Xác suất và Thống kê cho Khoa học Dữ liệu

Carlos Fernandez-Bà

lời nói đầu

Những ghi chú này được phát triển cho khóa học Xác suất và Thống kê cho Khoa học Dữ liệu tại Trung tâm Khoa học Dữ liệu ở NYU. Mục tiêu là cung cấp một cái nhìn tổng quan về các khái niệm cơ bản trong xác suất và thống kê từ các nguyên tắc đầu tiên. Tôi muốn cảm ơn Levent Sagun và Vlad Kobzar, những người đã trợ giảng cho khóa học, cũng như Brett Bernstein và David Rosenberg vì những gợi ý hữu ích của họ. Tôi cũng rất biết ơn tất cả các sinh viên của tôi về phản hồi của họ.

Trong khi viết những ghi chú này, tôi đã được Quỹ Khoa học Quốc gia hỗ trợ theo giải thưởng NSF DMS-1616340.

New York, tháng 8 năm 2017

nội dung

Chương 1

Lý thuyết xác suất cơ bản

Trong chương này, chúng tôi giới thiệu khuôn khổ toán học của lý thuyết xác suất, giúp chúng ta có thể suy luận về sự không chắc chắn một cách cơ bản bằng cách sử dụng lý thuyết tập hợp. Phụ lục A chứa một đánh giá về các khái niệm cơ bản của lý thuyết tập hợp.

1.1 Không gian xác suất

Mục tiêu của chúng tôi là xây dựng một khung toán học để biểu diễn và phân tích các hiện tượng không chắc chắn, chẳng hạn như kết quả của việc gieo xúc xắc, thời tiết ngày mai, kết quả của một trận đấu NBA, v.v. Để đạt được điều này, chúng tôi mô hình hóa hiện tượng quan tâm như một thử nghiệm với một số (có thể là vô hạn) kết quả loại trừ lẫn nhau.

Ngoại trừ những trường hợp đơn giản, khi số lượng kết quả ít, người ta thường suy luận về các tập hợp kết quả, được gọi là các sự kiện. Để định lượng khả năng kết quả của thử nghiệm thuộc về một sự kiện cụ thể như thế nào, chúng ta chỉ định một xác suất cho sự kiện đó. Chính thức hơn, chúng ta định nghĩa một thước đo (nhớ lại rằng thước đo là một hàm ánh xạ các tập hợp thành số thực) gán xác suất cho mỗi sự kiện quan tâm.

Chính thức hơn, thí nghiệm được đặc trưng bởi việc xây dựng một không gian xác suất.

Định nghĩa 1.1.1 (Không gian xác suất). Một không gian xác suất là một bộ ba (Ω, F, P) bao gồm:

- Không gian mẫu Ω chứa tất cả các kết quả có thể xảy ra của phép thử.
- Một tập hợp các sự kiện F, phải là một σ-đại số (xem Định nghĩa 1.1.2 bên dưới).
- Một phép đo xác suất P gán xác suất cho các sự kiện trong F (xem Định nghĩa 1.1.4 dưới).

Không gian mẫu có thể rời rạc hoặc liên tục. Ví dụ về các không gian mẫu rời rạc bao gồm các kết quả có thể xảy ra khi tung đồng xu, tỷ số của một trận bóng rổ, số người xuất hiện tại một bữa tiệc, v.v. Các không gian mẫu liên tục thường là các khoảng R hoặc R^n được sử dụng để mô hình hóa thời gian , vị trí, nhiệt độ, vv

Thuật ngữ σ-đại số được sử dụng trong lý thuyết đo lường để biểu thị một tập hợp các tập hợp thỏa mãn các điều kiện nhất định được liệt kê bên dưới. Đừng quá sợ hãi bởi nó. Đó chỉ là một cách nói phức tạp rằng nếu chúng ta gán xác suất cho một số sự kiện nhất định (ví dụ: trời sẽ mưa vào ngày mai hoặc trời sẽ mưa ngày mai tuyết rơi), chúng ta cũng cần gán xác suất cho các phần bù của chúng (tức là ngày mai trời sẽ không mưa hoặc ngày mai không có tuyết) và cho tập hợp của chúng (trời sẽ mưa hoặc tuyết vào ngày mai).

Định nghĩa 1.1.2 (σ-đại số). Một σ-đại số F là tập hợp các tập hợp trong Ω sao cho:

1. Nếu một tập hợp S F thì S c F.
2. Nếu các tập hợp S1, S2 F thì S1 S2 F. Điều này cũng đúng với các dãy vô hạn; nếu như S1, S2, . . . F thì ∞ i=1Si F.
3. Ω F .

Nếu không gian mẫu của chúng ta là rời rạc, thì một lựa chọn khả dĩ cho đại số σ là tập lũy thừa của không gian mẫu, bao gồm tất cả các tập phần tử có thể có trong không gian mẫu. Nếu chúng ta tung đồng xu và không gian mẫu là

$$\Omega := \{\text{ngửa}, \text{sấp}\} , \quad (1.1)$$

thì tập lũy thừa là một σ-đại số hợp lệ

$$F := \{\text{đầu hoặc đuôi}, \text{đầu}, \text{sấp}, \ \} , \quad (1.2)$$

trong đó là tập rỗng. Tuy nhiên, trong nhiều trường hợp σ-đại số không chứa mọi tập hợp kết quả có thể.

Ví dụ 1.1.3 (Cholesterol). Một bác sĩ quan tâm đến việc lập mô hình xác suất về mức cholesterol của bệnh nhân. Mỗi khi một bệnh nhân đến thăm, cô ấy đều kiểm tra mức cholesterol của họ. Ở đây thí nghiệm là xét nghiệm cholesterol, kết quả là mức cholesterol đo được và không gian mẫu Ω là đường thực dương. Bác sĩ chủ yếu quan tâm đến việc liệu bệnh nhân có cholesterol thấp, cao ở mức giới hạn hay cao. Sự kiện L (cholesterol thấp) bao gồm tất cả các kết quả dưới 200 mg/dL, sự kiện B (cholesterol cao ở ngưỡng giới hạn) bao gồm tất cả các kết quả từ 200 đến 240 mg/dL và sự kiện H (cholesterol cao) bao gồm tất cả các kết quả trên 240 mg /dL. Do đó, σ-đại số F của các sự kiện có thể xảy ra bằng

$$F := \{L \quad B \quad H, L \quad B, L \quad H, B \quad H, L, B, H, \ \} . \quad (1.3)$$

Các sự kiện là một phân vùng của không gian mẫu, giúp đơn giản hóa việc lấy đại số σ tương ứng.

Vai trò của thước đo xác suất P là định lượng khả năng chúng ta gặp phải từng sự kiện trong đại số σ. Theo trực giác, xác suất của một sự kiện A có thể được hiểu là phần nhỏ số lần mà kết quả của thử nghiệm nằm trong A, vì số lần lặp lại có xu hướng vô hạn. Theo đó, xác suất phải luôn luôn không âm. Ngoài ra, nếu hai biến cố A và B rời nhau (giao của chúng trống), thì

$$P(A \quad B) = \frac{\text{kết quả trong A hoặc B}}{\text{tổng}} \quad (1.4)$$

$$= \frac{\text{kết quả trong A + kết quả trong B}}{\text{tổng}} \quad (1.5)$$

$$= \frac{\text{kết quả trong A kết quả trong B +}}{\text{tổng}} \quad \frac{}{\text{tổng cộng}} \quad (1.6)$$

$$= P(A) + P(B). \quad (1.7)$$

Xác suất của sự kết hợp của các sự kiện rời rạc phải bằng tổng các xác suất riêng lẻ.
Ngoài ra, xác suất của toàn bộ không gian mẫu Ω phải bằng một, vì nó chứa tất cả các kết quả

$$p(\Omega) = \frac{\text{kết quả trong } \Omega}{\text{tổng}} \tag{1.8}$$

$$= \frac{\text{cộng}}{\text{tổng cộng}} \tag{1.9}$$

$$= 1. \tag{1.10}$$

Những điều kiện này là cần thiết để một biện pháp là một biện pháp xác suất hợp lệ.

Định nghĩa 1.1.4 (Độ đo xác suất). Một phép đo xác suất là một hàm được xác định trên các tập hợp trong σ-đại số F sao cho:

1. $P(S) \geq 0$ với mọi biến cố $S \quad F$.

Nếu các tập $S_1, S_2, . . . , S_n \quad F$ là rời nhau (tức là $S_i \cap S_j = \quad$ với $i = j$) thì 2.

$$P\left(\quad {}_{i=1}^{n} S_i \right) = \sum_{\text{tôi} = 1}^{N} p\,(S_i). \tag{1.11}$$

Tương tự, đối với một dãy vô hạn đếm được của các tập rời rạc $S_1, S_2, . . . \quad F$

$$P \lim_{n \to \infty} {}_{i=1}^{N} S_i = \lim_{n \to \infty} \sum_{\text{tôi} = 1}^{N} p\,(S_i). \tag{1.12}$$

3. $P(\Omega)=1$.

Hai tiên đề đầu tiên nắm bắt ý tưởng trực quan rằng xác suất của một sự kiện là một thước đo như khối lượng (hoặc chiều dài hoặc thể tích): giống như khối lượng của bất kỳ vật thể nào là không âm và tổng khối lượng của một số vật thể riêng biệt là tổng khối lượng của chúng , xác suất của bất kỳ sự kiện nào là không âm và xác suất hợp nhất của một số đối tượng rời rạc là tổng xác suất của chúng. Tuy nhiên, trái ngược với khối lượng, lượng xác suất trong một thử nghiệm không thể bị giới hạn. Nếu rất có thể ngày mai trời sẽ mưa, thì cũng không có khả năng là trời sẽ không mưa. Nếu xác suất của một sự kiện S lớn, thì xác suất của phần bù S^c phải nhỏ. Điều này được nắm bắt bởi tiên đề thứ ba, chuẩn hóa phép đo xác suất (và hàm ý rằng $P(S^c) = 1 \quad P(S)$).

Điều quan trọng cần nhấn mạnh là phép đo xác suất không ấn định xác suất cho các kết quả riêng lẻ, mà là cho các sự kiện trong đại số σ. Lý do cho điều này là khi số lượng các kết quả có thể xảy ra là vô hạn không thể đếm được, thì người ta không thể gán xác suất khác 0 cho tất cả các kết quả mà vẫn thỏa mãn điều kiện $P(\Omega) = 1$. Đây không phải là một tình huống kỳ lạ, nó xảy ra chẳng hạn trong ví dụ về cholesterol trong đó bất kỳ số thực dương nào cũng là một kết quả có thể xảy ra.
Trong trường hợp không gian mẫu rời rạc hoặc đếm được, đại số σ có thể bằng tập lũy thừa của không gian mẫu, nghĩa là chúng ta gán xác suất cho các sự kiện chỉ chứa một kết quả duy nhất (ví dụ: ví dụ tung đồng xu).

Ví dụ 1.1.5 (Cholesterol (tiếp theo)). Một thước đo xác suất hợp lệ cho Ví dụ 1.1.3 là

$$P(L) = 0{,}6, \ P(B) = 0{,}28, \ P(H) = 0{,}12. \tag{1.13}$$

Ví dụ, sử dụng các thuộc tính, chúng ta có thể xác định rằng P (B H) = 0,6 + 0,28 = 0,88.

Định nghĩa 1.1.4 có các hệ quả sau:

$$P(\ ) = \tag{1.14}$$

$$0, \ A \quad B \text{ nên } P(A) \leq P(B), \ P(A \tag{1.15}$$

$$B) = P(A) + P(B) \ \ P(A \cap B). \tag{1.16}$$

Chúng tôi bỏ qua các bằng chứng (hãy thử tự chứng minh chúng).

1.2 Xác suất có điều kiện

Xác suất có điều kiện là một khái niệm quan trọng trong mô hình xác suất. Nó cho phép chúng tôi cập nhật các mô hình xác suất khi thông tin bổ sung được tiết lộ. Xét một không gian xác suất (Ω, F, P) trong đó chúng ta thấy rằng kết quả của thí nghiệm thuộc về một sự kiện S F nhất định. Điều này rõ ràng ảnh hưởng đến khả năng xảy ra của bất kỳ sự kiện nào khác S F: chúng ta có thể loại trừ bất kỳ kết quả nào không thuộc về S. Xác suất cập nhật của mỗi sự kiện được gọi là xác suất có điều kiện của S cho trước S. Theo trực giác, xác suất có điều kiện có thể được hiểu là tỷ lệ các kết quả trong S cũng thuộc S ,

$$PS \ |S = \frac{\text{kết quả trong S và S}}{\text{kết quả trong S}} \tag{1.17}$$

$$= \frac{\text{kết quả trong S và S}}{\text{tổng cộng}} \ \frac{\text{tổng cộng}}{\text{kết quả trong S}} \tag{1.18}$$

$$= \frac{P\ (S \cap S)}{P\ (S)}, \tag{1.19}$$

trong đó chúng ta giả sử rằng P(S) = 0 (sau này chúng ta sẽ xử lý trường hợp khi S có xác suất bằng 0, thường xảy ra trong không gian xác suất liên tục). Định nghĩa khá trực quan: S bây giờ là không gian mẫu mới, vì vậy nếu kết quả là trong S thì nó phải thuộc S ∩S. Tuy nhiên, chỉ sử dụng xác suất của giao điểm sẽ đánh giá thấp khả năng xảy ra của S vì không gian mẫu đã bị giảm xuống S. Do đó, chúng tôi chuẩn hóa bằng xác suất của S. Để kiểm tra độ chính xác, chúng tôi có P (S|S) = 1 và nếu S và S rời nhau thì P(S |S) = 0.

Xác suất có điều kiện P (·|S) là một phép đo xác suất hợp lệ trong không gian xác suất (S, FS,P (·| S)), trong đó FS là một σ-đại số chứa giao của S và các tập hợp trong F. Để đơn giản hóa ký hiệu, khi chúng ta đặt điều kiện trên giao điểm của các tập hợp, chúng ta viết xác suất có điều kiện là

$$P\ (S|A,\ B,\ C) := P\ (S|A \cap B \cap C), \tag{1.20}$$

với mọi biến cố S, A, B, C.

Ví dụ 1.2.1 (Chuyến bay và cơn mưa). Sân bay JFK thuê bạn để ước tính mức độ đúng giờ của các chuyến bay đến bị ảnh hưởng bởi thời tiết. Bạn bắt đầu bằng cách xác định một không gian xác suất mà không gian mẫu là

$$\Omega = \{\text{trễ và mưa, muộn và không mưa, đúng giờ và mưa, đúng giờ và không mưa}\} \tag{1.21}$$

và σ-đại số là tập lũy thừa của Ω. Từ dữ liệu của các chuyến bay trước đây, bạn xác định rằng ước tính hợp lý cho phép đo xác suất của không gian xác suất là

$$P\ (\text{trễ, không mưa}) = \frac{2}{20}, \qquad P\ (\text{đúng giờ, không mưa}) = \frac{14}{20}, \tag{1.22}$$

$$P\ (\text{muộn,mưa}) = \frac{3}{20}, \qquad P\ (\text{đúng giờ, mưa}) = \frac{1}{20}. \tag{1.23}$$

Sân bay quan tâm đến khả năng chuyến bay bị trễ nếu trời mưa, vì vậy bạn xác định một điều kiện không gian xác suất mới cho sự kiện mưa. Không gian mẫu là tập hợp tất cả các kết quả sao cho có mưa, đại số σ là tập lũy thừa của {đúng giờ, trễ} và thước đo xác suất là P (·| mưa). Đặc biệt,

$$P\ (\text{muộn}|\text{mưa}) = \frac{P\ (\text{muộn, mưa})}{P\ (\text{mưa})} = \frac{20/3}{20/3 + 20/1} = \frac{3}{4} \tag{1.24}$$

và tương tự P (muộn|không mưa) = 1/8.

Xác suất có điều kiện có thể được sử dụng để tính toán giao điểm của một số sự kiện theo cách có cấu trúc. Theo định nghĩa, chúng ta có thể biểu diễn xác suất giao nhau của hai biến cố A, B F như sau,

$$P\ (A \cap B) = P\ (A)\ P\ (B|A) \tag{1.25}$$

$$= P(B)\ P(A|B). \tag{1.26}$$

Trong công thức này, P(A) được gọi là xác suất trước của A, vì nó nắm bắt thông tin chúng ta có về A trước khi bất kỳ điều gì khác được tiết lộ. Tương tự, P (A|B) được gọi là xác suất sau. Đây là những đại lượng cơ bản trong các mô hình Bayes, được thảo luận trong Chương 10.

Tổng quát hóa (1.25) cho một chuỗi các sự kiện sẽ đưa ra quy tắc chuỗi, cho phép biểu thị xác suất giao nhau của nhiều sự kiện dưới dạng xác suất có điều kiện. Chúng tôi bỏ qua bằng chứng, đây là một ứng dụng quy nạp đơn giản.

Định lý 1.2.2 (Quy tắc chuỗi). Cho (Ω, F, P) là một không gian xác suất và $S_1, S_2, \ldots$ một tập hợp các sự kiện trong F,

$$P\ (\cap_i S_i) = P\ (S_1)\ P\ (S_2|S_1)\ P\ (S_3|S_1 \cap S_2) \cdots \tag{1.27}$$

$$= P\ S_i \mid \cap i_{j=1} 1\ S_J\ . \tag{1.28}$$

Đôi khi, ước tính trực tiếp xác suất của một sự kiện nhất định có thể khó khăn hơn ước tính xác suất của nó dựa trên các sự kiện đơn giản hơn. Tập hợp các tập hợp khác nhau $A_1, A_2, \ldots$ sao cho $\Omega = \ iA_i$ được gọi là một phân hoạch của Ω. Quy luật tổng xác suất cho phép chúng ta gộp các xác suất có điều kiện lại với nhau, tính trọng số của chúng theo xác suất của các sự kiện riêng lẻ trong phân vùng, để tính xác suất của sự kiện quan tâm.

Định lý 1.2.3 (Quy luật xác suất toàn phần). Đặt (Ω, F, P) là một không gian xác suất và đặt tập các tập rời nhau $A1, A2, . . .$ F là một phân hoạch bất kỳ của Ω. Với mọi tập hợp S F

$$P(S) = P(S \cap Ai) \quad (1.29)$$

$$= P(Ai)\, P(S|Ai). \quad (1.30)$$

Bằng chứng. Đây là hệ quả trực tiếp của quy tắc dây chuyền và Tiên đề 2 trong Định nghĩa 1.1.4, vì $S = \ i\ S \cap Ai$ và các tập hợp $S \cap Ai$ là rời nhau. □

Ví dụ 1.2.4 (Thăm dì). Dì của bạn sẽ đến JFK vào ngày mai và bạn muốn biết khả năng chuyến bay của cô ấy đến đúng giờ là bao nhiêu. Từ ví dụ 1.2.1, bạn nhớ lại rằng

$$P(\text{muộn}|\text{mưa}) = 0{,}75, \ P(\text{muộn}|\text{không mưa}) = 0{,}125. \quad (1.31)$$

Sau khi xem một trang web về thời tiết, bạn xác định được rằng P (mưa) = 0,2.

Bây giờ, làm thế nào chúng ta có thể tích hợp tất cả các thông tin này? Các sự kiện mưa và không mưa là rời rạc và bao phủ toàn bộ không gian mẫu, vì vậy chúng tạo thành một phân vùng. Do đó, chúng ta có thể áp dụng định luật tổng xác suất để xác định

$$P(\text{muộn}) = P(\text{muộn}|\text{mưa})\, P(\text{mưa}) + P(\text{muộn}|\text{không mưa})\, P(\text{không mưa}) \quad (1.32)$$

$$= 0{,}75 \cdot 0{,}2 + 0{,}125 \cdot 0{,}8 = 0{,}25. \quad (1.33)$$

Vậy xác suất máy bay của dì bạn bị trễ là 1/4.

Điều quan trọng là phải nhận ra rằng nói chung $P(A|B) = P(B|A)$: hầu hết các cầu thủ ở NBA có lẽ sở hữu một quả bóng rổ (P (sở hữu bóng|NBA) lớn) nhưng hầu hết những người sở hữu bóng rổ đều không tham gia NBA (P (NBA|sở hữu bóng) nhỏ). Lý do là các xác suất trước đó rất khác nhau: P (NBA) nhỏ hơn nhiều so với P (sở hữu bóng). Tuy nhiên, có thể đảo ngược điều kiện xác suất, tức là tìm $P(A|B)$ từ $P(B|A)$, miễn là chúng ta tính đến các tiên nghiệm. Cái này hệ quả trực tiếp của định nghĩa xác suất có điều kiện được gọi là quy tắc Bayes.

Định lý 1.2.5 (Quy tắc Bayes). Đối với mọi sự kiện A và B trong không gian xác suất (Ω, F, P)

$$P(A|B) = \frac{P(A)\, P(B|A)}{P(B)}, \quad (1.34)$$

miễn là $P(B) > 0$.

Ví dụ 1.2.6 (Thăm dì (tiếp theo)). Bạn giải thích mô hình xác suất được mô tả trong Ví dụ 1.2.4 cho anh họ Marvin của bạn sống ở California. Một ngày sau, bạn nói với anh ấy rằng bạn dì đến muộn nhưng bạn không đề cập đến việc trời có mưa hay không. Sau khi cúp máy, Marvin muốn tính xác suất trời mưa. Nhớ lại rằng xác suất mưa là 0,2, nhưng vì dì của bạn đến muộn nên ông ấy nên cập nhật ước tính. Áp dụng quy tắc Bayes và định luật xác suất toàn phần:

$$P\,(\text{mưa}|\text{muộn}) = \frac{P\,(\text{muộn}|\text{mưa})\ P\,(\text{mưa})}{P\,(\text{muộn})} \tag{1.35}$$

$$= \frac{P\,(\text{muộn}|\text{mưa})\ P\,(\text{mưa})}{P\,(\text{muộn}|\text{mưa})\ P\,(\text{mưa}) + P\,(\text{muộn}|\text{không mưa})\ P\,(\text{không mưa})} \tag{1.36}$$

$$= \frac{0{,}75 \cdot 0{,}2}{0{,}75 \cdot 0{,}2 + 0{,}125 \cdot 0{,}8} = 0{,}6. \tag{1.37}$$

Đúng như dự đoán, xác suất trời mưa tăng lên với giả định rằng dì của bạn đang muộn.

1.3 Độc lập

Như đã thảo luận trong phần trước, xác suất có điều kiện định lượng mức độ mà kiến thức về sự xuất hiện của một sự kiện nhất định ảnh hưởng đến xác suất của một sự kiện khác. Trong một số trường hợp, nó không có gì khác biệt: các sự kiện là độc lập. Chính thức hơn, các sự kiện A và B là độc lập khi và chỉ khi

$$P(A|B) = P(A). \tag{1.38}$$

Định nghĩa này không hợp lệ nếu P(B) = 0. Định nghĩa sau đây bao trùm trường hợp này và ngược lại tương đương.

Định nghĩa 1.3.1 (Độc lập). Cho (Ω, F, P) là một không gian xác suất. Hai biến cố A, B F độc lập khi và chỉ khi

$$P(A \cap B) = P(A)\ P(B). \tag{1.39}$$

Ví dụ 1.3.2 (Đại hội). Chúng tôi xem xét một bộ dữ liệu tổng hợp phiếu bầu của các thành viên của Hạ viện Mỹ về hai vấn đề năm 1984 . Các vấn đề là chia sẻ chi [1] phí cho một nguồn nước dự án (vấn đề 1) và thông qua nghị quyết ngân sách (vấn đề 2). Chúng tôi mô hình hóa hành vi của các nghị sĩ theo xác suất, xác định một không gian mẫu trong đó mỗi kết quả là một chuỗi phiếu bầu. Chẳng hạn, một kết quả có thể xảy ra là vấn đề 1 = có, vấn đề 2 = không. Chúng tôi chọn σ-đại số là tập lũy thừa của không gian mẫu. Để ước tính phép đo xác suất liên quan đến các sự kiện khác nhau, chúng tôi chỉ tính phần nhỏ sự xuất hiện của chúng trong dữ liệu.

$$P\,(\text{vấn đề 1} = \text{có}) \approx \frac{\text{thành viên bỏ phiếu đồng ý cho vấn đề 1}}{\text{tổng số phiếu cho vấn đề 1}} \tag{1.40}$$

$$= 0{,}597, \tag{1.41}$$

$$P\,(\text{vấn đề 2} = \text{có}) \approx \frac{\text{thành viên bỏ phiếu đồng ý cho vấn đề 2}}{\text{tổng số phiếu cho vấn đề 2}} \tag{1.42}$$

$$= 0{,}417, \tag{1.43}$$

$$P\,(\text{vấn đề 1} = \text{có} \cap \text{vấn đề 2} = \text{có}) \approx \frac{\text{thành viên bỏ phiếu đồng ý cho vấn đề 1 và 2}}{\text{tổng số thành viên biểu quyết vấn đề 1 và 2}} \tag{1.44}$$

$$= 0{,}069. \tag{1.45}$$

1Dữ liệu có sẵn ở đây.

Dựa trên những dữ liệu này, chúng tôi có thể đánh giá liệu hành vi bỏ phiếu đối với hai vấn đề có phụ thuộc vào nhau hay không. Nói cách khác, nếu chúng tôi biết cách một thành viên bỏ phiếu cho vấn đề 1, điều này có cung cấp thông tin về làm thế nào họ bỏ phiếu cho vấn đề 2? Câu trả lời là có, vì

$$P\text{ (vấn đề 1 = có) } P\text{ (vấn đề 2 = có)} = 0{,}249 \tag{1.46}$$

rất khác với P (vấn đề 1 = có ∩ vấn đề 2 = có). Nếu một thành viên bỏ phiếu đồng ý cho vấn đề 1, họ ít có khả năng bỏ phiếu đồng ý cho vấn đề 2.

Tương tự, chúng ta có thể định nghĩa tính độc lập có điều kiện giữa hai biến cố khi biết biến cố thứ ba. A và B độc lập có điều kiện với C khi và chỉ khi

$$P(A|B, C) = P(A|C), \tag{1.47}$$

trong đó $P(A|B, C) := P(A|B \cap C)$. Theo trực giác, điều này có nghĩa là xác suất của A không bị ảnh hưởng bởi B có xảy ra hay không, miễn là C xảy ra.

Định nghĩa 1.3.3 (Độc lập có điều kiện). Cho (Ω, F, P) là một không gian xác suất. hai sự kiện A, B F độc lập có điều kiện với biến cố thứ ba C F khi và chỉ khi

$$P(A \cap B|C) = P(A|C)\, P(B|C). \tag{1.48}$$

Ví dụ 1.3.4 (Đại hội (tiếp theo)). Yếu tố chính quyết định cách các thành viên của đại hội bỏ phiếu là liên kết chính trị. Do đó, chúng tôi kết hợp nó vào mô hình xác suất trong Ví dụ 1.3.2. Mỗi kết quả hiện bao gồm các phiếu bầu cho vấn đề 1 và 2, cũng như liên kết của thành viên, ví dụ: vấn đề 1 = có, vấn đề 2 = không, liên kết = cộng hòa hoặc vấn đề 1 = không, vấn đề 2 = không, liên kết = dân chủ. Đại số σ là tập lũy thừa của không gian mẫu. chúng tôi một lần nữa ước tính các giá trị của phép đo xác suất liên quan đến các sự kiện khác nhau bằng cách sử dụng dữ liệu:

$$P\text{ (vấn đề 1 = có|cộng hòa)} \approx \frac{\text{đảng cộng hòa bỏ phiếu đồng ý cho vấn đề 1}}{\text{tổng số phiếu cộng hòa về vấn đề 1}} \tag{1.49}$$

$$= 0{,}134, \tag{1.50}$$

$$P\text{ (vấn đề 2 = có|cộng hòa)} \approx \frac{\text{đảng cộng hòa bỏ phiếu đồng ý về vấn đề 2}}{\text{tổng số phiếu cộng hòa về vấn đề 2}} \tag{1.51}$$

$$= 0{,}988, \tag{1.52}$$

$$P\text{ (vấn đề 1 = có} \cap \text{vấn đề 2 = có|đảng cộng hòa)} \approx \frac{\text{đảng cộng hòa bỏ phiếu đồng ý cho các vấn đề 1 và 2}}{\text{đảng cộng hòa bỏ phiếu về cả hai vấn đề}}$$

$$= 0{,}134. \tag{1.53}$$

Dựa trên những dữ liệu này, chúng tôi có thể đánh giá liệu hành vi bỏ phiếu đối với hai vấn đề có phụ thuộc vào với điều kiện thành viên là một người cộng hòa. Nói cách khác, nếu chúng ta biết một thành viên đã bỏ phiếu như thế nào về vấn đề 1 và họ là người cộng hòa, điều này có cung cấp thông tin về cách họ bỏ phiếu không về vấn đề 2? Câu trả lời là không, vì

$$P\text{ (vấn đề 1 = có|đảng cộng hòa) } P\text{ (vấn đề 2 = có|đảng cộng hòa)} = 0{,}133 \tag{1.54}$$

rất gần với P (vấn đề 1 = có ∩ vấn đề 2 = có|đảng cộng hòa). Các phiếu bầu gần như độc lập khi biết rằng thành viên đó là một người cộng hòa.

Như được gợi ý trong Ví dụ 1.3.2 và 1.3.4, tính độc lập không bao hàm tính độc lập có điều kiện hoặc ngược lại. Điều này được minh họa thêm bằng các ví dụ sau. Từ giờ trở đi, để đơn giản hóa ký hiệu, chúng tôi viết xác suất giao nhau của một số sự kiện sau hình thức

$$P(A, B, C) := P(A \cap B \cap C). \tag{1.55}$$

Ví dụ 1.3.5 (Độc lập có điều kiện không bao hàm tính độc lập có điều kiện). Marvin anh họ của bạn từ Bài tập 1.2.6 luôn phàn nàn về taxi ở New York. Từ nhiều lần đến thăm JFK, anh ấy đã tính toán rằng

$$P(\text{taxi} \mid \text{mưa}) = 0{,}1, \quad P(\text{taxi} \mid \text{không mưa}) = 0{,}6, \tag{1.56}$$

trong đó taxi biểu thị sự kiện tìm được một chiếc taxi miễn phí sau khi lấy hành lý của bạn. đưa ra các sự kiện mưa và không mưa, sẽ hợp lý khi lập mô hình các sự kiện máy bay đến muộn và taxi là độc lập có điều kiện

$$P(\text{taxi}, \text{muộn} \mid \text{mưa}) = P(\text{taxi} \mid \text{mưa})\, P(\text{muộn} \mid \text{mưa}), \tag{1.57}$$

$$P(\text{taxi}, \text{muộn} \mid \text{không mưa}) = P(\text{taxi} \mid \text{không mưa})\, P(\text{muộn} \mid \text{không mưa}). \tag{1.58}$$

Logic đằng sau điều này là sự sẵn có của taxi sau khi nhận hành lý của bạn phụ thuộc vào trời có mưa hay không, chứ không phải việc máy bay có đến muộn hay không (chúng tôi giả định rằng tính khả dụng không đổi trong suốt cả ngày). Giả định này có ngụ ý rằng các sự kiện là độc lập?

Nếu họ độc lập thì biết cô bạn đến trễ sẽ không cung cấp thông tin gì cho Marvin về sự sẵn có của taxi. Tuy nhiên,

$$\begin{aligned} P(\text{taxi}) &= P(\text{taxi}, \text{mưa}) + P(\text{taxi}, \text{không mưa}) \quad \text{(theo quy luật tổng xác suất)} && (1.59)\\ &= P(\text{taxi} \mid \text{mưa})\, P(\text{mưa}) + P(\text{taxi} \mid \text{không mưa})\, P(\text{không mưa}) = && (1.60)\\ &0{,}1 \cdot 0{,}2 + 0{,}6 \cdot 0{,}8 = 0{,}5, && (1.61) \end{aligned}$$

$$\begin{aligned} P(\text{taxi} \mid \text{muộn}) &= \frac{P(\text{taxi}, \text{muộn}, \text{mưa}) + P(\text{taxi}, \text{muộn}, \text{không mưa})}{P(\text{muộn})} \quad \text{(theo định luật xác suất toàn phần)}\\ &= \frac{P(\text{taxi} \mid \text{mưa})\, P(\text{muộn} \mid \text{mưa})\, P(\text{mưa}) + P(\text{taxi} \mid \text{không mưa})\, P(\text{muộn} \mid \text{không mưa})\, P(\text{không mưa})}{\text{Đĩa)}}\\ &= \frac{0{,}1 \cdot 0{,}75 \cdot 0{,}2 + 0{,}6 \cdot 0{,}125 \cdot 0{,}8}{0{,}25} = 0{,}3. && (1.62) \end{aligned}$$

$P(\text{taxi}) = P(\text{taxi} \mid \text{muộn})$ nên các sự kiện không độc lập. Điều này có ý nghĩa, vì nếu máy bay trễ, trời mưa nhiều khả năng sẽ khó tìm taxi hơn.

Ví dụ 1.3.6 (Độc lập không bao hàm tính độc lập có điều kiện). Sau khi nhìn vào của bạn mô hình xác suất từ Ví dụ 1.2.1 liên hệ của bạn tại JFK chỉ ra rằng sự chậm trễ thường gây ra bởi các vấn đề cơ khí trong máy bay. Bạn nhìn vào dữ liệu và xác định rằng

$$P(\text{vấn đề}) = P(\text{vấn đề} \mid \text{mưa}) = P(\text{vấn đề} \mid \text{không mưa}) = 0{,}1, \tag{1.63}$$

do đó, sự cố máy móc và mưa ở NYC là độc lập, điều này tạo nên sự trực quan giác quan. Sau khi phân tích thêm dữ liệu, bạn ước tính

$$P\text{ (muộn}|\text{có vấn đề)} = 0{,}7,\ P\text{ (muộn}|\text{không có vấn đề)} = 0{,}2,\ P\text{ (muộn}|\text{không có mưa, có vấn đề)} = 0{,}5.$$

Lần tới khi bạn đợi Marvin ở JFK, bạn bắt đầu tự hỏi về khả năng máy bay của anh ta đã có một số vấn đề máy móc. Không có thêm bất kỳ thông tin nào, xác suất này là 0,1. Đó là một ngày nắng ở New York, nhưng điều này không giúp được gì vì theo dữ liệu (và lẽ thường) vấn đề sự kiện và mưa là độc lập.

Đột nhiên họ thông báo rằng máy bay của Marvin bị trễ. Bây giờ, xác suất mà anh ấy máy bay có một vấn đề máy móc? Lúc đầu, bạn nghĩ rằng bạn có thể áp dụng quy tắc Bayes để tính toán P (vấn đề|muộn) = 0,28 như trong Ví dụ 1.2.6. Tuy nhiên, bạn không sử dụng thực tế là nó nhiều nắng. Điều này có nghĩa là mưa không phải là nguyên nhân gây ra sự chậm trễ, do đó, theo trực giác, một cơ học vấn đề nên có nhiều khả năng. Thực vậy,

$$
\begin{aligned}
P\text{ (vấn đề | muộn, không mưa)} &= \frac{P\text{ (trễ, không mưa, có vấn đề)}}{P\text{ (muộn, không mưa)}} && (1.64)\\
&= \frac{P\text{ (muộn}|\text{không mưa, có vấn đề) } P\text{ (không có mưa) } P\text{ (có vấn đề)}}{P\text{ (trễ}|\text{không mưa) } P\text{ (không mưa)}} && \text{(theo Quy tắc Chuỗi)}\\
&= \frac{0{,}5 \cdot 0{,}1}{0{,}125} = 0{,}4. && (1.65)
\end{aligned}
$$

Vì P (sự cố|muộn, không mưa) = P (sự cố|muộn) nên sự cố máy móc và mưa trong NYC không độc lập về mặt điều kiện do máy bay sự kiện đến muộn.

chương 2

Biến ngẫu nhiên

Các biến ngẫu nhiên là một công cụ cơ bản trong mô hình xác suất. Chúng cho phép chúng ta lập mô hình các đại lượng số không chắc chắn: nhiệt độ ở New York vào ngày mai, thời gian chuyến bay đến, vị trí của vệ tinh... Suy luận về các đại lượng đó theo xác suất cho phép chúng ta cấu trúc thông tin chúng ta có về chúng theo cách một cách có nguyên tắc.

2.1 Định nghĩa

Chính thức, chúng tôi định nghĩa một biến ngẫu nhiên là một hàm ánh xạ từng kết quả trong một không gian xác suất thành một số thực.

Định nghĩa 2.1.1 (Biến ngẫu nhiên). Cho một không gian xác suất $(\Omega, \mathcal{F}, P)$, biến ngẫu nhiên X là một hàm từ không gian mẫu Ω đến các số thực $\mathbb{R}$. Khi kết quả $\omega \in \Omega$ của thí nghiệm được tiết lộ, $X(\omega)$ tương ứng được biết như một hiện thực của biến ngẫu nhiên.

Nhận xét 2.1.2 (Định nghĩa chặt chẽ). Nếu chúng ta muốn hoàn toàn nghiêm ngặt, Định nghĩa 2.1.1 thiếu một số chi tiết. Xét hai không gian mẫu Ω_1 và Ω_2, và σ-đại số $\mathcal{F}_2$ của các tập hợp trong Ω_2. Khi đó, để X là một biến ngẫu nhiên, phải tồn tại một σ-đại số $\mathcal{F}_1$ trong Ω_1 sao cho với bất kỳ tập S nào trong $\mathcal{F}_2$ là ảnh nghịch đảo của S, được xác định bởi

$$X^{-1}(S) := \{\omega \mid X(\omega) \in S\}, \tag{2.1}$$

thuộc $\mathcal{F}_1$. Thông thường, chúng ta coi Ω_2 là số thực $\mathbb{R}$ và $\mathcal{F}_2$ là σ-đại số Borel, được định nghĩa là σ-đại số nhỏ nhất được xác định trên các số thực chứa tất cả các khoảng mở (điều đáng ngạc nhiên là có thể xây dựng các tập hợp số thực không thuộc σ-đại số này). Trong mọi trường hợp, với mục đích của các ghi chú này, Định nghĩa 2.1.1 là đủ (có thể tìm thêm thông tin về các cơ sở chính thức của xác suất trong bất kỳ cuốn sách nào về lý thuyết độ đo và lý thuyết xác suất nâng cao).

Ghi chú 2.1.3 (Ký hiệu). Ta thường ký hiệu các biến cố có dạng

$$\{X(\omega) \in S : \omega \in \Omega\} \tag{2.2}$$

đối với một số biến ngẫu nhiên X và một số đặt S là

$$\{X \in S\} \tag{2.3}$$

để giảm bớt ký hiệu, vì không gian xác suất cơ bản thường không có ý nghĩa một khi chúng tôi đã chỉ định các biến ngẫu nhiên quan tâm.

Một biến ngẫu nhiên định lượng sự không chắc chắn của chúng ta về đại lượng mà nó đại diện, chứ không phải giá trị mà nó cuối cùng sẽ nhận được sau khi kết quả được tiết lộ. Bạn đừng bao giờ nghĩ một biến ngẫu nhiên có một giá trị số cố định. Nếu kết quả được biết trước, thì điều đó quyết định việc thực hiện biến ngẫu nhiên. Để nhấn mạnh sự khác biệt giữa các biến ngẫu nhiên và việc thực hiện chúng, chúng tôi biểu thị cái trước bằng chữ in hoa (X, Y . . .) và cái sau bằng chữ thường (x, y, . . .). ,

Nếu chúng ta có quyền truy cập vào không gian xác suất (Ω, F, P) trong đó biến ngẫu nhiên được xác định, thì việc tính xác suất của một biến ngẫu nhiên X thuộc một tập hợp S nào đó rất đơn giản: đó là xác suất của[1] biến cố bao gồm tất cả các kết quả trong Ω mà X ánh xạ tới S,

$$P(X \quad S) = P(\{\omega \mid X(\omega) \quad S\}). \tag{2.4}$$

Tuy nhiên, hầu như chúng ta không bao giờ lập mô hình trực tiếp không gian xác suất, vì điều này đòi hỏi phải ước tính xác suất của mọi sự kiện có thể xảy ra trong đại số σ tương ứng. Như chúng tôi giải thích trong Phần 2.2 và 2.3, có nhiều phương pháp thực tế hơn để chỉ định các biến ngẫu nhiên, các phương pháp này tự động ngụ ý rằng tồn tại một không gian xác suất cơ sở hợp lệ. Sự tồn tại của không gian xác suất này đảm bảo rằng toàn bộ khuôn khổ hợp lý về mặt toán học, nhưng bạn không thực sự phải lo lắng về nó.

2.2 Biến ngẫu nhiên rời rạc

Các biến ngẫu nhiên rời rạc nhận các giá trị trên tập con hữu hạn hoặc vô hạn đếm được của R chẳng hạn như các số nguyên. Chúng được sử dụng để mô hình hóa các đại lượng số rời rạc: kết quả của một lần tung xúc xắc, tỷ số trong một trận đấu bóng rổ, v.v.

2.2.1 Hàm khối lượng xác suất

Để chỉ định một biến ngẫu nhiên rời rạc, chỉ cần xác định xác suất của từng giá trị mà nó có thể nhận. Trái ngược với trường hợp các biến ngẫu nhiên liên tục, điều này có thể kiểm soát được vì các giá trị này có thể đếm được theo định nghĩa.

Định nghĩa 2.2.1 (Hàm khối lượng xác suất). Đặt (Ω, F, P) là một không gian xác suất và X : Ω Z một biến ngẫu nhiên. Hàm khối lượng xác suất (pmf) của X được định nghĩa là

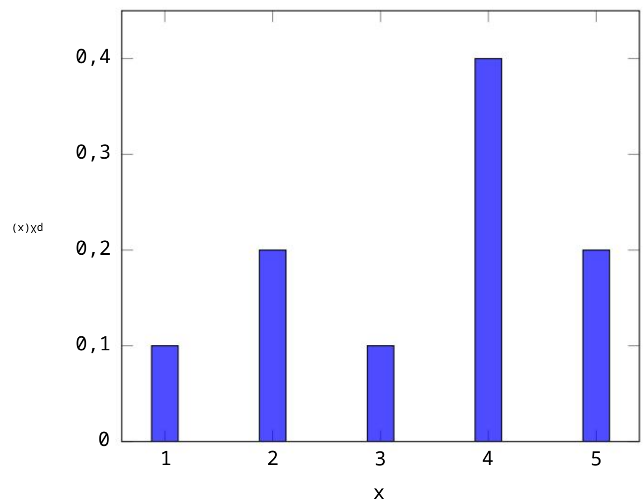
$$p_X(x) := P(\{\omega \mid X(\omega) = x\}). \tag{2.5}$$

Nói một cách dễ hiểu, $p_X(x)$ là xác suất để X bằng x.

Chúng ta thường nói rằng một biến ngẫu nhiên được phân phối theo một pmf nhất định.

Nếu ký hiệu khoảng rời rạc của X là D thì bộ ba D, 2 D, pX là không gian xác suất hợp lệ (nhớ lại rằng 2D là tập lũy thừa của D). Đặc biệt, px là thước đo xác suất hợp lệ

1Nói một cách chính xác, S cần phải thuộc về đại số Borel σ. Một lần nữa, về cơ bản, điều này bao gồm bất kỳ tập hợp con nào của số thực mà bạn sẽ gặp trong mô hình xác suất.



Hình 2.1: Hàm khối lượng xác suất của biến ngẫu nhiên X trong Ví dụ 2.2.2.

thỏa mãn

$$p_X(x) \geq 0 \text{ với mọi } x \in D, \tag{2.6}$$

$$\sum_{x \in D} p_X(x) = 1. \tag{2.7}$$

Điều ngược lại cũng đúng nếu một hàm xác định trên tập con đếm được D của các số thực là không âm và cộng lại thành một thì có thể hiểu là pmf của một biến ngẫu nhiên. Trên thực tế, trong thực tế, chúng tôi thường xác định các biến ngẫu nhiên rời rạc bằng cách chỉ định pmf của chúng.

Để tính xác suất để một biến ngẫu nhiên X nằm trong một tập hợp S nhất định, chúng ta lấy tổng của pmf trên tất cả các giá trị có trong S:

$$P(X \in S) = \sum_{x \in S} p_X(x). \tag{2.8}$$

Ví dụ 2.2.2 (Biến ngẫu nhiên rời rạc). Hình 2.1 cho thấy hàm khối lượng xác suất của một biến ngẫu nhiên rời rạc X (kiểm tra xem nó có cộng bằng một không). Để tính xác suất của X thuộc các tập hợp khác nhau ta áp dụng (2.8):

$$P(X \in \{1, 4\}) = p_X(1) + p_X(4) = 0{,}5, \tag{2.9}$$

$$P(X > 3) = p_X(4) + p_X(5) = 0{,}6. \tag{2.10}$$

2.2.2 Các biến ngẫu nhiên rời rạc quan trọng

Trong phần này, chúng tôi mô tả một số biến ngẫu nhiên rời rạc hữu ích cho xác suất người mẫu.

Bernoulli

Các biến ngẫu nhiên Bernoulli được sử dụng để mô hình hóa các thí nghiệm có hai kết quả có thể xảy ra. Theo quy ước, chúng ta thường biểu diễn một kết quả bằng 0 và kết quả kia bằng 1. ví dụ là tung một đồng xu bị thiên vị, sao cho xác suất xuất hiện mặt ngửa là p. Nếu chúng ta mã hóa đầu là 1 và đuôi là 0, khi đó kết quả tung đồng xu tương ứng với ngẫu nhiên Bernoulli biến có tham số p.

Định nghĩa 2.2.3 (Bernoulli). pmf của biến ngẫu nhiên Bernoulli với tham số $p \in [0, 1]$ được đưa ra bởi

$$p_X(0) = 1 - p, \quad p_X \tag{2.11}$$

$$(1) = p. \tag{2.12}$$

Một loại biến ngẫu nhiên Bernoulli đặc biệt là biến ngẫu nhiên chỉ thị của một biến cố. Cái này biến ngẫu nhiên đặc biệt hữu ích trong chứng minh.

Định nghĩa 2.2.4 (Chỉ báo). Cho $(\Omega, \mathcal{F}, P)$ là một không gian xác suất. Chỉ số biến ngẫu nhiên có thể của một sự kiện $S \in \mathcal{F}$ được định nghĩa là

$$1_S(\omega) = \begin{cases} 1, \text{ nếu } \omega \in S, \\ 0, \text{ ngược lại.} \end{cases} \tag{2.13}$$

Theo định nghĩa phân phối của một biến ngẫu nhiên chỉ thị là Bernoulli với tham số P(S).

hình học

Hãy tưởng tượng rằng chúng ta lấy một đồng xu thiên vị và lật nó cho đến khi chúng ta có được mặt ngửa. Nếu xác suất của có được mặt ngửa là p và các lần lật là độc lập thì xác suất phải lật k lần là

$$P(k \text{ lần lật}) = P(\text{Lật lần } 1 = \text{sấp}, \ldots, k - \text{Lần lật } 1 = \text{sấp, lần lật thứ } k = \text{ngửa}) \tag{2.14}$$

$$= P(\text{lần lật đầu tiên} = \text{sấp}) \cdots P(k - \text{lần lật thứ } 1 = \text{sấp})\, P(\text{lần lật thứ } k = \tag{2.15}$$

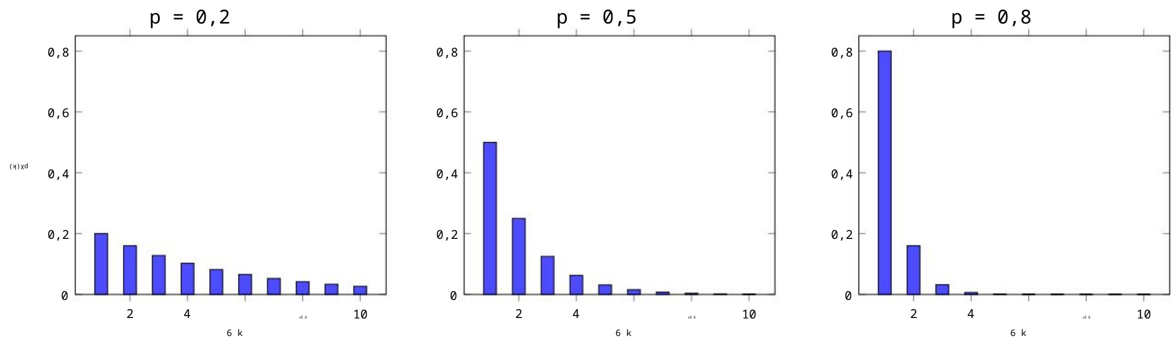
$$\text{ngửa}) = (1 - p)^{k-1} p. \tag{2.16}$$

Lập luận này có thể được áp dụng cho bất kỳ tình huống nào trong đó một thí nghiệm ngẫu nhiên với khả năng xác suất cố định p được lặp lại cho đến khi một kết quả cụ thể xảy ra, miễn là giả định độc lập được đáp ứng. Trong những trường hợp như vậy, số lần lặp lại được mô hình hóa như một biến ngẫu nhiên hình học.

Định nghĩa 2.2.5 (Hình học). pmf của một biến ngẫu nhiên hình học với tham số p là được cho bởi

$$p_X(k) = (1 - p)^{k-1} p, \quad k = 1, 2, \ldots \tag{2.17}$$

Hình 2.2 biểu diễn hàm khối lượng xác suất của các biến ngẫu nhiên hình học với các thông số. p càng lớn thì phân phối càng tập trung xung quanh các giá trị k nhỏ hơn.



Hình 2.2: Hàm khối lượng xác suất của ba biến ngẫu nhiên hình học với các tham số khác nhau.

nhị thức

Các biến ngẫu nhiên nhị thức cực kỳ hữu ích trong mô hình xác suất. Họ đã quen lập mô hình số kết quả tích cực của n thử nghiệm được lập mô hình là ngẫu nhiên Bernoulli độc lập các biến có cùng tham số. Ví dụ sau minh họa điều này với việc tung đồng xu.

Ví dụ 2.2.6 (Tung đồng xu). Nếu chúng ta tung một đồng xu bị thiên vị n lần, xác suất mà chúng ta có được chính xác k mặt ngửa nếu các lần lật là độc lập và xác suất mặt ngửa là p?

