (کارت j ام پیک باشد) $A_j =$

$$P(A_j) = \frac{13}{52}$$

$A_1^c = 39$ کارت غیر پیک

طبق قانون احتمال کل داریم:

$$P(A_2) = P(A_2|A_1)P(A_1) + P(A_2|A_1^c)P(A_1^c)$$

با جایگذاری مقادیر خواهیم داشت:

$$P(A_2) = \left(\frac{12}{51} \times \frac{13}{52}\right) + \left(\frac{13}{51} \times \frac{39}{52}\right)$$

تعداد پیک‌ها برابر است با مجموع متغیرهای نشانگر:

$$X = \sum_{j=1}^7 I_j$$

با گرفتن امید ریاضی از طرفین:

$$\Rightarrow E[X] = E\left[\sum_{j=1}^7 I_j\right] = \sum_{j=1}^7 E[I_j] = \sum_{j=1}^7 P(I_j) = \sum_{j=1}^7 \frac{13}{52} = \frac{7}{4}$$

نکات حل:

- تبدیل $E[\sum I_j]$ به $\sum E[I_j]$ به دلیل ویژگی خطی بودن امید است.

- تبدیل $E[I_j]$ به $P(I_j)$ به دلیل خاصیت متغیر نشانگر است.

ساختمان ۱۰ طبقه: ۱۲ نفر سوار آسانسور می‌شوند. هر یک به تصادف در طبقات ۱ تا ۱۰ پیاده می‌شوند. حساب کنید در میانگین، آسانسور چند بار متوقف می‌شود؟
 $X =$ متغیر تصادفی که تعداد توقف‌ها را نشان می‌دهد (توقف در طبقه یعنی لااقل یک نفر در آن طبقه پیاده شود).
 محاسبه $P(X = x)$ برای $x = 1, 2, \dots, 10$ ساده نیست (یعنی محاسبه تابع توزیع ساده نیست).
 برای محاسبه امید به تابع توزیع نیاز داریم، ولی اینجا ترفندی به کار می‌بریم (بدون نیاز به تابع توزیع) و از متغیر نشانگر استفاده می‌کنیم.
 $I_j = 1$ اگر لااقل ۱ نفر در طبقه j پیاده شود.

$$X = \sum_{j=1}^{10} I_j$$

(احتمال آنکه کسی در طبقه j پیاده نشود) $P[I_j = 1] = 1 -$

- احتمال آنکه کسی در طبقه j پیاده نشود = هر کدام از ۱۲ نفر در طبقه‌ای غیر از j پیاده شوند $= \left(\frac{9}{10}\right)^{12}$

- بنابراین:

$$P[I_j = 1] = 1 - \left(\frac{9}{10}\right)^{12}$$

$$E[X] = E\left[\sum_{j=1}^{10} I_j\right] = \sum_{j=1}^{10} E[I_j] = \sum_{j=1}^{10} P[I_j] = \sum_{j=1}^{10} \left(1 - \left(\frac{9}{10}\right)^{12}\right)$$

نکات حل:

- تبدیل $E[\sum I_j]$ به $\sum E[I_j]$ به دلیل ویژگی خطی بودن امید است.

- تبدیل $E[I_j]$ به $P[I_j]$ به دلیل خاصیت متغیر نشانگر است.

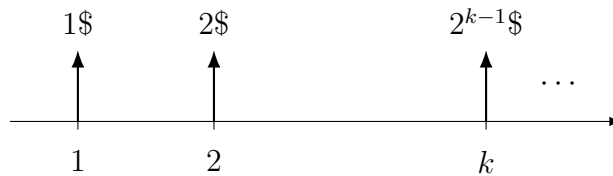
۳. پارادوکس سن پترزبورگ

در آغاز بازی، یک سکه را می‌اندازیم. اگر در پرتاب اول شیر بیاید، ۱ دلار می‌گیریم. اگر خط بیاید، دوباره پرتاب می‌کنیم. اگر اولین شیر در پرتاب k ام ظاهر شود، مقدار جایزه برابر 2^{k-1} دلار است.

مقادیر متغیر تصادفی جایزه به صورت زیر است:

$$X = 1, 2, \dots$$

نمایش جایزه‌ها روی محور پرتاب‌ها:



سؤال این است که برای انجام این بازی چه مقدار پول باید پرداخت کنیم. امید ریاضی جایزه برابر است با:

$$\begin{aligned} \overline{E}[X] &= \sum_{k=1}^{\infty} 2^{k-1} \mathbb{P}(k \text{ اولین شیر در پرتاب}) \\ &= \sum_{k=1}^{\infty} 2^{k-1} \left(\frac{1}{2}\right)^k \\ &= \sum_{k=1}^{\infty} \frac{1}{2} \\ &= \infty. \end{aligned}$$

اکنون فرض کنید طرف مقابل (کازینو) حداکثر 2^n دلار پول دارد. در این حالت، برای بازی یک سقف در نظر می‌گیریم: اگر اولین شیر در یکی از پرتاب‌های 1 تا n ظاهر شود، جایزه همان 2^{k-1} دلار است؛ و اگر در هر n پرتاب خط بیاید، 2^n دلار پرداخت می‌شود. در این صورت:

$$\begin{aligned} \overline{\mathbb{E}}[X] &= \sum_{k=1}^n 2^{k-1} \mathbb{P}(k \text{ اولین شیر در پرتاب}) + 2^n \mathbb{P}(n \text{ پرتاب خط بیاید در هر } n) \\ &= \sum_{k=1}^n \frac{1}{2} + 1 \\ &= \frac{n+2}{2}. \end{aligned}$$

مثلاً اگر $n = 10$ باشد، حداکثر متوسط پولی که برای انجام این بازی می‌توان پرداخت کرد برابر است با:

$$\mathbb{E}[X] = \frac{n+2}{2} = 6 \text{ دلار.}$$

۱. مسئله ی استخدام منشی

استخدام منشی: آژانس کاریابی هر روز یک کاندیدا می‌فرستد (نفر می‌فرستد n در کل). با او مصاحبه می‌کنید؛ یا او را رد می‌کنید و یا استخدام می‌کنید که در این صورت، فرد قبلی اخراج می‌شود. بهترین نیرو را می‌خواهیم.

C_i = هزینه ی مصاحبه , C_h = (اخراج قبلی) هزینه ی استخدام .

$$C_h \gg C_i.$$

الگوریتم نوشته شده روی تخته:

```
Hire_Assistant(n)
Best <- 0
for i <- 1 to n:
  if candid(i) is better than candid(Best)
    Best <- i
hire candid i
```

چگونه n نفر را تصادفی می‌کنیم؟

اگر m تعداد دفعات استخدام باشد، هزینه ی کل به صورت زیر نوشته شده است:

$$\text{total cost} = O(nC_i + mC_h), \quad m = \text{تعداد دفعات استخدام.}$$

$$nC_i + C_h \quad \text{بهترین هزینه}$$

$$nC_i + nC_h \quad \text{بدترین هزینه}$$

میانگین هزینه؟

در الگوریتم تصادفی، ورودی‌ها را ابتدا تصادفی می‌کنند و سپس به عنوان ورودی به الگوریتم می‌دهند.

۲. حل تصادفی مسئله‌ی استخدام

جواب: فرض کنید ورودی به صورت تصادفی ارسال می‌شود.

سؤال: میانگین تعداد استخدام چقدر است؟

متغیر تصادفی X را برابر تعداد کاندیداهای استخدام شده تعریف می‌کنیم:

$$X = \# \text{ of candidates hired.}$$

برای هر کاندیدای i ام، متغیر شاخص زیر را در نظر می‌گیریم:

$$X_i = I\{\text{candidate } i \text{ is hired}\}.$$

در نتیجه:

$$X = X_1 + \cdots + X_n, \quad \mathbb{E}[X] = ?$$

با استفاده از خطی بودن امید ریاضی و خاصیت متغیر شاخص داریم:

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\sum_{i=1}^n X_i\right] \\
&= \sum_{i=1}^n \mathbb{E}[X_i] \\
&= \sum_{i=1}^n \mathbb{P}(X_i = 1) \\
&= \sum_{i=1}^n \frac{1}{i} \\
&= \ln(n) + O(1) \\
&= O(\ln n).
\end{aligned}$$

دلیل رابطه ی $\mathbb{E}[X_i] = \frac{1}{i}$ این است که کاندیدای i ام تنها زمانی استخدام می شود که از میان i کاندیدای اول، بهترین فرد باشد. چون ترتیب ورود تصادفی است، بهترین فرد با احتمال برابر در هر یک از i جایگاه قرار می گیرد؛ بنابراین:

$$\mathbb{E}[X_i] = \mathbb{P}(X_i = 1) = \frac{1}{i}.$$

برای $i = 1$ کاندیدای اول حتماً استخدام می شود:

$$\mathbb{P}(X_1 = 1) = 1 = \frac{1}{1}.$$

برای $i = 2$ دو ترتیب ممکن وجود دارد:

$$(1, 2), \quad (2, 1),$$

و فقط در یکی از آن ها کاندیدای دوم بهترین فرد میان دو کاندیدای اول است؛ پس:

$$\mathbb{P}(X_2 = 1) = \frac{1}{2}.$$

برای $i = 3$ شش ترتیب ممکن عبارت اند از:

$$(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1).$$

در دو ترتیب $(1, 2, 3)$ و $(2, 1, 3)$ ، کاندیدای سوم بهترین فرد میان سه کاندیدای اول است. بنابراین:

$$\mathbb{P}(X_3 = 1) = \frac{2}{6} = \frac{1}{3} = \frac{1}{i}.$$

همین استدلال برای i نفر نیز برقرار است و احتمال استخدام کاندیدای i ام برابر $\frac{1}{i}$ است.

۱ توزیع نرمال چندمتغیره

فرض کنید بردار پارامتر $\vec{\theta} = (\mu, \sigma^2)$ باشد. توزیع مشترک دو متغیر تصادفی X, Y را نرمال دومتغیره در نظر می‌گیریم:

$$f(x, y) \sim N(\vec{\mu}, \Sigma),$$

که در آن $\vec{\mu}$ بردار میانگین و Σ ماتریس کوواریانس است. همچنین:

$$\Sigma^{-1} = \text{precision (ماتریس دقت)}.$$

توزیع شرطی

توزیع شرطی X به شرط Y نیز نرمال است، اما با میانگین و کوواریانس متفاوت (وابسته به Y):

$$f(x | y) \sim N(\theta_{\text{جدید}}).$$

توزیع حاشیه‌ای (Marginal)

توزیع حاشیه‌ای هر یک از متغیرها با انتگرال‌گیری روی متغیر دیگر به دست می‌آید:

$$f(x) \sim N(\theta) = \int_y f(x, y) dy,$$

$$f(y) \sim N(\theta) = \int_x f(x, y) dx.$$

۲ مسئله‌ی دو فرزند (Martin Gardner)

خانواده‌ای دو فرزند دارد. هر فرزند مستقل و با احتمال یکسان دختر یا پسر است (جنسیت فرزندان مستقل و هم‌احتمال فرض می‌شود). می‌خواهیم دو احتمال شرطی زیر را با هم مقایسه کنیم:

□ احتمال اینکه هر دو فرزند دختر باشند، به شرط اینکه بدانیم فرزند بزرگ‌تر دختر است؛

□ احتمال اینکه هر دو فرزند دختر باشند، به شرط اینکه بدانیم حداقل یکی از فرزندان دختر است.