Trước tiên chúng ta hãy xem xét một vấn đề đơn giản hơn: xác suất để lần đầu tiên nhận được k mặt ngửa là bao nhiêu và thì n k đuôi? Bằng sự độc lập, câu trả lời là

$$P\,(k \text{ mặt ngửa, sau đó } n \quad k \text{ mặt sấp}) \tag{2.18}$$

$$= P\,(\text{Lật đầu tiên = ngửa}, \ldots, \text{lật thứ } k = \text{ngửa}, \text{lật thứ } k+1 = \text{sấp}, \ldots, \text{lật thứ } n = \text{sấp}) =$$

$$P\,(\text{lần lật thứ 1 = ngửa}) \cdots P\,(\text{lần lật thứ } k = \text{ngửa})\; P\,(\text{lần lật thứ } k+1 = \text{sấp}) \cdots P\,(\text{lần lật thứ } n = \text{sấp})$$

$$= p^{k}\,(1-p)^{n\quad k}\,. \tag{2.19}$$

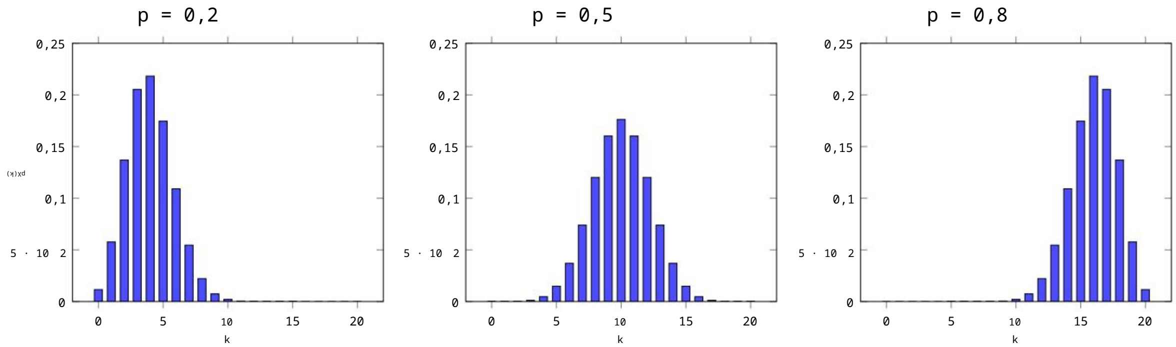
Lưu ý rằng lý do tương tự ngụ ý rằng đây cũng là xác suất để có được chính xác k đứng đầu trong bất kỳ thứ tự cố định. Xác suất lấy được chính xác k mặt ngửa là hợp của tất cả các mặt ngửa này sự kiện. Bởi vì những sự kiện này rời rạc (chúng ta không thể có được chính xác k mặt ngửa ở hai đơn đặt hàng đồng thời) chúng ta có thể thêm cá nhân của họ để tính xác suất của sự kiện của chúng ta quan tâm. Chúng tôi chỉ cần biết số lượng các đơn đặt hàng có thể. Bằng tổ hợp cơ bản, điều này được đưa ra bởi hệ số nhị thức được định $\binom{N}{k}$ nghĩa là

$$\binom{N}{k} := \frac{N!}{k!\,(n \quad k)!}. \tag{2.20}$$

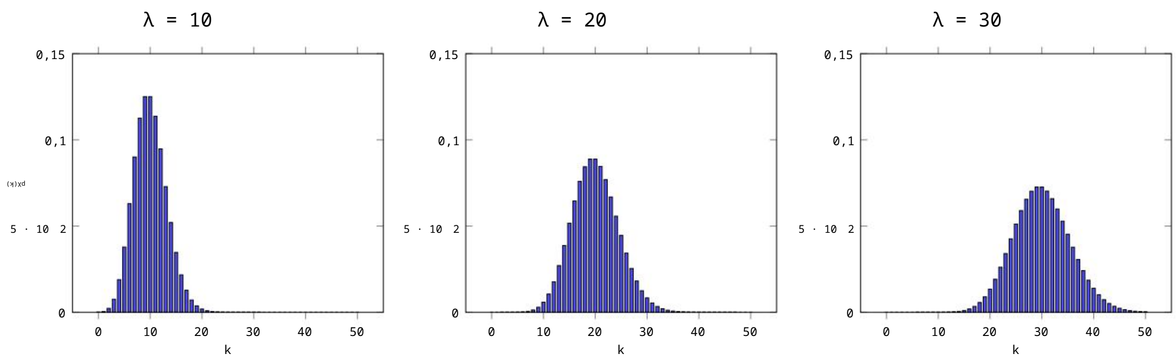
Chúng tôi kết luận rằng

$$P\,(k \text{ ra khỏi } n \text{ lần lật}) = \binom{N}{k} P^{k}\,(1-p)^{(n\quad k)}\,. \tag{2.21}$$

Biến ngẫu nhiên đại diện cho số mặt ngửa trong ví dụ được gọi là nhị thức biến ngẫu nhiên.



Hình 2.3: Hàm khối lượng xác suất của ba biến ngẫu nhiên nhị thức với các giá trị khác nhau của p và n=20.



Hình 2.4: Hàm khối lượng xác suất của ba biến ngẫu nhiên Poisson với các tham số khác nhau.

Định nghĩa 2.2.7 (Nhị thức). pmf của một biến ngẫu nhiên nhị thức với các tham số n và p được đưa ra bởi

$$p_X(k) = \binom{N}{k} p^k (1-p)^{(n-k)}, \quad k = 0, 1, 2, \ldots, N. \tag{2.22}$$

Hình 2.3 biểu diễn hàm khối lượng xác suất của các biến ngẫu nhiên nhị thức với các giá trị khác nhau của p.

Poisson

Chúng tôi thúc đẩy định nghĩa của biến ngẫu nhiên Poisson bằng một ví dụ.

Ví dụ 2.2.8 (Tổng đài). Một trung tâm cuộc gọi muốn mô hình hóa số lượng cuộc gọi họ nhận được hơn một ngày để quyết định thuê bao nhiêu người. Họ đưa ra các giả định sau:

1. Mỗi cuộc gọi diễn ra độc lập với mọi cuộc gọi khác.
2. Một cuộc gọi nhất định có xác suất xảy ra như nhau vào bất kỳ thời điểm nào trong ngày.
3. Các cuộc gọi diễn ra với tốc độ λ cuộc gọi mỗi ngày.

Machine Translated by Google

Trong Chương 5, chúng ta sẽ thấy rằng những giả định này xác định quy trình Poisson.

Mục đích của chúng tôi là tính xác suất nhận được chính xác k cuộc gọi trong một ngày nhất định. Để làm điều này, chúng ta rời rạc hóa ngày thành n khoảng thời gian, tính toán xác suất mong muốn giả sử mỗi khoảng thời gian là rất nhỏ và sau đó đặt $n \to \infty$.

Xác suất để một cuộc gọi xảy ra trong khoảng thời gian có độ dài 1/n là λ/n theo Giả định 2 và 3. Xác suất xảy ra m > 1 cuộc gọi là $(\lambda/n)^m$. Nếu n rất lớn thì xác suất này không đáng kể so với xác suất nhận được một hoặc 0 cuộc gọi trong khoảng thời gian, trên thực tế, nó có xu hướng bằng 0 khi chúng ta lấy giới hạn $n \to \infty$. Do đó, tổng số cuộc gọi xảy ra trong cả giờ có thể được xấp xỉ bằng số khoảng thời gian mà một cuộc gọi xảy ra, miễn là n đủ lớn. Vì một cuộc gọi xảy ra trong mỗi khoảng thời gian với cùng xác suất và các cuộc gọi xảy ra độc lập, số lượng cuộc gọi trong cả ngày có thể được mô hình hóa dưới dạng biến ngẫu nhiên nhị thức với các tham số n và $p := \lambda/n$.

Bây giờ chúng tôi tính toán phân phối các cuộc gọi khi các khoảng thời gian nhỏ tùy ý, tức là khi $n \to \infty$:

$$
\begin{aligned}
P(k \text{ cuộc gọi trong ngày}) &= \lim_{n\to\infty} P(k \text{ cuộc gọi trong } n \text{ khoảng thời gian nhỏ}) &(2.23)\\
&= \lim_{n\to\infty} \binom{N}{k} p^k (1-p)^{(n-k)} &(2.24)\\
&= \lim_{n\to\infty} \binom{N}{k} \left(\frac{\lambda}{N}\right)^k \left(1-\frac{\lambda}{N}\right)^{(n-k)} &(2.25)\\
&= \lim_{n\to\infty} \frac{n!\,\lambda^k}{k!\,(n-k)!\,(n-\lambda)^k}\left(1-\frac{\lambda}{N}\right)^N &(2.26)\\
&= \frac{\lambda^k e^{-\lambda}}{k!}. &(2.27)
\end{aligned}
$$

Bước cuối cùng bắt nguồn từ bổ đề sau được chứng minh trong Mục 2.7.1.

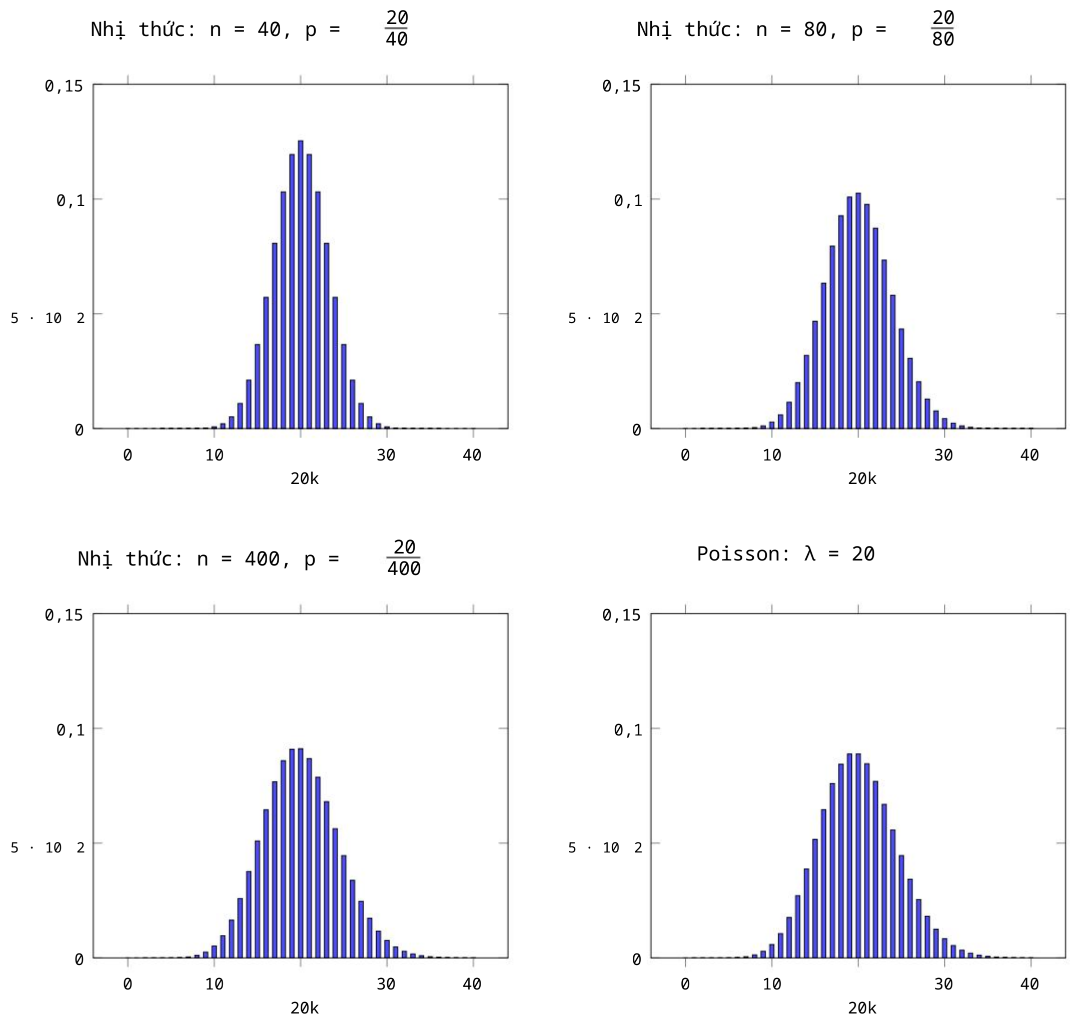
Bổ đề 2.2.9.

$$\lim_{n\to\infty} \frac{N!}{(n-k)!\,(n-\lambda)^k}\left(1-\frac{\lambda}{N}\right)^N = e^{-\lambda}. \tag{2.28}$$

Biến ngẫu nhiên với pmf mà chúng ta đã dẫn xuất trong ví dụ được gọi là biến ngẫu nhiên Poisson. Chúng được sử dụng để mô hình hóa các tình huống trong đó thỉnh thoảng có điều gì đó xảy ra với tốc độ không đổi: các gói đến bộ định tuyến Internet, động đất, tai nạn giao thông, v.v. Số lượng các sự kiện như vậy xảy ra trong một khoảng thời gian cố định tuân theo phân phối Poisson, miễn là như các giả định mà chúng tôi liệt kê trong ví dụ.

Định nghĩa 2.2.10 (Poisson). PMf của biến ngẫu nhiên Poisson với tham số λ được cho bởi

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots \tag{2.29}$$

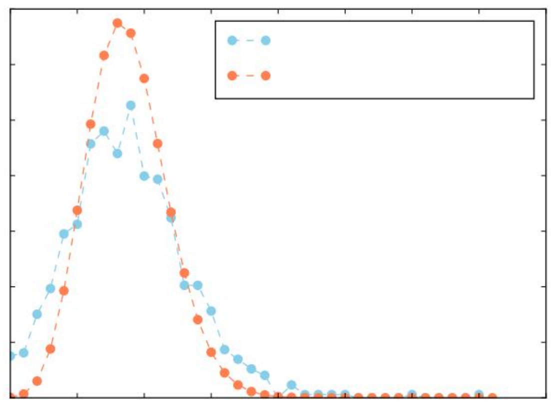


Hình 2.5: Sự hội tụ của pmf nhị thức với $p = \lambda/n$ thành một pmf Poisson của tham số λ khi n tăng lên.

Hình 2.4 biểu diễn hàm khối lượng xác suất của các biến ngẫu nhiên Poisson với các giá trị khác nhau của λ. Trong ví dụ 2.2.8, chúng ta chứng minh rằng khi n ∞ pmf của một biến ngẫu nhiên nhị thức với tham số n và λ/n có xu hướng pmf của Poisson với tham số λ (như chúng ta sẽ thấy sau trong nhiên, đây là một ví dụ về sự hội tụ trong phân phối). Hình 2.5 cho thấy một ví dụ về hiện tượng này về số lượng; hội tụ khá nhanh.

Bạn có thể cảm thấy hơi nghi ngờ về Ví dụ 2.2.8: xác suất nhận cuộc gọi chắc chắn thay đổi trong ngày và nó phải khác vào cuối tuần! Điều đó đúng, nhưng mô hình là thực sự rất hữu ích nếu chúng ta hạn chế sự chú ý của mình trong khoảng thời gian ngắn hơn. Trong Hình 2.6 , chúng tôi chỉ ra kết quả của việc lập mô hình số lượng cuộc gọi mà một trung tâm cuộc gọi ở Israel[2] nhận được trong khoảng thời gian bốn giờ (8 giờ tối đến nửa đêm) sử dụng biến ngẫu nhiên Poisson. Chúng tôi vẽ biểu đồ của số cuộc gọi nhận được trong khoảng thời gian đó trong hai tháng (tháng 9 và tháng 10 năm 1999) cùng với một pmf Poisson phù hợp với dữ liệu (chúng ta sẽ tìm hiểu cách điều chỉnh phân phối phù hợp với dữ liệu sau này trong khóa học). Mặc dù thực tế là các giả định của chúng ta không đúng, nhưng mô hình

[2]Dữ liệu có sẵn ở đây.



Hình 2.6: Trong màu xanh lam, chúng ta thấy biểu đồ số lượng cuộc gọi nhận được trong khoảng thời gian bốn giờ trong hai tháng tại một trung tâm cuộc gọi ở Israel. Một pmf Poisson gần đúng với phân phối dữ liệu được vẽ bằng màu cam.

tạo ra một sự phù hợp hợp lý tốt.

2.3 Biến ngẫu nhiên liên tục

Các đại lượng vật lý thường được mô tả tốt nhất là liên tục: nhiệt độ, khoảng thời gian, tốc độ, trọng lượng, v.v. Để mô hình hóa các đại lượng đó theo xác suất, chúng ta có thể rời rạc hóa miền của chúng và biểu diễn chúng dưới dạng các biến ngẫu nhiên rời rạc. Tuy nhiên, chúng tôi có thể không muốn kết luận của mình phụ thuộc vào cách chúng tôi chọn lưới rời rạc. Việc xây dựng một mô hình liên tục cho phép chúng tôi có được những hiểu biết sâu sắc có giá trị đối với các lưới đủ mịn mà không phải lo lắng về sự rời rạc.

Chính vì các miền liên tục mô hình hóa giới hạn khi các kết quả rời rạc có độ chi tiết nhỏ tùy ý, nên chúng ta không thể mô tả hành vi xác suất của một biến ngẫu nhiên liên tục bằng cách chỉ đặt các giá trị cho xác suất X bằng với các kết quả riêng lẻ, như cách chúng ta làm đối với các biến ngẫu nhiên rời rạc . Trên thực tế, chúng ta không thể gán xác suất khác 0 cho kết quả cụ thể của một đại lượng liên tục không chắc chắn. Điều này sẽ dẫn đến các kết quả rời rạc không thể đếm được với xác suất khác không. Tổng của vô số giá trị dương không thể đếm được là vô hạn, vì vậy xác suất hợp của chúng sẽ lớn hơn một, điều này không có ý nghĩa gì.

Nghiêm túc hơn, hóa ra chúng ta không thể xác định một độ đo xác suất hợp lệ trên tập lũy thừa của R (việc chứng minh điều này đòi hỏi phải có lý thuyết độ đo và nằm ngoài phạm vi của các ghi chú này). Thay vào đó, chúng ta xem xét các sự kiện bao gồm các hợp của các khoảng R. Các sự kiện như vậy tạo thành một σ-đại số được gọi là σ-đại số Borel. Đại số σ này đủ chi tiết để biểu diễn bất kỳ tập hợp nào mà bạn có thể quan tâm (hãy thử nghĩ về một tập hợp không thể biểu thị dưới dạng hợp đếm được của các khoảng), đồng thời cho phép xác định các độ đo xác suất hợp lệ trên tập hợp đó.

2.3.1 Hàm phân phối tích lũy

Để xác định một biến ngẫu nhiên trên Borel σ-đại số, chỉ cần xác định xác suất của biến ngẫu nhiên thuộc mọi khoảng có dạng $(-\infty, x)$ với mọi $x \in \mathbb{R}$.

Định nghĩa 2.3.1 (Hàm phân phối tích lũy). Cho (Ω, F, P) là một không gian xác suất và $X : \Omega \to \mathbb{R}$ một biến ngẫu nhiên. Hàm phân phối tích lũy (cdf) của X được định nghĩa là

$$F_X(x) := P(X \leq x). \tag{2.30}$$

Nói một cách dễ hiểu, $F_X(x)$ là xác suất để X nhỏ hơn x.

Lưu ý rằng hàm phân phối tích lũy có thể được xác định cho cả liên tục và rời rạc các biến ngẫu nhiên.

Bổ đề sau mô tả một số thuộc tính cơ bản của cdf. Bạn có thể tìm thấy bằng chứng trong Mục 2.7.2.

Bổ đề 2.3.2 (Tính chất của cdf). Với mọi biến ngẫu nhiên liên tục X

$$\lim_{x \to -\infty} F_X(x) = 0, \tag{2.31}$$

$$\lim_{x \to \infty} F_X(x) = 1, \tag{2.32}$$

$$F_X(b) \geq F_X(a) \text{ nếu } b > a, \text{ tức là } F_X \text{ không giảm.} \tag{2.33}$$

Để xem lý do tại sao cdf xác định hoàn toàn việc thu hồi biến ngẫu nhiên mà chúng tôi chỉ đang xem xét các tập hợp có thể biểu diễn dưới dạng hợp của các khoảng. Xác suất của biến ngẫu nhiên X thuộc một khoảng (a, b] được cho bởi

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) \tag{2.34}$$

$$= F_X(b) - F_X(a). \tag{2.35}$$

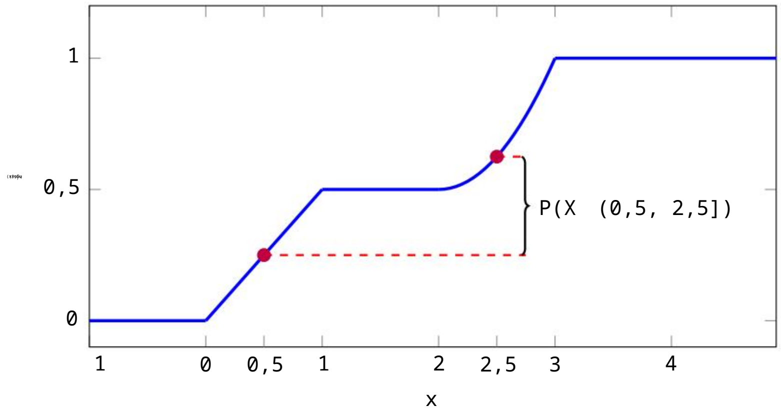
Nhận xét 2.3.3. Vì các điểm riêng lẻ có xác suất bằng không, nên với bất kỳ biến ngẫu nhiên liên tục nào X

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b). \tag{2.36}$$

Bây giờ, để tìm xác suất X thuộc bất kỳ tập hợp cụ thể nào, chúng ta chỉ cần phân rã nó thành các khoảng rời rạc và áp dụng (2.35), như minh họa bằng ví dụ sau.

Ví dụ 2.3.4 (Biến ngẫu nhiên liên tục). Xét biến ngẫu nhiên liên tục X với một cdf được cung cấp bởi

$$F_X(x) := \begin{cases} 0 & \text{cho } x < 0, \\ 0,5\,x & \text{cho } 0 \leq x \leq 1, \\ 0,5 & \text{cho } 1 \leq x \leq 2, \\ 0,5\left(1 + (x-2)^2\right) & \text{cho } 2 \leq x \leq 3, \\ 1 & \text{cho } x > 3. \end{cases} \tag{2.37}$$



Hình 2.7: Hàm phân phối tích lũy của biến ngẫu nhiên trong Ví dụ 2.3.4 và 2.3.7.

Hình 2.7 hiển thị cdf ở hình bên trái. Bạn có thể kiểm tra xem nó có thỏa mãn các thuộc tính trong Bổ đề 2.3.2. Để xác định xác suất X nằm trong khoảng từ 0,5 đến 2,5, chúng tôi áp dụng (2,35),

$$P(0{,}5 < X \leq 2{,}5) = F_X(2{,}5) - F_X(0{,}5) = 0{,}375, \tag{2.38}$$

như minh họa trong Hình 2.7.

2.3.2 Hàm mật độ xác suất

Nếu cdf của một biến ngẫu nhiên liên tục là khả vi, đạo hàm của nó có thể được hiểu là một hàm mật độ. Mật độ này sau đó có thể được tích hợp để có được xác suất ngẫu nhiên biến thuộc về một khoảng hoặc hợp các khoảng (và do đó thuộc bất kỳ tập Borel nào).

Định nghĩa 2.3.5 (Hàm mật độ xác suất). Cho $X : \Omega \to \mathbb{R}$ là biến ngẫu nhiên với cdf F_X. Nếu F_X khả vi thì hàm mật độ xác suất hoặc pdf của X được định nghĩa là

$$f_X(x) := \frac{dF_X(x)}{dx}. \tag{2.39}$$

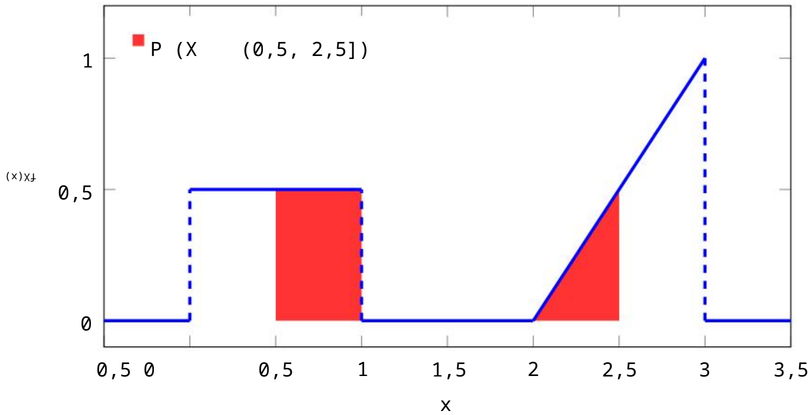
Theo trực giác, $f_X(x)\Delta$ là xác suất để X thuộc một khoảng có độ rộng Δ xung quanh x là $\Delta \to 0$. Theo định lý cơ bản của giải tích xác suất của biến ngẫu nhiên X thuộc một khoảng được cho bởi

$$P(a < X \leq b) = F_X(b) - F_X(a) \tag{2.40}$$

$$= \int_a^b f_X(x)\,dx. \tag{2.41}$$

Các tập hợp quan tâm của chúng ta thuộc về đại số σ Borel, và do đó có thể được phân tách thành các hợp của các khoảng, vì vậy chúng ta có thể thu được xác suất X thuộc bất kỳ tập S nào như vậy bằng cách lấy tích phân của nó pdf trên S

$$P(X \in S) = \int_S f_X(x)\,dx. \tag{2.42}$$



Hình 2.8: Hàm mật độ xác suất của biến ngẫu nhiên trong Ví dụ 2.3.4 và 2.3.7.

Đặc biệt, vì X thuộc R theo định nghĩa

$$\int_{-\infty}^{\infty} f_X(x)dx = P(X \in \mathbb{R}) = 1 . \tag{2.43}$$

Suy ra từ tính đơn điệu của cdf (2.33) mà pdf không âm

$$f_X(x) \geq 0, \tag{2.44}$$

vì ngược lại, chúng ta có thể tìm được hai điểm $x_1 < x_2$ mà $F_X(x_2) < F_X(x_1)$.

Ghi chú 2.3.6 (Bản pdf không phải là thước đo xác suất). PDF là một mật độ phải được tích hợp để mang lại một xác suất. Đặc biệt, nó không nhất thiết phải nhỏ hơn một (ví dụ: lấy a = 0 và b = 1/2 trong Định nghĩa 2.3.8 bên dưới).

Cuối cùng, giống như trong trường hợp các biến ngẫu nhiên rời rạc, chúng ta thường nói rằng một biến ngẫu nhiên là được phân phối theo một pdf hoặc cdf nhất định hoặc chúng tôi biết phân phối của nó. Lý do là rằng pmf, pdf hoặc cdf đủ để mô tả không gian xác suất cơ bản.

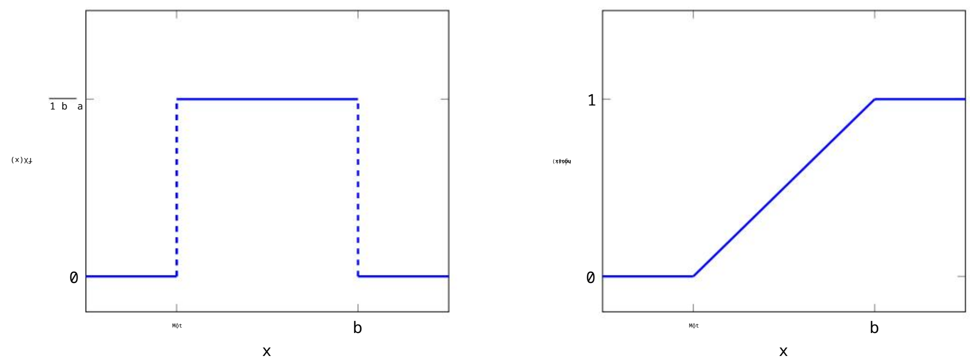
Ví dụ 2.3.7 (Biến ngẫu nhiên liên tục (tiếp theo)). Để tính toán pdf của ngẫu nhiên biến trong Ví dụ 2.3.4, chúng tôi phân biệt cdf của nó, để có được

$$f_X(x) = \begin{cases} 0 & \text{cho } x < 0, \\ 0{,}5 & \text{cho } 0 \leq x \leq 1 \\ 0 & \text{cho } 1 \leq x \leq 2 \\ x - 2 & \text{với } 2 \leq x \leq 3 \\ 0 & \text{cho } x > 3. \end{cases} \tag{2.45}$$

Hình 2.8 cho thấy bản pdf. Bạn có thể kiểm tra xem nó có tích hợp với một không. Để xác định xác suất rằng X nằm trong khoảng từ 0,5 đến 2,5, chúng ta chỉ cần lấy tích phân trên khoảng đó để có được câu trả lời tương tự như trong Ví dụ 2.3.4,

$$P(0{,}5 < X \leq 2{,}5) = \int_{0{,}5}^{2{,}5} f_X(x)\,dx \tag{2.46}$$

$$= \int_{0{,}5}^{1} 0{,}5\,dx + \int_{2}^{2{,}5} x - 2\,dx = 0{,}375. \tag{2.47}$$



Hình 2.9: Hàm mật độ xác suất (trái) và hàm phân phối tích lũy (phải) của biến ngẫu nhiên đều X.

Hình 2.8 minh họa rằng xác suất của một sự kiện bằng với diện tích dưới pdf khi chúng ta giới hạn nó trong tập con tương ứng của đường thực.

2.3.3 Các biến ngẫu nhiên liên tục quan trọng

Trong phần này, chúng tôi mô tả một số biến ngẫu nhiên liên tục hữu ích trong mô hình xác suất và thống kê.

Đồng phục

Một biến ngẫu nhiên thống nhất mô hình hóa một thí nghiệm trong đó mọi kết quả trong một khoảng thời gian liên tục đều có khả năng xảy ra như nhau. Kết quả là pdf không đổi trong khoảng thời gian. Hình 2.9 cho thấy pdf và cdf của một biến ngẫu nhiên thống nhất.

Định nghĩa 2.3.8 (Đồng phục). PDF của một biến ngẫu nhiên thống nhất với miền $[a, b]$, trong đó $b > a$ là các số thực, được cho bởi

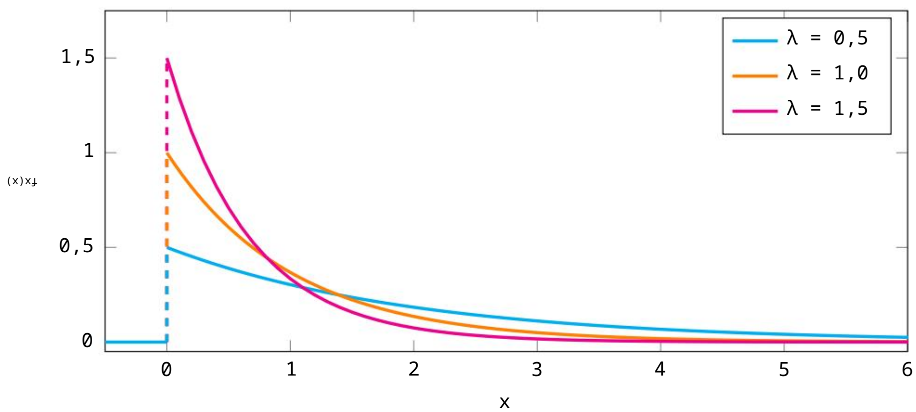
$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{nếu } a \le x \le b, \\ 0, & \text{ngược lại.} \end{cases} \tag{2.48}$$

số mũ

Các biến ngẫu nhiên hàm mũ thường được sử dụng để mô hình hóa thời gian trôi qua cho đến khi một sự kiện nhất định xảy ra. Các ví dụ bao gồm các hạt phóng xạ đang phân rã, các cuộc gọi điện thoại, động đất và nhiều thứ khác.

Định nghĩa 2.3.9 (Số mũ). PDF của biến ngẫu nhiên hàm mũ với tham số λ được cho bởi

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{nếu } x \ge 0, \\ 0, & \text{nếu không thì.} \end{cases} \tag{2.49}$$



Hình 2.10: Hàm mật độ xác suất của biến ngẫu nhiên hàm mũ với các tham số khác nhau.

Hình 2.10 cho thấy pdf của ba biến ngẫu nhiên hàm mũ với các tham số khác nhau. Để minh họa rằng tiềm năng của các phân phối theo hàm mũ để mô hình hóa dữ liệu thực, trong Hình 2.11 , chúng tôi vẽ biểu đồ về thời gian đến giữa các cuộc gọi tại cùng một trung tâm cuộc gọi ở Israel mà chúng tôi đã đề cập trước đó. Chi tiết hơn, thời gian giữa các lần đến là thời gian giữa các cuộc gọi liên tiếp xảy ra từ 8 giờ tối đến nửa đêm trong hai ngày vào tháng 9 năm 1999. Một mô hình hàm mũ khá phù hợp với dữ liệu.

Một thuộc tính quan trọng của biến ngẫu nhiên hàm mũ là nó không có bộ nhớ. Chúng tôi giải thích về thuộc tính này, được chia sẻ bởi phân bố hình học, trong Phần 2.4.

Gaussian hoặc bình thường

Biến ngẫu nhiên Gaussian hoặc bình thường được cho là biến ngẫu nhiên phổ biến nhất trong tất cả các xác suất và thống kê. Nó thường được sử dụng để mô hình hóa các biến có phân phối chưa biết trong khoa học tự nhiên. Điều này được thúc đẩy bởi thực tế là tổng của các biến ngẫu nhiên độc lập thường hội tụ về phân phối Gaussian. Hiện tượng này được mô tả bởi Định lý giới hạn trung tâm mà chúng ta sẽ thảo luận trong Chương 6.

Định nghĩa 2.3.10 (Gaussian). Bản pdf của biến ngẫu nhiên Gaussian hoặc thông thường với giá trị trung bình μ và độ lệch chuẩn σ được cho bởi

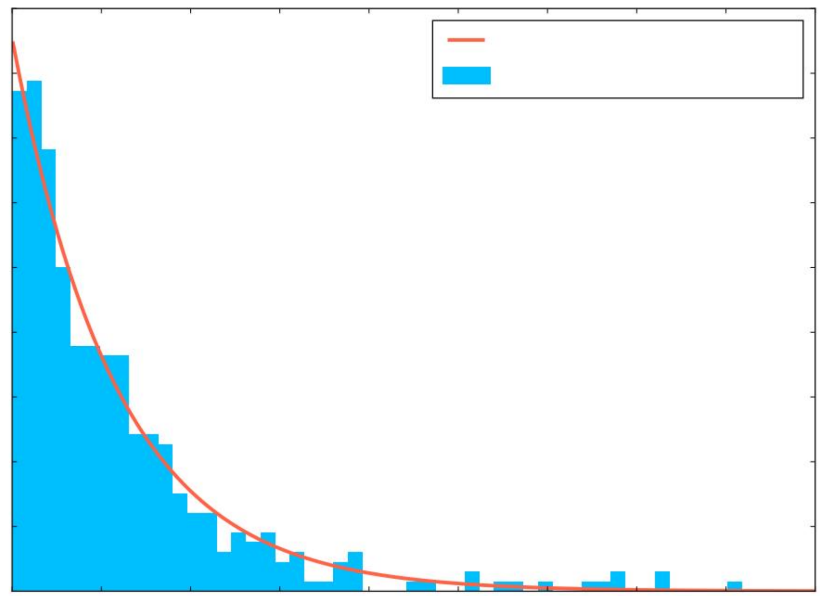
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (2,50)$$

Một phân phối Gauss với giá trị trung bình μ và độ lệch chuẩn σ thường được ký hiệu là $N(\mu, \sigma^2)$.

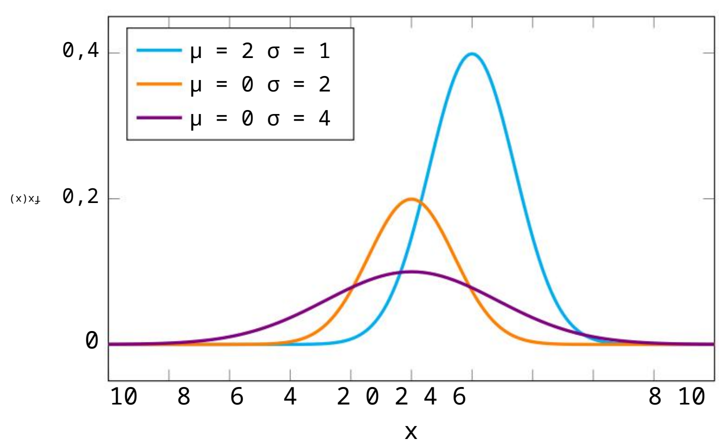
Chúng tôi cung cấp các định nghĩa chính thức về giá trị trung bình và độ lệch chuẩn của một biến ngẫu nhiên trong Chương 4. Hiện tại, bạn chỉ có thể coi chúng là các đại lượng tham số hóa Gaussian pdf.

Không rõ ràng ngay lập tức rằng bản pdf của Gaussian tích hợp thành một. Chúng tôi thiết lập điều này trong bổ đề sau.

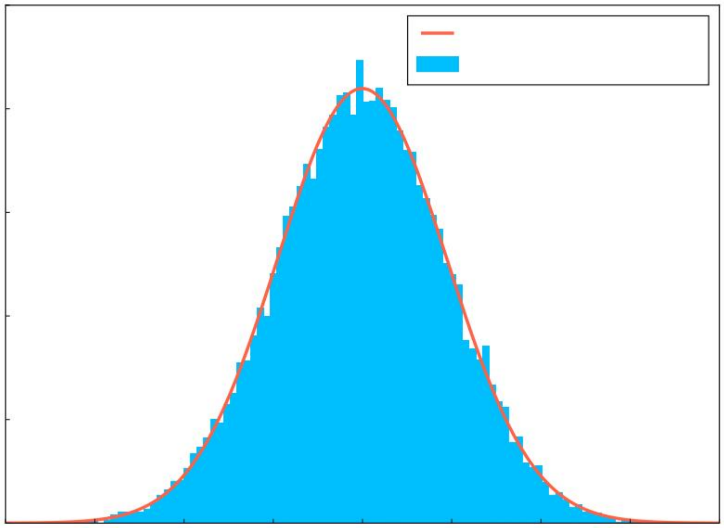
Bổ đề 2.3.11 (Chứng minh trong Mục 2.7.3). Bản pdf của một biến ngẫu nhiên Gaussian tích hợp với một.



Hình 2.11: Biểu đồ về thời gian giữa các lần đến của các cuộc gọi tại một trung tâm cuộc gọi ở Israel (màu đỏ) so với ước tính của nó bằng pdf hàm mũ.



Hình 2.12: Biến ngẫu nhiên Gaussian với các phương tiện và độ lệch chuẩn khác nhau.



Hình 2.13: Biểu đồ chiều cao của dân số 25.000 người (màu xanh lam) và giá trị gần đúng của nó bằng cách sử dụng phân phối Gaussian (màu cam).

Hình 2.12 cho thấy pdf của hai biến ngẫu nhiên Gaussian với các giá trị khác nhau của μ và σ. Hình 2.13 cho thấy biểu đồ chiều cao của một dân số 25.000 người và cách nó được xấp xỉ rất tốt bởi một biến ngẫu nhiên Gaussian[3] .

Một đặc điểm khó chịu của biến ngẫu nhiên Gaussian là cdf của nó không có nghiệm dạng đóng, trái ngược với các biến ngẫu nhiên đều và hàm mũ. Điều này làm phức tạp nhiệm vụ xác định xác suất mà một biến ngẫu nhiên Gaussian nằm trong một khoảng nhất định. Để giải quyết vấn đề này, chúng ta sử dụng thực tế là nếu X là một biến ngẫu nhiên Gauss với trung bình μ và độ lệch chuẩn σ, thì

$$\text{bạn} := \frac{X \quad \mu}{\sigma} \tag{2.51}$$

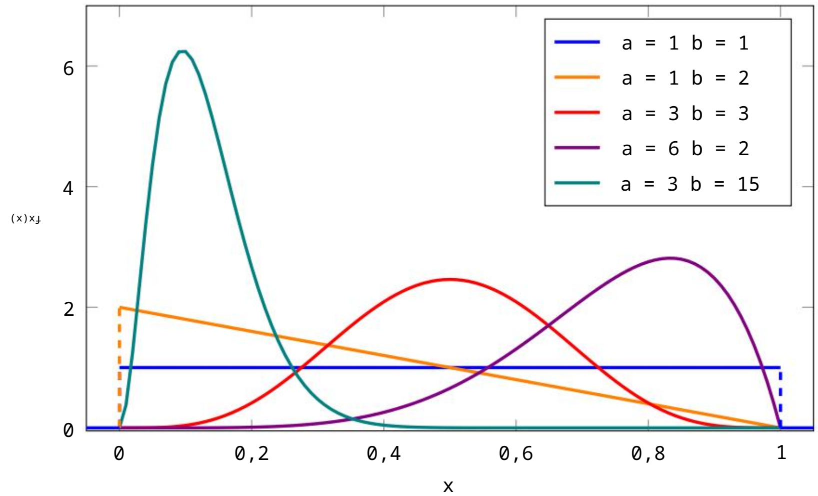
là một biến ngẫu nhiên Gaussian tiêu chuẩn, có nghĩa là giá trị trung bình của nó bằng 0 và độ lệch chuẩn của nó bằng một. Xem Bổ đề 2.5.1 để chứng minh. Điều này cho phép chúng ta biểu thị xác suất X nằm trong khoảng [a, b] theo cdf của một Gaussian tiêu chuẩn, mà chúng ta biểu thị bằng Φ,

$$P\,(X \quad [a,\, b]) = P\,\frac{X \quad \mu}{\sigma} \qquad \frac{\text{một} \quad \mu}{\sigma}, \frac{b \quad \mu}{\sigma} \tag{2.52}$$

$$= \Phi\,\frac{b \quad \mu}{\sigma} \qquad \Phi\,\frac{\text{một} \quad \mu}{\sigma}\ . \tag{2.53}$$

Miễn là chúng ta có thể đánh giá Φ, công thức này cho phép chúng ta xử lý các biến ngẫu nhiên Gaussian tùy ý. Để đánh giá Φ người ta thường sử dụng danh sách các giá trị được lập bảng, được tổng hợp bằng cách tính toán các tích phân tương ứng bằng số. Ngày nay, bạn chỉ có thể sử dụng Matlab, WolframAlpha, SciPy, v.v.

[3]Dữ liệu có sẵn ở đây.



Hình 2.14: Pdfs của biến ngẫu nhiên beta với các giá trị khác nhau của tham số a và b.

bản thử nghiệm

Các bản phân phối beta cho phép chúng tôi tham số hóa các bản phân phối liên tục không theo phương thức được hỗ trợ trên khoảng thời gian đơn vị. Điều này rất hữu ích trong thống kê Bayesian, như chúng ta sẽ thảo luận trong Chương 10.

Định nghĩa 2.3.12 (Phân phối beta). Bản pdf của bản phân phối beta có tham số a và b được định nghĩa là

$$f_\beta(\theta; a, b) := \begin{cases} \frac{\theta^{a-1}(1-\theta)^{b-1}}{\beta(a,b)}, & \text{nếu } 0 \leq \theta \leq 1, \\ 0 & \text{ngược lại,} \end{cases} \tag{2.54}$$

Ở đâu

$$(a, b) := \int_{bạn} u^{a-1}(1 - u)^{b-1}\, du. \tag{2,55}$$

β(a, b) là một hàm đặc biệt được gọi là hàm beta hay tích phân Euler loại một, hàm này phải được tính toán bằng số. Phân phối đều là một ví dụ về phân phối beta (trong đó a = 1 và b = 1). Hình 2.14 cho thấy bản pdf của một số bản phân phối beta khác nhau.

2.4 Điều hòa trên một sự kiện

Trong Phần 1.2 , chúng tôi giải thích cách sửa đổi thước đo xác suất của một không gian xác suất để kết hợp giả định rằng một sự kiện nhất định đã xảy ra. Trong phần này, chúng tôi xem xét tình huống này khi các biến ngẫu nhiên có liên quan. Cụ thể, chúng ta xem xét một biến ngẫu nhiên X có phân phối nhất định được biểu diễn bằng pmf, cdf hoặc pdf và giải thích phân phối của nó thay đổi như thế nào nếu chúng ta giả sử rằng X S, với bất kỳ tập S nào thuộc đại số σ Borel (hãy nhớ rằng điều này về cơ bản bao gồm bất kỳ bộ hữu ích nào bạn có thể nghĩ đến).

Nếu X rời rạc với pmf p_X, thì pmf có điều kiện của X cho $X \in S$ là

$$p_{X|X \in S}(x) := P(X = x | X \in S) \tag{2.56}$$

$$= \begin{cases} \frac{p_X(x)}{\sum_{s \in S} p_X(s)} & \text{nếu } x \in S \\ 0 & \text{khác.} \end{cases} \tag{2.57}$$

Đây là một pmf hợp lệ trong không gian xác suất mới giới hạn cho sự kiện $\{X \in S\}$.

Tương tự nếu X liên tục với pdf f_X, cdf có điều kiện của X với biến cố $X \in S$ là

$$F_{X|X \in S}(x) := P(X \le x | X \in S) \tag{2.58}$$

$$= \frac{P(X \le x, X \in S)}{P(X \in S)} \tag{2.59}$$

$$= \frac{\int_{u \le x, u \in S} f_X(u)\, du}{\int_{u \in S} f_X(u)\, du}, \tag{2.60}$$

một lần nữa theo định nghĩa của xác suất có điều kiện. Người ta có thể kiểm tra xem đây có phải là một cdf hợp lệ trong không gian xác suất mới hay không. Để có được pdf có điều kiện, chúng tôi chỉ cần phân biệt cdf này,

$$f_{X|X \in S}(x) := \frac{dF_{X|X \in S}(x)}{dx}. \tag{2.61}$$

Bây giờ chúng tôi áp dụng ý tưởng này để chỉ ra rằng các biến ngẫu nhiên hình học và hàm mũ là không có bộ nhớ.

Ví dụ 2.4.1 (Biến ngẫu nhiên hình học không nhớ). Chúng ta tung đồng xu liên tục cho đến khi có mặt ngửa, nhưng tạm dừng sau một vài lần tung (tức là mặt sấp). Chúng ta hãy giả sử rằng các lần tung là độc lập và có cùng độ lệch p (tức là xác suất xuất hiện mặt ngửa trong mỗi lần lật là p). Xác suất để có được mặt ngửa trong k lần tung nữa là bao nhiêu? Có lẽ đáng ngạc nhiên, nó chính xác giống như xác suất có được mặt ngửa sau k lần lật ngay từ đầu.

Để thiết lập điều này một cách chặt chẽ, chúng tôi tính toán pmf có điều kiện của một biến ngẫu nhiên hình học X dựa trên sự kiện $\{X > k_0\}$ (tức là k_0 đầu tiên là đuôi trong ví dụ của chúng tôi). Áp dụng (2.56) ta có

$$p_{X|X>k_0}(k) = \frac{p_X(k)}{\sum_{m=k_0+1}^{\infty} p_X(m)} \tag{2.62}$$

$$= \frac{(1-p)^{k-1}\, p}{\sum_{m=k_0+1}^{\infty} (1-p)^{m-1}\, p} \tag{2.63}$$

$$= (1-p)^{k-k_0-1}\, p \tag{2.64}$$

nếu $k > k_0$ và 0 nếu ngược lại. Chúng tôi đã sử dụng thực tế là chuỗi hình học

$$\sum_{m=k_0+1}^{\infty} \alpha^{tôi} = \frac{\alpha^{k_0+1}}{1-\alpha} \tag{2,65}$$

với mọi $\alpha < 1$.

Trong không gian xác suất mới nơi đếm bắt đầu từ $k_0 + 1$ pmf có điều kiện là biến ngẫu nhiên hình học có cùng tham số với biến ban đầu. K0 lần lật đầu tiên không ảnh hưởng đến tương lai, một khi nó được tiết lộ rằng chúng là mặt sấp.

Ví dụ 2.4.2 (Biến ngẫu nhiên hàm mũ không nhớ). Chúng ta hãy giả sử rằng thời gian đến giữa các email của bạn tuân theo phân phối hàm mũ (trong khoảng thời gian vài giờ, đây có thể là một ước tính gần đúng, hãy cho chúng tôi biết nếu bạn kiểm tra). Bạn nhận được một email. Thời gian cho đến khi bạn nhận được email tiếp theo được phân phối theo cấp số nhân với một tham số λ nhất định.
Không có email nào đến trong t_0 phút tiếp theo. Đáng ngạc nhiên là thời gian từ lúc đó cho đến khi bạn nhận được email tiếp theo lại được phân phối theo cấp số nhân với cùng một tham số, bất kể giá trị của t_0 là bao nhiêu. Cũng giống như biến ngẫu nhiên hình học, biến ngẫu nhiên hàm mũ không có bộ nhớ.

Hãy để chúng tôi chứng minh điều này một cách chặt chẽ. Chúng ta tính toán cdf có điều kiện của biến ngẫu nhiên hàm mũ T với tham số λ có điều kiện dựa trên biến cố $\{T > t_0\}$ đối với $t_0 > 0$ tùy ý bằng cách áp dụng (2.60)

$$F_{T|T>t_0}(t) = \frac{\int_{t_0}^{t} f_T(u)\,du}{\int_{t_0}^{\infty} f_T(u)\,du} \tag{2.66}$$

$$= \frac{e^{-\lambda t_0} - e^{-\lambda t}}{e^{-\lambda t_0}} \tag{2.67}$$

$$= 1 - e^{-\lambda(t-t_0)}. \tag{2.68}$$

Đạo hàm đối với t mang lại hàm mũ pdf $f_{T|T>t_0}(t) = \lambda e^{-\lambda(t-t_0)}$ bắt đầu từ t_0.

2.5 Hàm của biến ngẫu nhiên

Tính toán phân phối hàm của một biến ngẫu nhiên thường rất hữu ích trong mô hình xác suất. Ví dụ: nếu chúng ta lập mô hình dòng điện trong mạch sử dụng biến ngẫu nhiên X, chúng ta có thể quan tâm đến công suất $Y := rX^2$ tiêu tán trên một điện trở có điện trở xác định r. Nếu ta áp dụng hàm tất định $g : \mathbb{R} \to \mathbb{R}$ cho biến ngẫu nhiên X thì kết quả $Y := g(X)$ không phải là đại lượng tất định. Nhớ lại rằng các biến ngẫu nhiên là các hàm từ không gian mẫu Ω đến R. Nếu X ánh xạ các phần tử của Ω sang R thì Y cũng vậy vì $Y(\omega) = g(X(\omega))$.
Điều này có nghĩa là Y cũng là một biến ngẫu nhiên. Trong phần này, chúng tôi giải thích cách mô tả đặc điểm phân phối của Y khi biết phân phối của X.

Nếu X là rời rạc, thì việc tính pmf của g(X) từ pmf của X, (2.69) (2.70) (2.71) là rất đơn giản.

$$p_Y(y) = P(Y = y)$$

$$= P(g(X) = y)$$

$$= \sum_{\{x \mid g(x)=y\}} p_X(x).$$

Nếu X liên tục, thủ tục tinh vi hơn. Trước tiên, chúng tôi tính toán cdf của Y bằng cách áp dụng định nghĩa,

Năm tài chính $(y) = P(Y \leq y)$ (2.72)

$$= P(g(X) \leq y) \tag{2.73}$$

$$= \int_{\{x \mid g(x)\leq y\}} f_X(x)\,dx, \tag{2.74}$$

trong đó đẳng thức cuối cùng rõ ràng chỉ đúng nếu X có bản pdf. Sau đó, chúng tôi có thể lấy bản pdf của Y từ cdf của nó nếu nó khả vi. Ý tưởng này có thể được sử dụng để chứng minh một kết quả hữu ích về các biến ngẫu nhiên Gaussian.

Machine Translated by Google

Bổ đề 2.5.1 (Biến ngẫu nhiên Gauss). Nếu X là một biến ngẫu nhiên Gauss với giá trị trung bình μ và độ lệch chuẩn σ, thì

$$\text{bạn} := \frac{X \quad \mu}{\sigma} \tag{2,75}$$

là một biến ngẫu nhiên Gaussian tiêu chuẩn.

Bằng chứng. Ta áp dụng (2.74) để thu được

$$F_U(u) = P \left(\frac{X \quad \mu}{\sigma} \leq u \right) \tag{2.76}$$

$$= \int_{(x \quad \mu)/\sigma \leq u} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x \quad \mu)^2}{2\sigma^2}} dx \tag{2.77}$$

$$= \int_{\infty}^{\text{bạn}} \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw \text{ bằng phép đổi biến } w = \frac{x \quad \mu}{\sigma}. \tag{2,78}$$

Khác biệt đối với sản lượng u

$$f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\text{bạn}^2}{2}}, \tag{2,79}$$

vì vậy U thực sự là một biến ngẫu nhiên Gaussian tiêu chuẩn. □

2.6 Tạo biến ngẫu nhiên

Mô phỏng là một công cụ cơ bản trong mô hình xác suất. Mô phỏng kết quả của một mô hình yêu cầu lấy mẫu từ các biến ngẫu nhiên có trong mô hình đó. Chiến lược phổ biến nhất để tạo mẫu từ một biến ngẫu nhiên tách quá trình thành hai bước:

1. Sinh mẫu đều từ khoảng đơn vị [0, 1].

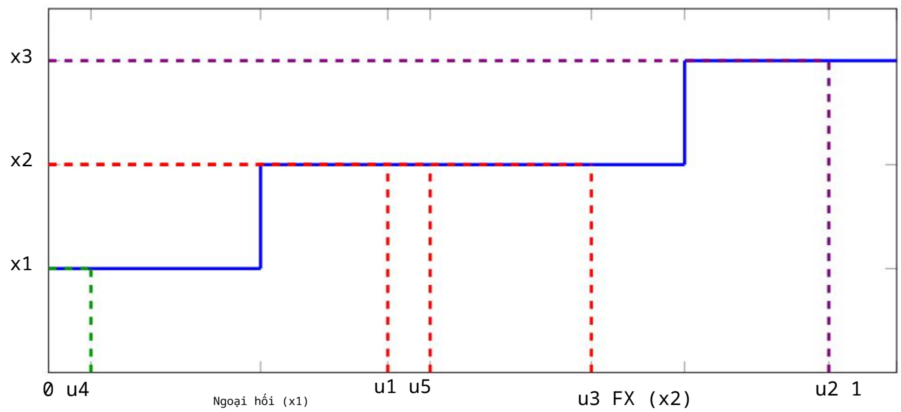
2. Biến đổi các mẫu đồng dạng để chúng có phân phối mong muốn.

Ở đây, chúng tôi tập trung vào bước thứ hai, giả sử rằng chúng tôi có quyền truy cập vào bộ tạo số ngẫu nhiên tạo ra các mẫu độc lập tuân theo phân phối đồng đều trong [0, 1]. Việc xây dựng các bộ tạo ngẫu nhiên thống nhất tốt là một vấn đề quan trọng, nằm ngoài phạm vi của các ghi chú này.

2.6.1 Lấy mẫu từ phân phối rời rạc

Cho X là biến ngẫu nhiên rời rạc với pmf p_X và U là biến ngẫu nhiên đều trong [0, 1]. Mục đích của chúng tôi là biến đổi một mẫu từ U để nó được phân phối theo p_X. Chúng tôi biểu thị các giá trị có xác suất khác không theo p_X bằng $x_1, x_2, \ldots$.

Đối với i cố định, giả sử rằng chúng ta chỉ định tất cả các mẫu của U trong một khoảng có độ dài $p_X(x_i)$ cho x_i. Khi đó xác suất để một mẫu đã cho từ U được gán cho x_i chính xác là $p_X(x_i)$!



Hình 2.15: Minh họa phương pháp tạo mẫu từ phân bố rời rạc tùy ý được mô tả trong Mục 2.6.1. cdf của một biến ngẫu nhiên rời rạc được hiển thị bằng màu xanh lam. Các mẫu u4 và u2 từ một phân bố đồng đều lần lượt được ánh xạ tới x1 và x3 , trong khi u1, u3 và u5 được ánh xạ tới x3.

Rất thuận tiện, khoảng đơn vị có thể được phân chia thành các khoảng có độ dài $p_X(x_i)$. Do đó, chúng tôi có thể tạo X bằng cách lấy mẫu từ U và đặt

$$x = \begin{cases} x_1 \text{ nếu } 0 \le U \le p_X(x_1), \\ x_2 \text{ nếu } p_X(x_1) \le U \le p_X(x_1) + p_X(x_2), \\ \cdots \\ x_i \text{ nếu } \sum_{j=1}^{i-1} p_X(x_j) \le U \le \sum_{j=1}^{i} p_X(x_j), \\ \cdots \end{cases} \tag{2.80}$$

Nhớ lại rằng cdf của một biến ngẫu nhiên rời rạc bằng

$$F_X(x) = P(X \le x) \tag{2.81}$$

$$= \sum_{x_i \le x} p_X(x_i), \tag{2.82}$$

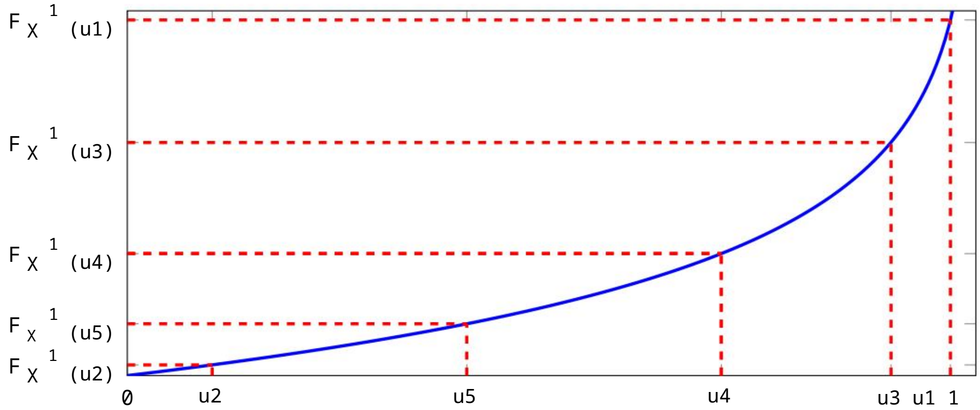
vì vậy thuật toán của chúng tôi rút gọn để lấy một mẫu u từ U và sau đó xuất ra x_i sao cho $F_X(x_{i-1}) \le u \le F_X(x_i)$. Điều này được minh họa trong Hình 2.15.

2.6.2 Lấy mẫu biến đổi nghịch đảo

Lấy mẫu biến đổi nghịch đảo cho phép lấy mẫu từ phân phối tùy ý với cdf đã biết bằng cách áp dụng biến đổi xác định cho các mẫu đồng nhất. Bằng trực giác, chúng ta có thể diễn giải nó như là sự tổng quát hóa của phương pháp trong Phần 2.6.1 đối với phân phối liên tục.

Thuật toán 2.6.1 (Lấy mẫu biến đổi nghịch đảo). Cho X là biến ngẫu nhiên liên tục với cdf F_X và U là biến ngẫu nhiên phân phối đều trong [0, 1] và không phụ thuộc vào X.

1. Lấy mẫu u của U .
2. Đặt $x := F_X^{-1}(u)$.



Hình 2.16: Các mẫu từ phân bố hàm mũ với tham số λ = 1 thu được bằng cách lấy mẫu biến đổi nghịch đảo như được mô tả trong Ví dụ 2.6.4. Các mẫu $u_1, \ldots, u_5$ được tạo ra từ một bộ đồng phục phân bổ.

Người đọc cẩn thận sẽ chỉ ra rằng F_X có thể không đảo ngược tại mọi thời điểm. Để tránh điều này vấn đề chúng tôi xác định nghịch đảo tổng quát của cdf là

$$F_X^{1}(u) := \text{tối thiểu}_x \{F_X(x) = u\}. \tag{2.83}$$

Hàm này được xác định rõ vì tất cả các cdf đều không giảm, vì vậy F_X bằng một hằng số c trong mọi khoảng $[x_1, x_2]$ mà nó không khả nghịch.

Bây giờ chúng tôi chứng minh rằng Thuật toán 2.6.1 hoạt động.

Định lý 2.6.2 (Lấy mẫu biến đổi nghịch đảo hoạt động). Sự phân phối của $Y = F_X^{1}(U)$ là giống như phân phối của X.

Bằng chứng. Chúng ta chỉ cần chứng minh rằng cdf của Y bằng F_X. Chúng ta có

$$\text{Năm tài chính}(y) = P(Y \leq y) \tag{2.84}$$
$$= P\left(F_X^{1}(U) \leq y\right) \tag{2,85}$$
$$= P(U \leq F_X(y)) \tag{2.86}$$
$$= \int_{u=0}^{\text{ngoại hối}(y)} du \tag{2.87}$$
$$= \text{Ngoại hối}(y), \tag{2.88}$$

ở bước (2.86) chúng ta phải tính đến việc chúng ta đang sử dụng nghịch đảo tổng quát của các cdf. Điều này được giải quyết bằng bổ đề sau được chứng minh trong Mục 2.7.4.

Bổ đề 2.6.3. Các sự kiện $F_X^{1}(U) \leq y$ và $\{U \leq F_X(y)\}$ là tương đương.

□

Ví dụ 2.6.4 (Lấy mẫu từ phân phối mũ). Cho X là một biến ngẫu nhiên hàm mũ với tham số λ. cdf FX (x) := 1 e $^{\lambda x}$ khả nghịch trong [0, ∞]. nghịch đảo của nó bằng

$$F_X^{1} \, 1 \, (u) = \overline{1}og \, \lambda \quad \frac{1}{1 \quad \text{bạn}} \quad . \tag{2.89}$$

F_X^{1} (U) là biến ngẫu nhiên hàm mũ với tham số λ theo Định lý 2.6.2. Hình 2.16 chỉ ra cách các mẫu của U được biến đổi thành các mẫu của X.

2.7 Bằng chứng

2.7.1 Chứng minh bổ đề 2.2.9

Đối với bất kỳ hằng số cố định c1 và c2

$$\text{limn} \; \infty \; \frac{n \quad c1}{n \quad c2} = 1, \tag{2,90}$$

để có thể

$$\text{limn} \; \infty \; \frac{N!}{(n \quad k)! \; (n \quad \lambda)^{k}} = \frac{N}{n \;\; \lambda} \cdot \frac{n \;\; 1}{n \;\; \lambda} \cdots \frac{n \;\; k+1}{n \;\; \lambda} = 1. \tag{2.91}$$

Kết quả theo sau từ sự đồng nhất tính toán cơ bản sau đây:

$$\text{limn} \; \infty \; 1 \quad \frac{\lambda}{N}^{N} \quad \lambda = e \, . \tag{2.92}$$

2.7.2 Chứng minh bổ đề 2.3.2

Để thiết lập (2.31)

$$\lim_{x \quad \infty} FX\ (x) = 1 \quad \lim_{x \quad \infty} P\ (X > x) \tag{2.93}$$

$$= 1 \quad P\ (X > 0) \quad \text{limn} \; \infty \sum_{\text{tôi}=0}^{N} P\ (\quad i \geq X > \quad (i + 1)) \tag{2,94}$$

$$= 1 \quad P \; \text{limn} \; \infty \; \{X > 0\} \quad {}_{\text{tôi}=0} \; \{ \quad i \geq X > \quad (i + 1)\} \tag{2,95}$$

$$n = 1 \quad P\ (\Omega) = 0. \tag{2,96}$$

Chứng minh của (2.32) xuất phát từ kết quả này. Đặt Y = X, sau đó

$$\text{limx} \; \infty \; FX\ (x) = \text{limx} \; \infty \; P\ (X \leq \tag{2,97}$$

$$x) = 1 \quad \text{limx} \; \infty \; P\ (X > x) \tag{2,98}$$

$$= 1 \quad \lim_{x \quad \infty} P\ (\quad X < x) \tag{2,99}$$

$$= 1 \quad \lim_{x \quad \infty} FY\ (x) = 1 \text{ theo } (2.32). \tag{2.100}$$

Cuối cùng, (2.33) đúng vì {X ≤ a} {X ≤ b}.

2.7.3 Chứng minh bổ đề 2.3.11

Kết quả là hệ quả của bổ đề sau.

Bổ đề 2.7.1.

$$\int_{-\infty}^{\infty} e^{-t^2}\, đt = \sqrt{\pi}. \tag{2.101}$$

Bằng chứng. Hãy để chúng tôi xác định

$$tôi = \int_{-\infty}^{\infty} e^{-x^2}\, dx. \tag{2.102}$$

Bây giờ lấy bình phương và đổi sang tọa độ cực,

$$tôi^2 = \int_{-\infty}^{\infty} e^{-x^2}\, dx \int_{-\infty}^{\infty} e^{-y^2}\, đê \tag{2.103}$$

$$= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} e^{-(x^2+y^2)}\, dxdy \tag{2.104}$$

$$= \int_{\theta=0}^{2\pi} \int_{r=-\infty}^{\infty} lại^{(-r^2)}\, d\theta dr \tag{2.105}$$

$$= \pi e^{(-r^2)} \Big]_{0}^{\infty} = \pi. \tag{2.106}$$

□

Để hoàn thành chứng minh, chúng ta sử dụng phép đổi biến $t = (x - \mu) / \sqrt{2}\sigma$.

2.7.4 Chứng minh Bổ đề 2.6.3

$F_X^{-1}(U) \leq y$ ngụ ý $\{U \leq F_X(y)\}$

Giả sử rằng $U > F_X(y)$, thì với mọi x, sao cho $F_X(x) = U$, $x > y$ vì cdf không gấp nếp. Cụ thể là $\min_x \{F_X(x) = U\} > y$.

$\{U \leq F_X(y)\}$ ngụ ý $F_X^{-1}(U) \leq y$

Giả sử rằng $\min_x \{F_X(x) = U\} > y$, sau đó $U > F_X(y)$ vì cdf không giảm. Bất đẳng thức là nghiêm ngặt vì $U = F_X(y)$ sẽ ngụ ý rằng y thuộc về $\{F_X(x) = U\}$, điều này không thể xảy ra vì chúng ta đang giả sử rằng nó nhỏ hơn giá trị nhỏ nhất của tập hợp đó.

Chương 3

Biến ngẫu nhiên đa biến

Các mô hình xác suất thường bao gồm nhiều đại lượng số không chắc chắn. Trong chương này, chúng tôi mô tả cách chỉ định các biến ngẫu nhiên để biểu diễn các đại lượng đó và các tương tác của chúng. Trong một số trường hợp, sẽ hợp lý khi nhóm các biến ngẫu nhiên này thành các vectơ ngẫu nhiên, mà chúng tôi viết bằng các chữ cái viết hoa có mũi tên ở trên cùng: $\vec{X}$. Việc thực hiện các vectơ ngẫu nhiên này được biểu thị bằng các chữ cái viết thường: $\vec{x}$.

3.1 Biến ngẫu nhiên rời rạc

Nhớ lại rằng các biến ngẫu nhiên rời rạc là các đại lượng số nhận các giá trị hữu hạn hoặc vô hạn đếm được. Trong phần này, chúng tôi giải thích cách thao tác với nhiều biến ngẫu nhiên rời rạc có chung một không gian xác suất.

3.1.1 Hàm khối xác suất chung

Nếu một số biến ngẫu nhiên rời rạc được xác định trên cùng một không gian xác suất, chúng ta chỉ định hành vi xác suất của chúng thông qua hàm khối lượng xác suất chung của chúng, là xác suất mà mỗi biến nhận một giá trị cụ thể.

Định nghĩa 3.1.1 (Hàm khối xác suất chung). Cho $X : \Omega \to R_X$ và $Y : \Omega \to R_Y$ là các biến ngẫu nhiên rời rạc (R_X và R_Y là các tập rời rạc) trên cùng một không gian xác suất (Ω, F, P). PMf chung của X và Y được định nghĩa là

$$p_{X,Y}(x, y) := P(X = x, Y = y) . \tag{3.1}$$

Nói cách khác, $p_{X,Y}(x, y)$ là xác suất của X và Y tương ứng bằng x và y.

Tương tự, pmf chung của một vectơ ngẫu nhiên rời rạc có kích thước n

$$\vec{X} := \begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_n \end{bmatrix} \tag{3.2}$$

với các mục $X_i : \Omega \to R_{X_i}$ ($R_1, \ldots, R_n$ đều là các tập rời rạc) thuộc cùng một không gian xác suất được định nghĩa là

$$p_{\vec{X}}(\vec{x}) := P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n). \tag{3.3}$$

Như trong trường hợp pmf của một biến ngẫu nhiên duy nhất, pmf chung là một phép đo xác suất hợp lệ R_{X^n} nếu chúng ta xem xét một không gian xác suất trong đó không gian mẫu [1] (hoặc $R_{X1} \times R_{X2} \cdots \times$ là $R_X \times R_Y$ trong trường hợp a vectơ ngẫu nhiên) và σ-đại số chỉ là tập lũy thừa của không gian mẫu. Điều này ngụ ý rằng pmf chung hoàn toàn đặc trưng cho các biến ngẫu nhiên hoặc vectơ ngẫu nhiên, chúng ta không cần phải lo lắng về không gian xác suất cơ bản.

Theo định nghĩa của phép đo xác suất, pmf chung phải không âm và tổng của nó trên tất cả các đối số có thể có của nó phải bằng một,

$$p_{X,Y}(x, y) \geq 0 \text{ với mọi } x \in R_X, y \in R_Y, \tag{3.4}$$

$$\sum_{x \in R_X} \sum_{y \in R_Y} p_{X,Y}(x, y) = 1. \tag{3.5}$$

Theo Luật xác suất toàn phần, pmf chung cho phép chúng ta có được xác suất X và Y thuộc bất kỳ tập hợp nào $S \subseteq R_X \times R_Y$,

$$P((X, Y) \in S) = P\left(\cup_{(x,y) \in S} \{X = x, Y = y\}\right) \text{ (hợp của các biến cố rời rạc)} \tag{3.6}$$

$$= \sum_{(x,y) \in S} P(X = x, Y = y) \tag{3.7}$$

$$= \sum_{(x,y) \in S} p_{X,Y}(x, y). \tag{3.8}$$

Các thuộc tính này cũng áp dụng cho các vectơ ngẫu nhiên (và các nhóm có nhiều hơn hai biến ngẫu nhiên). Đối với bất kỳ vectơ ngẫu nhiên X ,

$$p_X(x) \geq 0, \tag{3.9}$$

$$\sum_{x_1 \in R_1} \sum_{x_2 \in R_2} \cdots \sum_{x_n \in R_n} p_X(x) = 1. \tag{3.10}$$

Xác suất để X thuộc tập rời rạc $S \subseteq \mathbb{R}^N$ được đưa ra bởi

$$P_X(S) = \sum_{x \in S} p_X(x). \tag{3.11}$$

3.1.2 Định biên

Giả sử chúng ta có quyền truy cập vào pmf chung của một số biến ngẫu nhiên trong một không gian xác suất nhất định, nhưng chúng ta chỉ quan tâm đến hành vi của một trong số chúng. Để tính giá trị pmf của nó cho một giá trị cụ thể, chúng tôi cố định giá trị đó và tính tổng trên các biến ngẫu nhiên còn lại. Thật vậy, theo định luật xác suất toàn phần

$$p_X(x) = P(X = x) \tag{3.12}$$

$$= P\left(\cup_{y \in R_Y} \{X = x, Y = y\}\right) \text{ (hợp của các biến cố rời rạc)} \tag{3.13}$$

$$= \sum_{y \in R_Y} P(X = x, Y = y) \tag{3.14}$$

$$= \sum_{y \in R_Y} p_{X,Y}(x, y). \tag{3.15}$$

[1] Đây là tích Descartes của hai tập hợp, được xác định trong Phần A.2, chứa tất cả các cặp (x, y) có thể có trong đó $x \in R_X$ và $y \in R_Y$.

Khi pmf chung liên quan đến nhiều hơn hai biến ngẫu nhiên, đối số hoàn toàn giống nhau. Điều này được gọi là giảm biên so với các biến ngẫu nhiên khác. Trong ngữ cảnh này, pmf của một biến ngẫu nhiên duy nhất được gọi là pmf cận biên của nó. Bảng 3.1 cho thấy một ví dụ về một pmf chung và các pmf biên tương ứng.

Nếu chúng ta quan tâm đến việc tính toán pmf chung của một số mục trong một vectơ ngẫu nhiên, thay vì chỉ một, thì quá trình biên hóa về cơ bản là giống nhau. PMf một lần nữa thu được bằng cách tính tổng các mục còn lại. Cho $I \subseteq \{1, 2, \ldots, n\}$ là tập hợp con của $m < n$ phần tử của vectơ ngẫu nhiên n chiều X và X_I là vectơ con ngẫu nhiên tương ứng. Để tính pmf chung của X_I, chúng ta tính tổng tất cả các mục không có trong I, mà chúng ta ký hiệu là $\{j_1, j_2, \ldots, j_{n-m}\} := \{1, 2, \ldots, n\} / I$

$$p_{X_I}(x_I) = \sum_{x_{j_1} \in R_{j_1}} \sum_{x_{j_2} \in R_{j_2}} \cdots \sum_{x_{j_{n-m}} \in R_{j_{n-m}}} p_X(x). \tag{3.16}$$

3.1.3 Phân phối có điều kiện

Xác suất có điều kiện cho phép chúng ta cập nhật sự không chắc chắn của mình về các đại lượng trong một mô hình xác suất khi thông tin mới được tiết lộ. Phân phối có điều kiện của một biến ngẫu nhiên xác định hành vi của biến ngẫu nhiên khi chúng ta giả định rằng các biến ngẫu nhiên khác trong không gian xác suất lấy một giá trị cố định.

Định nghĩa 3.1.2 (Hàm khối xác suất có điều kiện). Hàm khối lượng xác suất có điều kiện của Y cho X, trong đó X và Y là các biến ngẫu nhiên rời rạc được xác định trên cùng một không gian xác suất, được cho bởi

$$p_{Y|X}(y|x) = P(Y = y|X = x) \tag{3.17}$$

$$= \frac{p_{X,Y}(x, y)}{p_X(x)} \quad \text{nếu } p_X(x) > 0 \tag{3.18}$$

và không được xác định khác.

Điều kiện pmf $p_{X|Y}(\cdot|y)$ đặc trưng cho sự không chắc chắn của chúng ta về X với điều kiện là sự kiện $\{Y = y\}$. Đối tượng này là một pmf hợp lệ của X, vì vậy nếu R_X là phạm vi của X

$$\sum_{x \in R_X} p_{X|Y}(x|y) = 1 \tag{3.19}$$

cho bất kỳ y nào mà nó được xác định rõ. Tuy nhiên, nó không phải là pmf cho Y. Đặc biệt, không có lý do gì để $\sum_{y \in R_Y} p_{X|Y}(x|y)$ cộng lại bằng một!

Bây giờ chúng ta xác định pmf có điều kiện chung của một số biến ngẫu nhiên (tương đương với một vectơ con của một vectơ ngẫu nhiên) với các biến ngẫu nhiên khác (hoặc các phần tử của vectơ ngẫu nhiên).

Định nghĩa 3.1.3 (pmf có điều kiện). PMf có điều kiện của vectơ con ngẫu nhiên rời rạc X_I, $I \subseteq \{1, 2, \ldots, n\}$, với một vectơ con khác X_J là

$$p_{X_I|X_J}(x_I|x_J) := \frac{p_X(x)}{p_{X_J}(x_J)}, \tag{3.20}$$

trong đó $\{j_1, j_2, \ldots, j_{n-m}\} := \{1, 2, \ldots, n\} / I$.

r

L

$p_{L,R}$	0	1
0	$\frac{14}{20}$	$\frac{1}{20}$
1	$\frac{2}{20}$	$\frac{3}{20}$

làm đn
$\frac{15}{20}$
$\frac{5}{20}$

$p_{L\|R}(\cdot\|0)$
$\frac{7}{8}$
$\frac{1}{8}$

$p_{L\|R}(\cdot\|1)$
$\frac{1}{4}$
$\frac{3}{4}$

p_R	$\frac{16}{20}$	$\frac{4}{20}$
$p_{R\|L}(\cdot\|0)$	$\frac{14}{15}$	$\frac{1}{15}$
$p_{R\|L}(\cdot\|1)$	$\frac{2}{5}$	$\frac{3}{5}$

Bảng 3.1: Các pmfs chung, cận biên và có điều kiện của các biến ngẫu nhiên L và R được xác định trong Ví dụ 3.1.5.

Các pmfs có điều kiện $p_{Y|X}(\cdot|x)$ và $p_{X_I|X_J}(\cdot|x_J)$ đó là các pmfs hợp lệ trong không gian xác suất trong $I|X_J$ $X = x$ hoặc $X_J = x_J$ tương ứng. Chẳng hạn, chúng phải không âm và cộng lại bằng một.

Từ định nghĩa của pmfs có điều kiện, chúng tôi rút ra quy tắc dây chuyền cho các biến ngẫu nhiên và vectơ rời rạc.

Bổ đề 3.1.4 (Quy tắc dây chuyền đối với biến ngẫu nhiên rời rạc và vectơ).

$$p_{X,Y}(x,y) = p_X(x)\, p_{Y|X}(y|x), \tag{3.21}$$

$$p_X(x) = p_{X_1}(x_1)\, p_{X_2|X_1}(x_2|x_1) \ldots p_{X_n|X_1,\ldots,X_{n-1}}(x_n|x_1,\ldots,x_{n-1}) \tag{3.22}$$

$$= \prod_{tôi=1}^{N} p_{X_i|X_{\{1,\ldots,i-1\}}}\left(x_i|x_{\{1,\ldots,i-1\}}\right), \tag{3.23}$$

trong đó thứ tự của các chỉ số trong vectơ ngẫu nhiên là tùy ý (bất kỳ thứ tự nào cũng hoạt động).

Ví dụ sau đây minh họa các định nghĩa về pmf cận biên và có điều kiện.

Ví dụ 3.1.5 (Chuyến bay và cơn mưa (tiếp theo)). Trong không gian xác suất được mô tả trong Ví dụ 1.2.1 , chúng tôi xác định một biến ngẫu nhiên

$$L = \begin{cases} 1 & \text{nếu máy bay đến muộn,} \\ 0 & \text{nếu ngược lại,} \end{cases} \tag{3.24}$$

để đại diện cho việc máy bay có bị trễ hay không. Tương tự,

$$R = \begin{cases} 1 & \text{trời mưa,} \\ 0 & \text{nếu không,} \end{cases} \tag{3.25}$$

đại diện cho dù trời có mưa hay không. Tương tự như vậy, các biến ngẫu nhiên này chỉ là các chỉ số $R = 1_{mưa}$ và $L = 1_{muộn}$. Bảng 3.1 cho thấy các pmfs chung, cận biên và có điều kiện của L và R.

3.2 Biến ngẫu nhiên liên tục

Các biến ngẫu nhiên liên tục cho phép chúng ta mô hình hóa các đại lượng liên tục mà không phải lo lắng về rời rạc hóa. Đổi lại, các công cụ toán học để thao tác với chúng phần nào phức tạp hơn so với trường hợp rời rạc.

3.2.1 Chung cdf và chung pdf

Như trong trường hợp các biến ngẫu nhiên liên tục đơn biến, chúng ta mô tả hành vi của một số biến ngẫu nhiên liên tục xác định trên cùng một không gian xác suất thông qua xác suất rằng chúng thuộc về các tập Borel (hoặc hợp tương đương của các khoảng). Trong trường hợp này, chúng ta đang xem xét các tập Borel nhiều chiều, là tích Descartes của Borel một chiều. bộ. Các tập Borel đa chiều có thể được biểu diễn dưới dạng hợp của các khoảng đa chiều hoặc siêu hình chữ nhật (được định nghĩa là tích Descartes của các khoảng một chiều). cdf chung tổng hợp xác suất mà các biến ngẫu nhiên thuộc về tích Descartes của các khoảng có dạng $(-\infty, r]$ với mọi $r \in R$.

Định nghĩa 3.2.1 (Hàm phân phối tích lũy chung). Cho (Ω, F, P) là một không gian xác suất và $X, Y : \Omega \to R$ biến ngẫu nhiên. cdf chung của X và Y được định nghĩa là

$$F_{X,Y}(x, y) := P(X \le x, Y \le y). \tag{3.26}$$

Nói cách khác, $F_{X,Y}(x, y)$ là xác suất để X và Y nhỏ hơn x và y tương ứng.

Cho $X : \Omega \to R^n$ là một vectơ ngẫu nhiên có kích thước n trên một không gian xác suất (Ω, F, P). khớp cdf của X được định nghĩa là

$$\text{Ngoại hối }(x) := P(X_1 \le x_1, X_2 \le x_2, \ldots, X_N \le x_n). \tag{3.27}$$

Nói cách khác, $F_X(x)$ là xác suất mà $X_i \le x_i$ với mọi $i = 1, 2, \ldots, N$.

Bây giờ chúng tôi ghi lại một số thuộc tính của cdf chung.

Bổ đề 3.2.2 (Tính chất của khớp cdf).

$$\lim_{x \to -\infty} F_{X,Y}(x, y) = 0, \tag{3.28}$$

$$\lim_{y \to -\infty} F_{X,Y}(x, y) = 0, \tag{3.29}$$

$$\lim_{x \to \infty, y \to \infty} F_{X,Y}(x, y) = 1, \tag{3.30}$$

$$F_{X,Y}(x_1, y_1) \le F_{X,Y}(x_2, y_2) \text{ nếu } x_2 \ge x_1, y_2 \ge y_1, \text{ tức là } F_{X,Y} \text{ không giảm.} \tag{3.31}$$

Bằng chứng. Chứng minh tuân theo cùng hướng với Bổ đề 2.3.2. □

cdf chung chỉ định hoàn toàn hành vi của các biến ngẫu nhiên tương ứng. Thật vậy, chúng ta có thể phân tách bất kỳ tập hợp Borel nào thành một hợp của các khoảng n chiều rời rạc và tính xác suất của chúng bằng cách đánh giá cdf chung. Hãy để chúng tôi minh họa điều này cho trường hợp hai biến:

$$P(x_1 \le X \le x_2, y_1 \le Y \le y_2) = P(\{X \le x_2, Y \le y_2\} \cap \{X > x_1\} \cap \{Y > y_1\}) \quad (3.32)$$

$$= P(X \le x_2, Y \le y_2) \quad P(X \le x_1, Y \le y_2) \quad (3.33)$$

$$P(X \le x_2, Y \le y_1) + P(X \le x_1, Y \le y_1) \quad (3.34)$$

$$= F_{X,Y}(x_2, y_2) \quad F_{X,Y}(x_1, y_2) \quad F_{X,Y}(x_2, y_1) + F_{X,Y}(x_1, y_1).$$

Điều này có nghĩa là, như trong trường hợp đơn biến, để xác định một vectơ ngẫu nhiên hoặc một nhóm các biến ngẫu nhiên, tất cả những gì chúng ta cần làm là xác định cdf chung của chúng. Chúng ta không phải lo lắng về không gian xác suất cơ bản.

Nếu khớp cdf khả vi, chúng ta có thể lấy đạo hàm của nó để thu được hàm mật độ xác suất chung của X và Y . Như trong trường hợp các biến ngẫu nhiên đơn biến, đây thường là cách thuận tiện hơn để xác định phân phối chung.

Định nghĩa 3.2.3 (Hàm mật độ xác suất chung). Nếu cdf chung của hai biến ngẫu nhiên X, Y khả vi thì pdf chung của chúng được định nghĩa là

$$f_{X,Y}(x, y) := \frac{2F_{X,Y}(x, y)}{x \ y}. \quad (3.35)$$

Nếu cdf chung của một vectơ ngẫu nhiên X là khả vi, thì pdf chung của nó được định nghĩa là

$$f_X(x) := \frac{nF_X(x)}{x_1 \ x_2 \cdots \ x_n}. \quad (3.36)$$

PDF chung nên được hiểu là mật độ n chiều, không phải là xác suất (ví dụ: nó có thể lớn hơn một). Trong trường hợp hai chiều,

$$\lim P(x \le X \le x + \ x, y \le Y \le y + \ y) = f_{X,Y}(x, y) \ x \ y. \ x \ 0, \ y \ 0 \quad (3.37)$$

Do tính đơn điệu của các cdf chung trong mọi biến, các pmf chung luôn luôn không âm.

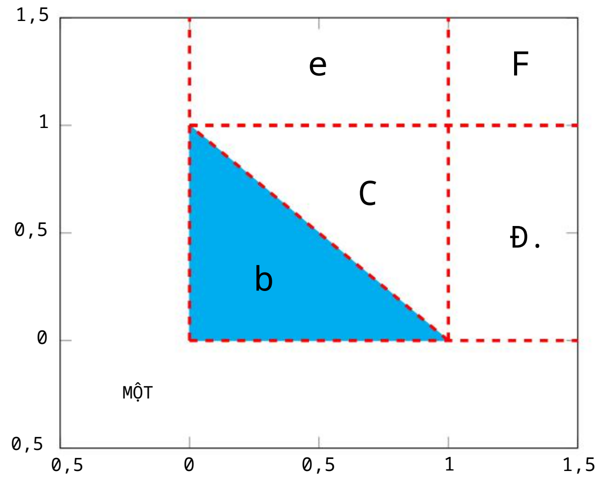
Bản pdf chung của X và Y cho phép chúng ta tính xác suất của bất kỳ tập Borel nào S R^2 qua tích phân trên S

$$P((X, Y) \ S) = \ _{(x,y) \ S} f_{X,Y}(x, y) \, dx \, dy. \quad (3.38)$$

Tương tự, pdf chung của vectơ ngẫu nhiên n chiều X cho phép tính xác suất X thuộc về một tập Borel tập S R^N,

$$P_X \ S = \ _{x \ S} f_X(x) dx. \quad (3.39)$$

Đặc biệt, nếu chúng ta tích hợp một pdf chung trên toàn bộ không gian R^n, sau đó nó phải tích hợp thành một theo Luật xác suất toàn phần.



Hình 3.1: Hồ tam giác trong Ví dụ 3.2.12.

Ví dụ 3.2.4 (Hồ tam giác). Một nhà sinh vật học đang theo dõi một con rái cá sống trong hồ. Cô ấy quyết định lập mô hình vị trí của con rái cá theo xác suất. Cái hồ có dạng tam giác như trong Hình 3.1, do đó chúng ta có thể biểu diễn nó bằng tập hợp

$$\text{Hồ} := \{x \mid x_1 \geq 0,\ x_2 \geq 0,\ x_1 + x_2 \leq 1\}\ . \tag{3.40}$$

Nhà sinh vật học không biết vị trí của con rái cá, vì vậy cô ấy lập mô hình vị trí như một véc tơ ngẫu nhiên X phân bố đều trên hồ. Nói cách khác, pdf chung của X là hằng số,

$$f_X(x) = \begin{cases} c & \text{nếu } x \\ \text{Lake}, 0 & \text{nếu không.} \end{cases} \tag{3.41}$$

Để tìm hằng số chuẩn hóa c, chúng tôi sử dụng thực tế là để trở thành một pdf chung hợp lệ fX nên tích hợp đến 1.

$$\int_{x_1=\infty}^{\infty} \int_{x_2=\infty}^{\infty} dx_1\, dx_2 = \int_{x_2=0}^{1} \int_{x_1=0}^{1\ x_2} dx_1\, dx_2 \tag{3.42}$$

$$= c \int_{x_2=0}^{1} (1 \quad x_2)\, dx_2 \tag{3.43}$$

$$= \frac{c}{2} = 1, \tag{3.44}$$

vậy c=2.

Bây giờ chúng ta tính cdf của X . FX (x) biểu thị xác suất con rái cá ở phía tây nam của điểm x. Việc tính toán cdf chung yêu cầu chia phạm vi thành các tập hợp được hiển thị trong Hình 3.1 và tích phân pdf chung.

Nếu x A thì FX (x) = 0 vì PX ≤ x = 0. Nếu (x) B thì

$$\text{Ngoại hối }(x) = \int_{u=0}^{x_2} \int_{v=0}^{x_1} 2\ \text{đv du} = 2x_1x_2. \tag{3.45}$$

Machine Translated by Google

Nếu x C,

$$\text{Ngoại hối } (x) = \int_{u=0}^{1\ \ x1} \int_{v=0}^{x1} 2\ \text{đv du} + \int_{u=1\ \ x1}^{x2} \int_{v=0}^{1\ \ u} 2\ \text{đv du} = 2x1 + 2x2 \quad x2 \quad {}_2x2 \quad 1_1 \tag{3.46}$$

Nếu x D,

$$FX\ (x) = PX\ 1 \leq x1,\ X\ 2 \leq x2 = PX\ 1 \leq 1,\ X \qquad 2 \leq x2 = FX\ (1,\ x2) = 2x2 \quad x2 \quad {}^2, \tag{3.47}$$

trong đó bước cuối cùng bắt đầu từ (3.46). Đổi chỗ x1 và x2, ta được FX (x) = 2x1 x2 x E theo cách 1 vì lập luận tương tự. Cuối cùng, với x F FX (x) = 1 vì PX 1 ≤ x1, X 2 ≤ x2 = 1.

Đặt mọi thứ lại với nhau,

$$\text{Ngoại hối } (x) = \begin{cases} 0 & \text{nếu } x1 < 0 \text{ hoặc } x2 < \\ 1, & 0, \text{ nếu } x1 \geq 0,\ x2 \geq 0,\ x1 + x2 \leq \\ 2x1x2,\ 2x1 + 2x2_2 \quad x2_1 \quad x2 & 1, \text{ nếu } x1 \leq 1,\ x2 \leq 1,\ x1 + x2 \geq 1, \\ 2x2 \quad x2_2, & \text{nếu } x1 \geq 1,\ 0 \leq x2 \leq 1, \\ 2x1 \quad x2_1, & \text{nếu } 0 \leq x1 \leq 1,\ x2 \geq 1, \\ 1, & \text{nếu } x1 \geq 1,\ x2 \geq 1. \end{cases} \tag{3.48}$$

3.2.2 Định biên

Bây giờ chúng ta thảo luận về cách mô tả các phân phối cận biên của các biến ngẫu nhiên riêng lẻ từ một cdf chung hoặc một pdf chung. Xét khớp cdf FX,Y (x, y). Khi x ∞ giới hạn của FX,Y (x, y) theo định nghĩa là xác suất của Y nhỏ hơn y, chính xác là cdf cận biên của Y . Chính thức hơn,

$$\lim_{x\ \infty} FX,Y\ (x,\ y) = \lim_{n\ \infty} P\ \left(\ {}_{\text{tôi} = 1}\ \{X \leq i,\ Y \leq y\}\right) \tag{3.49}$$

$$n = P \lim_{n\ \infty} \{X \leq n,\ Y \leq y\} \tag{3.50}$$

$$= P\ (Y \leq y) \tag{3.51}$$

$$= \text{Năm tài chính } (y). \tag{3.52}$$

Nếu các biến ngẫu nhiên có pdf chung, chúng ta cũng có thể tính toán cdf cận biên bằng cách lấy tích phân trên x

$$\text{Năm tài chính } (y) = P\ (Y \leq y) \tag{3.53}$$

$$= \int_{u=\ \infty}^{y} \int_{x=\ \infty}^{\infty} fX,Y\ (x,\ u)\ dx\ dy. \tag{3.54}$$

Vi phân phương trình sau đối với y, chúng ta thu được pdf cận biên của Y

$$fY\ (y) = \int_{x=\ \infty}^{\infty} fX,Y\ (x,\ y)\ dx. \tag{3.55}$$

Tương tự, pdf cận biên của véc tơ con XI có được của một vectơ ngẫu nhiên X được lập chỉ mục bởi I := {i1, i2, . . . , tôi} bằng cách lấy tích phân trên các thành phần còn lại {j1, j2, . . . , jn m} := {1, 2, . . . , n} /tôi,

$$\text{ngoại hối}_{\text{TÔI}} (x_I) = \int_{x_{j1}} \int_{x_{j2}} \cdots \int_{x_{jn\ m}} f_X (x)\, dx_{j1}\, dx_{j2} \cdots dx_{jn\ m}. \tag{3.56}$$

Ví dụ 3.2.5 (Hồ tam giác (tiếp theo)). Nhà sinh vật học quan tâm đến xác suất con rái cá ở phía nam của x1. Thông tin này được mã hóa trong cdf của véc tơ ngẫu nhiên, ta chỉ cần lấy giới hạn khi x2 ∞ để ra biên trên x2.

$$F_{X1} (x_1) = \begin{cases} 0 & \text{nếu } x_1 < 0, \\ 2x_1 \quad x_1^2 & \text{nếu } 0 \le x_1 \le 1, \\ 1 & \text{nếu } x_1 \ge 1. \end{cases} \tag{3.57}$$

Để có được pdf biên của X1, đại diện cho vĩ độ của vị trí của con rái cá, chúng tôi phân biệt cdf biên

$$f_{X1} (x_1) = \frac{dF_{X1} (x_1)}{dx_1} = \begin{cases} 2 (1 \quad x_1) & \text{nếu } 0 \le x_1 \le 1, \\ 0, & \text{ngược lại.} \end{cases} \tag{3.58}$$

Ngoài ra, chúng tôi có thể đã tích hợp pdf thống nhất chung trên x2 (chúng tôi khuyến khích bạn kiểm tra xem kết quả có giống nhau không).

3.2.3 Phân phối có điều kiện

Trong phần này, chúng ta thảo luận cách để có được phân phối có điều kiện của một biến ngẫu nhiên với thông tin về các biến ngẫu nhiên khác trong không gian xác suất. Để bắt đầu, chúng tôi xem xét trường hợp của hai biến ngẫu nhiên. Như trong trường hợp phân phối đơn biến, chúng ta có thể xác định cdf và pdf chung của hai biến ngẫu nhiên đã cho các sự kiện có dạng {(X, Y) S} cho bất kỳ Borel nào bằng cách áp dụng định nghĩa xác suất có điều kiện.
2 đặt trong R

Định nghĩa 3.2.6 (Cdf và pdf có điều kiện chung cho một sự kiện). Đặt X, Y là các biến ngẫu nhiên có chung pdf fX,Y và đặt S R² là bất kỳ tập Borel nào có xác suất khác 0, cdf và pdf có điều kiện của X và Y với sự kiện (X, Y) S được định nghĩa là

$$F_{X,Y |(X,Y)\ S} (x, y) := P (X \le x, Y \le y|(X, Y) \quad S) \tag{3.59}$$

$$= \frac{P (X \le x, Y \le y,(X, Y) \quad S)}{P ((X, Y) \quad S)} \tag{3.60}$$

$$= \frac{\int_{u\le x,v\le y,(u,v)\ S} f_{X,Y} (u, v)\, du\, dv}{\int_{(u,v)\ S} f_{X,Y} (u, v)\, du\, dv}, \tag{3.61}$$

$$f_{X,Y |(X,Y)\ S} (x, y) := \frac{\partial^2 F_{X,Y |(X,Y)\ S} (x, y)}{\partial x\, \partial y}. \tag{3.62}$$

Định nghĩa này chỉ đúng cho các sự kiện có xác suất khác không. Tuy nhiên, các sự kiện có dạng {X = x} có xác suất bằng 0 vì biến ngẫu nhiên là liên tục. Thật vậy, các

Machine Translated by Google

phạm vi của X là không đếm được, vì vậy xác suất của hầu hết mọi biến cố $\{X = x\}$ phải bằng 0, nếu không thì xác suất hợp của chúng sẽ không bị chặn.

Làm thế nào chúng ta có thể mô tả sự không chắc chắn của chúng ta về Y với X = x sau đó? Chúng tôi xác định một pdf có điều kiện ghi lại những gì chúng tôi đang cố gắng thực hiện trong giới hạn và sau đó tích hợp nó để có được một cdf có điều kiện.

Định nghĩa 3.2.7 (pdf và cdf có điều kiện). Nếu $F_{X,Y}$ khả vi, thì pdf có điều kiện của Y cho X được định nghĩa là

$$f_{Y|X}(y|x) := \frac{f_{X,Y}(x, y)}{f_X(x)} \quad \text{nếu } f_X(x) > 0 \tag{3.63}$$

và không được xác định khác.

cdf có điều kiện của Y đã cho X được định nghĩa là

$$F_{Y|X}(y|x) := \int_{u=-\infty}^{y} f_{Y|X}(u|x)\, du \text{ nếu } f_X(x) > 0 \tag{3.64}$$

và không được xác định khác.

Bây giờ chúng ta biện minh cho định nghĩa này, ngoài sự tương tự với (3.18). Giả sử rằng $f_X(x) > 0$. Hãy để chúng tôi viết định nghĩa của pdf có điều kiện dưới dạng các giới hạn. Chúng ta có

$$f_X(x) = \lim_{\Delta x \to 0} \frac{P(x \le X \le x + \Delta x)}{\Delta x}, \tag{3.65}$$

$$f_{X,Y}(x, y) = \lim_{\Delta x \to 0} \frac{1}{\Delta x} \frac{\partial P(x \le X \le x + \Delta x, Y \le y)}{\partial y}. \tag{3.66}$$

Điều này nghĩa là

$$\frac{f_{X,Y}(x, y)}{f_X(x)} = \lim_{\Delta x \to 0,\, \Delta y \to 0} \frac{1}{P(x \le X \le x + \Delta x)} \frac{\partial P(x \le X \le x + \Delta x, Y \le y)}{\partial y}. \tag{3.67}$$

Bây giờ chúng ta có thể viết cdf có điều kiện là

$$F_{Y|X}(y|x) = \int_{u=-\infty}^{y} \lim_{\Delta x \to 0,\, \Delta y \to 0} \frac{1}{P(x \le X \le x + \Delta x)} \frac{\partial P(x \le X \le x + \Delta x, Y \le u)}{\partial y}\, du \tag{3.68}$$

$$= \lim_{\Delta x \to 0} \frac{1}{P(x \le X \le x + \Delta x)} \int_{u=-\infty}^{y} \frac{\partial P(x \le X \le x + \Delta x, Y \le u)}{\partial y}\, du \tag{3.69}$$

$$= \lim_{\Delta x \to 0} \frac{P(x \le X \le x + \Delta x, Y \le y)}{P(x \le X \le x + \Delta x)} \tag{3.70}$$

$$= \lim_{\Delta x \to 0} P(Y \le y | x \le X \le x + \Delta x). \tag{3.71}$$

Do đó, chúng ta có thể giải thích cdf có điều kiện là giới hạn của cdf của Y tại y có điều kiện trên X thuộc một khoảng xung quanh x khi độ rộng của khoảng có xu hướng bằng không.

Nhận xét 3.2.8. Việc hoán đổi các giới hạn và tích phân như trong (3.69) nói chung không nhất thiết phải hợp lý. Trong trường hợp này, miễn là tích phân hội tụ và các đại lượng liên quan bị chặn.

Hệ quả trực tiếp của Định nghĩa 3.2.7 là quy luật dây chuyền đối với các biến ngẫu nhiên liên tục.

Bổ đề 3.2.9 (Quy tắc dây chuyền đối với biến ngẫu nhiên liên tục).

$$f_{X,Y}(x, y) = f_X(x)\, f_{Y|X}(y|x). \tag{3.72}$$

Áp dụng các ý tưởng tương tự như trong trường hợp hai biến, chúng tôi xác định phân phối có điều kiện của một vectơ con cho phần còn lại của vectơ ngẫu nhiên.

Định nghĩa 3.2.10 (PDF có điều kiện). PDF có điều kiện của vectơ con ngẫu nhiên X_I , I {1, 2, . . . , n}, với vectơ con X là {1,...,n}/I

$$f_{X_I|X_{\{1,...,n\}/I}}\left(x_I|x_{\{1,...,n\}/I}\right) := \frac{f_X(x)}{f_{X_{\{1,...,n\}/tôi}}\left(x_{\{1,...,n\}/I}\right)}. \tag{3.73}$$

Nó thường hữu ích để biểu diễn pdf chung của một vectơ ngẫu nhiên bằng cách đưa nó vào các tệp pdf có điều kiện bằng cách sử dụng quy tắc chuỗi cho các vectơ ngẫu nhiên.