فضای نمونه (با فرض ترتیب فرزند بزرگ‌تر، فرزند کوچک‌تر) به صورت زیر است و هر حالت با احتمال یکسان $\frac{1}{4}$ رخ می‌دهد:

$$\Omega = \{GG, GB, BG, BB\}$$

که در آن حرف اول جنسیت فرزند بزرگ‌تر و حرف دوم جنسیت فرزند کوچک‌تر است.

حالت اول: شرط، مشخص کردن فرزند بزرگ‌تر است

$$P(\text{both girls} \mid \text{elder is a girl}) = \frac{P(\text{both girls, elder is a girl})}{P(\text{elder is a girl})} = \frac{1/4}{1/2} = \frac{1}{2}$$

حالت دوم: شرط، وجود حداقل یک دختر است (بدون مشخص کردن کدام فرزند)

$$P(\text{both girls} \mid \text{at least one girl}) = \frac{P(\text{both girls, at least one girl})}{P(\text{at least one girl})} = \frac{1/4}{3/4} = \frac{1}{3}$$

۳ عدم تقارن احتمال شرطی و اثر «مشخص بودن» شرط

نکته‌ی مهم این است که خاصیت تقارن برای احتمال شرطی برقرار نیست، یعنی به طور کلی

$$P(A \mid B) \neq P(B \mid A),$$

و این یعنی اطلاع شرطی لزوماً به یک شکل روی دو رویداد اثر نمی‌گذارد.

با توجه به تقارن خود مسئله (بین فرزند بزرگ‌تر و کوچک‌تر)، داریم:

$$P(\text{both girls} \mid \text{younger is a girl}) = P(\text{both girls} \mid \text{elder is a girl}) = \frac{1}{2},$$

اما همان‌طور که دیدیم:

$$P(\text{both girls} \mid \text{elder is a girl}) \stackrel{?}{\neq} P(\text{both girls} \mid \text{at least one girl}) \quad \left(\frac{1}{2} \neq \frac{1}{3} \right).$$

چرا این دو با هم فرق دارند؟

□ مورد ۱ (شرط مشخص/Specific): شرط «فرزند بزرگ‌تر دختر است» یک فرزند مشخص (فرزند بزرگ‌تر) را تعیین می‌کند.

□ مورد ۲ (شرط غیرمشخص/Non-specific): شرط «حداقل یک دختر» به هیچ فرزند خاصی اشاره نمی‌کند و چند حالت را هم‌زمان پوشش می‌دهد (GB و BG), بنابراین اطلاعات کمتری نسبت به مورد اول در اختیار می‌گذارد.

قاعده‌ی کلی: هرچه اطلاعات موجود در شرط مشخص‌تر (*specific*) باشد، احتمال شرطی حاصل به $\frac{1}{2}$ نزدیک‌تر می‌شود؛ و هرچه اطلاعات غیرمشخص‌تر باشد، از $\frac{1}{2}$ دورتر می‌شود (در این مثال به $\frac{1}{3}$ رسیدیم).

۴ مثال: به‌روزرسانی باور با قاعده‌ی بیز (مثال دو سکه)

دو سکه داریم: سکه‌ی A که سالم است ($P(H) = \frac{1}{2}$) و سکه‌ی B که ناسالم است ($P(H) = \frac{3}{4}$). یکی از این دو سکه به‌طور تصادفی انتخاب و ۳ بار پرتاب می‌شود و نتیجه‌ی HHH مشاهده می‌شود. باور اولیه‌ی ما (پیشین *belief*) درباره‌ی اینکه کدام سکه انتخاب شده، برابر و ۵۰ درصد است:

$$P(A) = P(B) = 0.5.$$

سؤال: با مشاهده‌ی داده‌ی $D = HHH$, باور ما نسبت به اینکه سکه‌ی انتخاب‌شده سالم (A) باشد چگونه به‌روزرسانی (update) می‌شود؟
با قاعده‌ی بیز:

$$P(A \mid D) = \frac{P(D \mid A) P(A)}{P(D)} = \frac{P(D \mid A) P(A)}{P(D \mid A) P(A) + P(D \mid B) P(B)}.$$

با جای‌گذاری مقادیر:

$$P(A \mid D) = \frac{\left(\frac{1}{2}\right)^3 \times \frac{1}{2}}{\left(\frac{1}{2}\right)^3 \times \frac{1}{2} + \left(\frac{3}{4}\right)^3 \times \frac{1}{2}} \approx 0.23.$$

یعنی باور ما پیش از مشاهده‌ی داده (پیشین) برابر ۰.۵ بود و پس از مشاهده‌ی HHH (که با سکه‌ی ناسالم سازگارتر است) به حدود ۰.۲۳ کاهش یافت.

نکته: در این محاسبه، عملاً $P(D)$ (یعنی احتمال کل مشاهده‌ی HHH ، بدون توجه به اینکه کدام سکه انتخاب شده) نیز به صورت مجموع وزن دار روی هر دو فرضیه (A و B) محاسبه شد. در حالت کلی، محاسبه‌ی چنین توزیع حاشیه‌ای کار سختی است و فرم بسته (closed form) برای آن تنها در مثال‌های ساده‌ی خاص به دست می‌آید.

۵ تعریف odds و قاعده‌ی ضرب (Chain Rule)

تعریف odds یک پیشامد

برای پیشامد A ، odds آن به صورت زیر تعریف می‌شود:

$$\text{odds}(A) = \frac{P(A)}{P(A^c)}.$$

برای مثال اگر $P(A) = \frac{2}{3}$ آنگاه $\text{odds}(A) = 2$ (یا به عبارت دیگر «۲ به ۱ به نفع A »، معادل «۱ به ۲ به ضرر A »):

odds in favor of Event A are 2 to 1, 1 to 2 odds against A .

رابطه‌ی برعکس، یعنی به دست آوردن احتمال از روی odds:

$$P(A) = \frac{\text{odds}(A)}{1 + \text{odds}(A)}.$$

قاعده‌ی ضرب (چند رویداد)

احتمال اشتراک دو رویداد را می‌توان با استفاده از احتمال غیرشرطی و احتمال شرطی به دو صورت متقارن نوشت:

$$P(A, B) = P(A \cap B) = P(B) P(A | B) = P(A) P(B | A).$$

(نکته: چون دو رویداد داریم، $2! = 2$ ترتیب متفاوت برای بسط وجود دارد.)

به همین ترتیب، احتمال اشتراک n رویداد (قاعده‌ی زنجیره‌ای) به صورت زیر تعمیم می‌یابد:

$$P(A_1, A_2, \dots, A_n) = P(A_1) P(A_2 | A_1) P(A_3 | A_1, A_2) \cdots P(A_n | A_1, \dots, A_{n-1}),$$

که می‌توان آن را به $n!$ ترتیب مختلف نوشت. برای مثال با $n = 3$ رویداد، به $3! = 6$ صورت مختلف هم‌ارز می‌توان بسط داد:

$$P(A_1, A_2, A_3) = P(A_1)P(A_2 | A_1)P(A_3 | A_1, A_2) = P(A_2)P(A_3 | A_2)P(A_1 | A_2, A_3) = \cdots$$

۶ قاعده‌ی بیز به فرم odds و فرم کلی بیزی

قاعده‌ی بیز به شکل odds

اگر A و B پیشامدهایی با احتمال مثبت باشند ($P(\cdot) > 0$)، می‌توان odds پیشامد A به شرط B را چنین نوشت:

$$\underbrace{\frac{P(A | B)}{P(A^c | B)}}_{\text{posterior odds (پسین)}} = \underbrace{\frac{P(B | A)}{P(B | A^c)}}_{\text{likelihood ratio (نسبت درست‌نمایی)}} \times \underbrace{\frac{P(A)}{P(A^c)}}_{\text{prior odds (پیشین)}}.$$

یعنی: odds پسین = نسبت درست‌نمایی $\times$ odds پیشین.

فرم کلی قاعده‌ی بیز برای پارامتر

فرض کنید D داده‌های مشاهده‌شده و θ پارامتری باشد که می‌خواهیم برآورد کنیم. قاعده‌ی بیز به‌صورت زیر است:

$$P(\theta | D) = \frac{P(D | \theta) P(\theta)}{P(D)},$$

که در آن:

$P(\theta | D)$: پسین (posterior) - باور به‌روزشده درباره‌ی θ پس از دیدن داده؛

$P(\theta)$: پیشین (prior) - باور اولیه درباره‌ی θ پیش از دیدن داده؛

$P(D | \theta)$: درست‌نمایی (likelihood)؛

$P(D)$: حاشیه‌ای/ marginal (evidence) - احتمال کل مشاهده‌ی داده، صرف‌نظر از θ .

توزیع مشترک از حاصل‌ضرب درست‌نمایی و پیشین به دست می‌آید:

$$P(D | \theta) P(\theta) = P(D, \theta),$$

و مقدار حاشیه‌ای $P(D)$ با انتگرال‌گیری روی همه‌ی مقادیر ممکن θ حاصل می‌شود:

$$P(D) = \int_{\theta} P(D, \theta) d\theta.$$

نکته: هرچه درباره‌ی θ نامعین‌تر باشیم (یعنی θ ابعاد بیشتری داشته باشد)، محاسبه‌ی این انتگرال دشوارتر و پرهزینه‌تر می‌شود؛ در یادگیری ماشین بیزی معمولاً θ چندبُعدی است و محاسبه‌ی دقیق $P(D)$ (و نه فقط $\log P(D)$ که آن را *evidence* می‌نامند) عملاً بسیار پرهزینه یا غیرممکن است. فرم بسته‌ی (closed form) این محاسبات فقط در موارد خاص (مانند مثال بعدی) به‌سادگی امکان‌پذیر است.

problem – 4 – 1 until problem – 4 – 3

۱٪ از مردم دچار بیماری « w » می شوند. یک نفر بنام جورج مشکوک به این بیماری است، آزمایش او مثبت است یعنی جورج بیمار است. پیشامدهای زیر را در نظر بگیرید.

T = جورج بیمار است

D = آزمایش او مثبت است

فرض کنید صحت آزمایش = ۹۵٪ باشد. ”۹۵٪ *accurate*” معیارهای مختلفی برای اندازه گیری صحت داریم در این مثال فرض کنید صحت به صورت زیر تعریف شده

$$\text{Sensitivity} = \text{True positive Rate} \rightarrow P(D | T) = 0.95$$

$$\text{Specificity} = \text{True Negative Rate} \rightarrow P(D^c | T^c) = 0.95$$

احتمال اینکه جورج بیمار باشد را با توجه به نتیجه مشاهده شده (*observed data*) یا همون آزمایش مثبت جورج بدست آوریم:
احتمال پسین:

$$P(T | D) = \frac{P(D | T)P(T)}{P(D)}$$

$$P(T | D) = \frac{P(D | T)P(T)}{P(D | T)P(T) + P(D | T^c)P(T^c)} = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99} \approx 0.16$$

۱۶٪ احتمال اینکه جورج بیمار باشد، که خیلی خلاف شهود ما است، (آزمایش که ۹۵٪ موارد صحیح انجام شده). برای درک این رفتار خلاف شهود یک مثال می زنیم.

فرض کنید جامعه ما ۱۰۰۰۰ نفر است، ۱٪ از آن مریض است = ۱۰۰ نفر. اگر همه از این جامعه آزمایش بدهند، توقع داریم ۱۰۰ نفر مریض مشاهده کنیم

که با توجه به "صحت ۹۵٪" آزمایش، ۹۵ تست مثبت، ۵ تست منفی داریم. در ضمن ۹۹۰۰ نفر هم سالم هستند ولی با توجه به "صحت ۹۵٪" آزمایش،

$$0.95 \times 9900 \approx 9405$$

تست منفی و

$$0.05 \times 9900 \approx 495$$

تست مثبت داریم.

اگر روی تست مثبت تمرکز کنیم

$$TP \leftarrow 95 = \text{۹۵ نفر از گروه مریض است}$$

$$FP \rightarrow 495 = \text{۴۹۵ نفر از گروه سالم است}$$

یعنی اکثر کسانی که جواب آزمایش آنها مثبت است در واقع بیمار نیستند. که این ذهن ما را در مورد اینکه چرا جواب ۱۶٪ خلاف شهود است روشن می کند.

به طور خلاصه ۲ عامل ما را به عدد خلاف شهود ۱۶٪ می رساند:

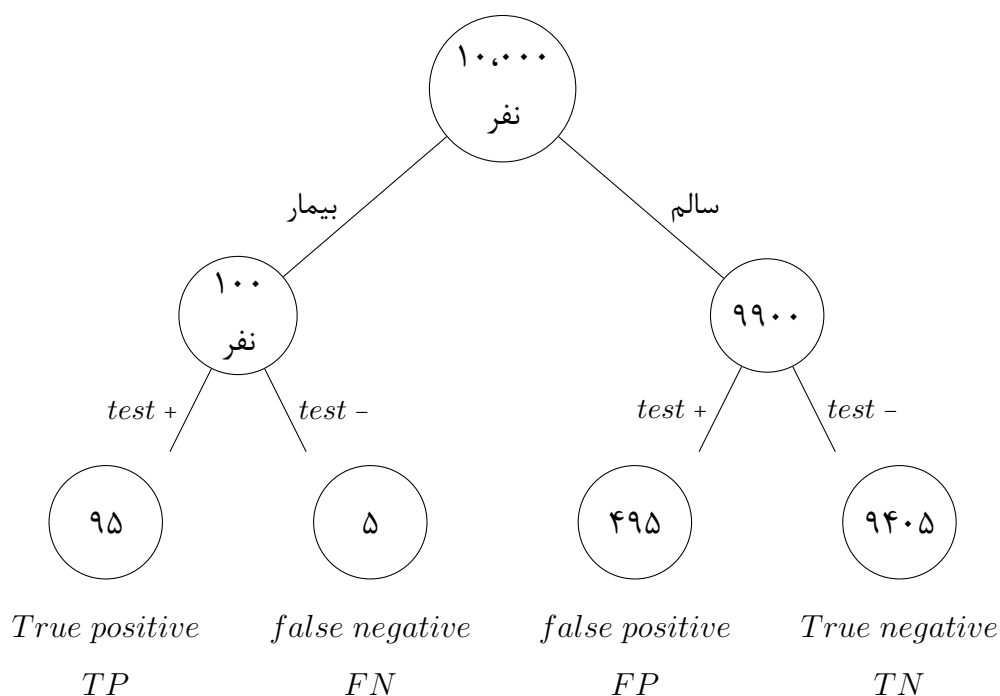
evidence from the test

our prior information about the prevalence of the disease

(اگر چه نتیجه آزمایش به نفع بیمار است ولی توجه داریم که این بیماری نادر است)

$$P(D | T)$$

reflects a balance between these 2 factors appropriately weighting the rarity of disease against the rarity of a mistaken test result.



conditional independence-1 ۷

① Events
Random variables
Random vectors
در آن بررسی

استقلال شرطی : Condo Ind.

شرایطی را می‌توانیم بگوییم که یک مشاهده (observation) در آن یقین (certainty) یک فرض (hypothesis) (تکذیب‌ناپذیریت) به عنوان یک حالت خاص فوق : احتمال شرطی یک فرض مشروط به مشاهده می‌اندر (صداها اطلاعات در غیر مفید) معادل احتمال بدون آن اطلاعات غیر مفید است. مثلاً $A = \text{فرض}$ و $B, C = \text{مشاهدات}$

استقلال شرطی به صورت زیر تعریف می‌شود:

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

احتمال A به شرط هر دو مشاهده B, C

در فوق A اثر B هیچ اثر در میان احتمال A به شرط C ندارد

B contributes nothing to the certainty of A

در اینجا می‌توانیم بگوییم که A و C مستقل شرطی هستند و B تاثیری ندارد

در این حالت سببیت (causality) تابع $f(y)$ و $g(y)$ که به هم پیوسته و وابسته هستند، تعریف می‌شود

$p(f|g, y) = p(f|y)$

در اینجا $f(y)$ بیان استقلال شرطی استفاده می‌کند و معادل است با:

شکل ۱: تصویر تخته مربوط به conditional independence-1

در این بخش موضوع استقلال شرطی برای پیشامدها مطرح شده است. ایده اصلی این است که یک مشاهده می‌تواند روی میزان یقین ما نسبت به یک فرض اثر بگذارد. روی تخته برای این سه مفهوم از عبارت های observation, certainty و hypothesis استفاده شده است.

فرض می‌کنیم A پیشامدی باشد که نقش فرض اصلی را دارد و B و C دو مشاهده باشند. اگر بعد از دانستن C ، اضافه شدن اطلاعات B هیچ تغییری در احتمال A ایجاد نکند، می‌نویسیم:

$$P(A | B, C) = P(A | C).$$

معنای این رابطه این است که وقتی C را می‌دانیم، مشاهده B دیگر اطلاعات تازه‌ای درباره رخ دادن A به ما نمی‌دهد. روی تخته این مفهوم با جمله زیر بیان شده است:

B contributes nothing to the certainty of A.

در این حالت گفته می‌شود که A و B به شرط C مستقل هستند و آن را به صورت زیر نشان می‌دهیم:

$$A \perp B | C.$$

پس این نماد به معنی استقلال معمولی A و B نیست. ممکن است A و B در حالت کلی وابسته باشند، ولی وقتی C معلوم شد، وابستگی بین آن‌ها از بین برود. به طور خلاصه، اگر داشته باشیم:

$$A \perp B | C,$$

آنگاه:

$$P(A | B, C) = P(A | C).$$

به صورت متقارن نیز می‌توان نوشت:

$$P(B | A, C) = P(B | C).$$

در انتهای تخته به ارتباط استقلال شرطی با بحث علیت یا causality اشاره شده است. برای مثال ممکن است دو کمیت $f(y)$ و $g(y)$ هر دو تحت تاثیر متغیر y باشند و به همین دلیل بدون دانستن y وابسته دیده شوند. اگر بعد از معلوم بودن y ، دانستن $g(y)$ اطلاعات اضافه‌ای درباره $f(y)$ ندهد، می‌توان نوشت:

$$P(f | g, y) = P(f | y).$$

این دقیقاً همان ایده استقلال شرطی است؛ یعنی متغیر شرطی می‌تواند علت مشترک یا اطلاعات مشترکی را که باعث وابستگی ظاهری شده است، مشخص کند.

۸ conditional independence-2

④ سرچرا: استقلال می‌دانستیم شرطی: بین آنها تا نظر بر قرار می‌کنیم.

اندازه احتمال P ، مجموعه پیشامد $\mathcal{I}$ ، فضای نمونه Ω ، $(\Omega, \mathcal{I}, P)$ = تعریف احتمال اصلی

اندازه احتمال جدید $q = P(\cdot | E)$ ، مجموعه پیشامد $\mathcal{I} \cap E$ ، فضای نمونه E ، $(E, \mathcal{I} \cap E, q)$ = تعریف احتمال شرطی محدود به پیشامد E

بیان زیر می‌گوید که هم اصل بر تعریف احتمال اصلی و هم اصل بر تعریف احتمال جدید برقرار است و راه آن:

آزمون پیشامد A ، B را در نظر می‌گیریم: A, B به شرط E مستقل شرطی هستند اگر:

بیان اول در درم صدق داشته باشد.