Bổ đề 3.2.11 (Quy tắc chuỗi cho vectơ ngẫu nhiên). PDF chung của một vectơ ngẫu nhiên X có thể được phân tách thành

$$f_X(x) = f_{X_1}(x_1)\, f_{X_2|X_1}(x_2|x_1) \ldots f_{X_n|X_1,...,X_{n-1}}(x_n|x_1, \ldots, x_{n-1}) \tag{3.74}$$

$$= \prod_{tôi=1}^{N} f_{X_i|X_{\{1,...,i-1\}}}\left(x_i|x_{\{1,...,i-1\}}\right). \tag{3,75}$$

Lưu ý rằng thứ tự là tùy ý, bạn có thể sắp xếp lại thứ tự các thành phần của vectơ theo bất kỳ cách nào bạn muốn.

Bằng chứng. Kết quả sau khi áp dụng định nghĩa của pdf có điều kiện theo cách đệ quy. □

Ví dụ 3.2.12 (Hồ tam giác (tiếp theo)). Nhà sinh vật học phát hiện rái cá từ bờ hồ. Cô ấy đang đứng ở phía tây của hồ ở vĩ độ $x_1 = 0{,}75$ nhìn về phía đông và con rái cá ở ngay trước mặt cô ấy. Do đó, con rái cá cũng ở vĩ độ $x_1 = 0{,}75$, nhưng cô ấy không thể nói được ở khoảng cách nào. Sự phân bố vị trí của rái cá với vĩ độ X_1 của nó được đặc trưng bởi pdf có điều kiện của kinh độ X_2 với X_1,

$$f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)} \tag{3.76}$$

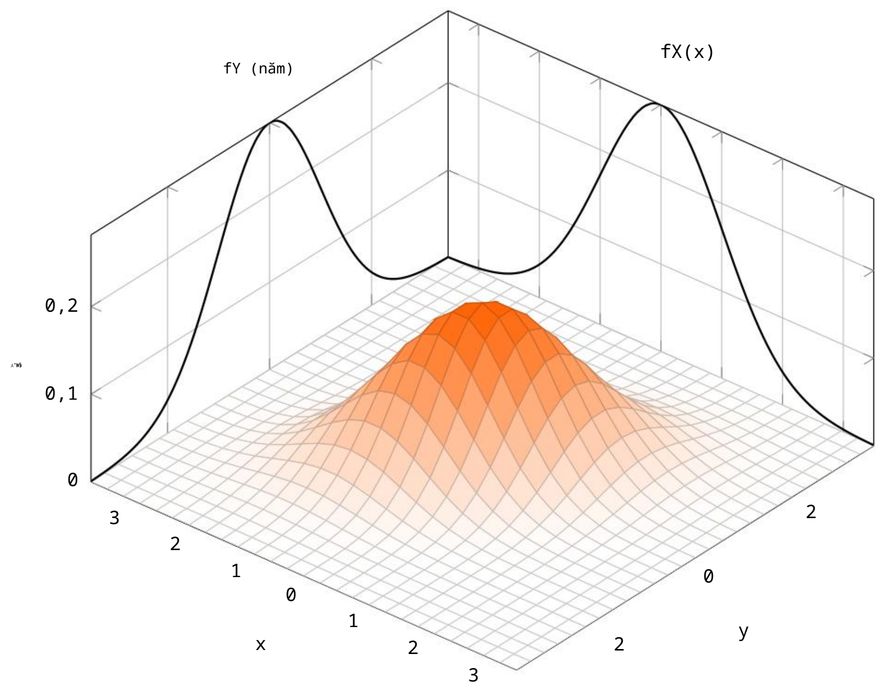
$$= \frac{1}{1 - x_1}, \quad 0 \le x_2 \le 1 - x_1. \tag{3.77}$$

Nhà sinh vật học quan tâm đến xác suất con rái cá ở gần cô ấy hơn x_2 . Xác suất này được đưa ra bởi cdf có điều kiện

$$F_{X_2|X_1}(x_2|x_1) = \int_{-\infty}^{x_2} f_{X_2|X_1}(u|x_1)\, du \tag{3.78}$$

$$= \frac{x_2}{1 - x_1}. \tag{3.79}$$

Xác suất để con rái cá đi xa ít hơn x_2 là $4x_2$ với $0 \le x_2 \le 1/4$.



Hình 3.2: PDF chung của biến ngẫu nhiên Gaussian hai biến (X, Y) cùng với pdf biên của X và Y .

3.2.4 Các vectơ ngẫu nhiên Gaussian

Các vectơ ngẫu nhiên Gaussian là một tổng quát hóa nhiều chiều của các biến ngẫu nhiên Gaussian. Chúng được tham số hóa bởi một vectơ và một ma trận tương ứng với ma trận trung bình và hiệp phương sai của chúng (chúng tôi định nghĩa các đại lượng này cho các biến ngẫu nhiên đa biến tổng quát trong Chương 4).

Định nghĩa 3.2.13 (Vectơ ngẫu nhiên Gauss). Một vectơ ngẫu nhiên Gaussian X là một vectơ ngẫu nhiên với pdf chung

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^N |\Sigma|}} \text{ điểm kinh nghiệm } \left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu) \right) \quad (3.80)$$

trong đó vectơ trung bình $\mu \in \mathbb{R}^n$ và ma trận hiệp phương sai Σ, đối xứng và xác định dương, tham số hóa phân bố. Một phân bố Gauss với trung bình μ và ma trận hiệp phương sai Σ thường được ký hiệu là N(μ, Σ).

Một tính chất cơ bản của các vectơ ngẫu nhiên Gaussian là việc thực hiện các phép biến đổi tuyến tính trên chúng luôn tạo ra các vectơ có phân bố chung cũng là Gaussian. Ta sẽ không chứng minh kết quả này một cách hình thức mà cách chứng minh tương tự như Bổ đề 2.5.1 (thực chất đây là một tổng quát hóa nhiều chiều của kết quả đó).

Định lý 3.2.14 (Phép biến đổi tuyến tính của vectơ ngẫu nhiên Gauss là Gaussian). Gọi X là một vectơ ngẫu nhiên Gaussian có chiều n với ma trận trung bình μ và hiệp phương sai Σ. Đối với bất kỳ ma trận $A \in \mathbb{R}^{m \times n}$ và $b \in \mathbb{R}^m$, $Y = AX + b$ là một vectơ ngẫu nhiên Gauss với trung bình $A\mu + b$ và ma trận hiệp phương sai $A\Sigma A^T$.

Một hệ quả tất yếu của kết quả này là pdf chung của vectơ con của vectơ ngẫu nhiên Gaussian cũng là một vectơ Gaussian.

Hệ quả 3.2.15 (Biên của vectơ ngẫu nhiên Gaussian là Gaussian). PDF chung của bất kỳ vectơ con nào của vectơ ngẫu nhiên Gaussian là Gaussian. Không mất tính tổng quát, giả sử rằng véc tơ con X bao gồm m phần tử đầu tiên của véc tơ ngẫu nhiên Gaussian,

$$z := \begin{bmatrix} X \\ Y \end{bmatrix}, \quad \text{với trung bình } \mu := \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} \tag{3.81}$$

và ma trận hiệp phương sai

$$\Sigma_Z = \begin{bmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{XY}^T & \Sigma_Y \end{bmatrix}. \tag{3.82}$$

Khi đó X là một vectơ ngẫu nhiên Gauss với trung bình μ_X và ma trận hiệp phương sai Σ_X.

Bằng chứng. Lưu ý rằng

$$x = \begin{bmatrix} I_m & 0_{m \times n-m} \\ 0_{n-m \times m} & 0_{n-m \times n-m} \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} I_m & 0_{m \times n-m} \\ 0_{n-m \times m} & 0_{n-m \times n-m} \end{bmatrix} Z, \tag{3.83}$$

trong đó $I_{m \times m}$ là một ma trận đơn vị và $0_{c \times d}$ đại diện cho một ma trận các số 0 có kích thước $c \times d$. R Kết quả sau đó suy ra từ Định lý 3.2.14. □

Hình 3.2 cho thấy pdf chung của biến ngẫu nhiên Gaussian hai biến số cùng với pdf biên của nó.

3.3 Phân phối chung của các biến rời rạc và liên tục

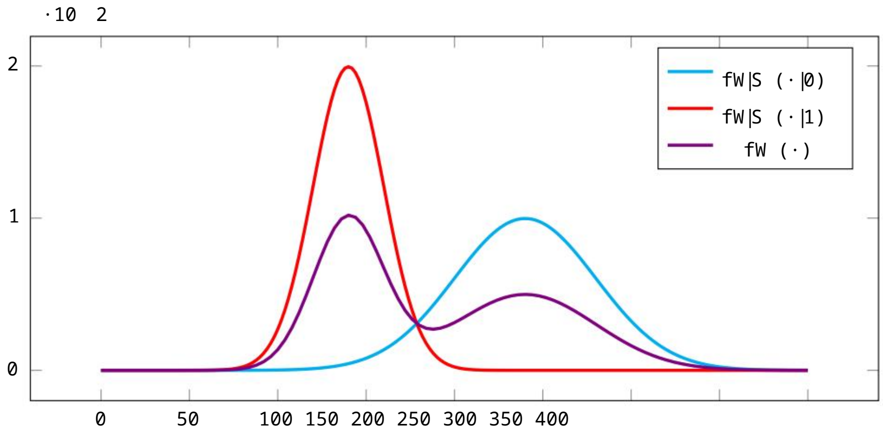
Các mô hình xác suất thường bao gồm cả các biến ngẫu nhiên rời rạc và liên tục. Tuy nhiên, pmf hoặc pdf chung của một biến ngẫu nhiên rời rạc và liên tục không được xác định rõ. Để chỉ định phân phối chung trong những trường hợp như vậy, chúng tôi sử dụng pmfs và pdf cận biên và có điều kiện của chúng. Giả sử rằng chúng ta có một biến ngẫu nhiên liên tục C và một biến ngẫu nhiên rời rạc D với phạm vi R_D. Chúng tôi xác định cdf và pdf có điều kiện của C cho D như sau.

Định nghĩa 3.3.1 (Cdf và pdf có điều kiện của biến ngẫu nhiên liên tục cho biến ngẫu nhiên rời rạc). Cho C và D là biến ngẫu nhiên liên tục và rời rạc xác định trên cùng một không gian xác suất. Khi đó, cdf và pdf điều kiện của C cho D có dạng

$$F_{C|D}(c|d) := P(C \leq c|d), \tag{3.84}$$

$$f_{C|D}(c|d) := \frac{dF_{C|D}(c|d)}{dc}. \tag{3.85}$$

Chúng tôi thu được cdf và pdf cận biên của C từ các cdf và pdf có điều kiện bằng cách tính tổng trọng số.



Hình 3.3: Phân phối trọng lượng có điều kiện và cận biên của các con gấu W trong Ví dụ 3.3.3.

Bổ đề 3.3.2. Cho $F_{C|D}$ và $f_{C|D}$ là cdf và pdf có điều kiện của biến ngẫu nhiên liên tục C cho trước một biến ngẫu nhiên rời rạc D. Sau đó,

$$F_C(c) = \sum_{d \in R_D} p_D(d)\, F_{C|D}(c|d), \tag{3.86}$$

$$f_C(c) = \sum_{d \in R_D} p_D(d)\, f_{C|D}(c|d). \tag{3.87}$$

Bằng chứng. Các sự kiện $\{D = d\}$ là một phân vùng của toàn bộ không gian xác suất (một trong số chúng phải xảy ra và tất cả chúng đều rời rạc), vì vậy

$$F_C(c) = P(C \le c) \tag{3.88}$$

$$= \sum_{d \in R_D} P(D = d)\, P(C \le c|d) \text{ theo Quy luật xác suất toàn phần} \tag{3.89}$$

$$= \sum_{d \in R_D} p_D(d)\, F_{C|D}(c|d). \tag{3.90}$$

Bây giờ, (3.87) theo sau bởi vi phân. □

Kết hợp pmf cận biên rời rạc với phân phối có điều kiện liên tục cho phép chúng ta xác định các mô hình hỗn hợp trong đó dữ liệu được rút ra từ một phân phối liên tục có các tham số là được chọn từ một tập hợp rời rạc. Nếu một Gaussian được sử dụng làm phân phối liên tục, điều này mang lại một Mô hình hỗn hợp Gaussian. Lắp các mô hình hỗn hợp Gaussian là một kỹ thuật phổ biến để phân cụm dữ liệu.

Ví dụ 3.3.3 (Grizzlies ở Yellowstone). Một nhà khoa học đang thu thập dữ liệu về những con gấu ở Yel lowstone. Nó chỉ ra rằng trọng lượng của con đực được mô hình hóa tốt bởi một Gaussian ngẫu nhiên biến với trung bình 240 kg và biến chuẩn 40 kg, trong khi trọng lượng của con cái là được mô hình hóa tốt bởi Gaussian với trung bình 140 kg và độ lệch chuẩn 20 kg. Có khoảng số nam và nữ như nhau.

Do đó, sự phân bố trọng số của tất cả các con gấu xám có thể được mô hình hóa bằng hỗn hợp Gauss bao gồm một biến ngẫu nhiên liên tục W để biểu thị trọng lượng và một biến ngẫu nhiên rời rạc S để biểu thị giới tính của những con gấu. S là Bernoulli với tham số 1/2, W với S = 0 (nam) là N(240, 1600) và W với S = 1 (nữ) là N(140, 400). Bởi (3.87) pdf của W do đó có dạng

$$f_W(w) = \sum_{s=0}^{1} p_S(s)\, f_{W|S}(w|s) \tag{3.91}$$

$$= \frac{1}{2\sqrt{2\pi}} \left(\frac{e^{-\frac{(w-240)^2}{3200}}}{40} + \frac{e^{-\frac{(w-140)^2}{800}}}{20} \right). \tag{3.92}$$

Hình 3.3 cho thấy các phân phối có điều kiện và cận biên của W.

Việc xác định pmf có điều kiện của biến ngẫu nhiên rời rạc D với biến ngẫu nhiên liên tục C là một thách thức vì xác suất của biến cố {C = c} bằng không. Chúng tôi thực hiện theo cách tiếp cận tương tự như trong Định nghĩa 3.2.7 và xác định pmf có điều kiện là một giới hạn.

Định nghĩa 3.3.4 (Pmf có điều kiện của biến ngẫu nhiên rời rạc cho biến ngẫu nhiên liên tục). Cho C và D là biến ngẫu nhiên liên tục và rời rạc xác định trên cùng một không gian xác suất. Sau đó, pmf có điều kiện của D cho C được định nghĩa là

$$p_{D|C}(d|c) := \lim_{\Delta \to 0} \frac{P(D = d, c \le C \le c + \Delta)}{P(c \le C \le c + \Delta)}. \tag{3.93}$$

Tương tự như Bổ đề 3.3.2, chúng ta thu được pmf biên của D từ các pmf có điều kiện bằng cách tính tổng trọng số.

Bổ đề 3.3.5. Đặt $p_{D|C}$ là pmf có điều kiện của biến ngẫu nhiên rời rạc D với biến ngẫu nhiên liên tục C. Sau đó,

$$p_D(d) = \int_{c=-\infty}^{\infty} f_C(c)\, p_{D|C}(d|c)\, dc. \tag{3.94}$$

Bằng chứng. Chúng tôi sẽ không đưa ra một bằng chứng chính thức mà là một lập luận trực quan có thể được đưa ra một cách chặt chẽ. Nếu chúng ta lấy một lưới các giá trị cho c nằm trên một lưới. . . , $c_{-1}, c_0, c_1, \ldots$ của chiều rộng Δ, sau đó

$$p_D(d) = \sum_{i=-\infty}^{\infty} P(D = d, c_i \le C \le c_i + \Delta) \tag{3,95}$$

theo Định luật xác suất toàn phần. Lấy giới hạn là $\Delta \to 0$ tổng trở thành tích phân và ta có

$$p_D(d) = \int_{c=-\infty}^{\infty} \lim_{\Delta \to 0} \frac{P(D = d, c \le C \le c + \Delta)}{\Delta}\, dc \tag{3.96}$$

$$= \int_{c=-\infty}^{\infty} \lim_{\Delta \to 0} \frac{P(c \le C \le c + \Delta)}{\Delta} \cdot \frac{P(D = d, c \le C \le c + \Delta)}{P(c \le C \le c + \Delta)}\, dc \tag{3.97}$$

$$= \int_{c=-\infty}^{\infty} f_C(c)\, p_{D|C}(d|c)\, dc. \tag{3.98}$$

vì $f_C(c) = \lim_{\Delta \to 0} \frac{P(c \le C \le c + \Delta)}{\Delta}$. □

Việc kết hợp các phân phối cận biên liên tục với các phân phối có điều kiện rời rạc đặc biệt hữu ích trong các mô hình thống kê Bayes, như được minh họa trong ví dụ sau (xem Chương 10 để biết thêm thông tin). Phân phối liên tục được sử dụng để định lượng độ không chắc chắn của chúng ta về tham số của phân phối rời rạc.

Ví dụ 3.3.6 (Tung đồng xu Bayesian). Chú của bạn cá với bạn mười đô la rằng tung đồng xu sẽ ra mặt ngửa. Bạn nghi ngờ rằng đồng xu bị thiên vị, nhưng bạn không chắc chắn ở mức độ nào. Để mô hình hóa sự không chắc chắn này, bạn biểu thị độ lệch dưới dạng biến ngẫu nhiên liên tục B với bản pdf sau:

$$f_B(b) = 2b \text{ với } b \in [0, 1]. \tag{3.99}$$

Bây giờ bạn có thể tính xác suất để đồng xu xuất hiện mặt ngửa được biểu thị bằng X bằng cách sử dụng Bổ đề 3.3.5. Dựa trên độ lệch B, kết quả của việc tung đồng xu là Bernoulli với tham số B.

$$p_X(1) = \int_{b=-\infty}^{\infty} f_B(b)\, p_{X|B}(1|b)\, db \tag{3.100}$$

$$= \int_{b=0}^{1} 2b^2\, db \tag{3.101}$$

$$= \frac{2}{3}. \tag{3.102}$$

Theo mô hình của bạn, xác suất để đồng xu xuất hiện mặt ngửa là 2/3.

Bổ đề sau cung cấp một tương tự cho quy tắc dây chuyền cho các biến ngẫu nhiên liên tục và rời rạc có phân phối đồng thời.

Bổ đề 3.3.7 (Quy tắc chuỗi đối với các biến ngẫu nhiên liên tục và rời rạc có phân phối chung). Cho C là biến ngẫu nhiên liên tục với pdf $f_{C|D}$ có điều kiện và D là biến ngẫu nhiên rời rạc với pmf $p_{D|C}$ có điều kiện. Sau đó,

$$p_D(d)\, f_{C|D}(c|d) = f_C(c)\, p_{D|C}(d|c). \tag{3.103}$$

Bằng chứng. Áp dụng các định nghĩa,

$$p_D(d)\, f_{C|D}(c|d) = \lim_{\Delta \to 0} P(D = d)\, \frac{P(c \le C \le c + \Delta | D = d)}{\Delta} \tag{3.104}$$

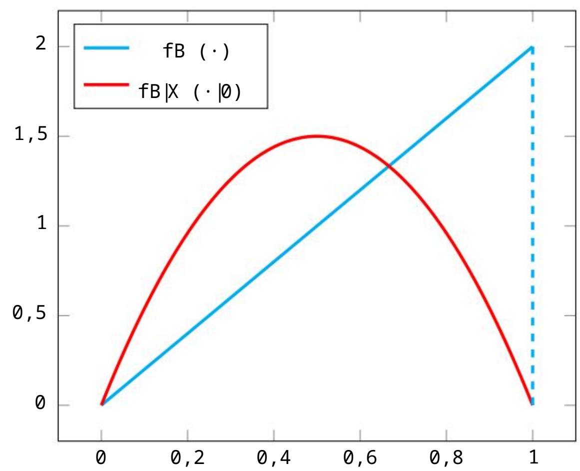
$$= \lim_{\Delta \to 0} \frac{P(D = d, c \le C \le c + \Delta)}{\Delta} \tag{3.105}$$

$$= \lim_{\Delta \to 0} \frac{P(c \le C \le c + \Delta)}{\Delta} \cdot \frac{P(D = d, c \le C \le c + \Delta)}{P(c \le C \le c + \Delta)} \tag{3.106}$$

$$= f_C(c)\, p_{D|C}(d|c). \tag{3.107}$$

□

Ví dụ 3.3.8 (Grizzlies ở Yellowstone (tiếp theo)). Nhà khoa học quan sát một con gấu bằng ống nhòm của cô ấy. Từ kích thước của chúng, cô ước tính rằng trọng lượng của nó là 180 kg. Xác suất mà con gấu là con đực là gì?



Hình 3.4: Phân phối biên và có điều kiện của độ chệch của trò tung đồng xu trong Ví dụ 3.3.9.

Ta áp dụng bổ đề 3.3.7 để tính

$$p_{S|W}(0|180) = \frac{p_S(0)\, f_{W|S}(180|0)}{f_W(180)} \tag{3.108}$$

$$= \frac{\frac{1}{40\text{ điểm kinh nghiệm}} \frac{602}{3200}}{\frac{1}{40\text{ điểm kinh nghiệm}} \frac{602}{3200} + \frac{1}{20\text{ điểm kinh nghiệm}} \frac{402}{800}} \tag{3.109}$$

$$= 0{,}545. \tag{3.110}$$

Theo mô hình xác suất, xác suất đó là nam là 0,545.

Ví dụ 3.3.9 (Tung đồng xu Bayesian (tiếp theo)). Đồng xu rơi vào mặt sấp. Bạn quyết định đề xuất phân phối thành kiến dựa trên thông tin này. Theo bổ đề 3.3.7

$$f_{B|X}(b|0) = \frac{f_B(b)\, p_{X|B}(0|b)}{p_X(0)} \tag{3.111}$$

$$= \frac{2b\,(1 \quad b)}{1/3} \tag{3.112}$$

$$= 6b(1 \quad b). \tag{3.113}$$

Tùy thuộc vào kết quả, bản pdf của xu hướng hiện được tập trung thay vì tập trung gần một như trước đây, như trong Hình 3.4.

3.4 Độc lập

Trong phần này, chúng tôi xác định tính độc lập và độc lập có điều kiện cho các biến ngẫu nhiên và vectơ.

3.4.1 Định nghĩa

Khi kiến thức về một biến ngẫu nhiên X không ảnh hưởng đến sự không chắc chắn của chúng ta về một biến khác biến ngẫu nhiên Y tạ nói X và Y độc lập. Về mặt hình thức, điều này được phản ánh bởi cdf cận biên và có điều kiện và pmf hoặc pdf có điều kiện phải bằng nhau, nghĩa là

$$\text{Năm tài chính } (y) = \text{Năm tài chính } |X\ (y|x) \tag{3.114}$$

Và

$$p_Y(y) = p_{Y|X}(y|x) \text{ hoặc } f_Y(y) = f_{Y|X}(y|x), \tag{3.115}$$

tùy thuộc vào việc biến là rời rạc hay liên tục, với bất kỳ x và bất kỳ y nào mà phân phối có điều kiện được xác định rõ. Tương tự, cdf chung và pmf có điều kiện hoặc pdf yếu tố vào bên lề.

Định nghĩa 3.4.1 (Biến ngẫu nhiên độc lập). Hai biến ngẫu nhiên X và Y độc lập với nhau khi và chỉ khi

$$F_{X,Y}(x, y) = F_X(x)\, F_Y(y), \text{ với mọi } (x, y) \quad R^2. \tag{3.116}$$

Nếu các biến rời rạc thì điều kiện sau tương đương

$$p_{X,Y}(x, y) = p_X(x)\, p_Y(y), \text{ với mọi } x \quad R_X,\ y \quad R_Y. \tag{3.117}$$

Nếu các biến liên tục có pdf chung và cận biên thì điều kiện sau là tương đương

$$f_{X,Y}(x, y) = f_X(x)\, f_Y(y), \text{ với mọi } (x, y) \quad R^2. \tag{3.118}$$

Bây giờ chúng ta mở rộng định nghĩa để giải thích cho một số biến ngẫu nhiên (hoặc tương đương với một số các mục trong một vectơ ngẫu nhiên) không cung cấp thông tin về nhau.

Định nghĩa 3.4.2 (Các biến ngẫu nhiên độc lập). N mục X1, X2, . . . , Xn một cách ngẫu nhiên vectơ X độc lập khi và chỉ khi

$$\text{Ngoại hối } (x) = \prod_{\text{tôi}=1}^{N} \text{Ngoại hối } (x_i), \tag{3.119}$$

tương đương với

$$p_X(x) = \prod_{\text{tôi}=1}^{N} p_{X_i}(x_i) \tag{3.120}$$

cho các vectơ rời rạc và

$$f_X(x) = \prod_{\text{tôi}=1}^{N} f_{X_i}(x_i) \tag{3.121}$$

đối với các vectơ liên tục, nếu tệp pdf chung tồn tại.

Ví dụ sau đây cho thấy tính độc lập theo cặp không có nghĩa là tính độc lập.

Ví dụ 3.4.3 (Độc lập theo cặp không có nghĩa là độc lập chung). Gọi X1 và X2 là kết quả của các lần tung đồng xu độc lập, không thiên vị. Gọi X3 là chỉ số của biến cố {X1 và X2 có cùng kết cục},

$$X_3 = \begin{cases} 1 \text{ nếu } X_1 = X_2, \\ 0 \text{ nếu } X_1 = X_2. \end{cases} \tag{3.122}$$

pmf của X3 là

$$p_{X_3}(1) = p_{X_1,X_2}(1,1) + p_{X_1,X_2}(0,0) = \frac{1}{2}, \tag{3.123}$$

$$p_{X_3}(0) = p_{X_1,X_2}(0,1) + p_{X_1,X_2}(1,0) = \frac{1}{2}. \tag{3.124}$$

X1 và X2 độc lập theo giả thiết. X1 và X3 độc lập vì

$$p_{X_1,X_3}(0,0) = p_{X_1,X_2}(0,1) = \frac{1}{4} = p_{X_1}(0)\, p_{X_3}(0), \tag{3.125}$$

$$p_{X_1,X_3}(1,0) = p_{X_1,X_2}(1,0) = \frac{1}{4} = p_{X_1}(1)\, p_{X_3}(0), \tag{3.126}$$

$$p_{X_1,X_3}(0,1) = p_{X_1,X_2}(0,0) = \frac{1}{4} = p_{X_1}(0)\, p_{X_3}(1), \tag{3.127}$$

$$p_{X_1,X_3}(1,1) = p_{X_1,X_2}(1,1) = \frac{1}{4} = p_{X_1}(1)\, p_{X_3}(1). \tag{3.128}$$

X2 và X3 cũng độc lập (lý do là như nhau).

Tuy nhiên, X1, X2 và X3 có độc lập không?

$$p_{X_1,X_2,X_3}(1,1,1) = P(X_1 = 1, X_2 = 1) = \frac{1}{4} \neq p_{X_1}(1)\, p_{X_2}(1)\, p_{X_3}(1) = \frac{1}{8}. \tag{3.129}$$

Chúng không phải, điều này hợp lý vì X3 là hàm của X1 và X2.

Tính độc lập có điều kiện chỉ ra rằng hai biến ngẫu nhiên không phụ thuộc vào nhau, miễn là biết thêm một biến ngẫu nhiên.

Định nghĩa 3.4.4 (Biến ngẫu nhiên độc lập có điều kiện). Hai biến ngẫu nhiên X và Y độc lập với biến ngẫu nhiên Z khác khi và chỉ khi

$$\text{ngoại hối}, _{Y \mid Z}(x,y|z) = F_{X|Z}(x|z)\, F_{Y|Z}(y \mid z), \text{ với mọi } (x,y) \in \mathbb{R}^2, \tag{3.130}$$

và bất kỳ z nào mà các cdf có điều kiện được xác định rõ. Nếu các biến rời rạc thì điều kiện sau tương đương

$$p_{X,Y \mid Z}(x, y \mid z) = p_{X \mid Z}(x|z)\, p_{Y|Z}(y \mid z), \text{ với mọi } x \in R_X, y \in R_Y, \tag{3.131}$$

và bất kỳ z nào mà pmfs có điều kiện được xác định rõ. Nếu các biến liên tục có pdf chung và cận biên thì điều kiện sau là tương đương

$$f_{X,Y \mid Z}(x, y \mid z) = f_{X \mid Z}(x|z)\, f_{Y|Z}(y \mid z), \text{ với mọi } (x,y) \in \mathbb{R}^2. \tag{3.132}$$

và bất kỳ z nào mà pmfs có điều kiện được xác định rõ.

Định nghĩa có thể được mở rộng cho điều kiện trên một số biến ngẫu nhiên.

Định nghĩa 3.4.5 (Biến ngẫu nhiên độc lập có điều kiện). Các thành phần của vectơ con X_I , $I \subseteq \{1, 2, . . . , n\}$ độc lập có điều kiện với một vectơ con khác X_J , $J \subseteq \{1, 2, . . . , n\}$, nếu và chỉ nếu

$$\text{ngoại hối}_{\text{Tôi}\,|X_J}(x_I|x_J) = \prod_{i \in I} F_{X_i|X_J}(x_i|x_J), \tag{3.133}$$

tương đương với

$$p_{X_{\text{Tôi}}|X_J}(x_I|x_J) = \prod_{i \in I} p_{X_i|X_J}(x_i|x_J) \tag{3.134}$$

cho các vectơ rời rạc và

$$f_{X_{\text{Tôi}}|X_J}(x_I|x_J) = \prod_{i \in I} f_{X_i|X_J}(x_i|x_J) \tag{3.135}$$

đối với các vectơ liên tục nếu tồn tại liên kết có điều kiện pdf.

Như đã thiết lập trong Ví dụ 1.3.5 và 1.3.6, tính độc lập không bao hàm tính độc lập có điều kiện hoặc ngược lại.

3.4.2 Sự phụ thuộc của biến trong mô hình xác suất

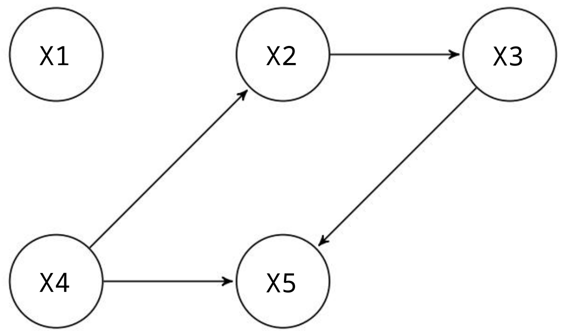
Một cân nhắc cơ bản khi thiết kế một mô hình xác suất là sự phụ thuộc giữa các biến khác nhau, tức là biến nào là độc lập hoặc độc lập có điều kiện với nhau. Mặc dù nghe có vẻ ngạc nhiên, nhưng nếu số lượng biến lớn, thì việc đưa ra một số giả định về tính độc lập có thể cần thiết để làm cho mô hình trở nên khả thi, ngay cả khi chúng ta biết rằng tất cả các biến đều phụ thuộc. Để minh họa điều này, hãy xem xét một mô hình bầu cử tổng thống Hoa Kỳ trong đó có 50 biến ngẫu nhiên, mỗi biến đại diện cho một tiểu bang. Nếu các biến chỉ nhận hai giá trị có thể (đại diện cho ứng cử viên nào giành được trạng thái đó), pmf chung của phân phối của chúng có $2^{50} - 1 \geq 10^{15}$ bậc tự do. Chúng tôi sẽ không thể lưu trữ pmf với tất cả bộ nhớ máy tính trên thế giới! Ngược lại, nếu chúng ta giả sử rằng tất cả các biến là độc lập, thì phân phối chỉ có 50 tham số miễn phí. Tất nhiên, đây không nhất thiết là một ý tưởng hay vì việc không thể hiện các phụ thuộc có thể ảnh hưởng nghiêm trọng đến độ chính xác dự đoán của một mô hình, như được minh họa trong Ví dụ 3.5.1 bên dưới. Đạt được sự cân bằng giữa tính khả thi và độ chính xác là một thách thức quan trọng trong mô hình xác suất.

Bây giờ chúng tôi minh họa cách cấu trúc phụ thuộc của các biến ngẫu nhiên trong mô hình xác suất có thể được khai thác để giảm số lượng tham số mô tả phân phối thông qua một hệ số thích hợp của pmf hoặc pdf chung của chúng. Xét ba biến ngẫu nhiên Bernoulli A, B và C. Nói chung, chúng ta cần $7 = 2^3 - 1$ tham số để mô tả pmf. Tuy nhiên, nếu B và C độc lập có điều kiện với A, chúng ta có thể thực hiện phân tích thành thừa số sau

$$p_{A,B,C} = p_A\, p_{B\,|\,\text{Một máy tính}\,|\,\text{MỘT}} \tag{3.136}$$

chỉ phụ thuộc vào năm tham số (một cho p_A và hai cho $p_{B\,|\,A}$ và $p_{C\,|\,B}$). Điều quan trọng cần lưu ý là có nhiều yếu tố có thể khác không khai thác các giả định phụ thuộc, chẳng hạn như

$$p_{A,B,C} = p_B\, p_{A\,|\,B}\, p_{C\,|\,A,B}. \tag{3.137}$$



Hình 3.5: Ví dụ về đồ họa tuần hoàn có hướng biểu diễn mô hình xác suất.

Đối với các mô hình xác suất lớn, điều quan trọng là tìm ra các hệ số làm giảm số lượng tham số càng nhiều càng tốt.

3.4.3 Mô hình đồ họa

Các mô hình đồ thị là một công cụ để mô tả cấu trúc phụ thuộc của các mô hình xác suất. Trong phần này, chúng tôi đưa ra một mô tả ngắn gọn về các mô hình đồ họa có hướng, còn được gọi là mạng Bayes. Các mô hình đồ thị vô hướng, được gọi là trường ngẫu nhiên Markov, nằm ngoài phạm vi của các ghi chú này. Chúng tôi giới thiệu người đọc quan tâm đến các văn bản nâng cao hơn về mô hình xác suất và học máy để xử lý sâu hơn các mô hình đồ họa.

Các đồ thị tuần hoàn có hướng, được gọi là DAG, có thể được hiểu là các sơ đồ biểu thị một hệ số hóa của pmf hoặc pdf chung của một mô hình xác suất. Để xác định một thừa số hợp lệ, các đồ thị bị ràng buộc không có bất kỳ chu trình nào (do đó có thuật ngữ chu kỳ). Mỗi nút trong DAG đại diện cho một biến ngẫu nhiên. Các cạnh giữa các nút biểu thị sự phụ thuộc giữa các biến. Hệ số tương ứng với một DAG chứa:

- PMf hoặc pdf cận biên của các biến tương ứng với tất cả các nút không có đầu vào cạnh.
- pmf hoặc pdf có điều kiện của các biến ngẫu nhiên còn lại đã cho cha của chúng. A là cha của B nếu có một cạnh có hướng từ (nút được gán cho) A đến (nút được gán cho) B.

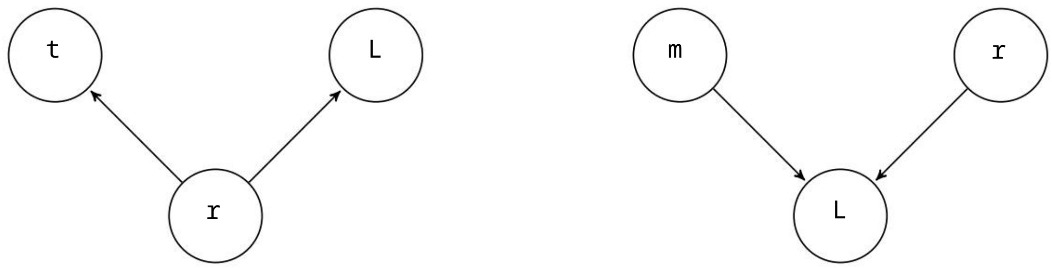
Để cụ thể hơn, hãy xem xét DAG trong Hình 3.5. Để đơn giản, chúng tôi biểu thị từng nút bằng cách sử dụng biến ngẫu nhiên tương ứng và giả sử rằng tất cả chúng đều rời rạc. Các nút X_1 và X_4 không có nút cha, do đó, hệ số của pmf chung bao gồm các pmf cận biên của chúng. Nút X_2 chỉ đi xuống từ X_4 , vì vậy chúng tôi bao gồm $p_{X_2 | X_4}$. Nút X_3 đi xuống từ X_2 , vì vậy chúng tôi bao gồm $p_{X_3 | X_2}$. Cuối cùng, nút X_5 đi xuống từ X_3 và X_4 , vì vậy chúng tôi bao gồm $p_{X_5 | X_3,X_4}$. Phép phân tích có dạng

$$p_{X_1,X_2,X_3,X_4,X_5} = p_{X_1} p_{X_4} p_{X_2 | X_4} p_{X_3 | X_2} p_{X_5 | X_3,X_4} . \tag{3.138}$$

Yếu tố này cho thấy một số giả định phụ thuộc. Theo quy tắc dây chuyền, một cấu trúc hợp lệ khác của pmf chung là

$$p_{X_1,X_2,X_3,X_4,X_5} = p_{X_1} p_{X_4 | X_1} p_{X_2 | X_1,X_4} p_{X_3 | X_1,X_2,X_4} p_{X_5 | X_1,X_2,X_3,X_4} . \tag{3.139}$$

$p_{L,R,T} = p_R \; p_{L \mid R} \; p_{T \mid r}$ $\qquad$ $p_{M,R,T} = p_M \; p_R \; p_{L \mid ÔNG}$



Hình 3.6: Các mô hình đồ họa có hướng tương ứng với các biến trong Ví dụ 1.3.5 và 1.3.6.

So sánh cả hai biểu thức, chúng ta thấy rằng X1 và tất cả các biến khác là độc lập, kể từ đó, v.v. X4 X4 cho X2 kể từ $p_{X3 \mid X2,X4} = p_{X3 \mid}$ Ngoài ra, X3 độc lập với $p_{X4 \mid X1} = p_{X4}$, $p_{X2 \mid X1,X4} = p_{X2 \mid X2}$. Các giả định phụ thuộc này có thể được đọc trực tiếp từ biểu đồ, sử dụng thuộc tính sau.

Định lý 3.4.6 (Tính chất Markov địa phương). Hệ số hóa của pmf hoặc pdf chung được đại diện bởi một DAG thỏa mãn thuộc tính Markov cục bộ: mỗi biến độc lập có điều kiện với các biến không phải con cháu của nó với tất cả các biến cha của nó. Đặc biệt, nếu nó không có cha mẹ, thì nó độc lập với những người không phải là con cháu của nó. Để rõ ràng, B không phải là con cháu của A nếu không có đường đi trực tiếp từ A đến B.

Bằng chứng. Cho Xi là một biến tùy ý. Chúng tôi biểu thị bằng XN tập hợp các con không phải con cháu của Xi , bởi XP là tập hợp cha mẹ và bằng XD tập hợp con cháu. Thừa số được biểu diễn bởi mô hình đồ thị có dạng

$$p_{X1,...,Xn} = p_{XN} \; p_{XP \mid XN} \; p_{Xi \mid XP} \; p_{XD \mid Xi} . \tag{3.140}$$

Theo quy tắc dây chuyền, một thừa số hợp lệ khác là

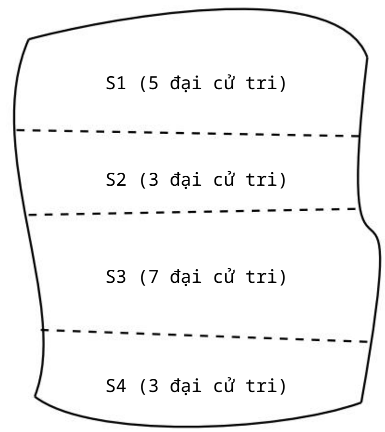
$$p_{X1,...,Xn} = p_{XN} \; p_{XP \mid XN} \; p_{Xi \mid XP ,XN} \; p_{XD \mid Xi,XP ,XN} . \tag{3.141}$$

So sánh cả hai biểu thức, chúng tôi kết luận rằng $p_{Xi \mid XP ,XN} = p_{Xi \mid XP}$ nên Xi phụ thuộc có điều kiện vào XN đã cho XP . □

Chúng tôi minh họa những ý tưởng này bằng cách hiển thị các DAG cho Ví dụ 1.3.5 và 1.3.6.

Ví dụ 3.4.7 (Mô hình đồ thị cho Ví dụ 1.3.5). Chúng tôi lập mô hình các sự kiện khác nhau trong Ví dụ 1.3.5 bằng cách sử dụng các biến ngẫu nhiên của chỉ báo. T đại diện cho việc có taxi hay không (T = 1) hay không (T = 0), L cho biết máy bay có bị trễ (L = 1) hay không (L = 0) và R cho biết trời mưa (R = 1) hay không (R = 0). Trong ví dụ này, T và L độc lập có điều kiện với R. Chúng ta có thể biểu diễn phép phân tích thành thừa số tương ứng bằng đồ thị bên trái của Hình 3.6.

Ví dụ 3.4.8 (Mô hình đồ thị cho Ví dụ 1.3.6). Chúng tôi lập mô hình các sự kiện khác nhau trong Ví dụ 1.3.6 bằng cách sử dụng các biến ngẫu nhiên chỉ báo. M đại diện cho một vấn đề cơ khí xảy ra



Hình 3.7: Quốc gia giả định được xem xét trong Ví dụ 3.4.9.

(M = 1) hoặc không (M = 0) và L và R giống như trong Ví dụ 3.4.7. Trong ví dụ, M và R độc lập, nhưng L phụ thuộc vào cả hai. Chúng ta có thể biểu diễn phép phân tích thành thừa số tương ứng bằng đồ thị bên phải của Hình 3.6.

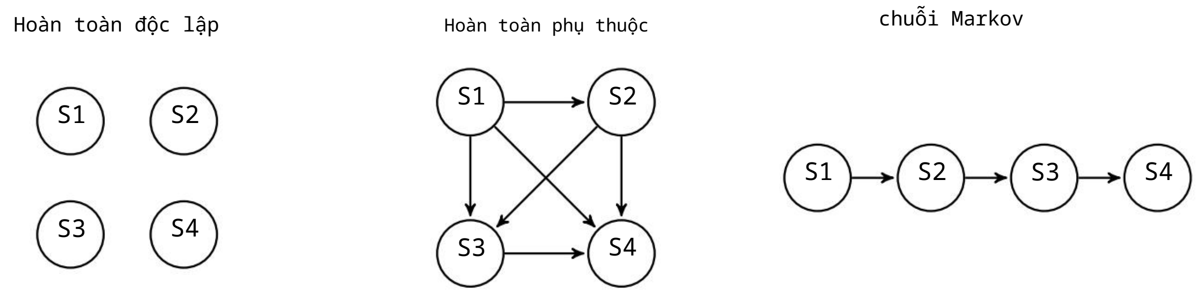
Ví dụ sau đây giới thiệu một lớp mô hình đồ họa quan trọng được gọi là chuỗi Markov, mà chúng ta sẽ thảo luận chi tiết trong Chương 7.

Ví dụ 3.4.9 (Bầu cử). Ở quốc gia trong Hình 3.7 , cuộc bầu cử tổng thống diễn ra theo hệ thống tương tự như ở Hoa Kỳ . Công dân bỏ phiếu bầu đại cử tri trong Đại cử tri đoàn. Mỗi bang được quyền có một số đại cử tri (ở Mỹ điều này thường giống với số đại biểu Quốc hội). Ở mỗi bang, các đại cử tri cam kết ủng hộ ứng cử viên giành chiến thắng ở bang đó. Mục tiêu của chúng tôi là mô hình hóa cuộc bầu cử theo xác suất. Chúng tôi giả định rằng chỉ có hai ứng cử viên A và B. Mỗi trạng thái được đại diện bởi một biến ngẫu nhiên Si

, 1 ≤ tôi ≤ 4,

$$\text{Sĩ} = \begin{cases} 1 & \text{nếu ứng cử viên A thắng bang i,} \\ 1 & \text{nếu ứng cử viên B thắng bang i.} \end{cases} \tag{3.142}$$

Một quyết định quan trọng cần đưa ra là giả định về tính độc lập của mô hình.
Hình 3.8 cho thấy ba tùy chọn khác nhau. Nếu chúng ta lập mô hình mỗi trạng thái là độc lập, thì chúng ta chỉ cần ước tính một tham số duy nhất cho mỗi trạng thái. Tuy nhiên, mô hình có thể không chính xác vì kết quả ở các bang có nhân khẩu học tương tự chắc chắn có liên quan. Một tùy chọn khác là ước tính pmf chung đầy đủ. Vấn đề là có thể khá khó khăn để tính toán các tham số. Chúng ta có thể ước tính pmfs cận biên của các tiểu bang riêng lẻ bằng cách sử dụng dữ liệu thăm dò ý kiến, nhưng xác suất có điều kiện khó ước tính hơn. Ngoài ra, đối với các mô hình lớn hơn, không thể xem xét các mô hình phụ thuộc hoàn toàn (ví dụ trong trường hợp bầu cử Hoa Kỳ , như đã đề cập trước đó). Một thỏa hiệp hợp lý có thể là mô hình hóa các trạng thái không liền kề như độc lập có điều kiện với các trạng thái nằm giữa chúng. Ví dụ, chúng ta giả định rằng kết quả của trạng thái 1 và 3 chỉ liên quan đến trạng thái 2. Mô hình đồ họa tương ứng, được mô tả ở bên phải của Hình 3.8, được gọi là chuỗi Markov. Nó tương ứng



Hình 3.8: Các mô hình đồ thị nắm bắt các giả định khác nhau về phân bố của các biến ngẫu nhiên được xem xét trong Ví dụ 3.4.9.

đến một thừa số của hình thức

$$p_{S1,S2,S3,S4} = p_{S1}\, p_{S2 \mid S1}\, p_{S3 \mid S2}\, p_{S4 \mid S3}. \tag{3.143}$$

Theo mô hình này, chúng ta chỉ cần lo lắng về việc ước tính xác suất có điều kiện theo cặp, trái ngược với pmf chung đầy đủ. Chúng ta thảo luận chi tiết về chuỗi Markov trong Chương 7.

Chúng tôi kết thúc phần này với một ví dụ liên quan đến các biến liên tục.

Ví dụ 3.4.10 (Sa mạc). Dani và Felix đang đi qua sa mạc ở Arizona. Họ lo ngại rằng xe của họ có thể bị hỏng và quyết định xây dựng một mô hình xác suất để đánh giá rủi ro. Họ lập mô hình thời gian cho đến khi chiếc ô tô bị hỏng dưới dạng một biến ngẫu nhiên hàm mũ T với tham số phụ thuộc vào trạng thái của động cơ M và trạng thái của con đường R. Ba đại lượng này được biểu diễn bằng các biến ngẫu nhiên trong cùng một không gian xác suất.

Thật không may, họ không biết trạng thái của động cơ là gì nên họ cho rằng nó đồng nhất giữa 0 (động cơ không có vấn đề gì) và 1 (động cơ gần chết). Tương tự, họ không có thông tin về con đường, vì vậy họ cũng cho rằng trạng thái của nó là một biến ngẫu nhiên thống nhất giữa 0 (con đường không có vấn đề gì) và 1 (con đường thật tệ). Ngoài ra, họ giả định rằng các trạng thái của đường và ô tô là độc lập và tham số của biến ngẫu nhiên hàm mũ biểu thị thời gian tính bằng giờ cho đến khi xảy ra sự cố bằng M + R.
Mô hình đồ thị tương ứng được thể hiện trong Hình 3.9 Để tìm

phân phối chung của các biến ngẫu nhiên, chúng tôi áp dụng quy tắc dây chuyền để thu được,

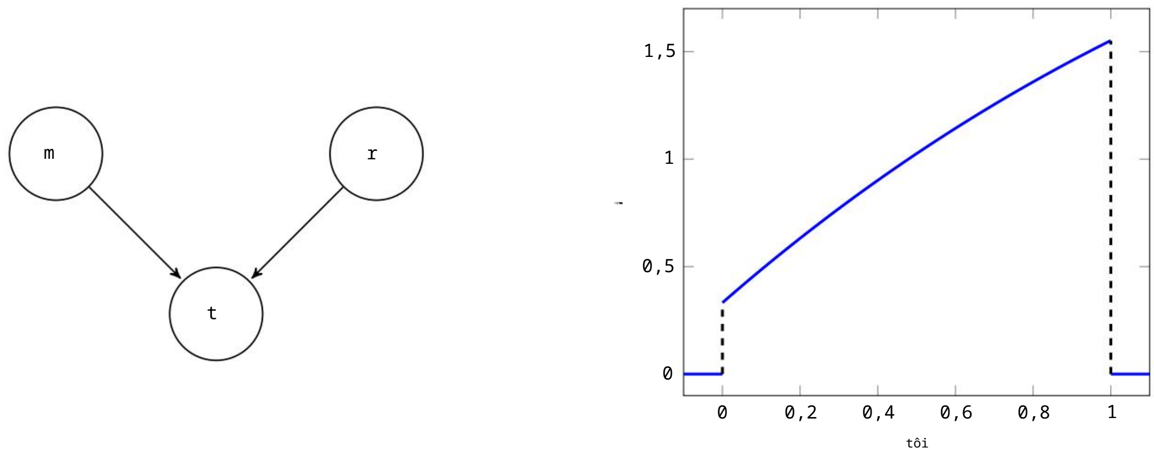
$$f_{M,R,T}(m, r, t) = f_M(m)\, f_{R|M}(r|m)\, f_{T|M,R}(t|m, r) \tag{3.144}$$

$$= f_M(m)\, f_R(r)\, f_{T|M,R}(t|m, r) \text{ (bởi sự độc lập của M và R)} \tag{3.145}$$

$$= \begin{cases} (m+r)\,e^{-(m+r)t} & \text{với } t \geq 0,\ 0 \leq m \leq 1,\ 0 \leq r \leq 1, \\ 0 & \text{nếu không thì.} \end{cases} \tag{3.146}$$

Lưu ý rằng chúng tôi bắt đầu với M và R vì chúng tôi biết phân phối cận biên của chúng, trong khi chúng tôi chỉ biết phân phối có điều kiện của T cho M và R.