؟ $P(A, B | E) = P(A | E) P(B | E)$

iff $\frac{P(A, B, E)}{P(E)} = \frac{P(A, E)}{P(E)} \times \frac{P(B, E)}{P(E)}$ تعریف احتمال شرطی

iff $P(A, B, E) = \frac{P(A, E) P(B, E)}{P(E)}$ ضرب $P(E)$ طرف دو طرف

iff $\frac{P(A, B, E)}{P(B, E)} = \frac{P(A, E)}{P(E)}$ تقسیم $P(B, E)$ طرف دو طرف

iff $P(A | B, E) = P(A | E)$ تعریف احتمال شرطی

شکل ۲: تصویر تخته مربوط به conditional independence-2

در این تصویر، استقلال شرطی از دید فضای احتمال بررسی شده است. فضای احتمال اولیه را به صورت زیر در نظر می‌گیریم:

$$(\Omega, \mathcal{I}, P),$$

که در آن Ω فضای نمونه، $\mathcal{I}$ مجموعه پیشامدهای قابل بررسی و P تابع احتمال است. حالا فرض می‌کنیم پیشامد E رخ داده است و $P(E) > 0$. بعد از شرط گذاشتن روی E ، فضای نمونه عملاً از Ω به E محدود می‌شود. مجموعه پیشامدهای جدید را می‌توان به صورت زیر نوشت:

$$\mathcal{I}_E = \{A \cap E \mid A \in \mathcal{I}\},$$

و تابع احتمال جدید نیز برابر است با:

$$q(A) = P(A | E).$$

بنابراین فضای احتمال شرطی را می‌توان به صورت زیر دید:

$$(E, \mathcal{I}_E, q).$$

اگر بخواهیم استقلال A و B را در این فضای شرطی بررسی کنیم، باید همان رابطه استقلال معمولی را با احتمال‌های شرطی بنویسیم:

$$P(A \cap B | E) = P(A | E)P(B | E).$$

این رابطه یکی از بیان‌های اصلی استقلال شرطی است. در نتیجه می‌نویسیم:

$$A \perp B | E.$$

روی تخته بیان دوم استقلال شرطی نیز نوشته شده است:

$$P(A | B, E) = P(A | E).$$

در ادامه نشان داده شده که این دو بیان معادل هستند. از رابطه ضربی شروع می‌کنیم:

$$P(A, B | E) = P(A | E)P(B | E).$$

با استفاده از تعریف احتمال شرطی داریم:

$$\frac{P(A, B, E)}{P(E)} = \frac{P(A, E)}{P(E)} \frac{P(B, E)}{P(E)}.$$

با ضرب دو طرف در $P(E)$ نتیجه می‌شود:

$$P(A, B, E) = \frac{P(A, E)P(B, E)}{P(E)}.$$

اگر $P(B, E) > 0$ باشد، دو طرف را بر $P(B, E)$ تقسیم می‌کنیم:

$$\frac{P(A, B, E)}{P(B, E)} = \frac{P(A, E)}{P(E)}.$$

سمت چپ برابر $P(A | B, E)$ و سمت راست برابر $P(A | E)$ است. پس:

$$P(A | B, E) = P(A | E).$$

بنابراین دو رابطه زیر یک مفهوم را بیان می‌کنند:

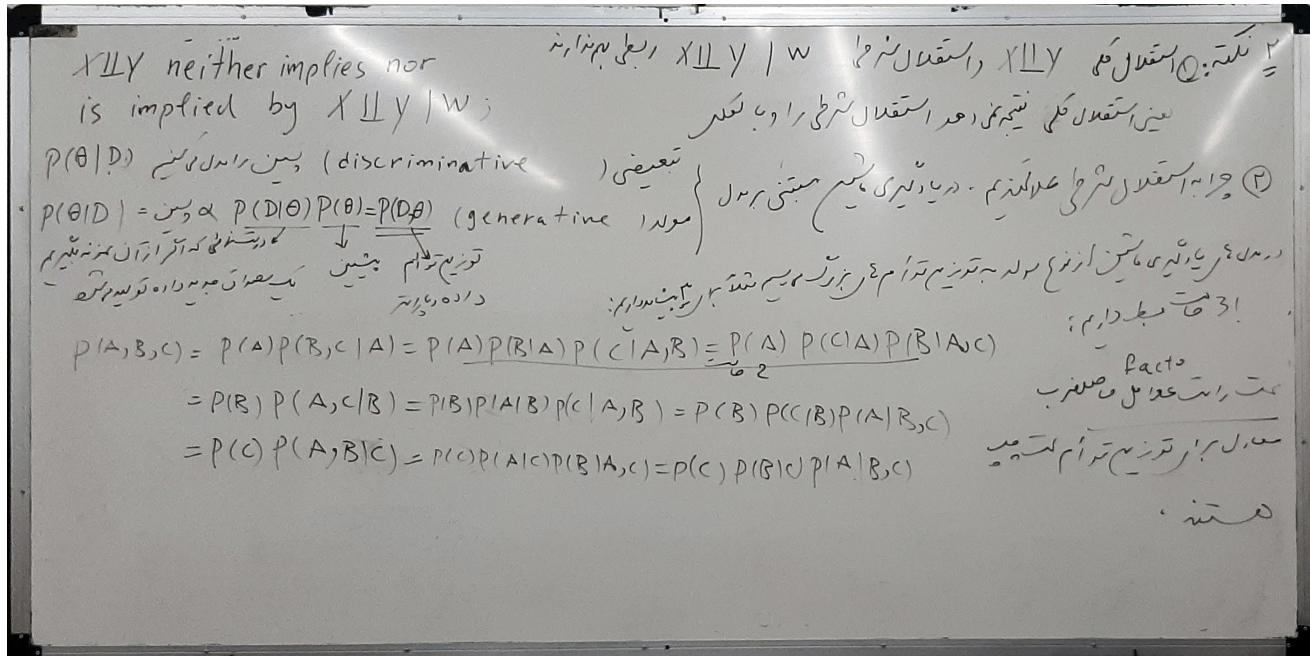
$$P(A \cap B | E) = P(A | E)P(B | E),$$

و:

$$P(A | B, E) = P(A | E).$$

رابطه اول شکل ضربی استقلال شرطی است و رابطه دوم می‌گوید بعد از دانستن E ، اطلاع از B احتمال A را تغییر نمی‌دهد.

conditional independence-3 9



شکل ۳: تصویر تخته مربوط به conditional independence-3

اولین نکته مهم این تصویر این است که استقلال معمولی و استقلال شرطی هیچ کدام لزوماً دیگری را نتیجه نمی دهند. روی تخته این موضوع به صورت زیر نوشته شده است:

$X \perp Y$ neither implies nor is implied by $X \perp Y \mid W$.

یعنی ممکن است X و Y مستقل باشند ولی با شرط گذاشتن روی W وابسته شوند. حالت برعکس هم ممکن است رخ دهد؛ یعنی X و Y در حالت کلی وابسته باشند ولی با معلوم شدن W مستقل شوند.

در ادامه به دو نوع مدل آماری اشاره شده است. در مدل discriminative مستقیماً توزیع شرطی پارامتر یا خروجی با داشتن داده مدل می شود:

$$P(\theta \mid D).$$

اما در مدل generative معمولاً توزیع مشترک داده و پارامتر ساخته می شود. با استفاده از رابطه بیز داریم:

$$P(\theta \mid D) \propto P(D \mid \theta)P(\theta) = P(D, \theta).$$

در این رابطه $P(D | \theta)$ نقش likelihood، عبارت $P(\theta)$ نقش prior و $P(D, \theta)$ نقش توزیع مشترک را دارد. ایده مدل مولد این است که مشخص کند داده چگونه از روی پارامترها تولید می شود.

بخش بعدی تخته مربوط به قاعده زنجیره ای برای سه پیشامد یا سه متغیر است. توزیع مشترک A ، B و C را می توان با ترتیب های مختلف تجزیه کرد. برای شروع از A داریم:

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

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

یا با جابه جا کردن ترتیب B و C :

$$P(A, B, C) = P(A)P(C | A)P(B | A, C).$$

اگر از B شروع کنیم:

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

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

و شکل دیگر آن:

$$P(A, B, C) = P(B)P(C | B)P(A | B, C).$$

اگر از C شروع کنیم:

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

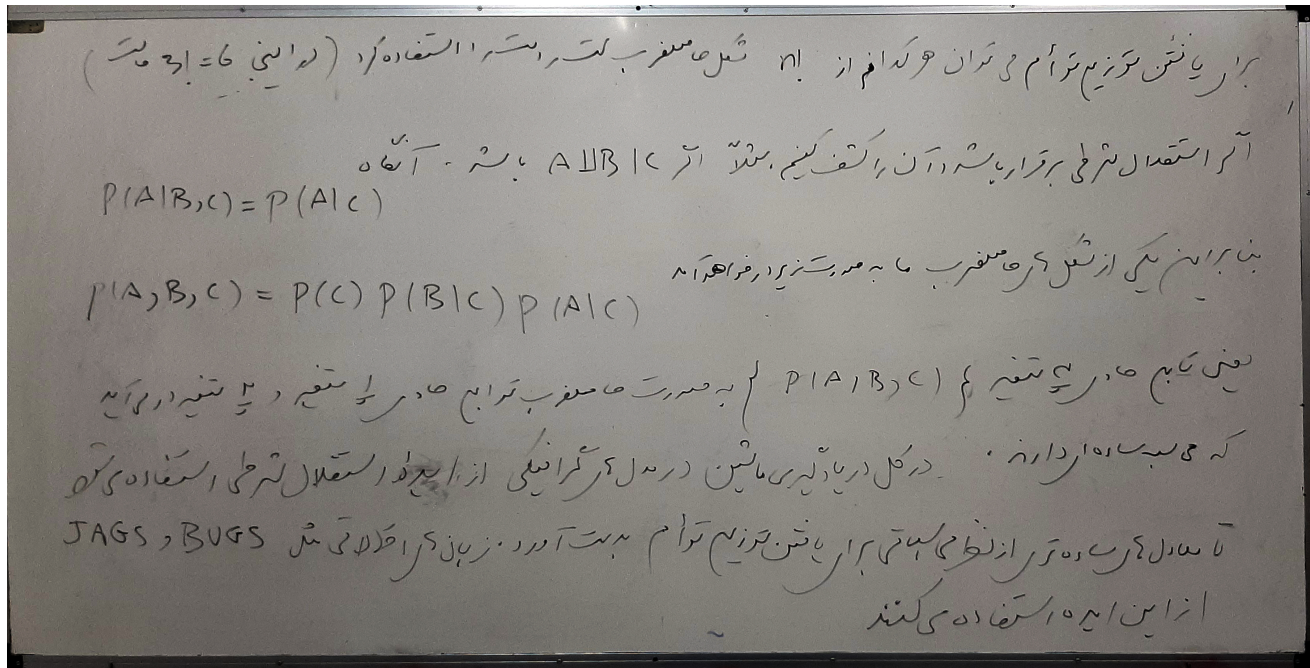
$$= P(C)P(A | C)P(B | A, C),$$

یا:

$$P(A, B, C) = P(C)P(B | C)P(A | B, C).$$

پس برای سه متغیر، شش ترتیب مختلف وجود دارد که همان 3! حالت ممکن است. همه این تجزیه ها از نظر مقدار با هم برابر هستند، ولی اگر بین بعضی متغیرها استقلال شرطی وجود داشته باشد، بعضی از این شکل ها ساده تر می شوند. دلیل اهمیت استقلال شرطی هم همین است، چون می تواند یک توزیع مشترک پیچیده را به چند عامل ساده تر تبدیل کند.

۱۰ conditional independence-4



شکل ۴: تصویر تخته مربوط به conditional independence-4

در ادامه بحث تجزیه توزیع مشترک، روی تخته اشاره شده است که برای n متغیر می‌توان از $n!$ ترتیب مختلف در قاعده زنجیره ای استفاده کرد. برای سه متغیر، این تعداد برابر است با:

$$3! = 6.$$

هر ترتیب یک فاکتورگیری متفاوت از توزیع مشترک می‌دهد، ولی همه آن‌ها یک توزیع را نمایش می‌دهند. حالا فرض می‌کنیم استقلال شرطی زیر برقرار باشد:

$$A \perp B | C.$$

طبق تعریف استقلال شرطی داریم:

$$P(A | B, C) = P(A | C).$$

یکی از تجزیه‌های قاعده زنجیره ای به صورت زیر است:

$$P(A, B, C) = P(C)P(B | C)P(A | B, C).$$

با استفاده از استقلال شرطی، عامل آخر ساده می شود و نتیجه می گیریم:

$$P(A, B, C) = P(C)P(B | C)P(A | C).$$

پس دانستن رابطه استقلال شرطی باعث می شود یک عامل که به دو متغیر شرطی وابسته بود، فقط به یک متغیر وابسته باشد. این موضوع هم فرمول را کوتاه تر می کند و هم تعداد احتمال ها یا پارامترهایی را که باید تعیین شوند کاهش می دهد.

برای مثال، اگر متغیرها گسسته باشند، تعیین کامل $P(A | B, C)$ معمولاً به مقدارهای مختلف هر دو متغیر B و C نیاز دارد. اما وقتی:

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

دیگر لازم نیست برای هر ترکیب از B و C یک احتمال جداگانه داشته باشیم. در نتیجه نمایش مدل فشرده تر می شود.

روی تخته به استفاده از این ایده در مدل های گرافیکی احتمالی و شبکه های بیزی یا Bayesian networks نیز اشاره شده است. در این مدل ها ساختار گراف مشخص می کند که کدام متغیرها با شرط روی متغیرهای دیگر مستقل هستند. سپس توزیع مشترک به حاصل ضرب چند توزیع شرطی کوچک تر تبدیل می شود. همچنین نام زبان ها یا ابزارهای BUGS و JAGS روی تخته آمده است. این ابزارها برای تعریف و محاسبه مدل های احتمالی بزرگ استفاده می شوند. دلیل امکان کار با چنین مدل هایی این است که با استفاده از استقلال شرطی، توزیع مشترک بزرگ به چند بخش ساده تر شکسته می شود و لازم نیست تمام حالت های توزیع مشترک به صورت مستقیم نوشته شوند.

۱۱ conditional independence-5

استقلال شرطی $X \perp\!\!\!\perp Y$ neither implies nor is implied by $X \perp\!\!\!\perp Y | W$ (استقلال کلی - طبیعی)

مثال: مستقل شرطی ولی درجه‌ای مستقل نیست: فرض کنید W مقدار ۰ یا ۱ را با احتمال ۰.۵ اختیار می‌کند

وقتی $W=0$ است X و Y مقدار ۰ یا ۱ را با احتمال ۰.۹۹ اختیار می‌کنند

وقتی $W=1$ است X و Y مقدار ۰ یا ۱ را با احتمال ۰.۰۱ اختیار می‌کنند

در اینجا $P(X=0)=0.5$ ولی $P(X=0|Y=0) > 0.5$ پس این دلیل می‌کند که اگر $Y=0$ باشد ضعیف‌تر احتمال می‌دهیم که $W=0$ باشد و در نتیجه برای حالت $W=0$ ضعیف‌تر احتمال می‌دهیم که $X=0$ باشد بنابراین $P(X=0|Y=0) \neq P(X=0)$ که یعنی $(*)$ برقرار نیست

ولی $X \perp\!\!\!\perp Y | W$ برقرار است چون وقتی W مقدار ۰ یا ۱ را با احتمال ۰.۵ اختیار می‌کند، $W=0$ یا $W=1$ پس از آن X و Y به طور مستقل رفتار می‌کنند

مثال ۲: برای استقلال شرطی ولی درجه‌ای مستقل نباشد. یک سکه ناهم‌دار داریم که ادس با احتمال ۰.۹ و H با احتمال ۰.۱ خط (T) در می‌آید و برعکس است. آیا انداختن سکه مستقل هستند جواب: خیر، اما احتمال زیر را محاسبه کنید تا در درجه ۱۱ مطمئن شوید $P(H_{11}) = P(Toss_{11} = H)$

نامبر تاکنن H یا H برده دو دفعه ۱۱ سکه پیاپی $\oplus \oplus P(Toss_{11} = H | First 10 is H)$

طبق احتمال $P(Toss_{11} = H) = P(A)P(H_{11}|A) + P(B)P(H_{11}|B) = 0.5 \times 0.9 + 0.5 \times 0.1 = 0.5$

توجه: $P(A,B) = P(A|B)P(B)$ استقلال کلی
 $P(A,B) = P(A)P(B)$ استقلال شرطی
 $P(A|B) = P(B)$ $(*)$

شکل ۵: تصویر تخته مربوط به conditional independence-5

در ابتدای این تصویر دوباره تفاوت استقلال معمولی و استقلال شرطی یادآوری شده است. برای دو پیشامد A و B همیشه داریم:

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

اگر A و B مستقل باشند، رابطه زیر برقرار است:

$$P(A, B) = P(A)P(B).$$

بنابراین، در صورت مثبت بودن $P(B)$ ، استقلال با رابطه زیر معادل است:

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

مثال اول: استقلال شرطی ولی وابستگی در حالت کلی

فرض می‌کنیم متغیر W مقدارهای صفر و یک را با احتمال برابر بگیرد:

$$P(W = 0) = P(W = 1) = \frac{1}{2}.$$

به شرط $W = 0$ ، متغیرهای X و Y مستقل از هم هستند و هر کدام مقدار صفر را با احتمال 0.99 و مقدار یک را با احتمال 0.01 می‌گیرند:

$$P(X = 0 \mid W = 0) = P(Y = 0 \mid W = 0) = 0.99.$$

به شرط $W = 1$ نیز X و Y مستقل هستند، ولی احتمال‌ها برعکس می‌شوند:

$$P(X = 0 \mid W = 1) = P(Y = 0 \mid W = 1) = 0.01.$$

پس در هر مقدار مشخص از W داریم:

$$X \perp Y \mid W.$$

حالا استقلال کلی X و Y را بررسی می‌کنیم. ابتدا:

$$P(X = 0) = P(W = 0)P(X = 0 \mid W = 0)$$

$$+ P(W = 1)P(X = 0 \mid W = 1)$$

$$= \frac{1}{2}(0.99) + \frac{1}{2}(0.01)$$

$$= 0.5.$$

اما وقتی می‌دانیم $Y = 0$ ، احتمال این که $W = 0$ بوده باشد زیاد می‌شود و در نتیجه احتمال $X = 0$ نیز بالا می‌رود. به طور دقیق:

$$P(X = 0, Y = 0) = \frac{1}{2}(0.99)^2 + \frac{1}{2}(0.01)^2 = 0.4901.$$

از طرفی:

$$P(Y = 0) = 0.5.$$

پس:

$$P(X = 0 \mid Y = 0) = \frac{0.4901}{0.5} = 0.9802.$$

بنابراین:

$$P(X = 0 \mid Y = 0) > P(X = 0),$$

و نتیجه می‌شود:

$$X \not\perp Y.$$

در این مثال، W یک علت یا عامل پنهان مشترک است. بعد از معلوم شدن W ، X و Y مستقل هستند، ولی اگر W را ندانیم، مشاهده Y درباره X اطلاعات می دهد.

مثال دوم: انتخاب یکی از دو سکه

دو سکه A و B در نظر می گیریم. برای سکه A احتمال شیر برابر 0.9 و احتمال خط برابر 0.1 است. برای سکه B این احتمال ها برعکس هستند:

$$P(H | A) = 0.9, \quad P(H | B) = 0.1.$$

یکی از دو سکه با احتمال برابر انتخاب می شود و همان سکه چند بار پرتاب می شود:

$$P(A) = P(B) = 0.5.$$

اگر H_i پیشامد شیر آمدن در پرتاب i ام باشد، برای هر پرتاب به تنهایی داریم:

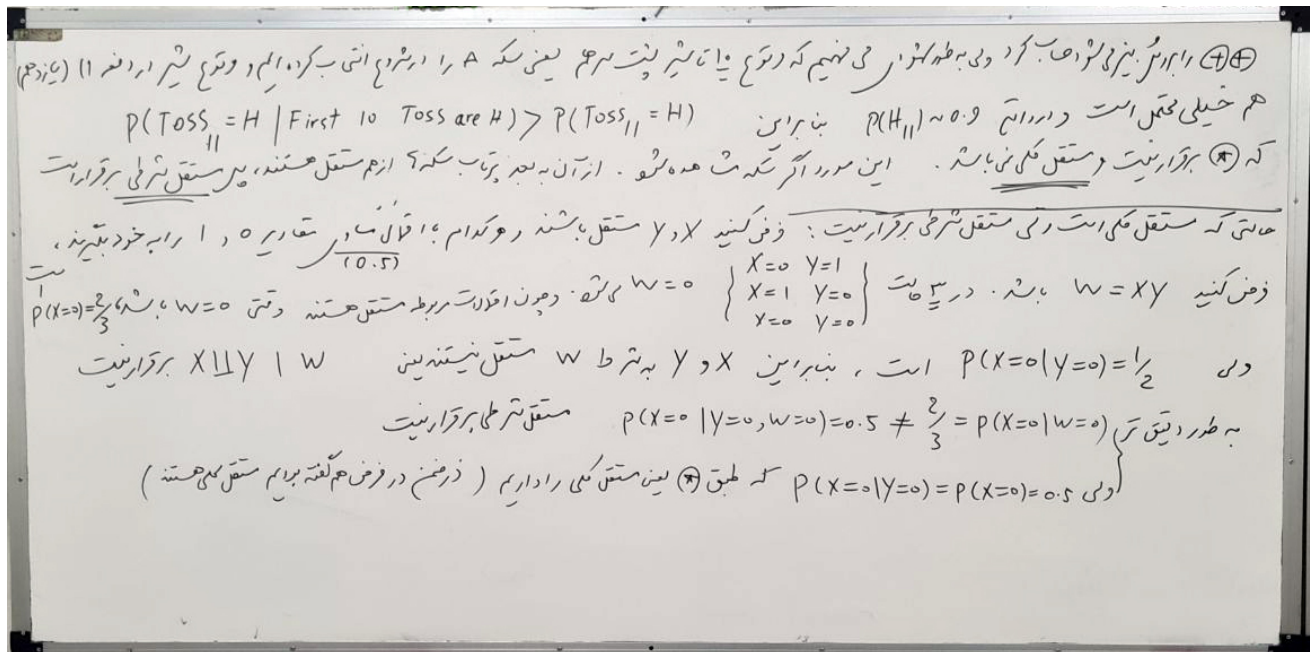
$$P(H_i) = P(A)P(H_i | A) + P(B)P(H_i | B)$$

$$= 0.5(0.9) + 0.5(0.1)$$

$$= 0.5.$$

ادامه این مثال در تصویر بعد بررسی شده است. نکته اصلی این است که مشاهده نتیجه پرتاب های قبلی، درباره سکه انتخاب شده اطلاعات می دهد و در نتیجه احتمال پرتاب های بعدی را تغییر می دهد.