Sau 15 phút, xe bị hỏng. Con đường có vẻ ổn, khoảng 0,2 trong thang đo mà họ xác định cho giá trị của R, vì vậy họ tự nhiên băn khoăn về tình trạng của động cơ. Cho họ



Hình 3.9: Hình bên trái là mô hình đồ thị biểu diễn các biến ngẫu nhiên trong Ví dụ 3.4.10. Biểu đồ bên phải hiển thị pdf có điều kiện của M với T = 0,25 và R = 0,2.

mô hình xác suất, sự không chắc chắn của chúng về động cơ với tất cả thông tin này được nắm bắt bởi phân phối có điều kiện của M cho trước T và R.

Để tính toán pdf có điều kiện, trước tiên chúng ta cần tính toán phân phối biên chung của T và R bằng cách trừ biên trên M. Để đơn giản hóa việc tính toán, chúng ta sử dụng bổ đề đơn giản sau.

Bổ đề 3.4.11. Với mọi hằng số $c > 0$,

$$\int_0^1 e^{-cx}\,dx = \frac{1-e^{-c}}{c}, \tag{3.147}$$

$$\int_0^1 x e^{-cx}\,dx = \frac{1-(1+c)\,e^{-c}}{c^2}. \tag{3.148}$$

Bằng chứng. Phương trình (3.147) thu được bằng cách sử dụng nguyên hàm của hàm mũ (chính nó), trong khi tích phân từng phần cho ra (3.148).

Chúng ta có

$$f_{R,T}(r,t) = \int_{m=0}^{1} f_{M,R,T}(m,r,t)\,dm \tag{3.149}$$

$$= e^{-tr}\left(\int_{m=0}^{1} m e^{-tm}\,dm + r\int_{m=0}^{1} e^{-tm}\,dm\right) \tag{3.150}$$

$$= e^{-tr}\left(\frac{1-(1+t)\,e^{-t}}{t^2} + \frac{r\left(1-e^{-t}\right)}{t}\right) \quad \text{bởi (3.147) và (3.148)} \tag{3.151}$$

$$= \frac{e^{-tr}}{t^2}\left(1+tr-e^{-t}(1+t+tr)\right), \tag{3.152}$$

với $t \geq 0$, $0 \leq r \leq 1$.

PDF có điều kiện của M cho T và R là

$$f_{M|R,T}(m|r,t) = \frac{f_{M,R,T}(m,r,t)}{f_{R,T}(r,t)} \tag{3.153}$$

$$= \frac{(m+r)\, t^2 e^{(1+r)t}}{\frac{e^{tr}}{tr}\, e^{t}\,(1+t+tr))} \tag{3.154}$$

$$= \frac{(m+r)t^2 e^{tm}}{1+tr \quad e^{t}\,(1+t+tr)}, \tag{3.155}$$

với $t \geq 0$, $0 \leq m \leq 1$, $0 \leq r \leq 1$. Thay các giá trị quan sát vào, pdf điều kiện bằng

$$f_{M|R,T}(m|0,2,\ 0,25) = \frac{(m+0,2)\ 0,25^2\ e^{0,25m}}{1+0,25\cdot 0,2 \quad e^{0,25}\,(1+0,25+0,25\cdot 0,2)} \tag{3.156}$$

$$= 1,66\ (m+0,2)\ e^{0,25m}. \tag{3.157}$$

với $0 \leq m \leq 1$ và bằng 0 nếu ngược lại. Bản pdf được vẽ trong Hình 3.9. Theo mô hình, có vẻ như tình trạng của động cơ không tốt.

3.5 Hàm một số biến ngẫu nhiên

pmf của một biến ngẫu nhiên $Y := g(X_1, \ldots, X_n)$ được định nghĩa là một hàm $g : \mathbb{R}^N \rightarrow \mathbb{R}$ của một số các biến ngẫu nhiên rời rạc $X_1, \ldots, X_n$ được cho bởi

$$p_Y(y) = \sum_{y=g(x_1,\ldots,x_n)} p_{X_1,\ldots,X_n}(x_1, \ldots, x_n). \tag{3.158}$$

Điều này suy trực tiếp từ (3.11). Nói cách khác, xác suất mà $g(X_1, \ldots, X_n) = y$ là tổng của pmf chung trên tất cả các giá trị có thể sao cho $y = g(x_1, \ldots, x_n)$.

Ví dụ 3.5.1 (Bầu cử). Trong Ví dụ 3.4.9, chúng ta đã thảo luận về một số mô hình khả thi cho cuộc bầu cử tổng thống cho một quốc gia có bốn bang. Hãy tưởng tượng rằng bạn đang cố gắng dự đoán kết quả của cuộc bầu cử bằng cách sử dụng dữ liệu thăm dò ý kiến từ các bang riêng lẻ. Mục tiêu là dự đoán kết quả của cuộc bầu cử, được đại diện bởi biến ngẫu nhiên

$$Ôi := \begin{cases} 1 \text{ nếu } \sum_{i=1}^{4} n_i S_i > 0, \\ 0 \text{ nếu không,} \end{cases} \tag{3.159}$$

trong đó n_i biểu thị số đại cử tri ở bang i (chú ý rằng tổng không bao giờ bằng 0).

Từ việc phân tích dữ liệu thăm dò ý kiến, bạn kết luận rằng xác suất mà ứng cử viên A thắng ở mỗi bang là 0,15. Nếu bạn cho rằng tất cả các trạng thái đều độc lập, thì điều này đủ để mô tả pmf chung. Bảng 3.2 liệt kê xác suất của tất cả các kết quả có thể xảy ra đối với mô hình này. Đến (3.158), chúng ta chỉ cần cộng các kết quả mà O = 1. Theo giả định độc lập hoàn toàn, xác suất ứng cử viên A thắng là 6%.

Bạn không hài lòng với kết quả bởi vì bạn nghi ngờ rằng kết quả ở các trạng thái khác nhau phụ thuộc rất nhiều. Từ các cuộc bầu cử trước đây, bạn xác định rằng xác suất có điều kiện của một

S1	S2	S3	S4	O	Có thể. (độc lập.) Prob.	(Markov)
-1	-1 -1	-1 0 0	,5220	1 0	0,0921 -1 0	0,6203
-1	-1 -1		0,0	921 0	,0163 0,0921	0,0687
-1	-1	1				0,0431
-1	-1	1	1	1		0,0332
-1	1	-1	-1 0			0,0431
-1	1	-1	1 0		0,0163	0,0048
-1	1	1	-1	1	0,0163	0,0208
-1	1	1	1	1	0,0029	0,0160
1	-1	-1 -1	0		0,0921	0,0687
1	-1	-1	1 0		0,0163	0,0077
1	-1	1	-1	1	0,0163	0,0048
1	-1	1	1	1	0,0029	0,0037
1	1	-1	-1 0		0,0163	0,0332
1	1	-1	1	1	0,0029	0,0037
1	1	1	-1	1	0,0029	0,0160
1	1	1	1	1	0,0005	0,0123

Bảng 3.2: Bảng giá trị phụ Ví dụ 3.5.1.

ứng cử viên chiến thắng một bang nếu họ thắng một bang liền kề thực sự rất cao. bạn kết hợp ước tính của bạn về các xác suất có điều kiện trong mô hình chuỗi Markov được mô tả bởi (3.143):

$$pS1\ (1) = 0{,}15, \tag{3.160}$$

$$pSi+1 \mid Si\ (1 \mid 1) = 0{,}435,\ 2 \leq i \leq \tag{3.161}$$

$$4,\ pSi+1 \mid Si\ (1 \mid 1) = 0{,}900\ 2 \leq i \leq 4. \tag{3.162}$$

Điều này có nghĩa là nếu ứng cử viên B giành được một bang, họ rất có khả năng giành được bang liền kề. Nếu như ứng cử viên A giành được một tiểu bang, cơ hội của họ để giành được một tiểu bang liền kề cao hơn đáng kể so với nếu họ không (nhưng vẫn thấp hơn ứng viên B). Theo mô hình này, xác suất cận biên mà ứng cử viên A thắng mỗi bang vẫn là 0,15. Bảng 3.2 liệt kê xác suất của tất cả các kết quả có thể xảy ra. Xác suất ứng cử viên A chiến thắng hiện là 11%, gần gấp đôi xác suất đó thu được theo mô hình hoàn toàn độc lập. Điều này cho thấy sự nguy hiểm của việc không tính đến sự phụ thuộc giữa các quốc gia, ví dụ có thể là một trong những lý do tại sao nhiều đã đánh giá thấp cơ hội của Donald Trump trong cuộc bầu cử năm 2016.

Mục 2.5 giải thích cách suy ra hàm phân phối của các biến ngẫu nhiên đơn biến theo đầu tiên tính toán cdf của họ và sau đó vi phân nó để lấy pdf của họ. Điều này trực tiếp mở rộng đến hàm ngẫu nhiên nhiều biến. Cho X, Y là các biến ngẫu nhiên xác định cùng xác suất không gian và đặt $U = g(X, Y)$ và $V = h(X, Y)$ cho hai hàm tùy ý $g, h : \mathbb{R}^2 \to \mathbb{R}$. Sau đó,

$$F_{U,V}(u, v) = P(U \leq u, V \leq v) \tag{3.163}$$

$$= P(g(X, Y) \leq u, h(X, Y) \leq v) \tag{3.164}$$

$$= \int_{\{(x,y) \,|\, g(x,y) \leq u, h(x,y) \leq v\}} f_{X,Y}(x, y)\, dx\, dy, \tag{3.165}$$

trong đó đẳng thức cuối cùng chỉ tồn tại nếu tồn tại bản pdf chung của X và Y. Bản pdf chung sau đó có thể là thu được bằng cách biệt hóa.

Định lý 3.5.2 (Pdf tổng của hai biến ngẫu nhiên độc lập). PDF của $Z = X + Y$ trong đó X và Y là các biến ngẫu nhiên độc lập bằng với tích chập của chúng tương ứng pdf f_X và f_Y,

$$f_Z(z) = \int_{u=-\infty}^{\infty} f_X(z - u)\, f_Y(u)\, du. \tag{3.166}$$

Bằng chứng. Đầu tiên chúng tôi lấy được cdf của Z

$$F_Z(z) = P(X + Y \leq z) \tag{3.167}$$

$$= \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{z-y} f_X(x)\, f_Y(y)\, dx\, dy \tag{3.168}$$

$$= \int_{y=-\infty}^{\infty} F_X(z - y)\, f_Y(y)\, dy. \tag{3.169}$$

Lưu ý rằng pdf chung của X và Y là sản phẩm của các pdf cận biên vì ngẫu nhiên các biến là độc lập. Bây giờ chúng tôi phân biệt cdf để lấy pdf. Lưu ý rằng điều này đòi hỏi một trao đổi của một toán tử giới hạn với một toán tử phân biệt và một trao đổi khác của một toán tử tích phân với một toán tử vi phân, điều này hợp lý vì các hàm liên quan là có giới hạn và có thể tích hợp được.

$$f_Z(z) = \frac{đ}{dz} \lim_{u \to \infty} \int_{y=-u}^{\text{bạn}} F_X(z - y)\, f_Y(y)\, dy \tag{3.170}$$

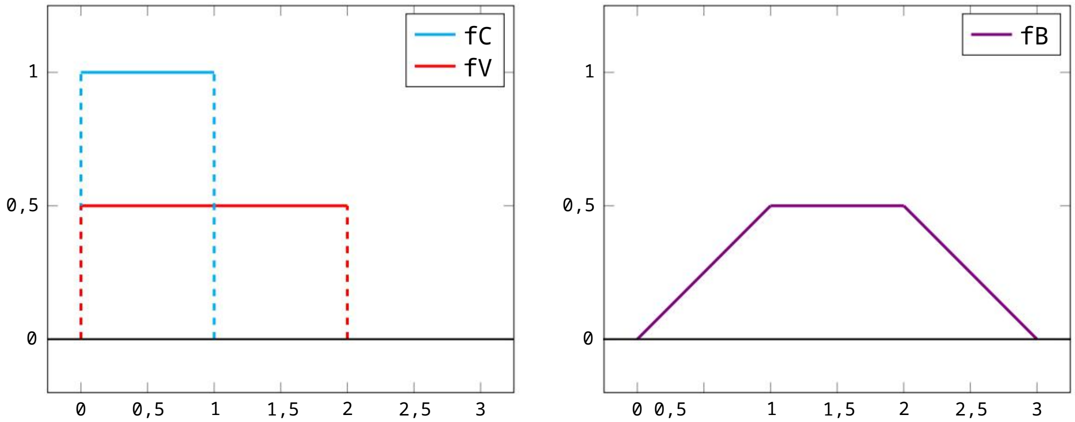
$$= \lim_{u \to \infty} \frac{đ}{dz} \int_{y=-u}^{\text{bạn}} F_X(z - y)\, f_Y(y)\, dy \tag{3.171}$$

$$= \lim_{u \to \infty} \int_{y=-u}^{\text{bạn}} \frac{đ}{dz} F_X(z - y)\, f_Y(y)\, dy \tag{3.172}$$

$$= \lim_{u \to \infty} \int_{y=-u}^{\text{bạn}} f_X(z - y)\, f_Y(y)\, dy. \tag{3.173}$$

□

Ví dụ 3.5.3 (Cà phê nhân). Một công ty sản xuất cà phê mua hạt từ hai công ty nhỏ ở địa phương nhà sản xuất ở Colombia và Việt Nam. Số lượng đậu họ có thể mua từ mỗi nhà sản xuất thay đổi tùy thuộc vào thời tiết. Công ty lập mô hình các đại lượng C và V này là độc lập biến ngẫu nhiên (giả sử rằng thời tiết ở Colombia độc lập với thời tiết ở Việt Nam) có phân bố đều lần lượt trong [0, 1] và [0, 2] (đơn vị là tấn).



Hình 3.10: Các hàm mật độ xác suất trong Ví dụ 3.5.3.

Bây giờ chúng ta tính pdf của tổng lượng hạt cà phê B := E+V áp dụng Định lý 3.5.2,

$$f_B(b) = \int_{u=-\infty}^{\infty} f_C(b-u)\, f_V(u)\, du \tag{3.174}$$

$$= \frac{1}{2}\int_{u=0}^{2} f_C(b-u)\, du \tag{3.175}$$

$$= \begin{cases} \frac{1}{2}\int_{u=0}^{b} du = \frac{b}{2} & \text{nếu } b \leq 1 \\ \frac{1}{2}\int_{u=b-1}^{b} du = \frac{1}{2} & \text{nếu } 1 \leq b \leq 2 \\ \frac{1}{2}\int_{u=b-1}^{2} du = \frac{3-b}{2} & \text{nếu } 2 \leq b \leq 3. \end{cases} \tag{3.176}$$

Bản pdf của B được hiển thị trong Hình 3.10.

3.6 Sinh biến ngẫu nhiên đa biến

Trong Phần 2.6 , chúng tôi xem xét vấn đề tạo các mẫu độc lập từ một tùy ý phân phối đơn biến. Giả sử rằng một thủ tục để đạt được điều này là có sẵn, chúng ta có thể sử dụng nó để lấy mẫu từ một phân phối đa biến tùy ý bằng cách tạo các mẫu từ phân phối có điều kiện.

Thuật toán 3.6.1 (Lấy mẫu từ phân phối đa biến). Cho $X_1, X_2, \ldots, X_n$ được các biến dom chạy thuộc cùng một không gian xác suất. Để tạo mẫu từ khớp của chúng phân phối chúng tôi lấy mẫu tuần tự từ các phân phối có điều kiện của chúng:

1. Lấy mẫu x_1 của X_1.

2. Với $i = 2, 3, \ldots, n$, lấy một mẫu x_i của X_i với biến cố $\{X_1 = x_1, \ldots, X_{i-1} = x_{i-1}\}$ bằng cách lấy mẫu từ $F_{X_i|X_1,\ldots,X_{i-1}}(\cdot|x_1, \ldots, x_{i-1})$.

Quy tắc dây chuyền ngụ ý rằng đầu ra $x_1, \ldots, x_n$ của quy trình này là các mẫu từ khớp phân phối của các biến ngẫu nhiên. Ví dụ sau xem xét vấn đề lấy mẫu từ một hỗn hợp các biến ngẫu nhiên hàm mũ.

Ví dụ 3.6.2 (Hỗn hợp cấp số nhân). Cho B là biến ngẫu nhiên Bernoulli với tham số p và X là biến ngẫu nhiên hàm mũ với tham số 1 nếu B = 0 và 2 nếu B = 1. Giả sử rằng chúng ta truy cập được hai mẫu độc lập u_1 và u_2 từ một phân bố đều trong [0 , 1].

Để lấy mẫu từ B và X:

1. Chúng ta đặt b := 1 nếu $u_1 \leq p$ và b := 0 nếu ngược lại. Điều này đảm bảo rằng b là một mẫu Bernoulli với tham số phù hợp.

2. Sau đó, chúng tôi thiết lập

$$x := \frac{1}{\lambda} \text{nhật ký} \frac{1}{1 \quad u_2} \tag{3.177}$$

trong đó λ := 1 nếu b = 0 và λ := 2 nếu b = 1. Theo Ví dụ 2.6.4 x được phân phối dưới dạng hàm mũ với tham số λ.

3.7 Lấy mẫu loại bỏ

Chúng tôi kết thúc chương bằng cách mô tả lấy mẫu từ chối, còn được gọi là phương pháp chấp nhận-từ chối, một quy trình thay thế để lấy mẫu từ các phân phối đơn biến. Lý do chúng tôi chuyển nó sang chương này là vì việc phân tích kỹ thuật này đòi hỏi sự hiểu biết về các biến ngẫu nhiên đa biến. Trước khi trình bày phương pháp, chúng tôi thúc đẩy nó bằng cách sử dụng các biến ngẫu nhiên rời rạc.

3.7.1 Lấy mẫu loại bỏ đối với các biến ngẫu nhiên rời rạc

Mục tiêu của chúng tôi là mô phỏng một biến ngẫu nhiên Y bằng cách sử dụng các mẫu từ một biến ngẫu nhiên khác X. Để đơn giản hóa việc trình bày, chúng tôi giả sử rằng pmfs p_X và p_Y của chúng có các giá trị khác không trong tập {1, 2, . . . , n} (dễ dàng tổng quát hóa cho các tập hợp rời rạc khác). Ý tưởng đằng sau việc lấy mẫu từ chối là chúng ta có thể chọn một tập hợp con các mẫu của X theo cách định hình lại phân phối của nó. Khi chúng tôi có được một mẫu X, chúng tôi quyết định chấp nhận hay từ chối nó với một xác suất nhất định. Xác suất phụ thuộc vào giá trị của mẫu x, nếu $p_X(x)$ lớn hơn nhiều so với $p_Y(x)$ thì có lẽ chúng ta nên từ chối nó hầu hết thời gian (nhưng không phải lúc nào cũng vậy!). Với mỗi $x \in \{1, 2, . . . , n\}$ ta định xác suất chấp nhận mẫu theo a_x.

Chúng tôi chỉ quan tâm đến việc phân phối các mẫu được chấp nhận. Về mặt toán học, pmf của các mẫu được chấp nhận bằng với pmf có điều kiện của X, với điều kiện là mẫu được chấp nhận,

$$p_{X \mid \text{Đã chấp nhận}}(x \mid \text{Đã chấp nhận}) = \frac{p_X(x)\, P(\text{Được chấp nhận} \mid X = x)}{\sum_{i=1}^{N} p_X(i)\, P(\text{Được chấp nhận} \mid X = i)} \quad \text{theo quy tắc Bayes} \tag{3.178}$$

$$= \frac{p_X(x)\, a_x}{\sum_{i=1}^{N} p_X(i)\, a_i}. \tag{3.179}$$

Chúng tôi muốn cố định các xác suất chấp nhận sao cho với mọi $x \in \{1, 2, . . . , N\}$

$$p_{X \mid \text{Được chấp nhận}}(x \mid \text{Được chấp nhận}) = p_Y(x). \tag{3.180}$$

Điều này có thể đạt được bằng cách sửa

$$rìu := \frac{p_Y(x)}{c p_X(x)}, \quad x \quad \{1, . . . , N\}, \tag{3.181}$$

cho bất kỳ hằng số c. Tuy nhiên, điều này sẽ không mang lại xác suất hợp lệ cho bất kỳ c tùy ý nào, vì ai có thể lớn hơn một! Để tránh vấn đề này, chúng ta cần

$$c \geq \max_{x \{1,. . . ,n\}} \frac{p_Y(x)}{p_X(x)}, \quad \text{với mọi } x \quad \{1, . . . , N\}. \tag{3.182}$$

Cuối cùng, chúng ta có thể sử dụng biến ngẫu nhiên thống nhất U trong khoảng từ 0 đến 1 để chấp nhận hoặc từ chối, chấp nhận từng mẫu x nếu U ≤ ax. Bạn có thể thắc mắc tại sao chúng ta không thể tạo Y trực tiếp từ U. Điều đó thực sự hiệu quả và đơn giản hơn nhiều; ở đây chúng ta chỉ trình bày trường hợp rời rạc như một phần giới thiệu sư phạm về trường hợp liên tục.

Thuật toán 3.7.1 (Lấy mẫu bác bỏ). Gọi X và Y là các biến ngẫu nhiên với pmfs p_X và p_Y sao cho

$$c \geq \max_{x \{1,. . . ,n\}} \frac{p_Y(x)}{p_X(x)} \tag{3.183}$$

với mọi x sao cho $p_Y(x)$ khác 0 và U là biến ngẫu nhiên phân phối đều trong [0, 1] và không phụ thuộc vào X.

1. Lấy mẫu y của X.
2. Lấy mẫu u của U .
3. Khai báo y là mẫu của Y nếu

$$u \leq \frac{p_Y(y)}{c \, p_X(y)}. \tag{3.184}$$

3.7.2 Lấy mẫu loại bỏ đối với các biến ngẫu nhiên liên tục

Ở đây chúng tôi chỉ ra rằng ý tưởng được trình bày trong phần trước có thể được áp dụng trong trường hợp liên tục. Mục tiêu là lấy các mẫu theo pdf f_Y mục tiêu bằng cách chọn các mẫu lấy được theo một pdf f_X khác. Như trong trường hợp rời rạc, chúng ta cần

$$f_Y(y) \leq c \, f_X(y) \tag{3.185}$$

với mọi y, trong đó c là một hằng số dương cố định. Nói cách khác, bản pdf của Y phải được giới hạn bởi phiên bản được chia tỷ lệ của bản pdf của X.

Thuật toán 3.7.2 (Lấy mẫu bác bỏ). Cho X là biến ngẫu nhiên với pdf f_X và U là biến ran dom phân bố đều trong [0, 1] và độc lập với X. Giả sử (3.185) đúng.

1. Lấy mẫu y của X.
2. Lấy mẫu u của U .

3. Khai báo y là mẫu của Y nếu

$$u \leq \frac{f_Y(y)}{c\, f_X(y)}. \tag{3.186}$$

Định lý sau đây chứng minh rằng các mẫu thu được bằng cách lấy mẫu bác bỏ có phân phối mong muốn.

Định lý 3.7.3 (Lấy mẫu bác bỏ hoạt động). Nếu giả định (3.185) đúng, thì các mẫu được tạo ra bằng cách lấy mẫu bác bỏ được phân phối theo f_Y .

Bằng chứng. Gọi Z là biến ngẫu nhiên được tạo ra bằng cách lấy mẫu từ chối. cdf của Z bằng

$$F_Z(y) = P\left(X \leq y \,\middle|\, U \leq \frac{f_Y(X)}{c\, f_X(X)}\right) \tag{3.187}$$

$$= \frac{P\left(X \leq y, U \leq \frac{f_Y(X)}{c\, f_X(X)}\right)}{P\left(U \leq \frac{f_Y(X)}{c\, f_X(X)}\right)}. \tag{3.188}$$

Để tính toán tử số, chúng tôi tích hợp pdf chung của U và X trên khu vực quan tâm

$$P\left(X \leq y, U \leq \frac{f_Y(X)}{c\, f_X(X)}\right) = \int_{x=-\infty}^{y} \int_{u=0}^{\frac{f_Y(x)}{c\, f_X(x)}} f_X(x)\, du\, dx \tag{3.189}$$

$$= \int_{x=-\infty}^{y} \frac{f_Y(x)}{c\, f_X(x)} f_X(x)\, dx \tag{3.190}$$

$$= \frac{1}{c} \int_{x=-\infty}^{y} f_Y(x)\, dx \tag{3.191}$$

$$= \frac{1}{c} \text{Năm tài chính (năm)}. \tag{3.192}$$

Mẫu số thu được theo cách tương tự

$$P\left(U \leq \frac{f_Y(X)}{c\, f_X(X)}\right) = \int_{x=-\infty}^{\infty} \int_{u=0}^{\frac{f_Y(x)}{c\, f_X(x)}} f_X(x)\, du\, dx \tag{3.193}$$

$$= \int_{x=-\infty}^{\infty} \frac{f_Y(x)}{c\, f_X(x)} f_X(x)\, dx \tag{3.194}$$

$$= \frac{1}{c} \int_{x=-\infty}^{\infty} f_Y(x)\, dx \tag{3.195}$$

$$= \frac{1}{c}. \tag{3.196}$$

Chúng tôi kết luận rằng

$$F_Z(y) = \text{Năm tài chính }(y), \tag{3.197}$$

vì vậy phương pháp tạo ra các mẫu từ phân phối của Y . □

Bây giờ chúng ta minh họa phương pháp này bằng cách áp dụng nó để tạo ra một biến ngẫu nhiên Gaussian từ một biến ngẫu nhiên hàm mũ và một biến ngẫu nhiên đồng nhất.

Ví dụ 3.7.4 (Tạo biến ngẫu nhiên Gaussian). Trong Ví dụ 2.6.4 , chúng ta đã học cách tạo biến ngẫu nhiên hàm mũ bằng cách sử dụng các mẫu từ phân phối đồng đều. Trong ví dụ này, chúng tôi sẽ sử dụng các mẫu từ phân phối hàm mũ để tạo biến ngẫu nhiên Gaussian tiêu chuẩn áp dụng lấy mẫu từ chối.

Bổ đề sau đây cho thấy rằng chúng ta có thể tạo một biến ngẫu nhiên Gaussian tiêu chuẩn Y bằng cách:

1. Tạo biến ngẫu nhiên H bằng pdf

$$f_H(h) := \begin{cases} \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{h^2}{2}\right) & \text{nếu } h \geq 0, \\ 0 & \text{nếu không thì.} \end{cases} \tag{3.198}$$

2. Tạo biến ngẫu nhiên S bằng 1 hoặc -1 với xác suất 1/2 chẳng hạn bằng cách áp dụng phương pháp được mô tả trong Mục 2.6.1.

3. Cài đặt Y := SH.

Bổ đề 3.7.5. Cho H là biến ngẫu nhiên liên tục với pdf cho bởi (3.198) và S là biến ngẫu nhiên rời rạc bằng 1 với xác suất 1/2 và 1 với xác suất 1/2. Biến ngẫu nhiên của Y := SH là một Gaussian chuẩn.

Bằng chứng. PDF có điều kiện của Y đã cho S được cho bởi

$$f_{Y|S}(y|1) = \begin{cases} f_H(y) & \text{nếu } y \geq 0, \\ 0 & \text{ngược lại,} \end{cases} \tag{3.199}$$

$$f_{Y|S}(y|-1) = \begin{cases} f_H(-y) & \text{nếu } y < 0, \\ 0 & \text{nếu không thì.} \end{cases} \tag{3.200}$$

Theo bổ đề 3.3.5 ta có

$$f_Y(y) = p_S(1)\, f_{Y|S}(y|1) + p_S(-1)\, f_{Y|S}(y|-1) \tag{3.201}$$

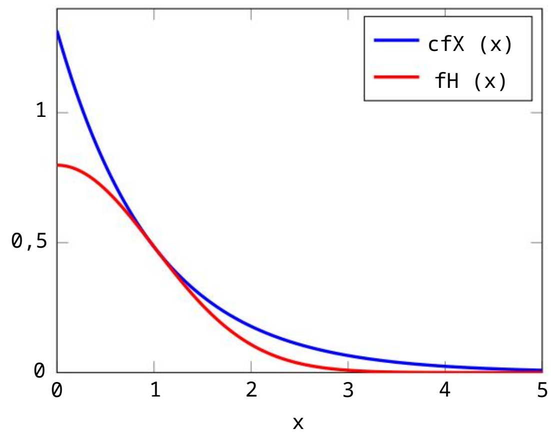
$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right). \tag{3.202}$$

Lý do tại sao chúng tôi giảm vấn đề để tạo H là vì pdf của nó chỉ khác 0 trên trục dương, điều này cho phép chúng tôi ràng buộc nó với pdf hàm mũ của một biến ngẫu nhiên hàm mũ X với tham số 1. Nếu chúng ta đặt $c := \sqrt{2e/\pi}$ thì $f_H(x) \leq c f_X(x)$ với mọi x, như minh họa trong Hình 3.11. Thực vậy,

$$\frac{f_H(x)}{f_X(x)} = \frac{\frac{2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)}{\exp(-x)} \tag{3.203}$$

$$= \sqrt{\frac{2e}{\pi}} \exp\left(-\frac{(x-1)^2}{2}\right) \tag{3.204}$$

$$\leq \sqrt{\frac{2e}{\pi}}. \tag{3.205}$$



Hình 3.11: Giới hạn trên pdf của phân phối mục tiêu trong Ví dụ 3.7.4.

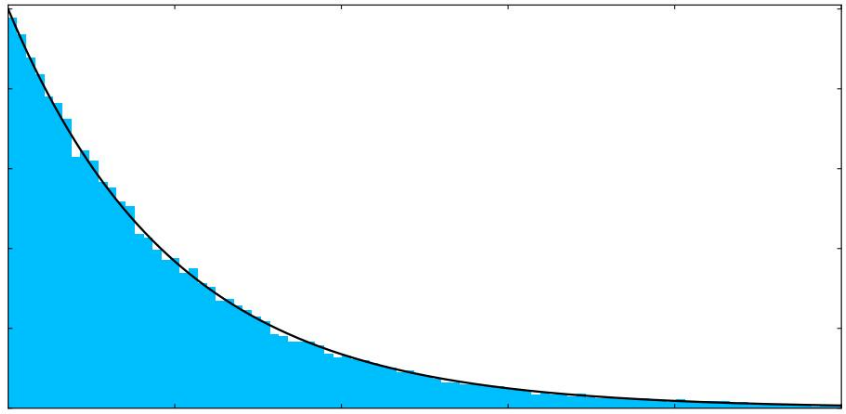
Bây giờ chúng ta có thể áp dụng lấy mẫu từ chối để tạo H. Các bước là

1. Lấy mẫu x từ biến ngẫu nhiên hàm mũ X với tham số một
2. Lấy mẫu u từ U phân bố đều trong [0, 1].
3. Chấp nhận x là mẫu của H nếu

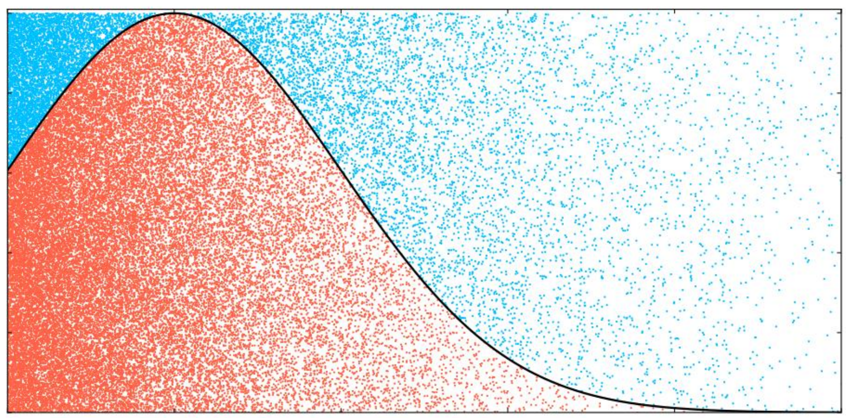
$$\text{bạn} \leq \text{kinh nghiệm} \; \frac{(x \quad 1)2}{2} \; . \tag{3.206}$$

Quy trình này được minh họa trong Hình 3.12. Cơ chế từ chối đảm bảo rằng cái được chấp nhận mẫu có phân phối đúng.

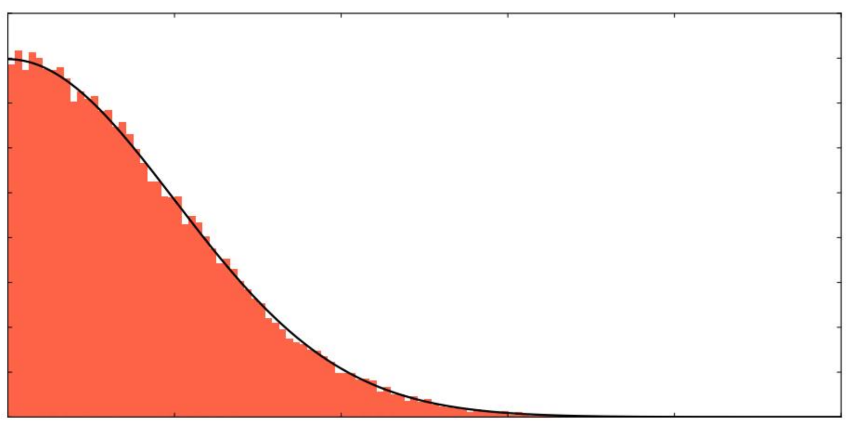
Biểu đồ của 50.000 mẫu iid từ X (f_X được hiển thị bằng màu đen)



Biểu đồ tán xạ của các mẫu từ X và các mẫu từ U (các mẫu được chấp nhận có màu đỏ, exp (x 1)2 /2 được hiển thị là bằng màu đen)



Biểu đồ của các mẫu được chấp nhận (f_H được hiển thị bằng màu đen)



Hình 3.12: Minh họa cách tạo 50.000 mẫu từ biến ngẫu nhiên H được xác định trong Ví dụ 3.7.4 thông qua lấy mẫu bác bỏ.

Chương 4

Kỳ vọng

Trong phần này, chúng tôi giới thiệu một số đại lượng mô tả hành vi của các biến ngẫu nhiên rất ngắn gọn. Giá trị trung bình là giá trị xung quanh đó phân phối của một biến ngẫu nhiên là trung tâm. Phương sai định lượng mức độ mà một biến ngẫu nhiên dao động xung quanh giá trị trung bình. Hiệp phương sai của hai biến ngẫu nhiên cho biết liệu chúng có xu hướng đi chệch khỏi giá trị trung bình theo cách tương tự hay không. Trong nhiều chiều, ma trận hiệp phương sai của một vectơ ngẫu nhiên mã hóa phương sai của nó theo mọi hướng có thể. Các đại lượng này không đặc trưng hoàn toàn cho sự phân bố của một biến ngẫu nhiên hoặc vectơ, nhưng chúng cung cấp một bản tóm tắt hữu ích về hành vi của chúng chỉ với một vài con số.

4.1 Toán tử kỳ vọng

Toán tử kỳ vọng cho phép chúng ta xác định giá trị trung bình, phương sai và hiệp phương sai một cách chặt chẽ. Nó ánh xạ một hàm của một biến ngẫu nhiên hoặc của một số biến ngẫu nhiên thành trung bình có trọng số theo pmf hoặc pdf tương ứng.

Định nghĩa 4.1.1 (Kỳ vọng đối với biến ngẫu nhiên rời rạc). Cho X là biến ngẫu nhiên rời rạc có khoảng R. Giá trị kỳ vọng của hàm $g(X)$, $g : \mathbb{R} \to \mathbb{R}$, của X là

$$E(g(X)) := \sum_{x \in R} g(x)\, p_X(x). \tag{4.1}$$

Tương tự, nếu X, Y đều là các biến ngẫu nhiên rời rạc có khoảng R_X và R_Y thì giá trị kỳ vọng của hàm $g(X, Y)$, $g : \mathbb{R}^2 \to \mathbb{R}$, của X và Y là

$$E(g(X, Y)) := \sum_{x \in R_X} \sum_{x \in R_Y} g(x, y)\, p_{X,Y}(x, y). \tag{4.2}$$

Nếu X là một vectơ ngẫu nhiên rời rạc n chiều, giá trị kỳ vọng của hàm $g(X)$, $g : \mathbb{R}^N \to \mathbb{R}$, của X là

$$E g(X) := \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} g(x)\, p_X(x). \tag{4.3}$$

Định nghĩa 4.1.2 (Kỳ vọng đối với biến ngẫu nhiên liên tục). Cho X là biến ngẫu nhiên liên tục. Giá trị kỳ vọng của hàm g(X), $g : \mathbb{R} \to \mathbb{R}$, của X là

$$E(g(X)) := \int_{x=-\infty}^{\infty} g(x)\, f_X(x)\, dx. \quad (4.4)$$

Tương tự, nếu X, Y đều là biến ngẫu nhiên liên tục thì giá trị kỳ vọng của hàm số g(X, Y), $g : \mathbb{R}^2 \to \mathbb{R}$, của X và Y là

$$E(g(X, Y)) := \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} g(x, y)\, f_{X,Y}(x, y)\, dx\, dy. \quad (4.5)$$

Nếu X là véc tơ ngẫu nhiên n chiều, giá trị kỳ vọng của hàm g(X), $g : \mathbb{R}^N \to \mathbb{R}$, của X là

$$E g(\vec{X}) := \int_{x_1=-\infty}^{\infty} \int_{x_2=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} g(\vec{x})\, f_{\vec{X}}(\vec{x})\, dx_1\, dx_2 \ldots dx_n \quad (4.6)$$

Trong trường hợp các đại lượng phụ thuộc vào cả biến ngẫu nhiên liên tục và biến ngẫu nhiên rời rạc, tích của các phân phối cận biên và có điều kiện đóng vai trò của pdf hoặc pmf chung.

Định nghĩa 4.1.3 (Kỳ vọng đối với biến ngẫu nhiên liên tục và rời rạc). Nếu C là biến ngẫu nhiên liên tục và D là biến ngẫu nhiên rời rạc có khoảng R_D xác định trên cùng một không gian xác suất thì giá trị kỳ vọng của hàm g(C,D) của C và D là

$$E(g(C, D)) := \int_{c=-\infty}^{\infty} \sum_{d \in R_D} g(c, d)\, f_C(c)\, p_{D|C}(d|c)\, dc \quad (4.7)$$

$$= \sum_{d \in R_D} \int_{c=-\infty}^{\infty} g(c, d)\, p_D(d)\, f_{C|D}(c|d)\, dc. \quad (4.8)$$

Giá trị kỳ vọng của một đại lượng nhất định có thể là vô hạn hoặc thậm chí không tồn tại nếu tổng hoặc tích phân tương ứng có xu hướng tiến tới vô cùng hoặc có giá trị không xác định. Điều này được minh họa bởi Ví dụ 4.1.4 và 4.2.2 dưới đây.

Ví dụ 4.1.4 (Nghịch lý St Petersburg). Một sòng bạc cung cấp cho bạn trò chơi sau. Bạn sẽ tung một đồng xu không thiên vị cho đến khi nó xuất hiện mặt ngửa và sòng bạc sẽ trả cho bạn 2 nghìn đô la trong đó k là số lần tung. Bạn sẵn sàng trả bao nhiêu để chơi?

Hãy để chúng tôi tính toán mức tăng dự kiến. Nếu các lần lật là độc lập, thì tổng số lần lật X là a . Mức biến ngẫu nhiên hình học nên $p_X(k) = 1/2^k$ tăng là 2^X có nghĩa là

$$E(\text{Tăng}) = \sum_{k=1}^{\infty} 2^k \cdot \frac{1}{2^k} = \infty. \quad (4.9)$$

Lợi nhuận kỳ vọng là vô hạn, nhưng vì bạn chỉ được chơi một lần nên số tiền bạn sẵn sàng trả có thể bị giới hạn. Điều này được gọi là nghịch lý St Petersburg.

Một tính chất cơ bản của toán tử kỳ vọng là nó tuyến tính.

Định lý 4.1.5 (Tính tuyến tính của kỳ vọng). Với mọi hằng số $a \in \mathbb{R}$, mọi hàm $g : \mathbb{R} \to \mathbb{R}$ và bất kỳ biến ngẫu nhiên liên tục hoặc rời rạc X

$$E(ag(X)) = aE(g(X)). \tag{4.10}$$

Với mọi hằng số $a, b \in \mathbb{R}$, mọi hàm số $g_1, g_2 : \mathbb{R}^N \to \mathbb{R}$ và bất kỳ liên tục hoặc rời rạc biến ngẫu nhiên X và Y

$$E(a\, g_1(X, Y) + b\, g_2(X, Y)) = aE(g_1(X, Y)) + bE(g_2(X, Y)). \tag{4.11}$$

Bằng chứng. Định lý suy ra ngay lập tức từ tính chất tuyến tính của tổng và tích phân. □

Tính tuyến tính của kỳ vọng làm cho việc tính toán kỳ vọng của các hàm tuyến tính của các biến ngẫu nhiên. Ngược lại, việc tính toán pdf hoặc pmf chung thường phức tạp hơn nhiều.

Ví dụ 4.1.6 (Hạt cà phê (tiếp theo từ Ví dụ 3.5.3)). Hãy để chúng tôi tính toán dự kiến tổng số đậu có thể mua. C đều trong [0, 1] nên E(C) = 1/2. V là thống nhất trong [0, 2] nên E(V) = 1. Theo tuyến tính của kỳ vọng

$$E(C + V) = E(C) + E(V) \tag{4.12}$$

$$= 1{,}5 \text{ tấn.} \tag{4.13}$$

Lưu ý rằng điều này đúng ngay cả khi hai đại lượng không độc lập.

Nếu hai biến ngẫu nhiên là độc lập, thì kỳ vọng của các yếu tố sản phẩm thành một sản phẩm của kỳ vọng.

Định lý 4.1.7 (Kỳ vọng của hàm các biến ngẫu nhiên độc lập). Nếu X, Y là các biến ngẫu nhiên độc lập được xác định trên cùng một không gian xác suất và $g, h : \mathbb{R} \to \mathbb{R}$ là đơn biến các hàm có giá trị thực, sau đó

$$E(g(X)\, h(Y)) = E(g(X))\, E(h(Y)). \tag{4.14}$$

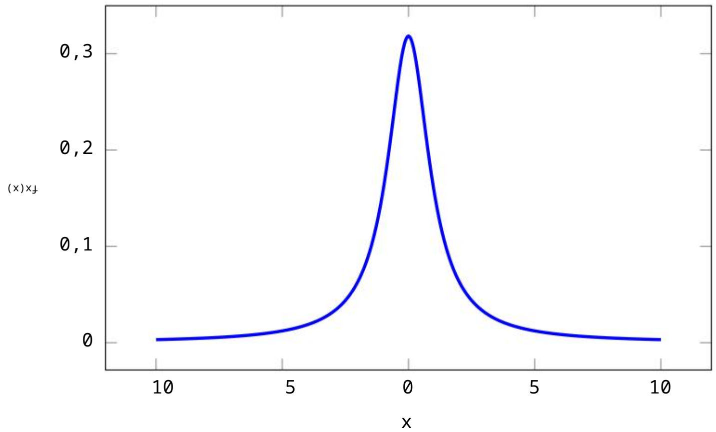
Bằng chứng. Ta chứng minh kết quả cho biến ngẫu nhiên liên tục, nhưng chứng minh cho biến ngẫu nhiên rời rạc biến về cơ bản là giống nhau.

$$E(g(X)\, h(Y)) = \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} g(x)\, h(y)\, f_{X,Y}(x, y)\, dx\, dy \tag{4.15}$$

$$= \int_{x=-\infty}^{\infty}\int_{y=-\infty}^{\infty} g(x)\, h(y)\, f_X(x)\, f_Y(y)\, dx\, dy \text{ độc lập} \tag{4.16}$$

$$= E(g(X))\, E(h(Y)). \tag{4.17}$$

□



Hình 4.1: Hàm mật độ xác suất của một biến ngẫu nhiên Cauchy.

4.2 Trung bình và phương sai

4.2.1 Ý nghĩa

Giá trị trung bình của một biến ngẫu nhiên bằng với giá trị kỳ vọng của nó.

Định nghĩa 4.2.1 (Trung bình). Giá trị trung bình hoặc thời điểm đầu tiên của X là giá trị kỳ vọng của X: E(X).

Bảng 4.1 liệt kê phương tiện của một số biến ngẫu nhiên quan trọng. Các dẫn xuất có thể được tìm thấy trong Phần 4.5.1. Như được minh họa trong Hình 4.3, giá trị trung bình là khối tâm của pmf hoặc pdf của biến ngẫu nhiên tương ứng.

Nếu phân phối của một biến ngẫu nhiên có đuôi rất nặng, nghĩa là xác suất biến ngẫu nhiên nhận các giá trị lớn giảm dần, thì giá trị trung bình của nó có thể là vô hạn. Đây là trường hợp của biến ngẫu nhiên đại diện cho mức tăng trong Ví dụ 4.1.4. Ví dụ sau đây cho thấy giá trị trung bình có thể không tồn tại nếu giá trị của tổng hoặc tích phân tương ứng không được xác định rõ.

Ví dụ 4.2.2 (Biến ngẫu nhiên Cauchy). PDF của biến ngẫu nhiên Cauchy, được thể hiện trong Hình 4.1, được cho bởi

$$f_X(x) = \frac{1}{\pi(1 + x^2)}. \tag{4.18}$$

Theo định nghĩa về giá trị kỳ vọng,

$$E(X) = \int_{-\infty}^{\infty} \frac{x}{\pi(1 + x^2)}\,dx = \int_{0}^{\infty} \frac{x}{\pi(1 + x^2)}\,dx - \int_{0}^{\infty} \frac{x}{\pi(1 + x^2)}\,dx. \tag{4.19}$$

Bây giờ, bằng cách đổi biến $t = x^2$,

$$\int_{0}^{\infty} \frac{x}{\pi(1 + x^2)}\,dx = \int_{0}^{\infty} \frac{1}{2\pi(1 + t)}\,dt = \lim_{t \to \infty} \frac{\log(1 + t)}{2\pi} = \infty, \tag{4.20}$$

nên E(X) không tồn tại, vì nó là hiệu của hai giới hạn có xu hướng tiến tới vô cùng.

Giá trị trung bình của một vectơ ngẫu nhiên được định nghĩa là vectơ được tạo bởi các thành phần của nó.

Định nghĩa 4.2.3 (Ý nghĩa của véc tơ ngẫu nhiên). Giá trị trung bình của một vectơ ngẫu nhiên X là

$$E(X) := \begin{bmatrix} EX_1 \\ EX_2 \\ \cdots \\ \text{BÁN TẠI}_N \end{bmatrix}. \qquad (4.21)$$

Như trong trường hợp đơn biến, giá trị trung bình có thể được hiểu là giá trị xung quanh đó phân phối của vectơ ngẫu nhiên là trung tâm.

Ngay lập tức từ tính tuyến tính của toán tử kỳ vọng trong một chiều mà toán tử trung bình là tuyến tính.

Định lý 4.2.4 (Trung bình biến đổi tuyến tính của một vectơ ngẫu nhiên). Đối với bất kỳ vectơ ngẫu nhiên X có kích thước n, bất kỳ ma trận $A \in R^{m\times n}$ và $b \in R^{tôi}$

$$E\left(AX + b\right) = AE(X) + b. \qquad (4.22)$$

Bằng chứng.

$$E\left(AX + b\right) = \begin{bmatrix} e\left(\sum_{tôi=1}^{N} A_{1i}X_i + b_1\right) \\ e\left(\sum_{i=1}^{n} A_{2i}X_{tôi} + b_2\right) \\ \cdots \\ e\left(\sum_{tôi=1}^{N} A_{mi}X_{tôi} + tỷ\right) \end{bmatrix} \qquad (4.23)$$

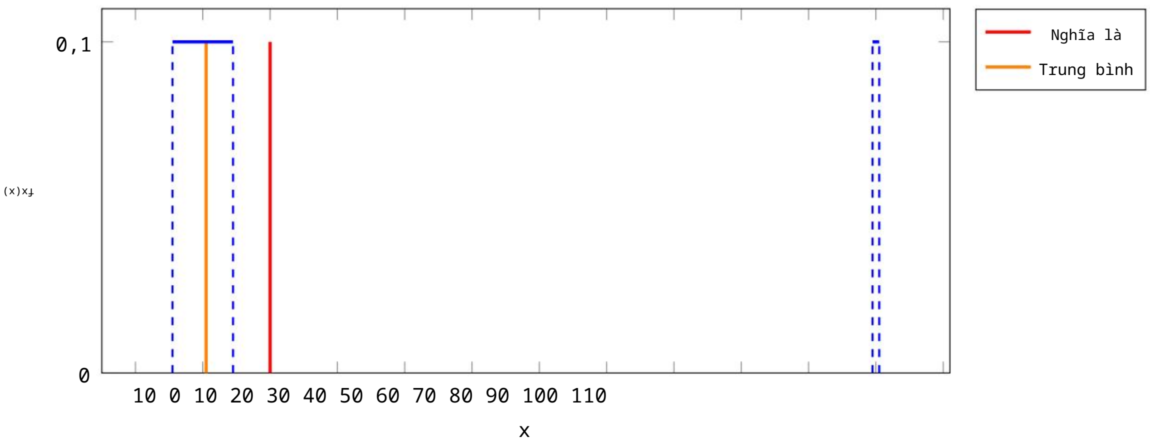
$$= \begin{bmatrix} \sum_{i=1}^{n} A_{1i}E(X_{tôi}) + b_1 \\ \sum_{tôi=1}^{N} A_{2i}E(X_{tôi}) + b_2 \\ \cdots \\ \sum_{tôi=1}^{N} A_{mi}E(X_i) + b_n \end{bmatrix} \quad \text{theo tuyến tính của kỳ vọng} \qquad (4.24)$$

$$= AE(X) + b. \qquad (4.25)$$

□

4.2.2 Trung bình

Giá trị trung bình thường được hiểu là đại diện cho một giá trị điển hình được lấy bởi biến ngẫu nhiên. Tuy nhiên, xác suất của một biến ngẫu nhiên bằng giá trị trung bình của nó có thể bằng không! Chẳng hạn, một biến ngẫu nhiên Bernoulli không thể bằng 0,5. Ngoài ra, giá trị trung bình có thể bị bóp méo nghiêm trọng bởi một tập con nhỏ các giá trị cực trị, như được minh họa trong Ví dụ 4.2.6 bên dưới. Trung vị là một đặc tính thay thế của một giá trị điển hình được lấy bởi biến ngẫu nhiên, được thiết kế để phù hợp hơn với các tình huống như vậy. Nó được định nghĩa là trung điểm của pmf hoặc pdf của biến ngẫu nhiên. Nếu biến ngẫu nhiên là liên tục, xác suất để nó lớn hơn hoặc nhỏ hơn số trung vị bằng 1/2.