۱۲ conditional independence-6



شکل ۶: تصویر تخته مربوط به conditional independence-6

در ابتدای این تصویر مثال دو سکه ادامه داده شده است. فرض می‌کنیم ده پرتاب اول همگی شیر آمده باشند. در این حالت احتمال این که سکه انتخاب شده همان سکه A باشد، بسیار زیاد می‌شود:

$$P(A | H_1, \dots, H_{10}) = \frac{(0.9)^{10}}{(0.9)^{10} + (0.1)^{10}}.$$

این مقدار تقریباً برابر یک است. پس احتمال شیر آمدن در پرتاب یازدهم برابر است با:

$$P(H_{11} | H_1, \dots, H_{10}) = 0.9P(A | H_1, \dots, H_{10})$$

$$+ 0.1P(B | H_1, \dots, H_{10}),$$

که مقدار آن تقریباً ۰.۹ می‌شود. بنابراین:

$$P(H_{11} | H_1, \dots, H_{10}) > P(H_{11}) = 0.5.$$

پس پرتاب‌ها در حالت کلی مستقل نیستند، چون نتیجه پرتاب‌های قبلی احتمال پرتاب بعدی را تغییر می‌دهد. با این حال، اگر بدانیم کدام سکه انتخاب شده است، پرتاب‌ها مستقل هستند. در نتیجه انتخاب سکه نقش متغیر شرطی را دارد و می‌توان گفت پرتاب‌ها به شرط نوع سکه مستقل هستند.

مثال سوم: استقلال کلی ولی نبود استقلال شرطی

در این مثال X و Y دو متغیر برنولی مستقل با پارامتر 0.5 هستند:

$$P(X = 0) = P(X = 1) = \frac{1}{2},$$

و:

$$P(Y = 0) = P(Y = 1) = \frac{1}{2}.$$

پس در حالت کلی:

$$X \perp Y.$$

برای مثال:

$$P(X = 0 | Y = 0) = P(X = 0) = 0.5.$$

حالا متغیر جدید زیر را تعریف می کنیم:

$$W = XY.$$

اگر $W = 0$ باشد، سه حالت ممکن برای (X, Y) وجود دارد:

$$(0, 0), \quad (0, 1), \quad (1, 0).$$

چون هر چهار حالت اولیه (X, Y) احتمال برابر دارند، بعد از شرط گذاشتن روی $W = 0$ ، این سه حالت نیز احتمال برابر دارند. در نتیجه:

$$P(X = 0 | W = 0) = \frac{2}{3}.$$

حالا علاوه بر $W = 0$ ، شرط $Y = 0$ را نیز اضافه می کنیم. در این حالت فقط دو زوج زیر ممکن هستند:

$$(0, 0), \quad (1, 0).$$

پس:

$$P(X = 0 | Y = 0, W = 0) = \frac{1}{2}.$$

این دو احتمال برابر نیستند:

$$P(X = 0 | Y = 0, W = 0) = \frac{1}{2} \neq \frac{2}{3} = P(X = 0 | W = 0).$$

بنابراین:

$$X \not\perp Y | W.$$

در این مثال X و Y در حالت کلی مستقل هستند، اما شرط گذاشتن روی حاصل ضرب آن‌ها باعث ایجاد وابستگی می‌شود. در نتیجه دو مثال آخر با هم نشان می‌دهند که هیچ کدام از دو رابطه زیر، رابطه دیگر را نتیجه نمی‌دهد:

$$X \perp Y \quad \text{و} \quad X \perp Y \mid W.$$

7 – conditional independence

F = اندازه پا

L = وسعت لغاتی که هر شخص می‌داند

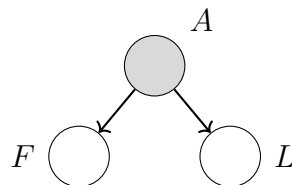
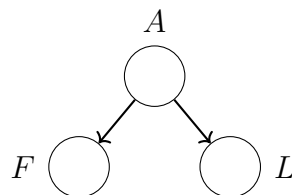
این دو به هم همبستگی دارند، یعنی اندازه پای کودک، کوچک است و دامنه لغات او محدود است و شخص بزرگسال که طبیعتاً در طول زمان اندازه پا رشد کرده، دامنه لغات او وسیع‌تر است پس F و L مستقل نیستند. اندازه پا، نشان از سن افراد دارد. پیشامد

A = سن یک نفر

بنابراین A ، سبب F و L است. (Cause) بنابراین

$$P(L \mid F) \neq P(L)$$

ولی اگر سن فرد را بدانیم (observe) که در ادبیات یادگیری ماشین یعنی «مشاهده کردیم» که در نمودار گرافیکی به صورت زیر است



اکنون ارتباطی بین اندازه پا (F) و دانسته لغات (L) نداریم و مستقل هستند یعنی

$$P(L | F, A) = P(L | A)$$

به این دلیل که در افراد با محدوده سنی مشابه، افراد با سبزه‌های مختلف داریم و دیگر اندازه پا دلیلی بر وسعت دامنه لغات فرد نمی باشد

۴. مسئله ی تطابق دِ مون مور – بخش اول

روی n کارت، شماره های 1 تا n نوشته شده است. کارت ها را به طور کامل بر می زنیم و سپس اعداد 1 تا n را به ترتیب اعلام می کنیم. هرگاه شماره ی کارت فراخوانده شده با عدد اعلام شده برابر باشد، تطابق رخ می دهد. سؤال، احتمال برنده شدن، یعنی احتمال رخ دادن حداقل یک تطابق است. پیشامد زیر را تعریف می کنیم:

$$A_i = \{ \text{باشد } i\text{ام، عدد } i \text{ در کارت فراخوانده شده ی } \}.$$

پس پیشامد حداقل یک تطابق برابر است با:

$$A_1 \cup A_2 \cup \dots \cup A_n.$$

برای محاسبه ی احتمال این پیشامد، از اصل شمول و عدم شمول استفاده می کنیم. برای یک کارت مشخص داریم:

$$\mathbb{P}(A_i) = \frac{1}{n}. \quad (1)$$

زیرا اگر کارت i ام در موقعیت i ام ثابت باشد، $(n-1)!$ حالت برای کارت های دیگر باقی می ماند و تعداد کل حالت ها $n!$ است؛ بنابراین:

$$\mathbb{P}(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

برای دو کارت مشخص i و j ، هر دو کارت را در موقعیت های خود ثابت می کنیم. در این صورت $(n-2)!$ حالت برای کارت های دیگر باقی می ماند؛ پس:

$$\mathbb{P}(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}. \quad (2)$$

برای سه کارت مشخص نیز به همین ترتیب داریم:

$$\mathbb{P}(A_i \cap A_j \cap A_k) = \frac{1}{n(n-1)(n-2)}.$$

در جمله ی اول اصل شمول و عدم شمول، $\binom{n}{1}$ انتخاب؛ در جمله ی دوم $\binom{n}{2}$ انتخاب؛ در جمله ی سوم $\binom{n}{3}$ انتخاب و به همین ترتیب تا n انتخاب وجود دارد.

۵. مسئله ی تطابق دِ مون مور – بخش دوم

با استفاده از اصل شمول و عدم شمول داریم:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{i=1}^n A_i\right) &= \binom{n}{1} \frac{1}{n} - \binom{n}{2} \frac{1}{n(n-1)} + \binom{n}{3} \frac{1}{n(n-1)(n-2)} \\ &\quad + \cdots + (-1)^{n+1} \frac{1}{n!}. \end{aligned}$$

بنابراین:

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1} \frac{1}{n!} \simeq 1 - \frac{1}{e}.$$

از بسط زیر استفاده شده است:

$$\frac{1}{e} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots = \frac{1}{2!} - \frac{1}{3!} + \cdots.$$

با رفتن $n \rightarrow \infty$ ، احتمال فوق تقریباً به $1 - \frac{1}{e}$ نزدیک می شود. در این سؤال، با داشتن تعداد زیادی کارت، اگر قرار باشد دست کم یک تطابق داشته باشیم، احتمال تقریباً $1 - \frac{1}{e}$ خواهد بود.

کد R

```

n <- 100                                # no. of cards

r <- replicate(
  10^4,
  sum(sample(n) == (1:n))              # shuffle; count matches
)

sum(r >= 1) / 10^4                      # proportion with a match

```

نمونه:

```
sum(sample(5) == (1:5))
```

سه حالت نمونه ی نوشته شده روی تخته:

$$(1, 3, 4, 2, 5) == (1, 2, 3, 4, 5) \implies (1, 0, 0, 0, 1) \implies 2,$$

$$(3, 1, 2, 4, 5) == (1, 2, 3, 4, 5) \implies (0, 0, 0, 1, 1) \implies 2,$$

$$(1, 3, 2, 4, 5) == (1, 2, 3, 4, 5) \implies (1, 0, 0, 1, 1) \implies 3.$$

نکته درباره ی نحو R:

```

sample(n, k) without replacement
sample(n) = sample(n, n) w/o replacement

```

دستور $\text{sample}(n, k)$ تعداد k نمونه ی تصادفی از اعداد 1 تا n ، بدون جایگذاری، تولید می کند. دستور $\text{sample}(n)$ نیز یک جایگشت تصادفی از اعداد 1 تا n است؛ تعداد این جایگشت های تصادفی $n!$ حالت است.

۶. استقلال شرطی به شرط E و E^c

$$E^c = \text{متمم } E.$$

فرض کنید دو نوع کلاس درس داریم:

$$G = \text{کلاس بد}, \quad G^c = \text{کلاس خوب}.$$

در کلاس خوب، اگر به طور جدی درس بخوانید، به طور حتمی نمره ی A می گیرید؛ ولی در کلاس بد، به طور شانسی هرکس می تواند نمره ی A بگیرد. پیشامدها را به صورت زیر تعریف می کنیم:

$$W = \{\text{به طور جدی درس خواندن}\},$$

$$A = \{\text{گرفته اید } A \text{ شما نمره ی } A\}.$$

در این مثال، W و A به شرط قرار گرفتن در کلاس بد از یکدیگر مستقل اند:

$$W \perp\!\!\!\perp A \mid G^c.$$

استقلال شرطی زیر برقرار نیست:

$$W \not\perp\!\!\!\perp A \mid G.$$

یعنی اگر به طور جدی درس بخوانید، شانس خوبی برای گرفتن نمره ی A وجود دارد. فرض کنید n نفر در یک اتاق هستند. احتمال آنکه لااقل ۲ نفر تاریخ تولد مشابه داشته باشند چقدر است؟ فرض: تولد در هر روز از ۳۶۵ روز سال احتمال برابر دارد. پیشامد A را به صورت زیر تعریف می کنیم:

$$A = \{\text{همه روزهای تولد متفاوت باشند}\}$$

فضای نمونه را S در نظر می گیریم. اگر روز تولد نفر i -ام را با x_i نشان دهیم، آنگاه:

$$S = \{(x_1, x_2, \dots, x_n) : x_i = 1, 2, \dots, 365\}$$

اندازه فضای نمونه برابر است با:

$$|S| = 365^n$$

جایگشت (Permutation) یعنی انتخاب‌های مرتب.

اندازه مجموعه A برابر است با تعداد جایگشت‌های n تایی از ۳۶۵ روز:

$$|A| = P_{365,n} = (365)(364) \dots (365 - n + 1)$$

فرمول بالا می‌گوید نفر اول در هر کدام از ۳۶۵ روز به دنیا می‌آید، نفر دوم در ۳۶۴ روز باقی‌مانده و به همین ترتیب. نکته: permutation با combination (ترکیب) فرق دارد.

پس:

$$\begin{aligned} Pr(A) &= \frac{|A|}{|S|} = \frac{365 \times 364 \times \dots \times (365 - n + 1)}{365 \times 365 \times \dots \times 365} \\ &= 1 \times \frac{364}{365} \times \frac{363}{365} \times \dots \times \frac{365 - n + 1}{365} \end{aligned}$$

توضیح:

□ نفر اول در هر کدام از ۳۶۵ روز می‌تواند به دنیا بیاید.

□ نفر دوم در ۳۶۴ روز باقی‌مانده به دنیا می‌آید و به همین ترتیب.

متمم پیشامد

هدف ما احتمال این است که حداقل دو نفر تاریخ تولد مشترک داشته باشند. پس از متمم استفاده می‌کنیم:

$$Pr(\bar{A}) = 1 - Pr(A)$$

بنابراین:

$$Pr(\bar{A}) = 1 - \left(1 \times \frac{364}{365} \times \cdots \times \frac{365 - n + 1}{365} \right)$$

نتایج مهم

اگر در اتاق $n = 366$ نفر باشند، حداقل ۲ نفر تاریخ تولد مشابه دارند:

$$n = 366 \implies Pr(\bar{A}) = 1$$

برای آنکه داشته باشیم:

$$Pr(\bar{A}) > \frac{1}{2}$$

باید:

$$n \geq 23$$

بنابراین:

$$n \geq 23 \implies Pr(\bar{A}) > \frac{1}{2}$$

و اگر $n = 50$ باشد:

$$Pr(\bar{A}) = 0.97$$

Chebychev's Inequality

نامساوی چبیشف

صورت قضیه

فرض کنید:

$$E(X) = \mu, \quad \text{Var}(X) = \sigma^2$$

آنگاه برای هر $\epsilon > 0$ داریم:

$$\Pr(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

اگر $\epsilon = k\sigma$ باشد، احتمال آنکه $k\sigma$ برابر از میانگین دور شویم را داریم:

$$\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

و نتیجه متمام:

$$\Rightarrow \Pr(|X - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2}$$

توضیح قضیه

چبیشف یک کران تقریبی برای آنکه احتمال انحراف X از میانگین خودش به میزان بیشتر از یک مقدار ثابت را بدست میدهد که بر حسب مقدار ثابت و واریانس محاسبه می‌شود. چبیشف همیشه قابل استفاده است (ولی تقریب بزرگی است)، به شرطی که واریانس موجود باشد.

مثال: توزیع یکنواخت

فرض کنیم:

$$X \sim U(0, 1)$$

یعنی برای $0 < x < 1$:

$$f(x) = 1$$

در حالت کلی اگر:

$$X \sim U(a, b)$$

آنگاه:

$$E(X) = \frac{a+b}{2}$$

$$\text{Var}(X) = \frac{(a-b)^2}{12}$$

با جایگذاری $a = 0$ و $b = 1$ داریم:

$$X \sim U(0, 1) \Rightarrow E(X) = \frac{1}{2}, \quad \text{Var}(X) = \frac{1}{12}$$

پس:

$$\sigma = \frac{1}{\sqrt{12}}$$

از قضیه چبیشف نتیجه می‌گیریم:

$$\Pr\left(\left|X - \frac{1}{2}\right| \geq \epsilon\right) \leq \frac{1}{12\epsilon^2}$$

محاسبه با $\epsilon = \frac{1}{4}$

اگر:

$$\epsilon = \frac{1}{4}$$

آنگاه:

$$\Pr\left(\left|X - \frac{1}{2}\right| \geq \frac{1}{4}\right) \leq \frac{1}{12(1/4)^2} = \frac{16}{12} = \frac{4}{3}$$

این کران بزرگ‌تر از ۱ است و عملاً اطلاعات مفیدی نمی‌دهد.

محاسبه با $\epsilon = k\sigma$

$$\epsilon = k\sigma, \quad \sigma^2 = \text{Var}(X)$$

$$k\sigma = \frac{k}{\sqrt{12}}$$

پس:

$$k = \frac{2\sqrt{3}}{4} = \frac{\sqrt{3}}{2}$$

$$k^2 = \frac{3}{4} \quad \Rightarrow \quad \frac{1}{k^2} = \frac{4}{3}$$

و نیز:

$$k\sigma = \frac{\sqrt{3}}{2} \times \frac{1}{\sqrt{12}} = \frac{\sqrt{3}}{2\sqrt{4 \times 3}} = \frac{1}{4} = \epsilon$$

کران چبیشف در مثال قبل مفید نبود و در بعضی موارد می‌توانیم آن را با مقدار واقعی مقایسه کنیم. مثلاً در همین مثال:

$$\begin{aligned} & \Pr \left(\left| X - \frac{1}{2} \right| \geq \frac{1}{4} \right) \\ &= 1 - \Pr \left(\left| X - \frac{1}{2} \right| < \frac{1}{4} \right) \\ &= 1 - \Pr \left(-\frac{1}{4} < X - \frac{1}{2} < \frac{1}{4} \right) \\ &= 1 - \Pr \left(\frac{1}{4} < X < \frac{3}{4} \right) \\ &= 1 - \int_{\frac{1}{4}}^{\frac{3}{4}} f(x) dx = 1 - \int_{\frac{1}{4}}^{\frac{3}{4}} 1 dx \end{aligned}$$

چون برای توزیع یکنواخت روی بازه $[0, 1]$ داریم:

$$f(x) = 1$$

پس:

$$= 1 - \frac{1}{2} = \frac{1}{2}$$

بنابراین:

$$\Pr \left(\left| X - \frac{1}{2} \right| \geq \frac{1}{4} \right) = \frac{1}{2}$$

Law of the unconscious statistician LOTUS

مقدار امید تابعی از X ، مثلاً $g(X)$ ، را گویند؛ که X دارای توزیع $f(x)$ است.

فرمول LOTUS

$$E(g(X)) = \begin{cases} \sum g(x)f(x), & X : \text{discrete} \\ \int_{\mathbb{R}} g(x)f(x) dx, & X : \text{continuous} \end{cases}$$

مثال‌ها

$$E(X^2) = \int_{\mathbb{R}} x^2 f(x) dx$$

$$E(\sin X) = \int_{\mathbb{R}} \sin x f(x) dx$$

نامساوی‌های معروف (inequalities)

□ مارکف (Markov)

□ چیشف (Chebyshev)

□ چرنوف (Chernoff)

1405 12 *khordad independence* 1

۱. استقلال: حالت بدیهی: هر ۲ پیشامد نامربوط، مستقل هستند.

مثلاً:

پیشامد

A = فردا در سیاره مشتری باران می بارد

L = در پرتاب سکه، خط بیاید

تعریف: A و B مستقل هستند اگر و فقط اگر (بعضی مراجع نوشته اند اگر)

$$Pr(A \cap B) = Pr(A)Pr(B)$$

مثلاً در مثال فوق اگر:

$$Pr(A) = 0.2$$

$$Pr(B) = 0.5$$

آنگاه

$$Pr(A \cap B) = 0.1$$

← پس مستقل هستند.

نکته: اگر $Pr(A) = 0$ باشد، پیشامد A مستقل از هر پیشامد دیگر است.

نکته مهم: برای استقلال، لازم نیست پیشامدها از نظر فیزیکی و مفهومی، نسبت به هم نامربوط باشند (به مثال زیر توجه کنید).

مثال: در پرتاب تاس ۶ وجهی،

$$A = \{2, 4, 6\}$$

(خال زوج)،

$$B = \{1, 2, 3, 4\}$$

(خال مشاهده شده عدد کمتر از ۵ باشد).

محاسبات این مثال (نوشته شده در سمت چپ و وسط):

$$A \cap B = \{2, 4\}$$

بنابراین:

$$Pr(A) = \frac{1}{2}$$

$$Pr(B) = \frac{4}{6} = \frac{2}{3}$$

$$Pr(A \cap B) = \frac{|A \cap B|}{|S|} = \frac{2}{6} = \frac{1}{3}$$

از طرفی:

$$Pr(A)Pr(B) = \frac{1}{2} \times \frac{2}{3} = \frac{1}{3}$$

مساوی شدند (یعنی $\frac{1}{3}$ $Pr(A \cap B) = Pr(A)Pr(B)$)

$\leftarrow A$ و B مستقل هستند.

تعبیر شهودی تر استقلال: قضیه: فرض کنید

$$Pr(B) > 0$$

آنگاه A و B مستقل هستند اگر و فقط اگر

$$Pr(A | B) = Pr(A)$$

اثبات:

$$\frac{Pr(A \cap B)}{Pr(B)} = Pr(A) \iff Pr(A \cap B) = Pr(A)Pr(B) \iff A \text{ و } B \text{ مستقل}$$

نکته مهم: اگر A و B مستقل باشند، احتمال وقوع A وابسته به وقوع یا عدم وقوع B نمی باشد.