Hình 4.2: PDF thống nhất trong [4,5, 4,5] [99,5, 100,5]. Giá trị trung bình là 10 và trung vị là 0,5.

Định nghĩa 4.2.5 (Trung vị). Trung vị của biến ngẫu nhiên rời rạc X là một số m sao cho

$$P(X \le m) \ge \frac{1}{2} \quad \text{và } P(X \ge m) \ge \frac{1}{2}. \tag{4.26}$$

Trung vị của biến ngẫu nhiên liên tục X là một số m sao cho

$$\text{Ngoại hối }(m) = \int_{\infty}^{\text{tôi}} f_X(x)\,dx = \frac{1}{2}. \tag{4.27}$$

Ví dụ sau đây minh họa độ chắc chắn của trung vị đối với sự hiện diện của một tập hợp con nhỏ các giá trị cực trị với xác suất khác không.

Ví dụ 4.2.6 (Trung bình so với trung vị). Xét biến ngẫu nhiên đều X có độ hỗ trợ [4,5, 4,5] [99,5, 100,5]. Giá trị trung bình của X bằng

$$E(X) = \int_{x=\ 4,5}^{4,5} x f_X(x)\,dx + \int_{x=99,5}^{100,5} x f_X(x)\,dx \tag{4.28}$$

$$= \frac{1}{10}\,\frac{99,5^{2}\;\;2\;\;100,5}{2} \tag{4.29}$$

$$= 10. \tag{4.30}$$

cdf của X giữa -4,5 và 4,5 bằng

$$\text{Ngoại hối }(m) = \int^{\text{tôi}} f_X(x)\,dx \tag{4.31}$$

$$= \frac{4,5\ m + 4,5}{10}. \tag{4.32}$$

Đặt giá trị này bằng 1/2 cho phép tính giá trị trung bình bằng 0,5. Hình 4.2 cho thấy pdf của X và vị trí của trung vị và trung bình. Trung bình cung cấp một thước đo thực tế hơn về trung tâm phân phối.

Biến ngẫu nhiên Tham số Phương sai trung bình			
Bernoulli	P	P	p (1 p)
hình học	P	$\frac{1}{\text{giờ chiều}}$	$\frac{1 \; p}{\text{giờ chiều}^2}$
nhị thức	n, p	np	np (1 p)
Poisson	λ	λ	λ
Đồng phục	một, b	$\frac{a+b \; 2}{}$	$\frac{(b \; a)^2}{12}$
số mũ	λ	$\frac{}{1 \; \lambda}$	$\frac{}{1 \; \lambda 2}$
Gaussian	μ, σ	μ	2σ _

Bảng 4.1: Trung bình và phương sai của các biến ngẫu nhiên chung, được trích dẫn trong Mục 4.5.1 của phụ lục.

4.2.3 Phương sai và độ lệch chuẩn

Giá trị kỳ vọng của bình phương của một biến ngẫu nhiên đôi khi được sử dụng để định lượng năng lượng của biến ngẫu nhiên.

Định nghĩa 4.2.7 (Thời điểm thứ hai). Bình phương trung bình hoặc khoảnh khắc thứ hai của một biến ngẫu nhiên X là giá trị kỳ vọng của X2 : E X2 .

Định nghĩa khái quát hóa các khoảnh khắc cao hơn, được định nghĩa là E (Xp) cho các số nguyên lớn hơn hai. Bình phương trung bình của hiệu giữa biến ngẫu nhiên và giá trị trung bình của nó được gọi là phương sai của giá trị ngẫu nhiên. Nó định lượng sự thay đổi của biến ngẫu nhiên xung quanh giá trị trung bình của nó và còn được gọi là thời điểm trung tâm thứ hai của phân phối. Căn bậc hai của đại lượng này là độ lệch chuẩn của biến ngẫu nhiên.

Định nghĩa 4.2.8 (Phương sai và độ lệch chuẩn). Phương sai của X là độ lệch bình phương trung bình so với giá trị trung bình

$$\text{Biến } (X) := E\,(X \quad E\,(X))2 \tag{4.33}$$

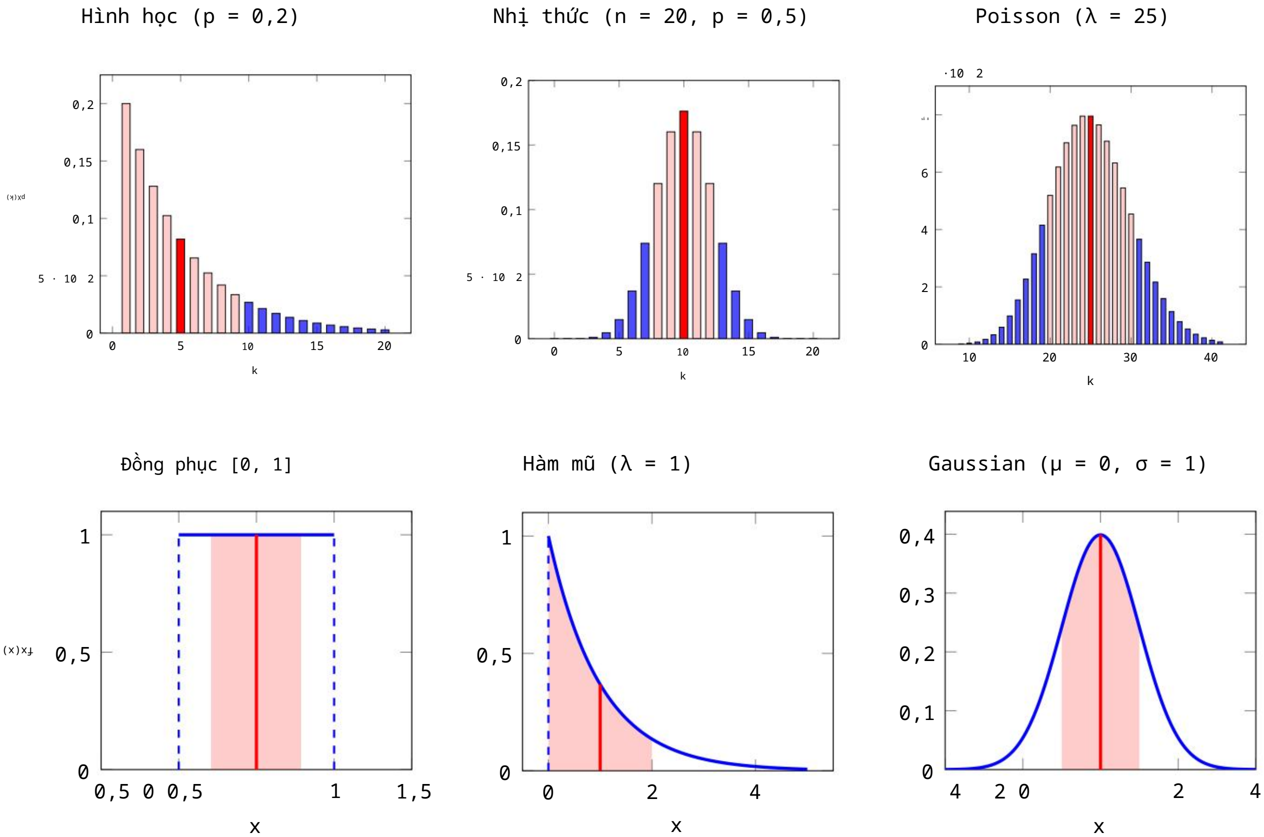
$$= E\,X2 \quad E^{2}\,(X). \tag{4.34}$$

Độ lệch chuẩn σX của X là

$$\sigma X := \sqrt{\text{Var}(X)}. \tag{4.35}$$

Chúng tôi đã tổng hợp phương sai của một số biến ngẫu nhiên quan trọng trong Bảng 4.1. Các dẫn xuất có thể được tìm thấy trong Phần 4.5.1. Trong Hình 4.3 , chúng tôi vẽ sơ đồ pmfs và pdfs của các biến ngẫu nhiên này và hiển thị phạm vi giá trị nằm trong một độ lệch chuẩn của giá trị trung bình.

Toán tử phương sai không phải là tuyến tính, nhưng có thể dễ dàng xác định phương sai của hàm tuyến tính của một biến ngẫu nhiên.



Hình 4.3: Pmfs của biến ngẫu nhiên rời rạc (hàng trên cùng) và pdfs của biến ngẫu nhiên liên tục (hàng dưới). Giá trị trung bình của biến ngẫu nhiên được đánh dấu màu đỏ. Các giá trị nằm trong một tiêu chuẩn độ lệch của giá trị trung bình được đánh dấu bằng màu hồng.

Bổ đề 4.2.9 (Phương sai của hàm tuyến tính). Với mọi hằng số a và b

$$\text{Biến}\,(a\,X + b) = a^2\,\text{Biến}\,(X). \tag{4.36}$$

Bằng chứng.

$$\text{Var}\,(a\,X + b) = E\,(a\,X + b \quad E\,(a\,X + b))2 \tag{4.37}$$

$$= E\,(a\,X + b \quad aE\,(X) \quad b)^2 \tag{4.38}$$

$$^{2} = \text{một}\; E\,(X \quad E\,(X))2 \tag{4.39}$$

$$^{2} = \text{một}\; \text{Biến}\,(X). \tag{4.40}$$

□

Kết quả này có ý nghĩa: Nếu chúng ta thay đổi tâm của biến ngẫu nhiên bằng cách thêm một hằng số, thì phương sai không bị ảnh hưởng vì phương sai chỉ đo độ lệch so với nghĩa là. Nếu chúng ta nhân một biến ngẫu nhiên với một hằng số, thì độ lệch chuẩn được tính theo tỷ lệ cùng một yếu tố.

4.2.4 Giới hạn xác suất sử dụng giá trị trung bình và phương sai

Trong phần này, chúng tôi giới thiệu hai bất đẳng thức cho phép mô tả hành vi của một giá trị ngẫu nhiên ở một mức độ nào đó chỉ từ việc biết giá trị trung bình và phương sai của nó. Đầu tiên là bất đẳng thức Markov, định lượng ý tưởng trực quan rằng nếu một biến ngẫu nhiên không âm và nhỏ thì xác suất để nó nhận các giá trị lớn phải nhỏ.

Định lý 4.2.10 (Bất đẳng thức Markov). Cho X là biến ngẫu nhiên không âm. Với mọi hằng số dương $a > 0$,

$$P(X \geq a) \leq \frac{\text{BÁN TẠI)}}{\text{Một}}. \tag{4.41}$$

Bằng chứng. Xem xét biến chỉ báo $1_{X \geq a}$. Chúng ta có

$$X \quad a\, 1_{X \geq a} \geq 0. \tag{4.42}$$

Cụ thể, kỳ vọng của nó là không âm (vì nó là tổng hoặc tích phân của một đại lượng không âm trên đường thực dương). Theo tính tuyến tính của kỳ vọng và việc $1_{X \geq a}$ là biến ngẫu nhiên Bernoulli với kỳ vọng $P(X \geq a)$ ta có

$$E(X) \geq aE\left(1_{X \geq a}\right) = aP(X \geq a). \tag{4.43}$$

□

Ví dụ 4.2.11 (Tuổi học sinh). Bạn nghe nói rằng độ tuổi trung bình của sinh viên NYU là 20, nhưng bạn biết khá nhiều sinh viên trên 30 tuổi. Bạn quyết định áp dụng bất đẳng thức Markov để giới hạn tỷ lệ sinh viên trên 30 tuổi bằng cách lập mô hình tuổi như một biến ngẫu nhiên không âm A.

$$P(A \geq 30) \leq \frac{E(A)}{30} = \frac{2}{3}. \tag{4.44}$$

Nhiều nhất hai phần ba số sinh viên trên 30 tuổi.

Như ví dụ minh họa 4.2.11, bất đẳng thức Markov có thể khá lỏng lẻo. Lý do là nó hầu như không sử dụng bất kỳ thông tin nào về phân phối của biến ngẫu nhiên.

Bất đẳng thức Chebyshev kiểm soát độ lệch của biến ngẫu nhiên so với giá trị trung bình của nó. Theo trực giác, nếu phương sai (và do đó là độ lệch chuẩn) nhỏ, thì xác suất để biến ngẫu nhiên ở xa giá trị trung bình của nó phải thấp.

Định lý 4.2.12 (Bất đẳng thức Chebyshev). Với mọi hằng số dương $a > 0$ và mọi biến ngẫu nhiên X có phương sai bị chặn,

$$P(|X \quad E(X)| \geq a) \leq \frac{\text{Biến }(X)}{\text{một}^2}. \tag{4.45}$$

Bằng chứng. Áp dụng bất đẳng thức Markov cho biến ngẫu nhiên $Y = (X \quad E(X))^2$ ta có kết quả.

□

Một hệ quả thú vị của bất đẳng thức Chebyshev cho thấy rằng nếu phương sai của một biến ngẫu nhiên bằng 0, thì biến ngẫu nhiên đó là một hằng số, hay nói một cách chính xác, xác suất mà nó đi chệch khỏi giá trị trung bình của nó là bằng không.

Hệ quả 4.2.13. Nếu Var(X)=0 thì P(X=E(X))=0.

Bằng chứng. Lấy bất kỳ > 0, theo bất đẳng thức Chebyshev

$$P(|X \quad E(X)| \geq) \leq \frac{\text{Biến }(X)}{2} = 0. \tag{4.46}$$

□

Ví dụ 4.2.14 (Tuổi học sinh (tiếp theo)). Bạn không hài lòng lắm với giới hạn của mình về số lượng sinh viên trên 30. Bạn phát hiện ra rằng độ lệch chuẩn của tuổi sinh viên thực ra chỉ là 3 năm. Áp dụng bất đẳng thức Chebyshev, điều này ngụ ý rằng

$$P(A \geq 30) \leq P(|A \quad E(A)| \geq 10) \tag{4.47}$$

$$\frac{\text{Biến số}}{(A) \leq 100} = \frac{9}{100}. \tag{4.48}$$

Vì vậy, thực tế ít nhất 91% sinh viên dưới 30 tuổi (và trên 10 tuổi).

4.3 Hiệp phương sai

4.3.1 Hiệp phương sai của hai biến ngẫu nhiên

Hiệp phương sai của hai biến ngẫu nhiên mô tả hành vi chung của chúng. Đó là giá trị kỳ vọng của sản phẩm giữa sự khác biệt của các biến ngẫu nhiên và phương tiện tương ứng của chúng. Theo trực giác, nó đo lường mức độ dao động của các biến ngẫu nhiên cùng nhau.

Định nghĩa 4.3.1 (Hiệp phương sai). Hiệp phương sai của X và Y là

$$\mathrm{Cov}(X, Y) := E((X \quad E(X))(Y \quad E(Y))) \tag{4.49}$$

$$= E(XY) \quad E(X)\,E(Y). \tag{4.50}$$

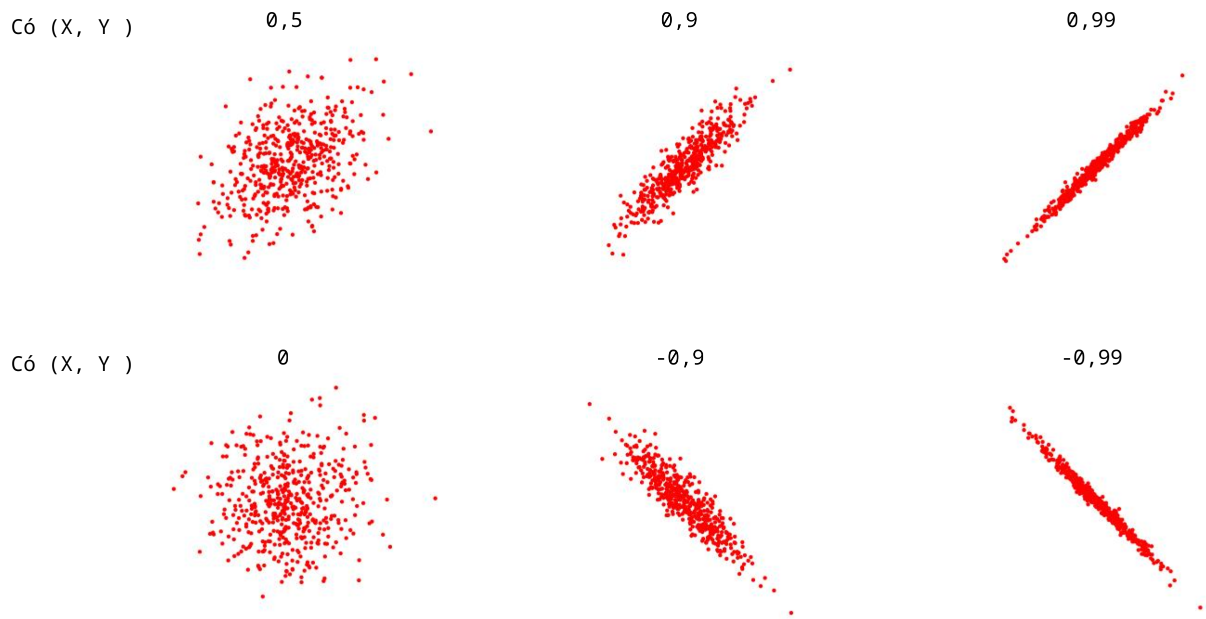
Nếu Cov(X, Y) = 0 thì X và Y không tương quan với nhau.

Hình 4.4 cho thấy các mẫu từ phân bố Gaussian hai biến với các hiệp phương sai khác nhau. Nếu hiệp phương sai bằng 0, thì pdf chung có dạng hình cầu. Nếu hiệp phương sai dương và lớn, thì pdf chung sẽ bị lệch để hai biến có xu hướng có giá trị tương tự nhau.
Nếu hiệp phương sai lớn và âm, thì hai biến sẽ có xu hướng có các giá trị tương tự nhưng ngược dấu.

Phương sai của tổng của hai biến ngẫu nhiên có thể được biểu diễn dưới dạng phương sai riêng lẻ và hiệp phương sai của chúng. Kết quả là, dao động của chúng củng cố lẫn nhau nếu hiệp phương sai là dương và triệt tiêu lẫn nhau nếu nó là âm.

Định lý 4.3.2 (Phương sai của tổng hai biến ngẫu nhiên).

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y). \tag{4.51}$$



Hình 4.4: Các mẫu từ vectơ Gaussian 2D (X, Y), trong đó X và Y là các biến ngẫu nhiên Gaussian tiêu chuẩn với giá trị trung bình bằng 0 và phương sai đơn vị, đối với các giá trị khác nhau của hiệp phương sai giữa X và Y .

Bằng chứng.

$$\operatorname{Var}(X + Y) = E(X + Y \quad E(X + Y))^2 \tag{4.52}$$

$$= E(X \quad E(X))^2 + E(Y \quad E(Y))^2 + 2E((X \quad E(X))(Y \quad E(Y)))$$

$$= \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X, Y). \tag{4.53}$$

□

Hệ quả tức thời là nếu hai biến ngẫu nhiên không tương quan với nhau, thì phương sai của tổng của chúng bằng tổng của các phương sai của chúng.

Hệ quả 4.3.3. Nếu X và Y không tương quan với nhau thì

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y). \tag{4.54}$$

Bổ đề và ví dụ sau đây cho thấy tính độc lập bao hàm sự không tương quan, nhưng sự không tương quan không phải lúc nào cũng bao hàm sự độc lập.

Bổ đề 4.3.4 (Độc lập hàm ý không tương quan). Nếu hai biến ngẫu nhiên là độc lập thì chúng không tương quan với nhau.

Bằng chứng. Theo Định lý 4.1.7, nếu X và Y độc lập

$$\operatorname{Cov}(X, Y) = E(XY) \quad E(X)\,E(Y) = E(X)\,E(Y) \quad E(X)\,E(Y) = 0. \tag{4.55}$$

□

Ví dụ 4.3.5 (Không tương quan không có nghĩa là độc lập). Cho X và Y là hai biến ngẫu nhiên Bernoulli độc lập với tham số 1/2. Xét các biến ngẫu nhiên

$$U = X + Y, \tag{4.56}$$
$$V = X \quad Y. \tag{4.57}$$

Lưu ý rằng

$$p_U(0) = P(X = 0, Y = 0) = \frac{1}{4}, \tag{4.58}$$

$$p_V(0) = P(X = 1, Y = 1) + P(X = 0, Y = 0) = \frac{1}{2}, \tag{4.59}$$

$$p_{U,V}(0, 0) = P(X = 0, Y = 0) = \frac{1}{4} = p_U(0)\, p_V(0) = \frac{1}{8}, \tag{4.60}$$

nên U và V không độc lập. Tuy nhiên, chúng không tương quan như

$$\mathrm{Cov}(U, V) = E(UV) \quad E(U)\,E(V) \tag{4.61}$$
$$= E((X + Y)(X \quad Y)) \quad E(X + Y)\,E(X \quad Y) \tag{4.62}$$
$$= E X^2 \quad EY^2 - E^2(X) + E^2(Y) = 0. \tag{4.63}$$

Đẳng thức cuối cùng đúng vì X và Y có cùng phân phối.

4.3.2 Hệ số tương quan

Hiệp phương sai không tính đến độ lớn của phương sai của các biến ngẫu nhiên liên quan. Hệ số tương quan Pearson thu được bằng cách chuẩn hóa hiệp phương sai bằng cách sử dụng độ lệch chuẩn của cả hai biến.

Định nghĩa 4.3.6 (Hệ số tương quan Pearson). Hệ số tương quan Pearson của hai biến ngẫu nhiên X và Y là

$$\rho_{X,Y} := \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}. \tag{4.64}$$

Hệ số tương quan giữa X và Y bằng hiệp phương sai giữa X/σ_X và Y/σ_Y. Hình 4.5 so sánh các mẫu biến ngẫu nhiên Gaussian hai chiều có cùng hệ số tương quan nhưng hiệp phương sai khác nhau và ngược lại.

Mặc dù có thể không rõ ràng ngay lập tức, nhưng độ lớn của hệ số tương quan bị giới hạn bởi một vì hiệp phương sai của hai biến ngẫu nhiên không thể vượt quá tích của các độ lệch chuẩn của chúng. Một giải thích hữu ích về hệ số tương quan là nó định lượng mức độ X và Y có liên quan tuyến tính. Trên thực tế, nếu nó bằng 1 hoặc -1 thì một trong các biến là hàm tuyến tính của biến kia! Tất cả điều này suy ra từ bất đẳng thức Cauchy-Schwarz. Bằng chứng ở Mục 4.5.3.

Định lý 4.3.7 (Bất đẳng thức Cauchy-Schwarz). Đối với mọi biến ngẫu nhiên X và Y được xác định trên cùng một không gian xác suất

$$|E(XY)| \leq \sqrt{E(X^2)\,E(Y^2)}. \tag{4.65}$$

$\sigma_Y = 1$, Cov(X, Y) = 0,9, $\rho_{X,Y} = 0,9$ | $\sigma_Y = 3$, Cov(X, Y) = 0,9, $\rho_{X,Y} = 0,3$ | $\sigma_Y = 3$, Cov(X, Y) = 2,7, $\rho_{X,Y} = 0,9$

Hình 4.5: Các mẫu từ vectơ Gaussian 2D (X, Y), trong đó X là biến ngẫu nhiên Gaussian tiêu chuẩn với giá trị trung bình bằng 0 và phương sai đơn vị, đối với các giá trị khác nhau của độ lệch chuẩn σ_Y của Y (có nghĩa là bằng 0) và của hiệp phương sai giữa X và Y .

Giả sử $EX^2 = 0$,

$$E(XY) = \sqrt{E(X^2)\,E(Y^2)} \quad Y = \frac{\sqrt{E(Y^2)}}{E(X^2)} X, \tag{4.66}$$

$$E(XY) = \sqrt{E(X^2)\,E(Y^2)} \quad Y = \frac{\sqrt{E(Y^2)}}{E(X^2)} X. \tag{4.67}$$

Hệ quả 4.3.8. Đối với mọi biến ngẫu nhiên X và Y ,

$$\mathrm{Cov}(X, Y) \leq \sigma_X \sigma_Y . \tag{4.68}$$

Tương tự, hệ số tương quan Pearson thỏa mãn

$$|\rho_{X,Y}| \leq 1, \tag{4.69}$$

bằng nhau khi và chỉ khi tồn tại quan hệ tuyến tính giữa X và Y

$$|\rho_{X,Y}| = 1 \quad Y = c X + d. \tag{4.70}$$

Ở đâu

$$c := \begin{cases} \frac{\sigma_Y}{\sigma_X} & \text{nếu } \rho_{X,Y} = \\ -\frac{\sigma_Y}{\sigma_X} & 1, \text{ nếu } \rho_{X,Y} = \quad 1, \end{cases} \qquad d := E(Y) \quad cE(X). \tag{4.71}$$

Bằng chứng. Cho phép

$$U := X \quad E(X), \tag{4.72}$$

$$V := Y \quad E(Y). \tag{4.73}$$

Từ định nghĩa của phương sai và hệ số tương quan,

$$EU^2 = \text{Biến } (X), \tag{4,74}$$

$$\text{xe điện}^2 = \text{Biến } (Y) \tag{4,75}$$

$$\rho_{X,Y} = \frac{\text{E (Tia cực tím)}}{\sqrt{E(U^2)\,E(V^2)}}. \tag{4.76}$$

Kết quả bây giờ có được từ việc áp dụng Định lý 4.3.7 cho U và V . □

4.3.3 Ma trận hiệp phương sai của véc tơ ngẫu nhiên

Ma trận hiệp phương sai của một vectơ ngẫu nhiên ghi lại sự tương tác giữa các thành phần của vectơ. Nó chứa phương sai của từng thành phần trong đường chéo và hiệp phương sai giữa các thành phần khác nhau trong các đường chéo ngoài.

Định nghĩa 4.3.9. Ma trận hiệp phương sai của một vectơ ngẫu nhiên X được định nghĩa là

$$\Sigma_X := \begin{pmatrix} \text{biến}\,(X_1) & \text{Cov}\,(X_1, X_2) & \cdots & \text{Cov}\,(X_1, X_N) \\ \text{Cov}\,(X_2, X_1) & \text{biến}\,(X_2) & \cdots & \text{Cov}\,(X_2, X_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}\,(X_N, X_1) & \text{Cov}\,(X_N, X_2) & \cdots & \text{Var}\,(X_N) \end{pmatrix} \tag{4.77}$$

$$= E\left(XX^t\right) - E(X)E(X)^t. \tag{4.78}$$

Lưu ý rằng nếu tất cả các mục của một vectơ không tương quan với nhau, thì ma trận hiệp phương sai của nó là đường chéo.

Từ Định lý 4.2.4 , chúng ta thu được một biểu thức đơn giản cho ma trận hiệp phương sai của phép biến đổi tuyến tính của một vectơ ngẫu nhiên.

Định lý 4.3.10 (Ma trận hiệp phương sai sau một phép biến đổi tuyến tính). Cho X là một vectơ ngẫu nhiên có chiều n với ma trận hiệp phương sai Σ. Với bất kỳ ma trận $A \in \mathbb{R}^{m\times n}$ và $b \in \mathbb{R}^m$,

$$\Sigma_{AX+b} = A\Sigma_X A^t. \tag{4.79}$$

Bằng chứng.

$$\Sigma_{AX+b} = E\left((AX+b)(AX+b)^t\right) - E(AX+b)\,E(AX+b)^t \tag{4.80}$$

$$= AE\left(XX^t\right)\text{MỘT}^t + bE(X)^t\,\text{MỘT}^t + AE(X)\,b^t + bb^t$$

$$- AE(X)E(X)^t\,\text{MỘT}^t - AE(X)\,b^t - bE(X)^t\,\text{MỘT}^t - bb^t \tag{4.81}$$

$$= A\left(E\left(XX^t\right) - E(X)E(X)^t\right)\text{MỘT}^t \tag{4.82}$$

$$= A\Sigma_X A^t. \tag{4.83}$$

□

Một hệ quả trực tiếp của kết quả này là chúng ta có thể dễ dàng giải mã phương sai của vectơ ngẫu nhiên theo bất kỳ hướng nào từ ma trận hiệp phương sai. Về mặt toán học, phương sai của vectơ ngẫu nhiên theo hướng của vectơ đơn vị v bằng với phương sai của hình chiếu của nó lên v.

Hệ quả 4.3.11. Cho v là một vectơ đơn vị,

$$\text{biến}\left(v^t X\right) = v^t \Sigma_X v. \tag{4.84}$$

Xem xét sự phân rã riêng của ma trận hiệp phương sai của một vectơ ngẫu nhiên n chiều X

$$\Sigma_X = U \Lambda U^t \tag{4.85}$$

$$= \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}^t, \tag{4.86}$$

trong đó các giá trị riêng được sắp xếp $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. Ma trận hiệp phương sai là đối xứng theo định nghĩa, do đó theo Định lý B.7.1 các vectơ riêng $u_1, u_2, \ldots, u_n$ có thể được chọn là song phương hoặc vuông góc. Các vectơ riêng này và các giá trị riêng này hoàn toàn đặc trưng cho phương sai của vectơ ngẫu nhiên theo các hướng khác nhau. Định lý là hệ quả trực tiếp của Hệ quả 4.3.11 và Định lý B.7.2.

Định lý 4.3.12. Cho X là một vectơ ngẫu nhiên có chiều n với ma trận hiệp phương sai Σ_X . Các phân tích riêng của Σ_X cho bởi (4.86) thỏa mãn

$$\lambda_1 = \text{tối đa}_{||v||_2=1} \text{biến } v^T X, \tag{4.87}$$

$$u_1 = \text{đối số tối đa}_{||v||_2=1} \text{biến } v^t X, \tag{4.88}$$

$$\lambda_k = \text{tối đa}_{||v||_2=1, v \perp u_1,\ldots,u_{k-1}} \text{Var } v^T X, \tag{4.89}$$

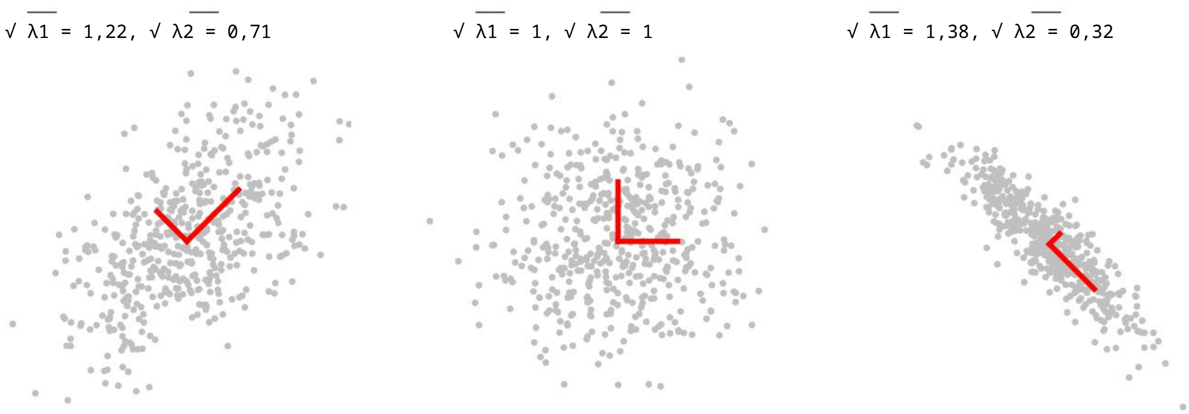
$$= \arg \text{tối đa}_{||v||_2=1, v \perp u_1,\ldots,u_{k-1}} \text{Var } v^T X. u_k \tag{4,90}$$

Nói cách khác, u_1 là hướng của phương sai tối đa. Vectơ riêng u_2 tương ứng với giá trị riêng lớn thứ hai λ_2 là hướng biến thiên cực đại trực giao với u_1. TRONG chung, véc tơ riêng u_k tương ứng với giá trị riêng lớn thứ k λ_k tiết lộ hướng của độ biến thiên lớn nhất trực giao với $u_1, u_2, \ldots, u_{k-1}$. Cuối cùng, u_n là hướng của phương sai tối thiểu. Hình 4.6 minh họa điều này bằng một ví dụ, trong đó n = 2. Khi chúng ta thảo luận trong Chương 8, phân tích thành phần chính- một phương pháp phổ biến cho việc học không giám sát và giảm kích thước- áp dụng nguyên tắc tương tự để xác định hướng biến đổi của một bộ dữ liệu.

Để kết thúc phần này, chúng tôi mô tả một thuật toán để biến đổi các mẫu từ một vectơ ngẫu nhiên để chúng có ma trận hiệp phương sai theo quy định. Quá trình biến đổi các mẫu không tương quan cho mục đích này được gọi là tô màu vì các mẫu không tương quan thường được được mô tả là tiếng ồn trắng. Như chúng ta sẽ thấy trong phần tiếp theo, tô màu cho phép mô phỏng Các vectơ ngẫu nhiên Gaussian.

Thuật toán 4.3.13 (Tô màu các mẫu không tương quan). Cho x là một thực thể từ một n-chiều vectơ ngẫu nhiên với ma trận hiệp phương sai I. Để tạo các mẫu có ma trận hiệp phương sai Σ, chúng tôi:

1. Tính toán phân tích riêng $\Sigma = U \Lambda U^t$.



Hình 4.6: Các mẫu từ các vectơ ngẫu nhiên Gaussian hai biến với các ma trận hiệp phương sai khác nhau được hiển thị bằng màu xám. Các vectơ riêng của ma trận hiệp phương sai được vẽ bằng màu đỏ. Mỗi cái được chia tỷ lệ theo mái vuông của giá trị riêng tương ứng λ1 hoặc λ2.

2. Đặt $y := U\sqrt{\Lambda}x$, trong đó $\sqrt{\Lambda}$ là ma trận đường chéo chứa các căn bậc hai của hàm riêng các giá trị của Σ,

$$\sqrt{\Lambda} := \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{bmatrix}. \tag{4.91}$$

Theo Định lý 4.3.10, ma trận hiệp phương sai của $Y := U\sqrt{\Lambda}x$ thực sự bằng Σ.

$$\Sigma_Y = U\sqrt{\Lambda}\Sigma_X\sqrt{\Lambda}^T U^T \tag{4.92}$$

$$= U\sqrt{\Lambda}I\sqrt{\Lambda}^T U^T \tag{4.93}$$

$$= \Sigma. \tag{4.94}$$

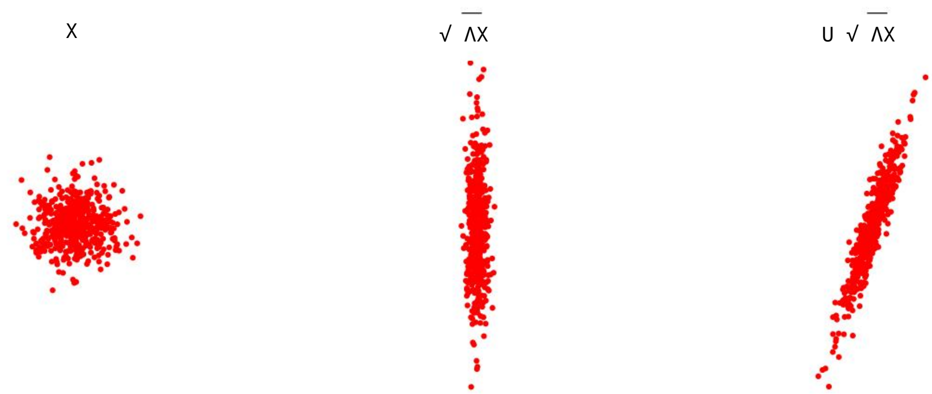
Hình 4.7 minh họa hai bước tô màu trong 2D: Đầu tiên, các mẫu được kéo giãn theo các giá trị riêng của Σ và sau đó chúng được xoay để sắp xếp chúng với các véc tơ riêng tương ứng.

4.3.4 Các vectơ ngẫu nhiên Gaussian

Chúng tôi chủ yếu sử dụng các vectơ Gaussian để trực quan hóa các thuộc tính khác nhau của toán tử hiệp phương sai. Trái ngược với các vectơ ngẫu nhiên khác, vectơ ngẫu nhiên Gauss hoàn toàn được xác định bởi giá trị trung bình và ma trận hiệp phương sai của chúng. Một hệ quả quan trọng là nếu các mục của vectơ ngẫu nhiên Gaussian không tương quan thì chúng cũng độc lập lẫn nhau.

Bổ đề 4.3.14 (Không tương quan ngụ ý tính độc lập lẫn nhau đối với các vectơ ngẫu nhiên Gaussian). Nếu tất cả các thành phần của vectơ ngẫu nhiên Gaussian X không tương quan với nhau, điều này có nghĩa là chúng độc lập lẫn nhau.

Bằng chứng. Tham số Σ của pdf chung của vectơ ngẫu nhiên Gaussian là ma trận hiệp phương sai của nó (người ta có thể xác minh điều này bằng cách áp dụng định nghĩa hiệp phương sai và tích hợp). Nếu tất cả các thành phần



Hình 4.7: Khi chúng ta tô màu các mẫu không tương quan hai chiều (trái), đầu tiên ma trận đường chéo $\sqrt{\Lambda}$ kéo dài chúng theo các hướng khác nhau theo các giá trị riêng của ma trận hiệp phương sai mong muốn (ở giữa) và sau đó U xoay chúng sao cho chúng thẳng hàng với các vectơ riêng tương ứng (phải).

thì không tương quan

$$\Sigma_X = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix}, \tag{4,95}$$

trong đó σ_i là độ lệch chuẩn của thành phần thứ i. Bây giờ, nghịch đảo của ma trận đường chéo này chỉ là

$$\Sigma_X^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_N^2} \end{bmatrix}, \tag{4.96}$$

và định thức của nó là $|\Sigma| = \prod_{i=1}^{N} \sigma_i^2$ để có thể

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^N |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right) \tag{4.97}$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)}\sigma_i} \exp\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right) \tag{4,98}$$

$$= \prod_{i=1}^{N} f_{X_i}(x_i). \tag{4,99}$$

Vì các yếu tố pdf chung thành một sản phẩm của các biên, nên các thành phần này đều độc lập lẫn nhau. □

Thuật toán sau đây tạo các mẫu từ một vectơ ngẫu nhiên Gaussian với ma trận hiệp phương sai và trung bình tùy ý bằng cách tô màu (và căn giữa) một vectơ gồm các mẫu độc lập từ phân phối Gaussian tiêu chuẩn.

Thuật toán 4.3.15 (Tạo vectơ ngẫu nhiên Gaussian). Để lấy mẫu từ một vectơ ngẫu nhiên Gaussian n chiều với ma trận trung bình μ và hiệp phương sai Σ, chúng tôi:

1. Tạo một vectơ x chứa n mẫu Gaussian tiêu chuẩn độc lập.
2. Tính toán phân tích riêng $\Sigma = U\Lambda U^t$.
3. Đặt $y := U\sqrt{\Lambda}x + \mu$, trong đó $\sqrt{\Lambda}$ được xác định bởi (8.20).

Thuật toán chỉ căn giữa và tô màu vectơ ngẫu nhiên $Y := U\sqrt{\Lambda}X + \mu$. Theo tuyến tính của kỳ vọng, ý nghĩa của nó là

$$E(Y) = U\sqrt{\Lambda}E(X) + \mu \tag{4.100}$$

$$= \mu \tag{4.101}$$

vì giá trị trung bình của X bằng không. Lập luận tương tự được sử dụng trong phương trình (4.94) cho thấy ma trận hiệp phương sai của X là Σ. Vì tô màu và định tâm là các phép toán tuyến tính, theo Định lý 3.2.14 Y là Gaussian với ma trận trung bình và hiệp phương sai mong muốn. Ví dụ, trong Hình 4.7 , các mẫu được tạo ra là Gaussian. Đối với các vectơ ngẫu nhiên không phải Gaussian, việc tô màu sẽ sửa đổi ma trận hiệp phương sai, nhưng không nhất thiết phải bảo toàn phân phối.

4.4 Kỳ vọng có điều kiện

Kỳ vọng có điều kiện là một công cụ hữu ích để thao tác với các biến ngẫu nhiên. Thật không may, nó có thể hơi khó hiểu (như chúng ta thấy bên dưới, đây là một biến ngẫu nhiên không phải là một kỳ vọng!). Xét một hàm g gồm hai biến ngẫu nhiên X và Y . Kỳ vọng của g có điều kiện dựa trên sự kiện X = x với bất kỳ giá trị cố định nào x có thể được tính bằng cách sử dụng pmf hoặc pdf có điều kiện của Y cho trước X.

$$E(g(X, Y)|X = x) = \sum_{y \in R} g(x, y)\, p_{Y|X}(y|x), \tag{4.102}$$

nếu Y là rời rạc và có phạm vi R, trong khi

$$E(g(X, Y)|X = x) = \int_{y=-\infty}^{\infty} g(x, y)\, f_{Y|X}(y|x)\, dy, \tag{4.103}$$

nếu Y liên tục.

Lưu ý rằng $E(g(X, Y)|X = x)$ thực sự có thể được hiểu là một hàm của x vì nó ánh xạ mọi giá trị của x thành một số thực. Điều này cho phép xác định kỳ vọng có điều kiện của g(X, Y) cho X như sau.

Định nghĩa 4.4.1 (Kỳ vọng có điều kiện). Kỳ vọng có điều kiện của g(X, Y) cho trước X là

$$E(g(X, Y)|X) := h(X), \tag{4.104}$$

Ở đâu

$$h(x) := E(g(X, Y)|X = x). \tag{4.105}$$

Machine Translated by Google

Coi chừng định nghĩa khó hiểu, kỳ vọng có điều kiện thực sự là một biến ngẫu nhiên!

Một trong những ứng dụng chính của kỳ vọng có điều kiện là áp dụng kỳ vọng lặp cho tính toán Các giá trị dự kiến. Ý tưởng là giá trị kỳ vọng của một số lượng nhất định có thể được biểu thị bằng kỳ vọng của kỳ vọng có điều kiện của đại lượng.

Định lý 4.4.2 (Kỳ vọng lặp). Đối với bất kỳ biến ngẫu nhiên X và Y và bất kỳ hàm nào $g : R^2 \quad R$

$$E(g(X, Y)) = E(E(g(X, Y)|X)). \tag{4.106}$$

Bằng chứng. Ta chứng minh kết quả cho biến ngẫu nhiên liên tục, chứng minh cho biến ngẫu nhiên rời rạc các biến và đối với các đại lượng phụ thuộc vào cả biến ngẫu nhiên liên tục và rời rạc, là gần như giống hệt nhau. Để làm cho lời giải thích rõ ràng hơn, chúng tôi xác định

$$h(x) := E(g(X, Y)|X = x) \tag{4.107}$$

$$= \int_{y= \quad \infty}^{\infty} g(x, y)\, f_{Y|X}(y|x)\, dy. \tag{4.108}$$

Hiện nay,

$$E(E(g(X, Y)|X)) = E(h(X)) \tag{4.109}$$

$$= \int_{x= \quad \infty}^{\infty} h(x)\, f_X(x)\, dx \tag{4.110}$$

$$= \int_{x= \quad \infty}^{\infty} \int_{y= \quad \infty}^{\infty} f_X(x)\, f_{Y|X}(y|x)\, g(x, y)\, dy\, dx \tag{4.111}$$

$$= E(g(X,Y)). \tag{4.112}$$

□

Kỳ vọng lặp lại cho phép thu được kỳ vọng của các đại lượng phụ thuộc vào một số đại lượng rất dễ dàng nếu chúng ta có quyền truy cập vào các phân phối cận biên và có điều kiện. chúng tôi minh họa này với một số ví dụ được lấy từ các chương trước.

Ví dụ 4.4.3 (Sa mạc (tiếp theo từ Ví dụ 3.4.10)). Hãy để chúng tôi tính toán thời gian trung bình tại mà chiếc xe bị hỏng, tức là giá trị trung bình của T. Bằng kỳ vọng lặp

$$E(T) = E(E(T|M, R)) \tag{4.113}$$

$$= E\left(\frac{1}{M + R}\right) \quad \text{bởi vì T là hàm mũ khi có điều kiện trên M và R} \tag{4.114}$$

$$= \int_0^1 \int_0^1 \frac{1}{m + r}\, dm\, dr \tag{4.115}$$

$$= \int_0^1 \log(r + 1) \quad \log(r)\, dr \tag{4.116}$$

$$= \log 4 \approx 1{,}39 \quad \text{tích hợp từng phần.} \tag{4.117}$$

Ví dụ 4.4.4 (Grizzlies ở Yellowstone (tiếp theo từ Ví dụ 3.3.3)). Hãy tính trọng lượng trung bình của một con gấu ở Yosemite. Bằng kỳ vọng lặp đi lặp lại

$$E(W) = E(E(W|S)) \tag{4.118}$$

$$= \frac{E(W|S=0) + E(W|S=1)}{2} \tag{4.119}$$

$$= 180\text{kg}. \tag{4.120}$$

Ví dụ 4.4.5 (Tung đồng xu Bayesian (tiếp theo từ Ví dụ 3.3.6). Hãy tính giá trị trung bình của kết quả tung đồng xu X. Bằng kỳ vọng lặp

$$E(X) = E(E(X|B)) \tag{4.121}$$

$$= E(B) \text{ vì X là Bernoulli khi có điều kiện trên B} \tag{4.122}$$

$$= \int_0^1 2b^2\, db \tag{4.123}$$

$$= \frac{2}{3}. \tag{4.124}$$

4.5 Bằng chứng

4.5.1 Đạo hàm của trung bình và phương sai trong Bảng 4.1

Bernoulli

$$E(X) = p_X(1) = p, \tag{4.125}$$

$$E\left(X^2\right) = p_X(1), \tag{4.126}$$

$$\text{Biến}(X) = E\left(X^2\right) - E^2(X) = p(1-p). \tag{4.127}$$

hình học

Để tính giá trị trung bình của một biến ngẫu nhiên hình học, chúng ta cần xử lý một chuỗi hình học. Theo Bổ đề 4.5.3 trong Mục 4.5.2 dưới đây, ta có:

$$E(X) = \sum_{k=1}^{\infty} k\, p_X(k) \tag{4.128}$$

$$= \sum_{k=1}^{\infty} kp(1-p)^{k-1} \tag{4.129}$$

$$= \frac{p}{1-p} \sum_{k=1}^{\infty} k(1-p)^k = \frac{1}{p}. \tag{4.130}$$

Machine Translated by Google

Để tính giá trị bình phương trung bình, chúng ta áp dụng Bổ đề 4.5.4 trong cùng phần:

$$X2 = \sum_{k=1}^{\infty} k^2 p_X(k) \quad (4.131)$$

$$= \sum_{k=1}^{\infty} k^2 p(1-p)^{k-1} \quad (4.132)$$

$$= \frac{p}{1-p} \sum_{k=1}^{\infty} k^2 (1-p)^k \quad (4.133)$$

$$= \frac{2-p}{p^2}. \quad (4.134)$$

nhị thức

Như trong Ví dụ 2.2.6, chúng ta có thể biểu diễn một biến ngẫu nhiên nhị thức với tham số n và p là tổng của n biến ngẫu nhiên Bernoulli độc lập B1, B2, . . . với tham số p

$$X = \sum_{i=1}^{n} B_i . \quad (4.135)$$

Vì giá trị trung bình của các biến ngẫu nhiên Bernoulli là p, theo tính tuyến tính của kỳ vọng

$$E(X) = \sum_{i=1}^{n} E(B_i) = np. \quad (4.136)$$

Lưu ý rằng $E\left(B_i^2\right) = p$ và $E(B_iB_j) = p^2$ bởi sự độc lập, vì vậy

$$E\left(X^2\right) = E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} B_iB_j\right) \quad (4.137)$$

$$= \sum_{i=1}^{n} E\left(B_i^2\right) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E(B_iB_j) = np + n(n-1)p^2. \quad (4.138)$$

Poisson

Từ phép tính ta có

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}, \quad (4.139)$$

đó là khai triển chuỗi Taylor của hàm mũ. Điều này nghĩa là

$$E(X) = \sum_{k=1}^{\infty} k p_X(k) \quad (4.140)$$

$$= \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} \quad (4.141)$$

$$= e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^{m+1}}{m!} = \lambda, \quad (4.142)$$

Và

$$E(X^2) = \sum_{k=1}^{\infty} k^2 p_X(k) \tag{4.143}$$

$$= \sum_{k=1}^{\infty} \frac{k\lambda^k e^{-\lambda}}{(k-1)!} \tag{4.144}$$

$$= e^{-\lambda} \sum_{k=1}^{\infty} \frac{(k-1)\lambda^k}{(k-1)!} + \frac{\lambda^k}{(k-1)!} \tag{4.145}$$

$$= e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^{m+2}}{m!} + \sum_{m=1}^{\infty} \frac{\lambda^{m+1}}{m!} = \lambda^2 + \lambda. \tag{4.146}$$

Đồng phục

Chúng tôi áp dụng định nghĩa về giá trị kỳ vọng cho các biến ngẫu nhiên liên tục để có được

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx = \int_a^b \frac{x}{b-a} dx \tag{4.147}$$

$$= \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}. \tag{4.148}$$

Tương tự,

$$E(X^2) = \int_a^b \frac{x^2}{b-a} dx \tag{4.149}$$

$$= \frac{b^3 - a^3}{3(b-a)} \tag{4.150}$$

$$= \frac{a^2 + ab + b^2}{3}. \tag{4.151}$$

số mũ

Áp dụng tích phân từng phần,

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx \tag{4.152}$$

$$= \int_0^{\infty} x\lambda e^{-\lambda x} dx \tag{4.153}$$

$$= \left. x e^{-\lambda x} \right]_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda}. \tag{4.154}$$

Tương tự,

$$E(X^2) = \int_0^{\infty} x^2 \lambda e^{-\lambda x} dx \tag{4.155}$$

$$= \left. x^2 e^{-\lambda x} \right]_0^{\infty} + 2\int_0^{\infty} x e^{-\lambda x} dx = \frac{2}{\lambda^2}. \tag{4.156}$$

Gaussian

Ta áp dụng phép đổi biến $t = (x - \mu)/\sigma$.