بررسی کنید در پرتاب ۲ تاس، پیشامد مجموع ۲ تاس ۱۱ باشد، مستقل است. اگر مجموع ۲ تاس ۸ یا ۹ یا ۱۰ باشد چطور؟

احتمال پیشامد ناممکن چیست؟ (Impossible Events)

جواب:

اگر

$$A = \{\} \rightarrow Pr(A) = 0 \quad \text{مجموعه تهی} = \{\} = \emptyset = \text{پیشامد ناممکن}$$

چه تفاوتی با پیشامدی که احتمال آن صفر است دارد؟

جواب: برعکس فوق صحیح نیست؛ یعنی اگر $Pr(A) = 0$ باشد نمی‌توان گفت $A = \{\}$. مثلاً A انتخاب یک عدد با احتمال z در بازه‌ی 0 و 1 باشد. که در اینجا $Pr(A) = 0$ است ولی $A \neq \{\}$ چون هر عدد بین 0 و 1 می‌تواند باشد.

در احتمال شرطی $P(A|B)$ ، اگر احتمال وقوع B صفر باشد، چه اتفاقی می‌افتد؟

جواب:

مخرج کسر باید مخالف صفر باشد. بنابراین فرض می‌کنند $Pr(B) > 0$ (و حالت مساوی حذف شده است)

اگر دو مجموعه A و B از هم جدا یا Disjoint باشند، احتمالی که روی این دو مجموعه تعریف می‌شود یعنی پیشامدهای A, B را می‌توان به صورت زیر حساب کرد.

مثال:

اگر سکه‌ای را بیندازیم تا اولین شیر رخ دهد، مجموعه S یا Ω برای آن چیست؟

$$S = \{H, TH, TTH, TTTH, \dots\}$$

جواب:

پیشامدهای از هم جدا را به صورت مقابل تعریف می‌کنیم

$$A_1 = \{H\}, A_2 = \{TH\}, A_3 = \{TTH\}, A_4 = \{TTTH\}, \dots$$

می‌دانیم احتمال وقوع S (یعنی $\Pr(S)$) ۱ است. بنابراین داریم:

$$1 = \Pr(S) = \Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i)$$

اصل شمول عدم شمول (Principle of inclusion-exclusion)

برای دو پیشامد A و B دلخواه داریم:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

برای سه پیشامد A و B و C دلخواه داریم:

$$\Pr(A \cup B \cup C) = \Pr(A) + \Pr(B) + \Pr(C)$$

$$- \Pr(A \cap B) - \Pr(A \cap C) - \Pr(B \cap C)$$

$$+ \Pr(A \cap B \cap C)$$

برای حالت کلی n پیشامد $A_1, A_2, \dots, A_n$ داریم

$$\Pr(A_1 \cup \dots \cup A_n) \quad J \text{ Event Set}$$

مثال:

مردم شهر ۷۵ درصد می‌دوند (پیشامد J)، ۲۰ درصد بستنی در دست دارند (پیشامد I)، و ۶۰ درصد موسیقی گوش می‌دهند (پیشامد M) همچنین ۱۵ درصد (I, J) ، ۳۰ درصد (M, J) ، ۱۰ درصد (M, I) و ۵ درصد هر ۳ مورد را دارند. احتمال اینکه یک شهروند به تصادف انتخاب شود که

حداقل یکی از ۳ مورد بالا را داشته باشد چقدر است:

$$\begin{aligned}
 P(J \cup I \cup M) &= P(J) + P(I) + P(M) \\
 &\quad - P(J \cap I) - P(J \cap M) - P(I \cap M) \\
 &\quad + P(J \cap I \cap M) \\
 &= 0.75 + 0.2 + 0.4 - 0.15 - 0.3 - 0.1 + 0.05 = 0.85
 \end{aligned}$$

مثال:

فرض کنید: ۴۰ درصد احتمال دارد هوا سردتر شود. (پیشامد C)
 ۱۰ درصد شانس بارش و سردتر شدن هوا. (پیشامد $C \cup R$)
 ۸۰ درصد شانس بارش یا سردتر شدن هوا. (پیشامد $C \cap R$)
 احتمال بارش چیست؟

$$P(R) = P(R \cup C) - P(C) + P(R \cap C) = 0.8 - 0.4 + 0.1 = 0.5$$

قضیه:

$$A, B \text{ مستقل} \rightarrow A, \bar{B}$$

اثبات:

$$\Pr(A) = \Pr(A \cap \bar{B}) + \Pr(A \cap B) \text{ So That}$$

$$\Pr(A \cap B) = \Pr(A) - \Pr(A \cap \bar{B})$$

$$= \Pr(A) - \Pr(A) \Pr(B)$$

$$= \Pr(A)[1 - \Pr(B)] = \Pr(A)\Pr(\bar{B})$$

استقلال (independence) و از هم جدا بودن (disjoint) را باهم اشتباه نگیرید.

قضیه: اگر $Pr(A) > 0, Pr(B) > 0$ آنگاه A و B همزمان نمی‌توانند مستقل و از هم جدا باشند.

اثبات:

$A \cap B = \emptyset \rightarrow Pr(A \cap B) = 0 < Pr(A)Pr(B)$ یعنی A و B از هم جدا باشند یعنی اشتراک آن‌ها تهی است

پس A و B مستقل نیستن، به طور مشابه استقلال دلالت بر از هم جدا بودن ندارد.

تمرین: مباحث مربوط به استقلال حالتی که بیش از ۲ پیشامد داریم را مطالعه نمایید. راهنمایی: سه پیشامد A و B و C مستقل هستند اگر:

$$a) Pr(A \cap B \cap C) = Pr(A)Pr(B)Pr(C)$$

b) همه زوج‌های ممکن هم مستقل باشند یعنی:

$$Pr(A \cap B) = Pr(A)Pr(B)$$

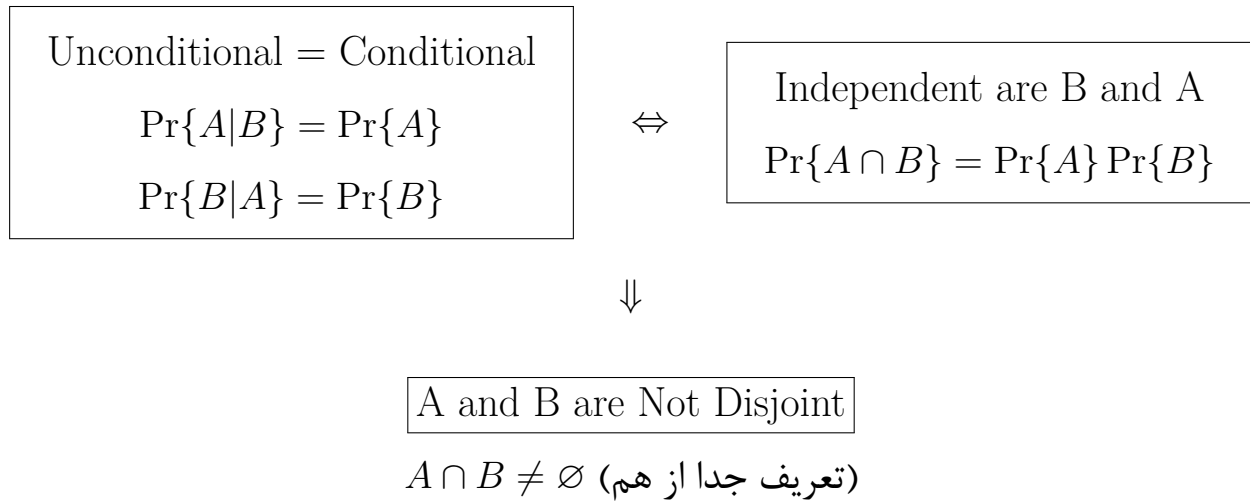
$$Pr(A \cap C) = Pr(A)Pr(C)$$

$$Pr(B \cap C) = Pr(B)Pr(C)$$

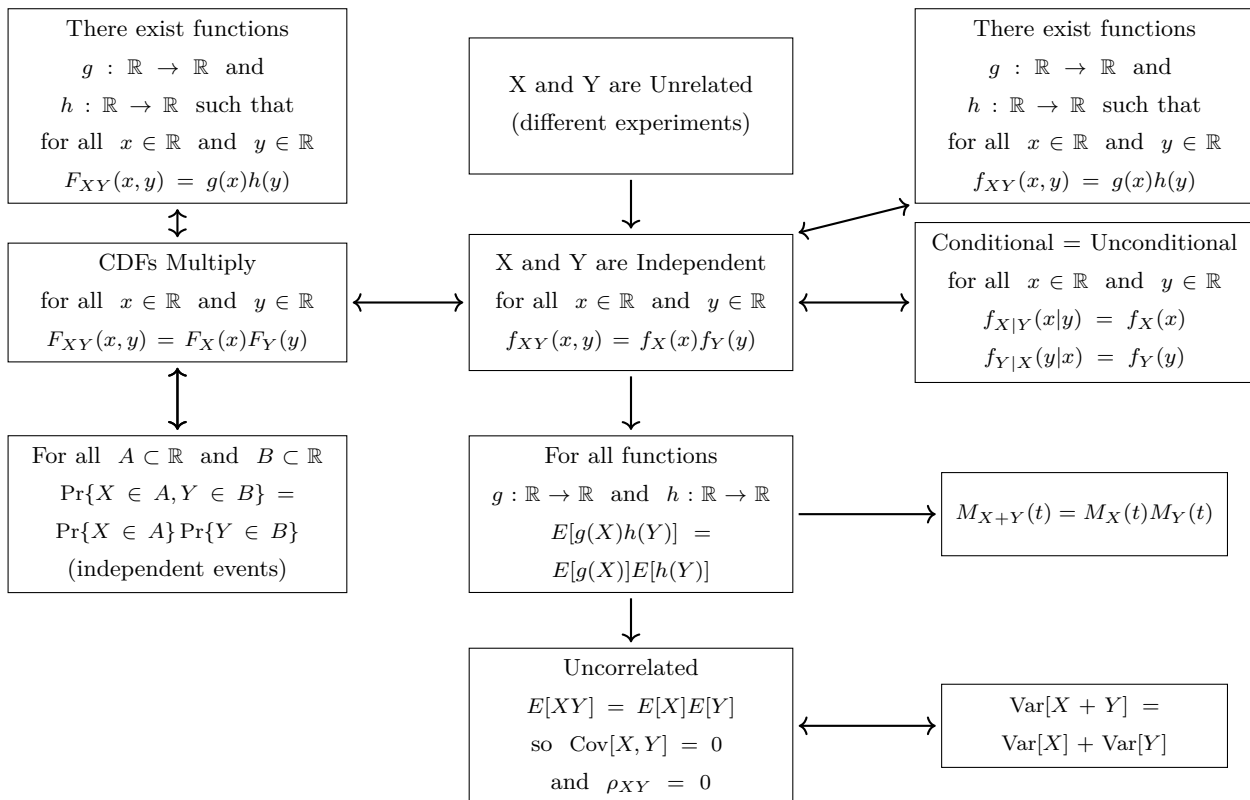
Events

A and B are Unrelated (different experiments)





Random Variables



Random Variables