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx \tag{4.157}$$

$$= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \tag{4.158}$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}}\, dt + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\, dt \tag{4.159}$$

$$= \mu, \tag{4.160}$$

trong đó bước cuối cùng xuất phát từ thực tế là tích phân của một hàm lẻ có giới hạn trên một khoảng đối xứng bằng không.

Áp dụng phép đổi biến $t = (x - \mu)/\sigma$ và lấy tích phân từng phần ta được

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx \tag{4.161}$$

$$= \int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \tag{4.162}$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}}\, dt + \frac{2\mu\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2}}\, dt + \frac{\mu^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\, dt \tag{4.163}$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \left(\left[-t e^{-\frac{t^2}{2}} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\, dt \right) + \mu^2 \tag{4.164}$$

$$= \sigma^2 + \mu^2. \tag{4.165}$$

4.5.2 Chuỗi hình học

Bổ đề 4.5.1. Với mọi $\alpha = 0$ và mọi số nguyên n_1 và n_2

$$\sum_{k=n_1}^{n_2} \alpha^k = \frac{\alpha^{n_1} - \alpha^{n_2+1}}{1-\alpha}. \tag{4.166}$$

Hệ quả 4.5.2. Nếu $0 < a < 1$

$$\sum_{k=0}^{\infty} \alpha^k = \frac{\alpha}{1-\alpha}. \tag{4.167}$$

Bằng chứng. Chúng ta chỉ cần nhân tổng với thừa số $(1-\alpha)/(1-\alpha)$ hiển nhiên bằng một,

$$\alpha^{n_1} + \alpha^{n_1+1} + \cdots + \alpha^{n_2 - 1} + \alpha^{n_2} = \frac{1-\alpha}{1-\alpha}\left(\alpha^{n_1} + \alpha^{n_1+1} + \cdots + \alpha^{n_2 - 1} + \alpha^{n_2}\right) \tag{4.168}$$

$$= \frac{\alpha^{n_1} - \alpha^{n_1+1} + \alpha^{n_1+1} + \cdots - \alpha^{n_2} + \alpha^{n_2} - \alpha^{n_2+1}}{1-\alpha}$$

$$= \frac{\alpha^{n_1} - \alpha^{n_2+1}}{1-\alpha} \tag{4.169}$$

□

Bổ đề 4.5.3. Với $0 < a < 1$

$$\sum_{k=1}^{\infty} k\alpha^k = \frac{\alpha}{(1-\alpha)^2}. \tag{4.170}$$

Bằng chứng. Theo Hệ quả 4.5.2,

$$\sum_{k=0}^{\infty} \alpha^k = \frac{1}{1-\alpha}. \tag{4.171}$$

Vì giới hạn bên trái hội tụ nên ta có thể phân biệt hai vế để thu được

$$\sum_{k=0}^{\infty} k\alpha^{k-1} = \frac{1}{(1-\alpha)^2}. \tag{4.172}$$

□

Bổ đề 4.5.4. Với $0 < a < 1$

$$\sum_{k=1}^{\infty} k^2\alpha^k = \frac{a(1+a)}{(1-\alpha)^3}. \tag{4.173}$$

Bằng chứng. Theo bổ đề 4.5.3,

$$\sum_{k=1}^{\infty} k\alpha^k = \frac{\alpha}{(1-\alpha)^2}. \tag{4.174}$$

Vì giới hạn bên trái hội tụ nên ta có thể phân biệt hai vế để thu được

$$\sum_{k=1}^{\infty} k^2\alpha^{k-1} = \frac{1+\alpha}{(1-\alpha)^3}. \tag{4.175}$$

□

4.5.3 Chứng minh Định lý 4.3.7

Nếu $E X^2 = 0$ thì $X = 0$ theo Hệ quả 4.2.13 $X = 0$ với xác suất là một, nghĩa là $E(XY) = 0$ và do đó đẳng thức đúng trong (4.65). Điều này cũng đúng nếu $EY^2 = 0$. Bây giờ giả sử rằng $E X^2 \neq 0$ và $EY^2 \neq 0$. Hãy xác định các hằng số $a = \sqrt{E(Y^2)}$ và $b = \sqrt{E(X^2)}$. Theo tuyến tính của kỳ vọng,

$$E(aX + bY)^2 = a^2 E X^2 + b^2 E Y^2 + 2abE(XY) \tag{4.176}$$

$$= 2E X^2 EY^2 + \sqrt{E(X^2)E(Y^2)}E(XY), \tag{4.177}$$

$$E(aX - bY)^2 = a^2 E X^2 + b^2 E Y^2 - 2abE(XY) \tag{4.178}$$

$$= 2E X^2 EY^2 - \sqrt{E(X^2)E(Y^2)}E(XY). \tag{4.179}$$

Kỳ vọng của một đại lượng không âm là khác không vì tích phân hoặc tổng của một đại lượng không âm là không âm. Do đó, vế trái của (4.176) và (4.178) không âm nên (B.117) và (B.118) đều không âm, điều này suy ra (4.65).

Hãy chứng minh (B.21) bằng cách chứng minh cả hai

hàm ý. ( ). Giả sử E(XY) = $\overline{\quad E (X2) E(Y 2)}$. Khi đó (B.117) bằng 0 nên

$$EE\ \overline{(X2)}X + E\ \overline{(X2)}Y^{2} = 0, \quad (4.180)$$

mà theo Hệ quả 4.2.13 có nghĩa là E $\overline{(Y 2)}$X = E $\overline{(X2)}$Y với xác suất là một.

2) ( ). Giả sử Y = $\frac{E(Y}{E(X2)}$ X. Sau đó, người ta có thể dễ dàng kiểm tra rằng (B.117) bằng 0, mà hàm ý E(XY) = $\overline{\quad E (X2) E(Y 2)}$.

Chứng minh của (B.22) gần như giống hệt (dùng (4.176) thay cho (B.117)).

Chương 5

Quy trình ngẫu nhiên

Các quy trình ngẫu nhiên, còn được gọi là quy trình ngẫu nhiên, được sử dụng để mô hình hóa các đại lượng không chắc chắn phát triển theo thời gian: quỹ đạo của một hạt, giá dầu, nhiệt độ ở New York, nợ quốc gia của Hoa Kỳ , v.v. chúng tôi giới thiệu một khung toán học cho phép suy luận theo xác suất về những đại lượng như vậy.

5.1 Định nghĩa

Chúng tôi biểu thị các quá trình ngẫu nhiên bằng cách sử dụng dấu ngã trên chữ hoa X. Đây không phải là ký hiệu chuẩn, nhưng chúng tôi muốn nhấn mạnh sự khác biệt với các biến ngẫu nhiên và vectơ ngẫu nhiên.
Chính thức, một quá trình ngẫu nhiên X là một hàm ánh xạ các phần tử trong không gian mẫu Ω thành các hàm có giá trị thực.

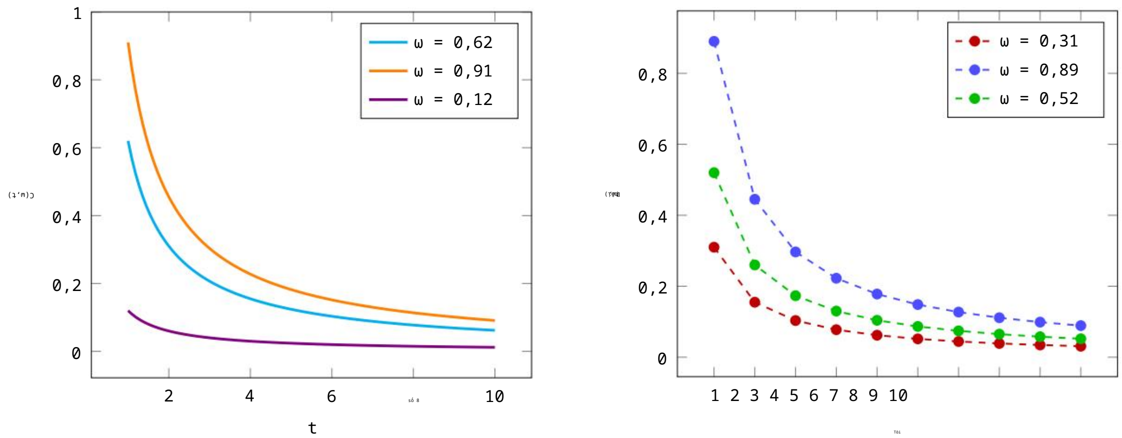
Định nghĩa 5.1.1 (Quá trình ngẫu nhiên). Cho một không gian xác suất (Ω, F, P), một quá trình ngẫu nhiên X là một hàm ánh xạ mỗi phần tử ω trong không gian mẫu Ω thành một hàm X (ω, ·) : T R, trong đó T là một biến rời rạc hoặc liên tục bộ.

Có hai cách giải thích cho X (ω, t):

- Nếu cố định ω thì X(ω, t) là một hàm xác định của t được gọi là một thực thể của quá trình ngẫu nhiên.
- Nếu cố định t thì X(ω, t) là biến ngẫu nhiên, thường ta chỉ ký hiệu là X(t).

Do đó, chúng ta có thể giải thích X là một tập hợp vô hạn các biến ngẫu nhiên được lập chỉ mục bởi t. Tập hợp các giá trị khả dĩ mà biến ngẫu nhiên X(t) có thể nhận với t cố định được gọi là không gian trạng thái của quá trình ngẫu nhiên. Các quá trình ngẫu nhiên có thể được phân loại theo biến đánh chỉ mục hoặc theo không gian trạng thái của chúng.

- Nếu biến chỉ số t được xác định trên R hoặc trên một khoảng bán vô hạn (t0, ∞) đối với một số t0 R thì X là quá trình ngẫu nhiên liên tục theo thời gian.
- Nếu biến chỉ số t được xác định trên một tập hợp rời rạc, thường là các số nguyên hoặc số tự nhiên, thì X là một quá trình ngẫu nhiên thời gian rời rạc. Trong những trường hợp như vậy, chúng tôi thường sử dụng một chữ cái khác với t, chẳng hạn như i, làm biến chỉ mục.



Hình 5.1: Hiện thực hóa quá trình ngẫu nhiên thời gian liên tục (trái) và thời gian rời rạc (phải) được định nghĩa trong Ví dụ 5.1.2.

- Nếu X(t) là biến ngẫu nhiên rời rạc với mọi t thì X là quá trình ngẫu nhiên trạng thái rời rạc. Nếu biến ngẫu nhiên rời rạc nhận một số hữu hạn các giá trị giống nhau với mọi t, thì X là một quá trình ngẫu nhiên hữu hạn trạng thái.

- Nếu X(t) là biến ngẫu nhiên liên tục với mọi t thì X là biến ngẫu nhiên trạng thái liên tục quá trình.

Lưu ý rằng có các quá trình ngẫu nhiên thời gian rời rạc trạng thái liên tục và các quá trình ngẫu nhiên thời gian liên tục trạng thái rời rạc. Bất kỳ sự kết hợp là có thể.

Không gian xác suất cơ bản (Ω, F, P) được đề cập trong định nghĩa xác định hoàn toàn hành vi ngẫu nhiên của quá trình ngẫu nhiên. Về nguyên tắc, chúng ta có thể chỉ định các quy trình ngẫu nhiên bằng cách xác định (1) không gian xác suất (Ω, F, P) và (2) ánh xạ gán hàm cho từng phần tử của Ω, như được minh họa trong ví dụ sau. Cách xác định các quá trình ngẫu nhiên này chỉ có thể xử lý được đối với các trường hợp rất đơn giản.

Ví dụ 5.1.2 (Vũng nước). Bob yêu cầu Mary lập mô hình xác suất. Khi vũng nước được hình thành, nó chứa một lượng nước được phân bố đều trong khoảng từ 0 đến 1 gallon.
Thời gian trôi qua, nước bốc hơi. Sau khoảng thời gian t thì lượng nước còn lại ít hơn t lần lượng nước ban đầu.

Mary lập mô hình nước trong vũng nước như một quá trình ngẫu nhiên thời gian liên tục ở trạng thái liên tục C. Không gian mẫu cơ sở là (0, 1), đại số σ là đại số Borel σ tương ứng (tất cả các hợp đếm được có thể có của các khoảng trong (0, 1)) và độ đo xác suất là độ đo xác suất thống nhất trên (0, 1). Đối với một phần tử cụ thể trong không gian mẫu ω (0, 1)

$$C(\omega, t) := \frac{\omega}{t}, \quad t \quad [1, \infty), \qquad (5.1)$$

trong đó đơn vị của t là ngày trong ví dụ này. Hình 6.1 cho thấy các hiện thực khác nhau của quá trình ngẫu nhiên. Mỗi nghiệm là một hàm xác định trên [1, ∞).

Bob chỉ ra rằng anh ấy chỉ quan tâm đến trạng thái của vũng nước mỗi ngày, chứ không phải bất cứ lúc nào t. Mary quyết định đơn giản hóa mô hình bằng cách sử dụng ngẫu nhiên thời gian rời rạc trạng thái liên tục quá trình D. Không gian xác suất cơ bản hoàn toàn giống như trước đây, nhưng chỉ số thời gian bây giờ là rời rạc. Đối với một phần tử cụ thể trong không gian mẫu ω (0, 1)

$$D(\omega, i) := \frac{\omega}{i}, \quad \text{tôi} = 1, 2, \ldots \tag{5.2}$$

Hình 6.1 cho thấy các hiện thực khác nhau của quá trình ngẫu nhiên liên tục. Lưu ý rằng mỗi lần thực hiện chỉ là một chuỗi rời rạc xác định.

Nhớ lại rằng giá trị của quá trình ngẫu nhiên tại một thời điểm cụ thể là một biến ngẫu nhiên. Do đó, chúng ta có thể mô tả hành vi của quá trình tại thời điểm đó bằng cách tính toán phân phối của biến ngẫu nhiên tương ứng. Tương tự, chúng ta có thể xem xét phân phối chung của quy trình được lấy mẫu tại n thời điểm cố định. Điều này được đưa ra bởi phân phối thứ n của ngẫu nhiên quá trình.

Định nghĩa 5.1.3 (phân phối bậc n). Phân phối cấp n của một quá trình ngẫu nhiên X là phân phối chung của các biến ngẫu nhiên X (t1), X (t2), . . . , X (tn) với n mẫu bất kỳ {t1, t2, . . . , tn} của chỉ số thời gian t.

Ví dụ 5.1.4 (Vũng nước (tiếp theo)). cdf bậc nhất của C(t) trong Ví dụ 5.1.2 là

$$F_{C(t)}(x) := P\left(C(t) \le x\right) \tag{5.3}$$

$$= P(\omega \le tx) \tag{5.4}$$

$$= \begin{cases} \int_{u=0}^{tx} du = tx & \text{nếu } 0 \le x \le \frac{1}{t}, \\ 1 & \text{nếu } x > \frac{1}{t}, \\ 0 & \text{nếu } x < 0. \end{cases} \tag{5.5}$$

Chúng tôi có được pdf bậc nhất bằng cách lấy vi phân.

$$f_{C(t)}(x) = \begin{cases} t & \text{nếu } 0 \le x \le \frac{1}{t}, \\ 0 & \text{nếu không thì.} \end{cases} \tag{5.6}$$

Nếu phân phối bậc n của một quá trình ngẫu nhiên là bất biến dịch chuyển, thì quá trình đó được gọi là dừng nghiêm ngặt hoặc mạnh.

Định nghĩa 5.1.5 (Quy trình dừng nghiêm ngặt/mạnh mẽ). Một quá trình là dừng theo nghĩa chặt chẽ hoặc mạnh mẽ nếu với mọi n ≥ 0 nếu chúng ta chọn n mẫu t1, t2, . . . , tn và độ dời τ bất kỳ của các biến ngẫu nhiên X (t1), X (t2), . . . , X (tn) có cùng phân phối khớp với X(t1 + τ), X(t2 + τ), . . . , X (tn + τ).

Các quá trình ngẫu nhiên trong Ví dụ 5.1.2 rõ ràng là không dừng hoàn toàn vì pdf và pmf bậc nhất của chúng không giống nhau tại mọi điểm. Một ví dụ quan trọng của các quá trình tĩnh nghiêm ngặt là các trình tự phân bố đồng nhất độc lập, được trình bày trong Phần 5.3.

Như trong trường hợp các biến ngẫu nhiên và véc-tơ ngẫu nhiên, việc xác định không gian xác suất cơ sở để xác định một quy trình ngẫu nhiên thường không thực tế lắm, ngoại trừ trường hợp rất đơn giản.

các trường hợp như trong Ví dụ 5.1.2. Lý do là rất khó để đưa ra một không gian xác suất làm phát sinh phân phối lãi suất bậc n cho trước. May mắn thay, chúng ta cũng có thể chỉ định một quá trình ngẫu nhiên bằng cách chỉ định trực tiếp phân phối bậc n của nó cho tất cả các giá trị của n = 1, 2, . . . Điều này hoàn toàn đặc trưng cho quá trình ngẫu nhiên. Hầu hết các quy trình ngẫu nhiên được mô tả trong chương này, ví dụ như các chuỗi độc lập giống hệt nhau, chuỗi Markov, quy trình Poisson và quy trình Gaussian, được chỉ định theo cách này.

Cuối cùng, các quá trình ngẫu nhiên cũng có thể được xác định bằng cách biểu diễn chúng dưới dạng hàm của các quá trình ngẫu nhiên khác. Một hàm $Y := g(X)$ của một quá trình ngẫu nhiên X cũng là một quá trình ngẫu nhiên, vì nó ánh xạ bất kỳ phần tử ω nào trong không gian mẫu Ω thành một hàm $Y(\omega, \cdot) := g(X(\omega, \cdot))$. Trong Phần 5.6 , chúng tôi xác định bước đi ngẫu nhiên theo cách này.

5.2 Hàm trung bình và tự hiệp phương sai

Như trong trường hợp biến ngẫu nhiên và vectơ ngẫu nhiên, toán tử kỳ vọng cho phép rút ra các đại lượng tóm tắt hành vi của quá trình ngẫu nhiên. Giá trị trung bình của vectơ ngẫu nhiên là giá trị trung bình của $X(t)$ tại bất kỳ thời điểm t cố định nào.

Định nghĩa 5.2.1 (Trung bình). Giá trị trung bình của một quá trình ngẫu nhiên là hàm

$$\mu_X(t) := E X(t) . \tag{5.7}$$

Lưu ý rằng giá trị trung bình là một hàm xác định của t. Hiệp phương sai của một quá trình ngẫu nhiên là một hàm xác định khác bằng hiệp phương sai của $X(t_1)$ và $X(t_2)$ đối với hai điểm t_1 và t_2 bất kỳ . Nếu chúng ta đặt $t_1 := t_2$, thì hiệp phương sai bằng với phương sai tại t_1.

Định nghĩa 5.2.2 (Tự hiệp phương sai). Hiệp phương sai của một quá trình ngẫu nhiên là hàm

$$R_X(t_1, t_2) := \operatorname{Cov}\left(X(t_1), X(t_2)\right) . \tag{5.8}$$

Đặc biệt,

$$R_X(t, t) := \operatorname{Var} X(t) . \tag{5.9}$$

Theo trực giác, hiệp phương sai định lượng mối tương quan giữa quá trình tại hai thời điểm khác nhau. Nếu mối tương quan này chỉ phụ thuộc vào khoảng cách giữa hai điểm, thì quá trình được gọi là dừng theo nghĩa rộng.

Định nghĩa 5.2.3 (Quá trình theo nghĩa rộng/cố định yếu). Một quá trình là dừng theo nghĩa rộng hoặc yếu nếu giá trị trung bình của nó không đổi

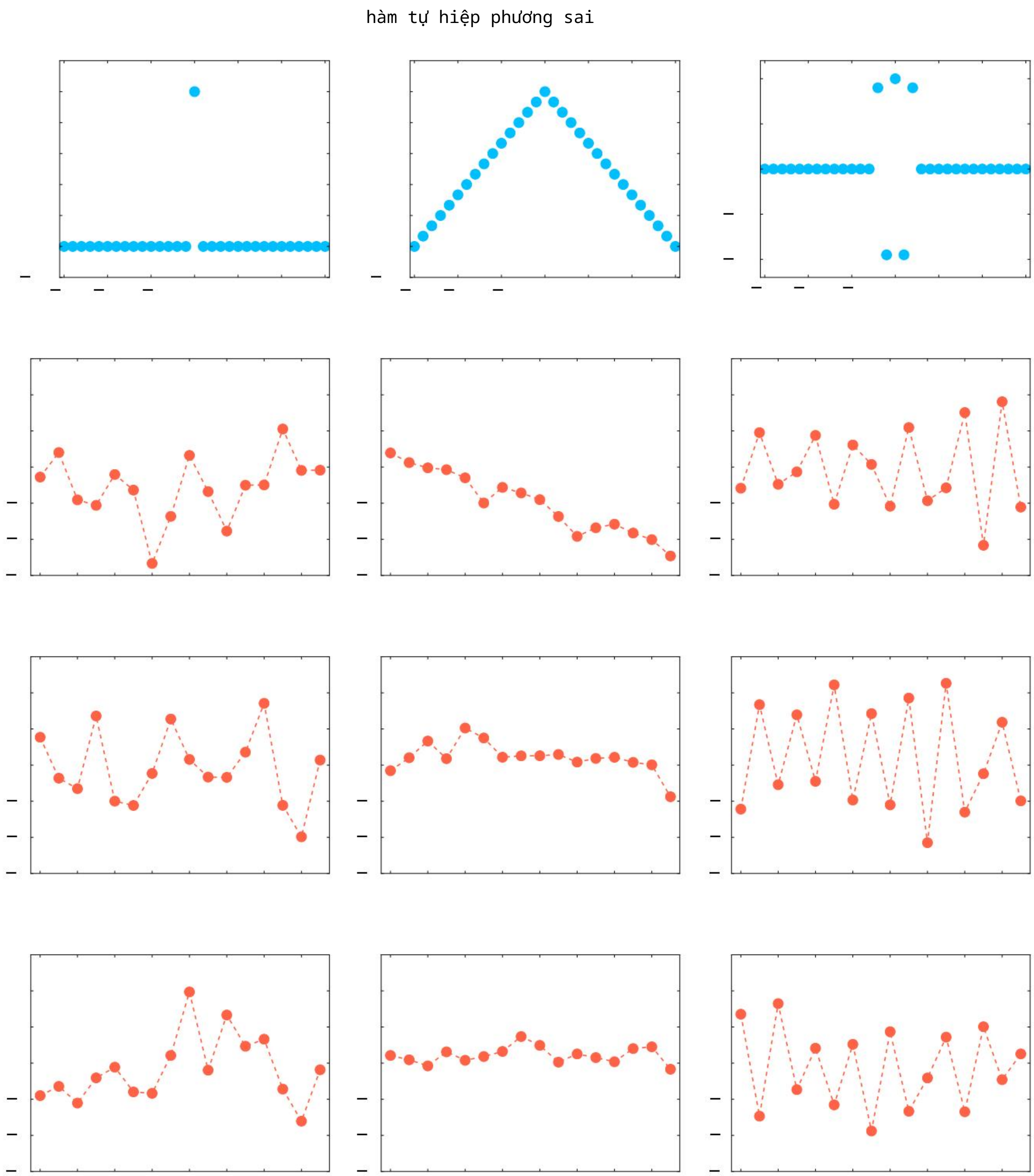
$$\mu_X(t) := \mu \tag{5.10}$$

và hàm autocovariance của nó là shift bất biến, tức là

$$R_X(t_1, t_2) := R_X(t_1 + \tau, t_2 + \tau) \tag{5.11}$$

cho bất kỳ t_1 và t_2 và bất kỳ dịch chuyển τ . Đối với các quá trình ổn định yếu, hiệp phương sai thường được biểu thị dưới dạng một hàm của sự khác biệt giữa hai điểm thời gian,

$$R_X(s) := R_X(t, t + s) \text{ với mọi } t. \tag{5.12}$$



Hình 5.2: Hiện thực hóa (ba hàng dưới cùng) của các quy trình Gauss với giá trị trung bình bằng 0 và các hàm tự đồng biến được hiển thị trên hàng trên cùng.

Machine Translated by Google

Lưu ý rằng bất kỳ quá trình dừng nghiêm ngặt nào nhất thiết phải dừng yếu vì các phân bố bậc nhất và bậc hai của nó là bất biến dịch chuyển.

Hình 5.2 cho thấy một số quá trình ngẫu nhiên cố định với các hàm tự hiệp phương sai khác nhau. Nếu hàm tự hiệp phương sai bằng không ở mọi nơi ngoại trừ tại điểm gốc, thì các giá trị của các quá trình ngẫu nhiên tại các điểm khác nhau là không tương quan. Điều này dẫn đến sự dao động thất thường. Khi hệ số tự động ở các thời điểm lân cận cao, quá trình ngẫu nhiên quỹ đạo trở nên mượt mà hơn. Tự tương quan cũng có thể tạo ra hành vi có cấu trúc hơn, như trong cột bên phải của hình. Trong ví dụ đó X(i) tương quan nghịch với hai hàng xóm X(i 1) và X(i+1), nhưng tương quan thuận với X(i 2) và X(i+2). Điều này dẫn đến biến động định kỳ nhanh chóng.

5.3 Các trình tự phân bố đồng nhất độc lập

Chuỗi (iid) phân phối đồng nhất độc lập X là một quá trình ngẫu nhiên thời gian rời rạc trong đó X(i) có cùng phân phối cho bất kỳ i cố định nào và X(i1), X (i2), . . . , X (in) độc lập lẫn nhau đối với n chỉ số cố định bất kỳ và n ≥ 2 bất kỳ . Nếu X(i) là một biến ngẫu nhiên rời rạc (hoặc tương đương không gian trạng thái của quá trình ngẫu nhiên là rời rạc), thì ta ký hiệu pmf liên quan đến phân phối của mỗi mục theo pX . Bản pdf này đặc trưng hoàn toàn cho quá trình ngẫu nhiên, vì với n chỉ số bất kỳ i1, i2, . . . , trong và bất kỳ n:

$$p_{X(i_1),X(i_2),...,X(i_n)}(x_{i_1}, x_{i_2}, \ldots, x_{i_n}) = \prod_{tôi=1}^{N} p_X(x_i). \tag{5.13}$$

Lưu ý rằng phân phối không thay đổi nếu chúng ta thay đổi mọi chỉ số theo cùng một lượng, vì vậy quá trình này hoàn toàn dừng.

Tương tự, nếu X(i) là biến ngẫu nhiên liên tục, thì ta ký hiệu pdf liên quan đến phân phối bởi fX . Với n chỉ số bất kỳ i1, i2, . . . , trong và bất kỳ n nào chúng ta có

$$f_{X(i_1),X(i_2),...,X(i_n)}(x_{i_1}, x_{i_2}, \ldots, x_{i_n}) = \prod_{tôi=1}^{N} f_X(x_i). \tag{5.14}$$

Hình 5.3 cho thấy một số nhận thức từ các chuỗi iid tuân theo phân bố hình học và đồng nhất.

Giá trị trung bình của một chuỗi ngẫu nhiên iid là hằng số và bằng giá trị trung bình của phân phối liên quan của nó, mà chúng ta biểu thị bằng μ,

$$\mu_X(i) := E X(i) \tag{5.15}$$

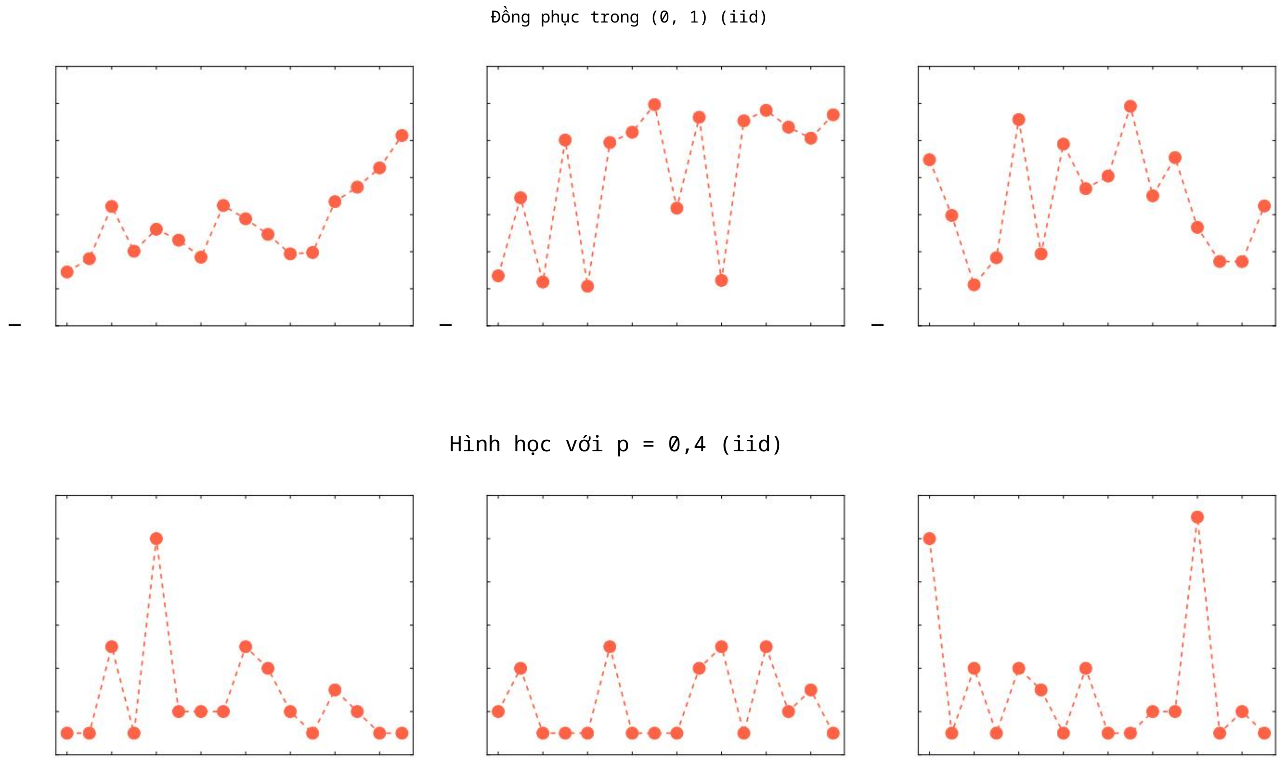
$$= \mu. \tag{5.16}$$

Chúng ta hãy biểu thị phương sai của phân phối liên quan đến chuỗi iid bằng σ hàm hiệp ^{2} . Tự động phương sai được cho bởi

$$R_X(i, j) := E X(i) X(j) \quad E X(i) E X(j) \tag{5.17}$$

$$= \begin{cases} \sigma^2, \\ 0. \end{cases} \tag{5.18}$$

Điều này không có gì ngạc nhiên, X(i) và X(j) độc lập với mọi i = j nên chúng cũng không tương quan với nhau.



Hình 5.3: Sự thể hiện của một dãy thống nhất iid trong (0, 1) (hàng đầu tiên) và một dãy hình học iid với tham số p = 0,4 (hàng thứ hai).

5.4 Quá trình Gaussian

Một quá trình ngẫu nhiên X là Gaussian nếu bất kỳ tập mẫu nào là một vectơ ngẫu nhiên Gaussian.

Một quy trình Gaussian X được đặc trưng đầy đủ bởi hàm trung bình μ_X và hàm tự biến của nó R_X. Với mọi $t_1, t_2, \ldots, t_n$ và bất kỳ $n \geq 1$, vectơ ngẫu nhiên

$$X := \begin{bmatrix} X(t_1) \\ X(t_2) \\ \cdots \\ X(t_n) \end{bmatrix} \tag{5.19}$$

là một vectơ ngẫu nhiên Gaussian có nghĩa là

$$\mu_X := \begin{bmatrix} \mu_X(t_1) \\ \mu_X(t_2) \\ \cdots \\ \mu_X(t_n) \end{bmatrix} \tag{5.20}$$

và ma trận hiệp phương sai

$$\Sigma_X := \begin{bmatrix} R_X(t_1, t_1) & R_X(t_1, t_2) & \cdots & R_X(t_1, t_n) \\ R_X(t_1, t_2) & R_X(t_2, t_2) & \cdots & R_X(t_2, t_n) \\ \vdots & \vdots & \ddots & \vdots \\ R_X(t_2, t_n) & R_X(t_2, t_n) & \cdots & R_X(t_n, t_n) \end{bmatrix} \tag{5.21}$$

Hình 5.2 cho thấy việc thực hiện một số quy trình Gaussian rời rạc với các hàm tự hiệp phương sai khác nhau. Lấy mẫu từ một quy trình ngẫu nhiên Gaussian rút gọn lại để lấy mẫu một vectơ ngẫu nhiên Gaussian với ma trận trung bình và hiệp phương sai thích hợp.

Thuật toán 5.4.1 (Tạo quá trình ngẫu nhiên Gaussian). Để lấy mẫu từ quy trình dom chạy Gaussian với hàm trung bình μ_X và hàm tự hiệp phương sai Σ_X tại n điểm $t_1, . . .$, chúng tôi:

1. Tính vectơ trung bình μ_X cho bởi (5.20) và ma trận hiệp phương sai Σ_X cho bởi (5.21).

2. Tạo n mẫu độc lập từ Gaussian tiêu chuẩn.

3. Tô màu các mẫu theo Σ_X và căn giữa chúng xung quanh μ_X , như được mô tả trong Algo rithm 4.3.15.

5.5 Quá trình Poisson

Trong ví dụ 2.2.8 , chúng ta thúc đẩy định nghĩa biến ngẫu nhiên Poisson bằng cách lấy phân phối của số lượng các sự kiện xảy ra trong một khoảng thời gian cố định trong các điều kiện sau:

1. Mỗi sự kiện xảy ra độc lập với mọi sự kiện khác.

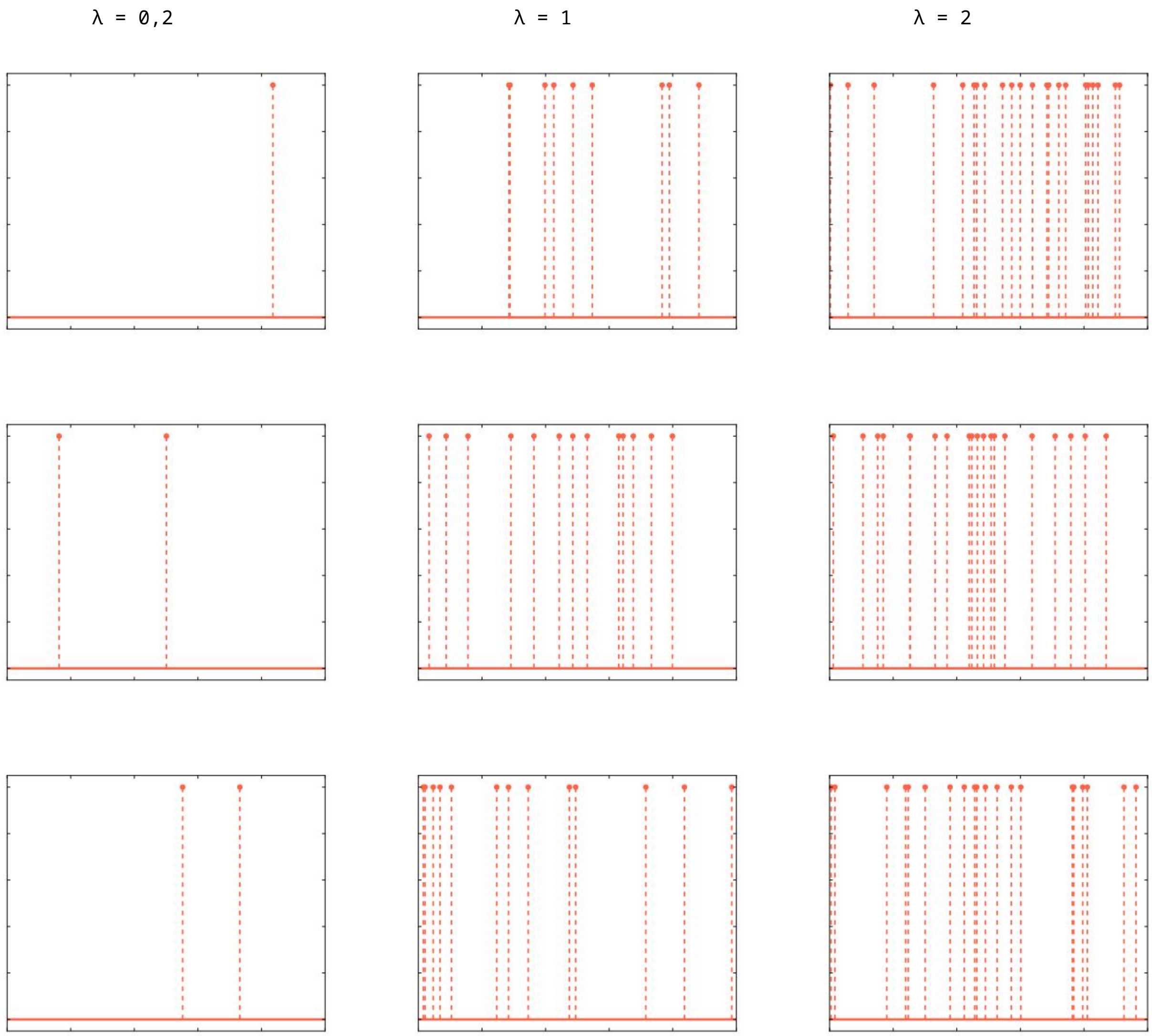
2. Các sự kiện xảy ra đồng nhất.

3. Các sự kiện xảy ra với tốc độ λ sự kiện trên mỗi khoảng thời gian.

Bây giờ chúng ta giả sử rằng các điều kiện này tồn tại trong khoảng bán vô hạn $[0, \infty)$ và xác định một quá trình ngẫu nhiên N đếm các sự kiện. Để rõ ràng $N(t)$ là số lượng các sự kiện xảy ra giữa 0 và t.

Bằng cách lập luận tương tự như trong Ví dụ 2.2.8, phân phối của biến ngẫu nhiên $N(t_2)$ $N(t_1)$, đại diện cho số biến cố xảy ra giữa t_1 và t_2, là một biến ngẫu nhiên Poisson với tham số $\lambda(t_2$ $t_1)$. Điều này giữ cho bất kỳ t_1 và t_2. Ngoài ra, các biến ngẫu nhiên $N(t_2)$ $N(t_1)$ và $N(t_4)$ $N(t_3)$ độc lập cũng như các khoảng $[t_1, t_2]$ và (t_3, t_4) không trùng nhau bởi Điều kiện 1. A Quá trình Poisson là một quá trình ngẫu nhiên liên tục trạng thái rời rạc thỏa mãn hai tính chất này.

Các quy trình Poisson thường được sử dụng để mô hình hóa các sự kiện như động đất, cuộc gọi điện thoại, sự phân rã của các hạt phóng xạ, gai thần kinh, v.v. Hình 2.6 cho thấy một ví dụ về một tình huống thực tế trong đó số lượng cuộc gọi nhận được tại một trung tâm cuộc gọi gần đúng như một Poisson quá trình (miễn là chúng tôi chỉ xem xét một vài giờ). Lưu ý rằng ở đây chúng tôi đang sử dụng từ sự kiện để chỉ một điều gì đó xảy ra, chẳng hạn như sự xuất hiện của một email, thay vì một tập hợp trong một không gian mẫu, nghĩa là nó thường có ở những nơi khác trong các ghi chú này.



Hình 5.4: Các sự kiện tương ứng với việc thực hiện quy trình Poisson N cho các giá trị khác nhau của tham số λ. N(t) bằng số biến cố tính đến thời điểm t.

Định nghĩa 5.5.1 (Quá trình Poisson). Một quá trình Poisson với tham số λ là một trạng thái rời rạc quá trình ngẫu nhiên liên tục N sao cho

1. $N(0)=0$.
2. Với mọi $t_1 < t_2 < t_3 < t_4$ $N(t_2)$ $N(t_1)$ là biến ngẫu nhiên Poisson với tham số $\lambda(t_2 \quad t_1)$.
3. Với mọi $t_1 < t_2 < t_3 < t_4$, các biến ngẫu nhiên $N(t_2)$ $N(t_1)$ và $N(t_4)$ $N(t_3)$ là độc lập.

Bây giờ chúng ta kiểm tra xem quá trình ngẫu nhiên có được xác định rõ ràng hay không, bằng cách chứng minh rằng chúng ta có thể lấy được pmf chung của N tại bất kỳ n điểm nào $t_1 < t_2 < \dots < t_n$ với mọi $n \geq 0$. Để giảm là nhẹ ký hiệu, đặt $p_{\tilde{\lambda}, x}$ giá trị pmf của biến ngẫu nhiên Poisson với tham số $\tilde{\lambda}$ tại x, nghĩa là

$$p_{\tilde{\lambda}, x} := \frac{\tilde{\lambda}^x e^{-\tilde{\lambda}}}{x!}. \tag{5.22}$$

Chúng ta có

$$p_{N(t_1),\dots,N(t_n)}(x_1, \dots, x_n) \tag{5.23}$$

$$= P\left(N(t_1) = x_1, \dots, N(t_n) = x_n\right) \tag{5.24}$$

$$= P\left(N(t_1) = x_1, N(t_2) - N(t_1) = x_2 - x_1, \dots, N(t_n) - N(t_{n-1}) = x_n - x_{n-1}\right) \tag{5.25}$$

$$= P\left(N(t_1) = x_1\right) P\left(N(t_2) - N(t_1) = x_2 - x_1\right) \dots P\left(N(t_n) - N(t_{n-1}) = x_n - x_{n-1}\right)$$

$$= p_{\lambda t_1, x_1}\, p_{\lambda(t_2 - t_1), x_2 - x_1} \dots p_{\lambda(t_n - t_{n-1}), x_n - x_{n-1}}. \tag{5.26}$$

Nói một cách đơn giản, chúng ta đã biểu thị sự kiện $N(t_i) = x_i$ với $1 \leq i \leq n$ theo các biến ngẫu nhiên $N(t_1)$ và $N(t_i) - N(t_{i-1})$, $2 \leq i \leq n$, là các biến ngẫu nhiên Poisson độc lập với các tham số lần lượt là λt_1 và $\lambda(t_i - t_{i-1})$.

Hình 5.4 cho thấy một số chuỗi sự kiện tương ứng với việc thực hiện quy trình Poisson N đối với các giá trị khác nhau của tham số λ (N(t) bằng số lượng sự kiện cho đến thời điểm t). Điều thú vị là thời gian đến của các sự kiện, tức là thời gian giữa các sự kiện liền kề nhau, luôn có cùng một phân phối: đó là một biến ngẫu nhiên hàm mũ.

Bổ đề 5.5.2 (Thời gian tới hạn của một quá trình Poisson là cấp số nhân). Đặt T biểu thị thời gian giữa hai sự kiện liền kề trong quy trình Poisson với tham số λ. T là biến ngẫu nhiên hàm mũ với tham số λ.

Bằng chứng nằm trong Mục 5.7.1 của phụ lục. Hình 2.11 cho thấy thời gian đến giữa các cuộc gọi điện thoại tại một trung tâm cuộc gọi thực sự được mô hình hóa theo cấp số nhân.

Bổ đề 5.5.2 gợi ý rằng để mô phỏng quá trình Poisson, tất cả những gì chúng ta cần làm là lấy mẫu từ một phân bố hàm mũ.

Thuật toán 5.5.3 (Tạo quá trình ngẫu nhiên Poisson). Để lấy mẫu từ một quá trình ngẫu nhiên Poisson với tham số λ, chúng tôi:

1. Tạo mẫu độc lập từ biến ngẫu nhiên hàm mũ với tham số λ t_1, t_2, t_3, . . .
2. Đặt các sự kiện của quá trình Poisson xảy ra tại t_1, $t_1 + t_2$, $t_1 + t_2 + t_3$, . . .

Hình 5.4 được tạo ra theo cách này. Để xác nhận rằng thuật toán cho phép lấy mẫu từ quy trình Poisson, chúng ta sẽ phải chứng minh rằng quy trình kết quả thỏa mãn các điều kiện trong Định nghĩa 5.5.1. Đây thực sự là trường hợp, nhưng chúng tôi bỏ qua bằng chứng.

Bổ đề sau đây suy ra các hàm trung bình và tự hiệp phương sai của một quá trình Poisson được chứng minh trong Mục 5.7.2.

Bổ đề 5.5.4 (Trung bình và tự hiệp phương sai của một quá trình Poisson). Giá trị trung bình và tự hiệp phương sai của một quá trình Poisson bằng

$$EX(t) = \lambda t, \tag{5.27}$$

$$R_X(t_1, t_2) = \lambda \min\{t_1, t_2\}. \tag{5.28}$$

Giá trị trung bình của quá trình Poisson không phải là hằng số và tự hiệp phương sai của nó không phải là bất biến dịch chuyển, vì vậy quá trình này không dừng nghiêm ngặt cũng như không cố định theo nghĩa rộng.

Ví dụ 5.5.5 (Động đất). Số trận động đất có cường độ ít nhất là 3 độ Richter xảy ra ở bán đảo San Francisco được mô hình hóa bằng quy trình Poisson với tham số 0,3 trận động đất/năm. Xác suất để không có trận động đất nào trong mười năm tới và sau đó có ít nhất một trận động đất trong hai mươi năm tiếp theo là bao nhiêu?

Chúng tôi xác định quy trình Poisson X với tham số 0,3 để mô hình hóa vấn đề. Số trận động đất trong 10 năm tới, tức là $X(10)$, là một biến ngẫu nhiên Poisson với tham số $0,3 \cdot 10 = 3$. Các trận động đất trong 20 năm sau, $X(30)\ X(10)$, là Poisson với tham số $0,3 \cdot 20 = 6$. Hai biến ngẫu nhiên là độc lập vì các khoảng không trùng nhau.

$$P_X(10) = 0, X(30) \geq 1 = P_X(10) = 0, X(30)\ X(10) \geq 1 \tag{5.29}$$

$$= P_X(10) = 0\ P_X(30)\ \ X(10) \geq 1 \tag{5.30}$$

$$= P_X(10) = 0\ \ 1\ P_X(30)\ X(10) = 0 \tag{5.31}$$

$$3 = e\ \ 1\ \ e^{6} = 4,97\ 10\ \ 2\ . \tag{5.32}$$

Xác suất là 4,97%.

5.6 Bước đi ngẫu nhiên

Bước đi ngẫu nhiên là một quá trình ngẫu nhiên trong thời gian rời rạc mô hình hóa một chuỗi các bước theo các hướng ngẫu nhiên. Để chỉ định chính thức một bước đi ngẫu nhiên, trước tiên chúng ta xác định một chuỗi iid các bước S sao cho

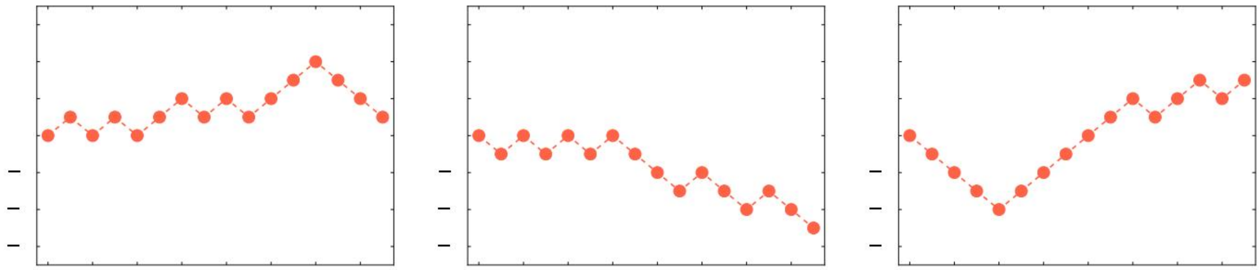
$$S(i) = \begin{cases} +1 & \text{với xác suất } \frac{1}{2} \\ 1 & \text{với xác suất } \frac{1}{2}. \end{cases} \tag{5.33}$$

Chúng tôi định nghĩa bước đi ngẫu nhiên X là quá trình ngẫu nhiên thời gian rời rạc ở trạng thái rời rạc

$$X(tôi) := \begin{cases} 0 & \text{với } i = 0, \\ \sum_{j=1} S(j) & \text{với } i = 1, 2, \ldots \end{cases} \tag{5.34}$$

Chúng tôi đã chỉ định X là một hàm của chuỗi iid, vì vậy nó được xác định rõ. Hình 5.5 cho thấy một số hiện thực của bước đi ngẫu nhiên.

X là đối xứng (có cùng xác suất thực hiện bước dương và bước âm) và bắt đầu tại gốc tọa độ. Thật dễ dàng để xác định các biến thể trong đó bước đi không đối xứng và bắt đầu



Hình 5.5: Thực hiện bước đi ngẫu nhiên được xác định trong Phần 5.5.

tại một điểm khác. Cũng có thể khái quát hóa các không gian có chiều cao hơn – chẳng hạn để mô hình hóa các quá trình ngẫu nhiên trên bề mặt 2D.

Chúng tôi rút ra pmf bậc nhất của bước đi ngẫu nhiên trong bổ đề sau, được chứng minh trong Mục 5.7.3 của phụ lục.

Bổ đề 5.6.1 (Pmf bậc nhất của bước đi ngẫu nhiên). PMf bậc nhất của bước đi ngẫu nhiên X là

$$p_{X(i)}(x) = \begin{cases} \binom{\text{tôi}}{\frac{\text{tôi}+x}{2}} \frac{1}{2^{\text{tôi}}} & \text{nếu } i + x \text{ chẵn và } i \le x \le i \\ 0 & \text{nếu không thì.} \end{cases} \tag{5.35}$$

Phân phối bậc nhất của bước đi ngẫu nhiên rõ ràng phụ thuộc vào thời gian, vì vậy quá trình ngẫu nhiên không hoàn toàn dừng. Theo bổ đề sau, giá trị trung bình của bước đi ngẫu nhiên là không đổi (nó bằng 0). Tuy nhiên, hiệp phương sai không phải là bất biến dịch chuyển, vì vậy quá trình này cũng không phải là ổn định yếu.

Bổ đề 5.6.2 (Trung bình và tự hiệp phương sai của bước đi ngẫu nhiên). Giá trị trung bình và hiệp phương sai của bước đi ngẫu nhiên X là

$$\mu_X(i) = 0, \tag{5.36}$$

$$R_X(i,j) = \min\{i,j\}. \tag{5.37}$$

Bằng chứng.

$$\mu_X(i) := E X(i) \tag{5.38}$$

$$= E\left(\sum_{j=1}^{\text{tôi}} S(j)\right) \tag{5.39}$$

$$= \sum_{j=1}^{\text{tôi}} E S(j) \quad \text{theo tuyến tính của kỳ vọng} \tag{5.40}$$

$$= 0. \tag{5.41}$$

$$R_X(i,j) := E\left(X(i)\,X(j)\right) - E\left(X(i)\right)E\left(X(j)\right) \tag{5.42}$$

$$= E\left(\sum_{k=1}^{i}\sum_{l=1}^{j} S(k)\,S(l)\right) \tag{5.43}$$

$$= E\left(\sum_{k=1}^{\text{tối thiểu}\{i,j\}} S(k)^2 + \sum_{k=1}^{i}\sum_{\substack{l=1\\ l\neq k}}^{j} S(k)\,S(l)\right) \tag{5.44}$$

$$= \sum_{k=1}^{\text{tối thiểu}\{i,j\}} 1 + \sum_{k=1}^{i}\sum_{\substack{l=1\\ l\neq k}}^{j} E\,S(k)\,E\,S(l) \tag{5.45}$$

$$= \text{tối thiểu}\{i,j\}, \tag{5.46}$$

trong đó (5.45) suy ra từ tuyến tính của kỳ vọng và tính độc lập. □

Phương sai của X tại i bằng $R_X(i,i) = i$, nghĩa là độ lệch chuẩn của bước đi ngẫu nhiên có tỷ lệ là $\sqrt{i}$.

Ví dụ 5.6.3 (Người đánh bạc). Một con bạc đang chơi trò chơi sau. Một đồng xu công bằng được lật tuần tự. Mỗi khi kết quả là mặt ngửa, con bạc sẽ thắng một đô la, mỗi khi nó rơi xuống mặt sấp, cô ấy sẽ mất một đô la. Chúng ta có thể lập mô hình số tiền mà con bạc kiếm được (hoặc bị mất) như một bước đi ngẫu nhiên, miễn là các lần lật là độc lập. Điều này cho phép chúng tôi ước tính rằng số tiền kiếm được kỳ vọng bằng 0 hoặc xác suất mà người đánh bạc kiếm được 6 đô la trở lên sau 10 lần lật đầu tiên là

$$P(\text{con bạc tăng \$6 trở lên}) = p_{X(10)}(6) + p_{X(10)}(8) + p_{X(10)}(10) \tag{5.47}$$

$$= \binom{10}{8}\frac{1}{2^{10}} + \binom{10}{9}\frac{1}{2^{10}} + \frac{1}{2^{10}} \tag{5.48}$$

$$= 5{,}47 \cdot 10^{-2}. \tag{5.49}$$

5.7 Bằng chứng

5.7.1 Chứng minh bổ đề 5.5.2

Chúng tôi bắt đầu bằng cách lấy cdf của T,

$$F_T(t) := P(T \leq t) \tag{5.50}$$

$$= 1 - P(T > t) \tag{5.51}$$

$$= 1 - P(\text{không có biến cố nào trong khoảng thời gian } t) \tag{5.52}$$

$$= 1 - e^{-\lambda t} \tag{5.53}$$

bởi vì số điểm trong một khoảng có độ dài t tuân theo phân bố Poisson với tham số λt. Phân biệt chúng tôi kết luận rằng

$$f_T(t) = \lambda e^{-\lambda t}. \tag{5.54}$$

5.7.2 Chứng minh bổ đề 5.5.4

Theo định nghĩa, số sự kiện giữa 0 và t được phân phối dưới dạng biến ngẫu nhiên Poisson với tham số λ t và do đó giá trị trung bình của nó bằng λ t.

Hiệp phương sai bằng

$$R_X(t_1, t_2) := E X(t_1) X(t_2) \quad E X(t_1) E X(t_2) \tag{5.55}$$

$$= E X(t_1) X(t_2) \quad \lambda^2 t_1 t_2. \tag{5.56}$$

Theo giả thiết X (t1) và X (t2) X (t1) độc lập sao cho

$$E X(t_1) X(t_2) = E X(t_1) X(t_2) \quad X(t_1) + X(t_1)^2 \tag{5.57}$$

$$= E X(t_1) E X(t_2) \quad X(t_1) + E X(t_1)^2 \tag{5.58}$$

$$= \lambda^2 t_1 (t_2 \quad t_1) + \lambda t_1 + \lambda^2 t_1^2 \tag{5.59}$$

$$= \lambda^2 t_1 t_2 + \lambda t_1. \tag{5.60}$$

5.7.3 Chứng minh bổ đề 5.6.1

Hãy để chúng tôi xác định số bước dương S+ mà bước đi ngẫu nhiên thực hiện. Với các giả định về S, đây là một biến ngẫu nhiên nhị thức với các tham số i và 1/2. Số bước âm là S := i S+. Để X(i) bằng x, chúng ta cần có số bước ròng bằng x, nghĩa là

$$x = S_+ \quad S \tag{5.61}$$

$$= 2S_+ \quad i. \tag{5.62}$$

Điều này có nghĩa là S+ phải bằng $\frac{tôi+x}{2}$. Chúng tôi kết luận rằng

$$p_{X(i)}(i) = P\left(\sum_{j=0}^{tôi} S(i) = x\right) \tag{5.63}$$

$$= \binom{tôi}{\frac{tôi+x}{2}} \frac{1}{2^{tôi}} \quad \text{nếu } \frac{i+x}{2} \text{ là số nguyên trong khoảng từ 0 đến i.} \tag{5.64}$$

Chương 6

Hội tụ các quá trình ngẫu nhiên

Trong chương này chúng ta nghiên cứu sự hội tụ của các quá trình ngẫu nhiên rời rạc. Điều này cho phép mô tả hai hiện tượng cơ bản trong ước lượng thống kê và mô hình xác suất: định luật số lớn và định lý giới hạn trung tâm.

6.1 Các loại hội tụ

Ta hãy nhắc lại nhanh khái niệm về sự hội tụ của dãy xác định các số thực x1, x2, . . .Chúng ta có

$$\lim_{i \to \infty} x_i = x \tag{6.1}$$

nếu xi tùy ý gần với x khi chỉ số i tăng. Chính thức hơn, dãy hội tụ đến x nếu với mọi > 0 tồn tại một chỉ số i0 sao cho với mọi chỉ số i lớn hơn i0 ta có |xi x| < .

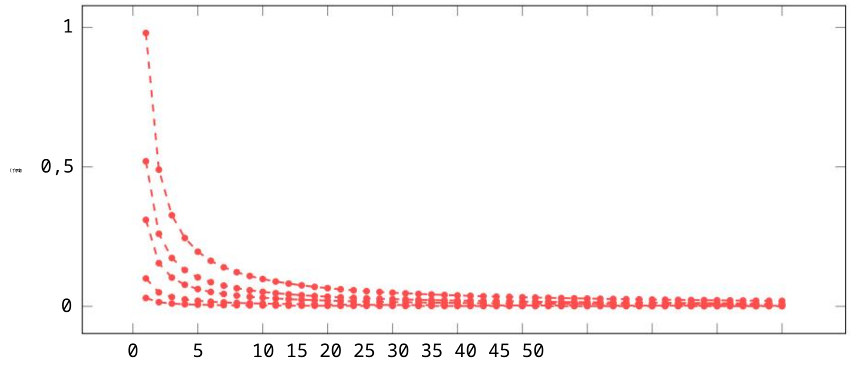
Hãy nhớ lại rằng bất kỳ sự thực hiện nào của quá trình ngẫu nhiên thời gian rời rạc X (ω, i) trong đó chúng ta cố định kết quả ω là một chuỗi xác định. Do đó, có thể đạt được sự hội tụ của các hiện thực như vậy với một số cố định bằng cách tính toán giới hạn tương ứng. Tuy nhiên, nếu chúng ta xem xét chính quá trình ngẫu nhiên thay vì một hiện thực và chúng ta muốn xác định xem cuối cùng nó có hội tụ với một biến ngẫu nhiên X hay không, thì sự hội tụ tất định không còn có ý nghĩa nữa. Trong phần này, chúng tôi mô tả một số định nghĩa thay thế về sự hội tụ, cho phép mở rộng khái niệm này sang các đại lượng ngẫu nhiên.

6.1.1 Hội tụ với xác suất một

Xét một quá trình ngẫu nhiên rời rạc X và một biến ngẫu nhiên X xác định trên cùng một không gian xác suất. Nếu ta cố định một phần tử ω của không gian mẫu Ω thì X(i, ω) là dãy xác định và X(ω) là hằng số. Do đó, có thể xác minh xem X (i, ω) có hội tụ tối thiểu đến X (ω) hay không khi i ∞ đối với giá trị cụ thể đó của ω. Trên thực tế, chúng ta có thể đặt câu hỏi: xác suất để điều này xảy ra là bao nhiêu? Nói chính xác, đây sẽ là xác suất mà nếu chúng ta vẽ ω chúng ta có

$$\lim_{i \to \infty} X(i, \omega) = X(\omega). \tag{6.2}$$

Nếu xác suất này bằng 1 thì ta nói rằng X(i) hội tụ đến X với xác suất bằng 1.



Hình 6.1: Sự hội tụ về 0 của quá trình ngẫu nhiên rời rạc D được định nghĩa trong Ví dụ 5.1.2.

Định nghĩa 6.1.1 (Hội tụ với xác suất một). Một vectơ ngẫu nhiên rời rạc X hội tụ với xác suất một đến một biến ngẫu nhiên X thuộc cùng một không gian xác suất (Ω, F, P) nếu

$$P\left(\omega \mid \omega \in \Omega, \lim_{i \to \infty} X(\omega, i) = X(\omega)\right) = 1. \tag{6.3}$$

Nhớ lại rằng nói chung không gian mẫu Ω rất khó xác định và thao tác một cách tường minh, ngoại trừ những trường hợp rất đơn giản.

Ví dụ 6.1.2 (Puddle (tiếp theo từ Ví dụ 5.1.2)). Chúng ta hãy xem xét quá trình ngẫu nhiên rời rạc D được định nghĩa trong Ví dụ 5.1.2. Nếu chúng ta cố định $\omega \in (0, 1)$

$$\lim_{i \to \infty} D(\omega, i) = \lim_{i \to \infty} \frac{\omega}{i} \tag{6.4}$$

$$= 0. \tag{6.5}$$

Hóa ra các nhận thức có xu hướng bằng 0 đối với tất cả các giá trị có thể có của ω trong không gian mẫu. Điều này ngụ ý rằng D hội tụ về 0 với xác suất là 1.

6.1.2 Hội tụ theo bình phương trung bình và theo xác suất

Để xác minh sự hội tụ với xác suất, chúng tôi cố định kết quả ω và kiểm tra xem các thực hiện tương ứng của quá trình ngẫu nhiên có hội tụ một cách xác định hay không. Một quan điểm khác là cố định biến chỉ số i và xem xét biến ngẫu nhiên X(i) gần như thế nào với một biến ngẫu nhiên X khác khi chúng ta tăng i.

Một thước đo khả dĩ về khoảng cách giữa hai biến ngẫu nhiên là bình phương trung bình của sự khác biệt của chúng. Nếu $E((X - Y)^2) = 0$ thì X = Y với xác suất bằng bất đẳng thức Chebyshev.
Độ lệch bình phương trung bình giữa X(i) và X là một đại lượng xác định (một số), vì vậy chúng ta có thể đánh giá sự hội tụ của nó là $i \to \infty$. Nếu nó hội tụ về 0 thì ta nói dãy ngẫu nhiên hội tụ bình phương trung bình.

Định nghĩa 6.1.3 (Sự hội tụ trong bình phương trung bình). Một quá trình ngẫu nhiên rời rạc X hội tụ trong bình phương trung bình của một biến ngẫu nhiên X thuộc cùng một không gian xác suất nếu

$$\lim_{i \infty} EX \quad X(i) \qquad ^2 = 0. \tag{6.6}$$

Ngoài ra, ta có thể xét xác suất để X(i) cách X bởi một cố định nào đó > 0. Nếu với bất kỳ , dù nhỏ đến đâu, xác suất này hội tụ về 0 khi i ∞ thì ta nói dãy ngẫu nhiên hội tụ trong xác suất.

Định nghĩa 6.1.4 (Sự hội tụ trong xác suất). Một quá trình ngẫu nhiên rời rạc X hội tụ khả năng xác suất với một biến ngẫu nhiên X khác thuộc cùng một không gian xác suất nếu với mọi > 0

$$\lim_{i \infty} PX \quad X(i) > = 0. \tag{6.7}$$

Lưu ý rằng như trong trường hợp hội tụ theo bình phương trung bình, giới hạn trong định nghĩa này là xác định, vì nó là giới hạn của xác suất, vốn chỉ là các số thực.

Như một hệ quả trực tiếp của bất đẳng thức Markov, sự hội tụ trong bình phương trung bình bao hàm sự hội tụ trong xác suất.

Định lý 6.1.5. Sự hội tụ trong bình phương trung bình ngụ ý sự hội tụ trong xác suất.

Bằng chứng. Chúng ta có

$$\lim_{i \infty} PX \quad X(i) > = \lim_{\text{tôi} \infty} PX \quad X(i) \qquad ^2 > \; ^2 \tag{6.8}$$

$$\leq \lim_{\text{tôi} \infty} \frac{EX \quad X(i) \qquad ^2}{^2} \quad \text{bằng bất đẳng thức Markov} \tag{6.9}$$

$$= 0, \tag{6.10}$$

nếu dãy hội tụ theo bình phương trung bình. □

Hóa ra sự hội tụ theo xác suất cũng bao hàm sự hội tụ theo xác suất. Hội tụ về xác suất không có nghĩa là hội tụ theo bình phương trung bình hoặc ngược lại. Sự khác biệt giữa ba loại hội tụ này không quan trọng lắm đối với mục đích của khóa học này.

6.1.3 Hội tụ trong phân phối

Trong một số trường hợp, một quá trình ngẫu nhiên X˜ không hội tụ đến giá trị của bất kỳ biến ngẫu nhiên nào, nhưng cdf của X˜(i) hội tụ điểm đến cdf của biến ngẫu nhiên X khác. Trong trường hợp đó, các giá trị thực của X˜ (i) và X không nhất thiết gần nhau nhưng trong giới hạn chúng có cùng phân phối. Trong trường hợp này, ta nói rằng X hội tụ trong phân phối của X.

Định nghĩa 6.1.6 (Sự hội tụ trong phân phối). Một quá trình ngẫu nhiên X hội tụ trong phân phối tới một biến ngẫu nhiên X thuộc cùng một không gian xác suất nếu

$$\lim_{i \infty} F_{X(i)}(x) = F_X(x) \tag{6.11}$$

với mọi x R trong đó FX liên tục.

Lưu ý rằng sự hội tụ trong phân phối là một khái niệm yếu hơn nhiều so với sự hội tụ với xác suất một, theo bình phương trung bình hoặc theo xác suất. Nếu một quá trình ngẫu nhiên rời rạc X hội tụ về một biến ngẫu nhiên X trong phân phối, điều này chỉ có nghĩa là khi i trở nên lớn thì phân phối của X (i) có xu hướng về phân phối của X, chứ không phải giá trị của hai biến ngẫu nhiên gần nhau. Tuy nhiên, hội tụ theo xác suất (và do đó hội tụ với xác suất một hoặc bình phương trung bình) không hàm ý hội tụ trong phân phối.

Ví dụ 6.1.7 (Định thức hội tụ Poisson). Chúng ta hãy định nghĩa một quá trình ngẫu nhiên rời rạc X(i) sao cho phân phối của X(i) là nhị thức với các tham số i và p := λ/i. X(i) và X(j) là độc lập với i = j, đặc trưng hoàn toàn cho phân bố cấp n của quá trình với mọi n > 1. Xét biến ngẫu nhiên Poisson X với tham số λ độc lập với X˜ (i) cho mọi i. Bạn có mong đợi các giá trị của X và X˜(i) gần nhau khi i ∞ không?

KHÔNG! Trên thực tế thậm chí X˜(i) và X˜(i+1) nói chung sẽ không gần nhau. Tuy nhiên, X˜ hội tụ trong phân phối tới X, như được thiết lập trong Ví dụ 2.2.8:

$$\lim_{i \to \infty} p_{X(i)}(x) = \lim_{tôi \to \infty} \binom{tôi}{x} p^x (1-p)^{(i-x)} \tag{6.12}$$

$$= \frac{\lambda^x e^{-\lambda}}{x!} \tag{6.13}$$

$$= p_X(x). \tag{6.14}$$

6.2 Luật số lớn

Hãy để chúng tôi xác định mức trung bình của một quá trình ngẫu nhiên rời rạc.

Định nghĩa 6.2.1 (Trung bình động). Trung bình động hoặc chạy A của một quá trình ngẫu nhiên rời rạc X, được xác định cho i = 1, 2, . . . (tức là 1 là điểm bắt đầu), bằng

$$A(i) := \frac{1}{tôi} \sum_{j=1}^{tôi} X(j). \tag{6.15}$$

Hãy xem xét một chuỗi iid. Một cách giải thích rất tự nhiên cho đường trung bình động là nó là ước tính thời gian thực của giá trị trung bình. Trên thực tế, theo thuật ngữ thống kê, trung bình động là giá trị trung bình mẫu của quá trình đến thời điểm i (giá trị trung bình mẫu được định nghĩa trong Chương 8). Định luật số lớn chứng minh rằng giá trị trung bình thực sự hội tụ về giá trị trung bình của chuỗi iid.

Định lý 6.2.2 (Luật số lớn yếu). Cho X là một quá trình ngẫu nhiên rời rạc iid có bị chặn. Khi đó trung bình A của nghĩa là $\mu_X := \mu$ sao cho phương sai của X(i) σ theo bình ^{2} X hội tụ phương trung bình của μ.

Bằng chứng. Đầu tiên, chúng ta thiết lập rằng giá trị trung bình của A(i) là hằng số và bằng μ,

$$E A(i) = E\left(\frac{1}{i}\sum_{j=1}^{i} X(j)\right) \tag{6.16}$$

$$= \frac{1}{i}\sum_{j=1}^{i} E X(j) \tag{6.17}$$

$$= \mu. \tag{6.18}$$

Do giả định độc lập, phương sai quy mô tuyến tính trong i. Nhớ lại rằng đối với các biến ngẫu nhiên độc lập, phương sai của tổng bằng tổng của các phương sai,

$$\text{Biến } A(i) = \text{Biến}\left(\frac{1}{i}\sum_{j=1}^{i} X(j)\right) \tag{6.19}$$

$$= \frac{1}{i^2}\sum_{j=1}^{i} \text{Biến } X(j) \tag{6.20}$$

$$= \frac{\sigma^2}{i}. \tag{6.21}$$

Chúng tôi kết luận rằng

$$\lim_{i\to\infty} E\left(\left(A(i) - \mu\right)^2\right) = \lim_{i\to\infty} E\left(\left(A(i) - E A(i)\right)^2\right) \quad \text{bởi (6.18)} \tag{6.22}$$

$$= \lim_{i\to\infty} \text{Biến } A(i) \tag{6.23}$$

$$= \lim_{i\to\infty} \frac{\sigma^2}{i} \quad \text{bởi (6.21)} \tag{6.24}$$

$$= 0. \tag{6.25}$$

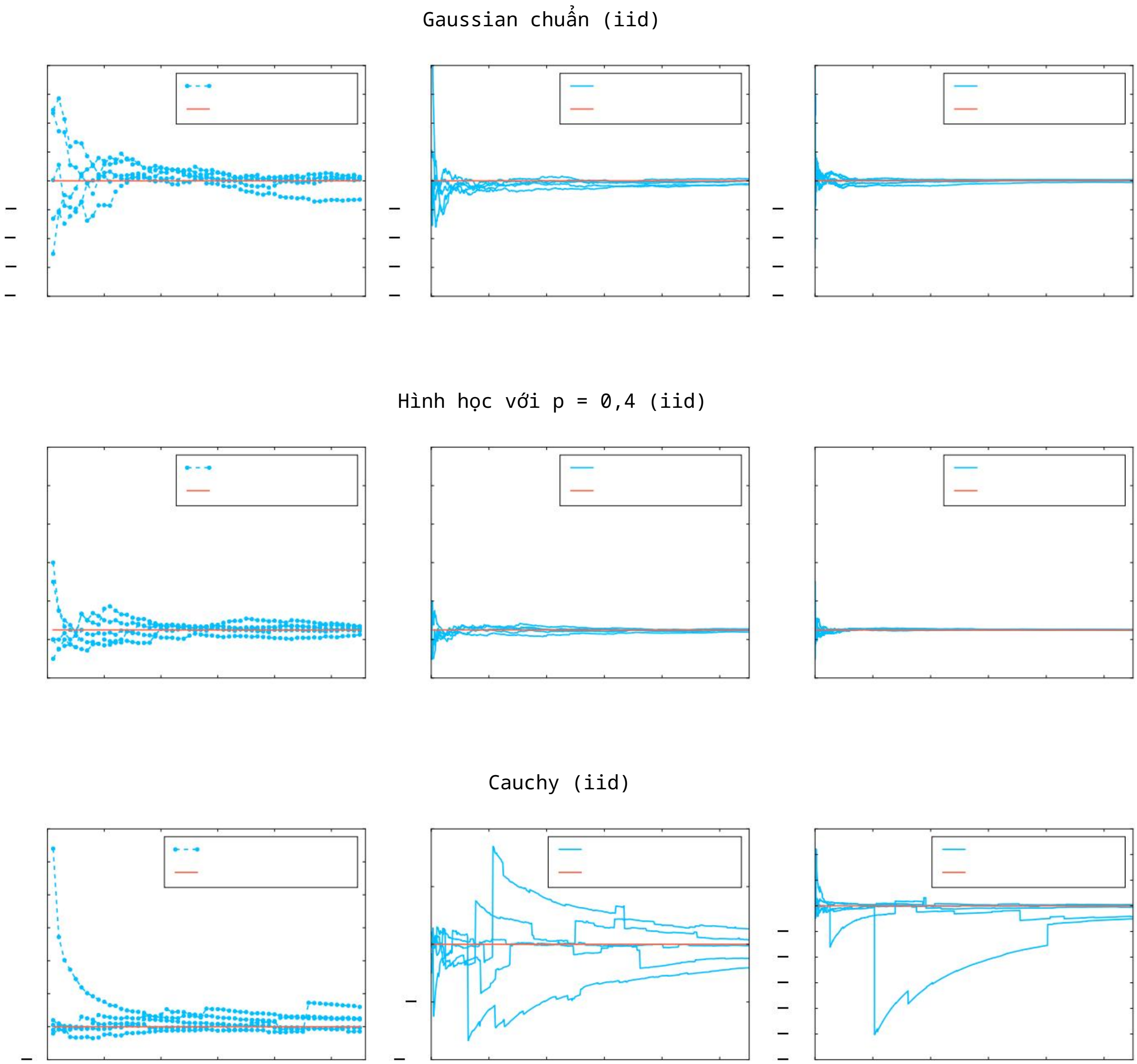
□

Theo Định lý 6.1.5, giá trị trung bình cũng hội tụ về giá trị trung bình của chuỗi iid theo xác suất. Trên thực tế, người ta cũng có thể chứng minh sự hội tụ với xác suất bằng giả thiết tương tự. Kết quả này được gọi là luật mạnh của số lớn, nhưng bằng chứng nằm ngoài phạm vi của những ghi chú này. Chúng tôi giới thiệu người đọc quan tâm đến các văn bản nâng cao hơn về lý thuyết xác suất.

Hình 6.2 cho thấy mức trung bình của việc thực hiện một số trình tự iid. Khi chuỗi iid là Gaussian hoặc hình học, chúng ta quan sát thấy sự hội tụ với giá trị trung bình của phân phối, tuy nhiên khi chuỗi là Cauchy thì trung bình động phân kỳ . Lý do là, như trong Ví dụ 4.2.2, phân phối Cauchy không có giá trị trung bình được xác định rõ ràng! Theo trực giác, các giá trị cực trị có xác suất không thể bỏ qua theo phân phối Cauchy, do đó, theo thời gian, chuỗi iid nhận các giá trị có độ lớn rất lớn và điều này ngăn đường trung bình động hội tụ.

6.3 Định lý giới hạn trung tâm

Trong phần trước, chúng ta đã xác định rằng trung bình động của một chuỗi các biến ngẫu nhiên iid hội tụ về giá trị trung bình của phân phối của chúng (miễn là giá trị trung bình được xác định rõ và



Hình 6.2: Hiện thực đường trung bình động của dãy Gauss chuẩn iid (trên), dãy hình học iid với tham số p = 0,4 (giữa) và dãy Cauchy iid (dưới).

phương sai là hữu hạn). Trong phần này, chúng tôi mô tả đặc điểm phân phối của trung bình A(i) khi i tăng. Nó chỉ ra rằng A hội tụ tới một biến ngẫu nhiên Gaussian trong phân phối, điều này rất hữu ích trong thống kê như chúng ta sẽ thấy sau này.

Kết quả này, được gọi là định lý giới hạn trung tâm, biện minh cho việc sử dụng phân phối Gaussian để mô hình hóa dữ liệu là kết quả của nhiều yếu tố độc lập khác nhau. Ví dụ, sự phân bố chiều cao hoặc cân nặng của những người trong một quần thể nhất định thường có dạng Gaussian– như minh họa trong Hình 2.13– bởi vì chiều cao và cân nặng của một người phụ thuộc vào nhiều yếu tố khác nhau và gần như độc lập với nhau. Trong nhiều ứng dụng xử lý tín hiệu, nhiễu được mô hình hóa tốt khi có phân phối Gaussian vì lý do tương tự.

Định lý 6.3.1 (Định lý giới hạn trung tâm). Cho X là một iid quá trình ngẫu nhiên rời rạc với trung bình $\mu_X := \mu$ sao cho phương sai của X(i) σ^2 bị chặn. Quá trình ngẫu nhiên $\sqrt{n}(\overline{A} - \mu)$, tương ứng với trung bình trượt định tâm và chia tỷ lệ của X, hội tụ trong phân phối tới một biến ngẫu nhiên Gauss với trung bình 0 và phương sai σ^2.

Bằng chứng. Bằng chứng về kết quả đáng chú ý này nằm ngoài phạm vi của những ghi chú này. Nó có thể được tìm thấy trong bất kỳ văn bản nâng cao nào về lý thuyết xác suất. Tuy nhiên, chúng tôi vẫn muốn cung cấp một số trực giác về lý do tại sao định lý đúng. Định lý 3.5.2 chứng minh rằng pdf của tổng hai biến ngẫu nhiên độc lập bằng tích chập của pdf riêng lẻ của chúng. Điều tương tự cũng xảy ra đối với các biến ngẫu nhiên rời rạc: pmf của tổng bằng với tích chập của pmfs, miễn là các biến ngẫu nhiên là độc lập.

Nếu mỗi phần tử của dãy iid có pdf f, thì pdf của tổng i phần tử đầu tiên có thể thu được bằng cách tổ hợp f với chính nó i lần

$$f_{\sum_{j=1}^{i} X(j)}(x) = (f * f * \cdots * f)(x). \tag{6.26}$$

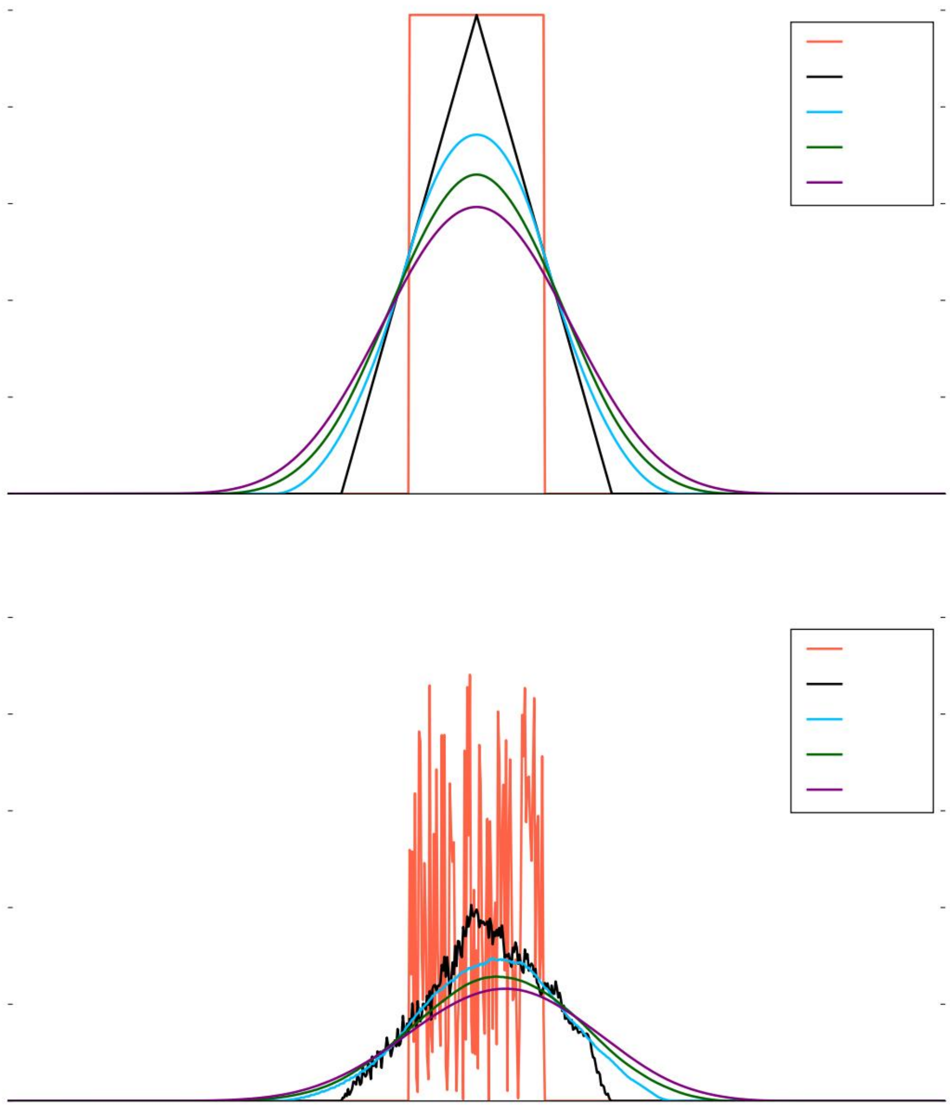
Nếu dãy có trạng thái rời rạc và mỗi phần tử có pmf p, pmf của tổng i phần tử đầu tiên có thể thu được bằng cách xoắn p với chính nó i lần

$$p_{\sum_{j=1}^{i} X(j)}(x) = (p * p * \cdots * p)(x). \tag{6.27}$$

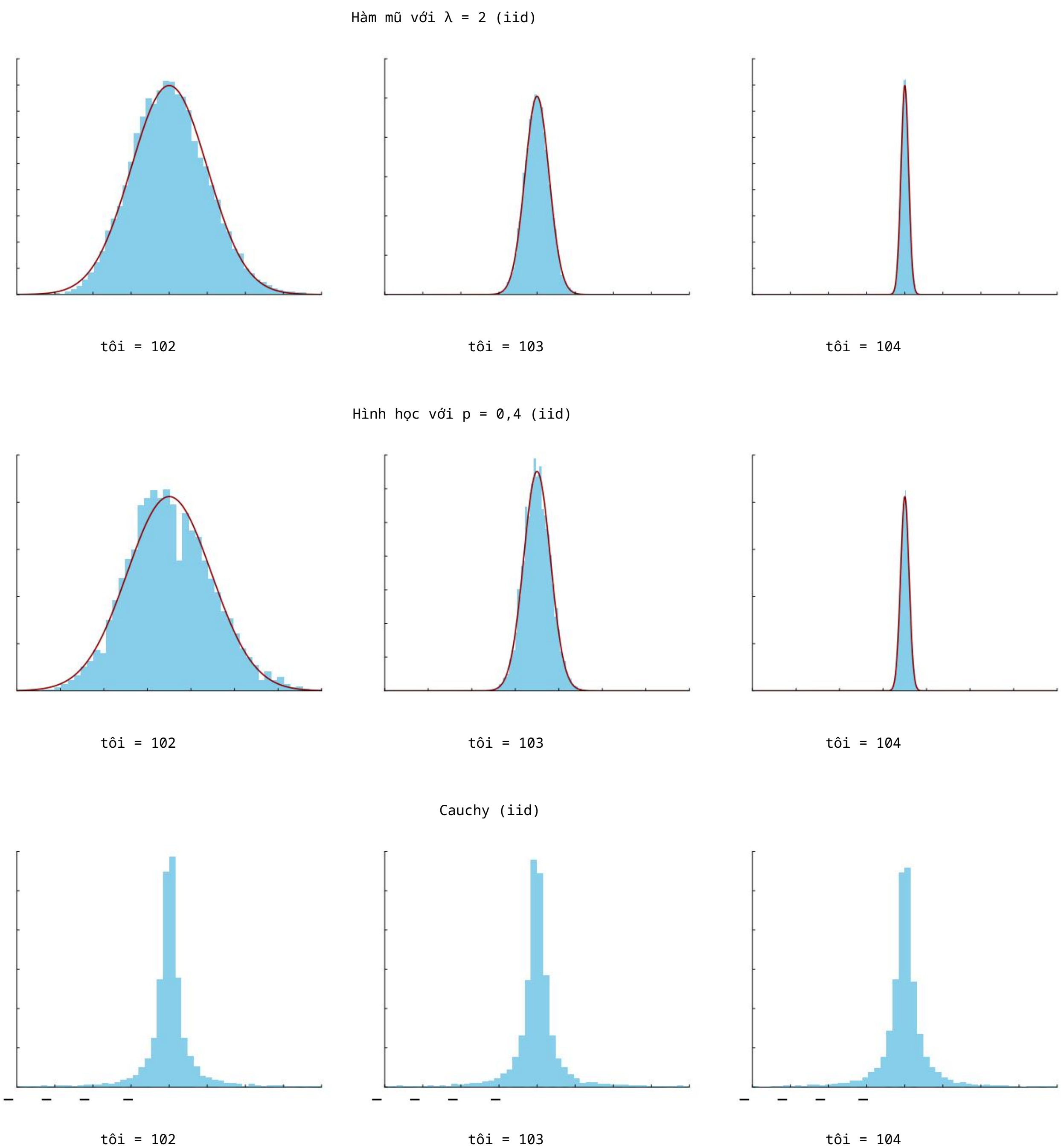
Chuẩn hóa bởi i chỉ dẫn đến việc chia tỷ lệ kết quả của tích chập, do đó, pmf hoặc pdf của trung bình di chuyển A là kết quả của các tích chập lặp đi lặp lại của một hàm cố định. Các kết cấu này có tác dụng làm mịn, cuối cùng biến đổi pmf/pdf thành Gaussian! Chúng tôi chỉ ra điều này bằng số trong Hình 6.3 cho hai phân phối rất khác nhau: phân phối đồng đều và phân phối rất bất thường. Cả hai đều hội tụ về hình dạng giống như Gaussian chỉ sau 3 hoặc 4 lần tích chập. Định lý giới hạn trung tâm làm cho điều này trở nên chính xác, chứng minh rằng hình dạng của pmf hoặc pdf trở thành tiệm cận Gaussian. □

Trong thống kê, định lý giới hạn trung tâm thường được viện dẫn để biện minh cho việc xử lý các giá trị trung bình như thể chúng có phân phối Gaussian. Ý tưởng là đối với n đủ lớn $\sqrt{n}(\overline{A} - \mu)$ xấp xỉ Gaussian với trung bình 0 và phương sai σ^2 ngụ ý rằng A xấp xỉ Gaussian với trung bình μ và , phương sai σ^2/n. Điều quan trọng cần nhớ là chúng tôi đã không thiết lập điều này một cách nghiêm ngặt. Tốc độ hội tụ sẽ phụ thuộc vào sự phân bố cụ thể của các phần tử trong chuỗi iid.

Trong thực tế hội tụ thường rất nhanh. Hình 6.4 cho thấy sự phân bố theo kinh nghiệm của đường trung bình động của một chuỗi cấp số nhân và một chuỗi iid hình học. Trong cả hai trường hợp, giá trị gần đúng thu được từ lý thuyết giới hạn trung tâm là rất chính xác ngay cả đối với trung bình 100 mẫu. Các



Hình 6.3: Kết quả của việc gộp hai phân phối khác nhau với chính chúng nhiều lần. Các hình dạng nhanh chóng trở nên giống Gaussian.



Hình 6.4: Phân bố thực nghiệm của đường trung bình động của một dãy Gaussian chuẩn iid (trên cùng), một dãy hình học iid với tham số p = 0,4 (ở giữa) và một dãy Cauchy iid (dưới cùng). Phân phối theo kinh nghiệm được tính toán từ 104 mẫu trong mọi trường hợp. Đối với hai hàng đầu tiên, ước tính được cung cấp bởi định lý giới hạn trung tâm được vẽ bằng màu đỏ.

Machine Translated by Google

hình cũng chỉ ra rằng đối với dãy Cauchy iid, phân phối của đường trung bình động không trở thành Gaussian, điều này không mâu thuẫn với định lý giới hạn trung tâm vì phân phối không có giá trị trung bình được xác định rõ. Để kết thúc phần này, chúng ta rút ra một xấp xỉ hữu ích đối với phân bố nhị thức bằng cách sử dụng định lý giới hạn trung tâm.

Ví dụ 6.3.2 (Xấp xỉ Gauss đối với phân bố nhị thức). Cho X có phân phối nhị thức với tham số n và p sao cho n lớn. Việc tính toán xác suất X nằm trong một khoảng nhất định yêu cầu tính tổng pmf của nó trên tất cả các giá trị trong khoảng đó. Ngoài ra, chúng ta có thể đạt được một xấp xỉ nhanh bằng cách sử dụng thực tế là đối với n lớn, phân phối của một biến ngẫu nhiên nhị thức xấp xỉ Gaussian. Thật vậy, chúng ta có thể viết X là tổng của n biến ngẫu nhiên Bernoulli độc lập với tham số p,

$$X = \sum_{tôi=1}^{N} bí \,. \tag{6.28}$$

Giá trị trung bình của Bi là p và phương sai của nó là p (1 - p). Theo định lý giới hạn trung tâm, $\frac{1}{n}X$ xấp xỉ là Gaussian với trung bình p và phương sai p (1 p) /n. Tương tự, theo Bổ đề 2.5.1, X xấp xỉ Gaussian với np trung bình và phương sai np(1 - p).

Giả sử rằng một cầu thủ bóng rổ thực hiện mỗi cú đánh mà cô ấy thực hiện với xác suất p = 0,4. Nếu chúng ta giả sử rằng mỗi lần bắn là độc lập, xác suất để cô ấy thực hiện được hơn 420 lần trong số 1000 lần là bao nhiêu? Chúng ta có thể lập mô hình các bức ảnh được thực hiện dưới dạng nhị thức X với các tham số 1000 và 0,4. câu trả lời chính xác là

$$P(X \geq 420) = \sum_{x=420}^{1000} p_X(x) \tag{6.29}$$

$$= \sum_{x=420}^{1000} \binom{1000}{x} 0{,}4^{x}\, 0{,}6^{(n\ x)} \tag{6.30}$$

$$= 10{,}4\ 10\ \ 2\,. \tag{6.31}$$

Nếu chúng ta áp dụng xấp xỉ Gaussian, theo Bổ đề 2.5.1 X lớn hơn 420 cũng giống như một Gaussian U tiêu chuẩn lớn hơn $\frac{420\ \ \mu}{\sigma}$ trong đó μ và σ là giá trị trung bình và độ lệch chuẩn của X, bằng np = 400 và $\sqrt{np(1\ \ p)}$ = 15,5 tương ứng.

$$P\,(X \geq 420) \approx P\left(\sqrt{np\,(1\ \ p)}\,U + np \geq 420\right) \tag{6.32}$$

$$= P\,(U \geq 1{,}29) \tag{6.33}$$

$$= 1\ \ \Phi\,(1{,}29) \tag{6.34}$$

$$= 9{,}85\ 10\ \ 2\,. \tag{6.35}$$

6.4 Mô phỏng Monte Carlo

Mô phỏng là một công cụ mạnh trong xác suất và thống kê. Các mô hình xác suất thường quá phức tạp để chúng ta rút ra các nghiệm dạng đóng của phân phối hoặc kỳ vọng của các đại lượng quan tâm, như chúng ta làm trong các bài tập về nhà.

Ví dụ, hãy tưởng tượng rằng bạn thiết lập một mô hình xác suất để xác định xác suất thắng một trò chơi solitaire. Nếu các thẻ được xáo trộn tốt, xác suất này bằng

$$P\ (\text{Thắng}) = \frac{\text{Số hoán vị dẫn đến chiến thắng}}{\text{Tổng số}}. \tag{6.36}$$

Vấn đề là việc xác định những hoán vị nào dẫn đến chiến thắng là rất khó nếu không thực sự chơi trò chơi để xem kết quả. Làm điều này cho mọi hoán vị có thể là khó tính toán, vì có 52! ≈ 8 1067 trong số đó. Tuy nhiên, có một cách đơn giản để tính gần đúng xác suất quan tâm: mô phỏng một số lượng lớn trò chơi và ghi lại phần nào dẫn đến chiến thắng. Trò chơi solitaire chính là thứ đã truyền cảm hứng cho Stanislaw Ulam đề xuất các phương pháp dựa trên mô phỏng, được gọi là phương pháp Monte Carlo (tên mã, lấy cảm hứng từ Sòng bạc Monte Carlo ở Monaco), trong bối cảnh nghiên cứu vũ khí hạt nhân vào những năm 1940: Những suy nghĩ và nỗ lực đầu tiên mà tôi thực hiện để thực hành (Phương pháp

Monte Carlo) được gợi ý bởi một câu hỏi đến với tôi vào năm 1946 khi tôi đang dưỡng bệnh và chơi solitaires. Câu hỏi đặt ra là cơ hội để một ván bài Canfield được bày ra với 52 lá bài sẽ ra thành công là bao nhiêu? Sau khi dành nhiều thời gian cố gắng ước tính chúng bằng các phép tính tổ hợp thuần túy, tôi tự hỏi liệu một phương pháp thực tế hơn "tư duy trừu tượng" có thể không phải là đặt nó ra nói một trăm lần và chỉ đơn giản là quan sát và đếm số lượt chơi thành công.

Điều này đã có thể hình dung được khi bắt đầu kỷ nguyên mới của máy tính nhanh, và tôi nghĩ ngay đến các vấn đề về khuếch tán neutron và các câu hỏi khác của vật lý toán học, và nói chung hơn là làm thế nào để thay đổi các quá trình được mô tả bởi các phương trình vi phân nhất định thành một dạng tương đương có thể hiểu được như một chuỗi các hoạt động ngẫu nhiên. Sau đó, tôi mô tả ý tưởng này với John von Neumann, và chúng tôi bắt đầu lập kế hoạch tính toán thực

tế.[1] Phương pháp Monte Carlo sử dụng mô phỏng để ước tính các đại lượng khó tính toán chính xác. Trong phần này, chúng ta xem xét vấn đề tính gần đúng xác suất của một sự kiện E, như trong trò chơi solitaire ví dụ.

Thuật toán 6.4.1 (xấp xỉ Monte Carlo). Để tính gần đúng xác suất của biến cố E,

chúng tôi:

1. Tạo n mẫu độc lập từ chức năng chỉ báo 1E liên quan đến sự kiện:
   I1, I2, . . . , Trong.

2. Tính giá trị trung bình của n mẫu

$$A(n) := \frac{1}{n} \sum_{i=1}^{N} \text{tôi} \tag{6.37}$$

   đó là ước tính cho xác suất của E

Xác suất quan tâm có thể được hiểu là kỳ vọng của hàm chỉ báo 1E liên quan đến sự kiện,

$$E(1E\ ) = P(E). \tag{6.38}$$

Theo luật số lớn, ước lượng A hội tụ thành xác suất thực khi n ∞. Ví dụ sau minh họa sức mạnh của kỹ thuật đơn giản này.

[1] http://en.wikipedia.org/wiki/Monte_Carlo_method#History

kết quả trò chơi			Thứ hạng			xác suất
1-2	1-3	2-3	R1	R2	R3	
1	1	2	1	2	3	1/6
1	1	3	1	3	2	1/6
1	3	2	1	1	1	12/1
1	3	3	2	3	1	12/1
2	1	2	2	1	3	1/6
2	1	3	1	1	1	1/6
2	3	2	3	1	2	12/1
2	3	3	3	2	1	12/1

Chức năng có thể xảy ra tập trung

	R1	R2	R3
1	7/12		5/12
1/2 2	1/4		1/4
1/4 3	1/6	1/4	1/3

Bảng 6.1: Bảng bên trái cho thấy tất cả các kết quả có thể xảy ra trong một giải đấu gồm ba đội ($m = 3$), xếp hạng kết quả cho mỗi đội và xác suất tương ứng. Bảng bên phải hiển thị pmf thứ hạng của mỗi đội.

Ví dụ 6.4.2 (Giải bóng rổ). Trong một giải đấu bóng rổ nội bộ m đội chơi với nhau mỗi mùa một lần. Các đội được sắp xếp theo kết quả trước đây của họ: đội 1 là tốt nhất và đội m tệ nhất. Chúng tôi lập mô hình xác suất đội i đánh bại đội j, với $1 \leq i < j \leq m$ BẰNG

$$P \text{ (đội j thắng đội i)} := j \, \frac{1}{i+1}. \tag{6.39}$$

Đội tốt nhất đánh bại đội thứ hai với xác suất 1/2 và đội thứ ba với xác suất 2/3, thứ hai đánh bại thứ ba với xác suất 1/2, thứ tư với xác suất 2/3 và thứ năm với xác suất 3/4, v.v. Chúng tôi giả định rằng kết quả của các trò chơi khác nhau là độc lập.

Vào cuối mùa giải, sau khi mỗi đội đã chơi với mọi đội khác, các đội được xếp hạng theo số trận thắng của họ. Nếu một số đội có cùng số trận thắng, thì họ chia sẻ cùng một thứ hạng. Ví dụ: nếu hai đội có nhiều trận thắng nhất, cả hai đều có thứ hạng 1 và đội tiếp theo có thứ hạng 3. Mục tiêu là tính toán phân phối thứ hạng cuối cùng của mỗi đội trong giải đấu mà chúng tôi mô hình hóa dưới dạng các biến ngẫu nhiên $R_1, R_2, \ldots, R_m$. Chúng ta có tất cả thông tin để tính pmf chung của các biến ngẫu nhiên này bằng cách áp dụng luật tổng xác suất. Như thể hiện trong Bảng 6.1 với $m = 3$, tất cả những gì chúng ta cần làm là liệt kê tất cả các khả năng có thể kết quả của các trò chơi và tính tổng xác suất của các kết quả dẫn đến một kết quả cụ thể thứ hạng.

Thật không may, số lượng các kết quả có thể xảy ra tăng lên đáng kể với m. Số lượng trò chơi bằng $m(m-1)/2$, do đó, kết quả có thể xảy ra là 10 $2^{m(m-1)/2}$. Khi chỉ có đội, số này lớn hơn 10^{13}. Tính phân phối chính xác của thứ hạng cuối cùng cho do đó, các giải đấu không phải là rất nhỏ đòi hỏi rất nhiều tính toán. May mắn thay, Al gorithm 6.4.1 đưa ra một giải pháp thay thế dễ xử lý hơn: Chúng ta có thể lấy mẫu một số lượng lớn các mùa n bằng cách mô phỏng mỗi trò chơi dưới dạng biến ngẫu nhiên Bernoulli với tham số được cho bởi phương trình (6.39) và ước tính pmf bằng cách sử dụng phần nhỏ thời gian mà mỗi đội kết thúc ở mỗi vị trí. Mô phỏng toàn bộ mùa giải chỉ yêu cầu lấy mẫu $m(m-1)/2$ trò chơi, điều này có thể được thực hiện rất nhanh.

Bảng 6.2 minh họa cách tiếp cận Monte Carlo cho $m = 3$. Xấp xỉ khá thô nếu chúng tôi chỉ sử dụng $n = 10$ mùa mô phỏng, nhưng trở nên rất chính xác khi $n = 2.000$. Hình 6.5

kết quả trò chơi			Thứ hạng		
1-2	1-3	2-3	R1	R2	R3
1	3	2	1	1	1
1	1	3	1	3	2
2	1	2	2	1	3
2	3	2	3	1	2
2	1	3	1	1	1
1	1	2	1	2	3
2	1	3	1	1	1
2	3	2	3	1	2
	1	2	1	2	3
1 2	3	2	3	1	2

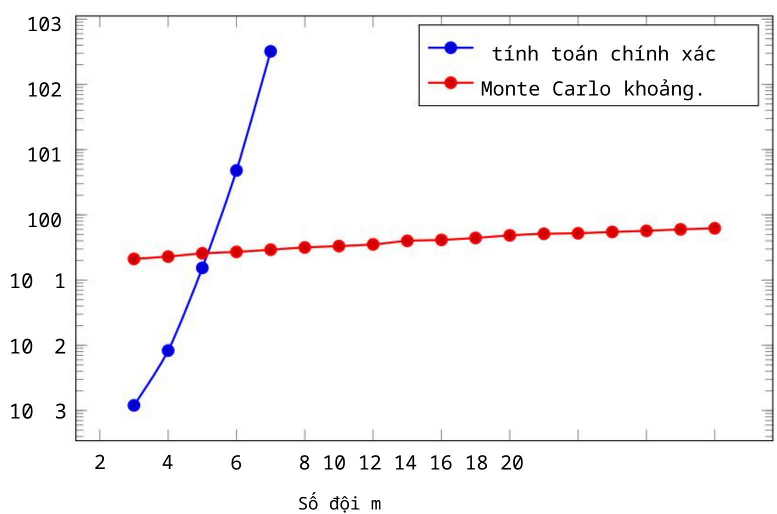
Ước tính pmf (n = 10)

	R1	R2	R3
1	0,6 (0,583)	0,7 (0,5)	0,3 (0,417)
2	0,1 (0,25)	0,2 (0,25)	0,4 (0,25)
3	0,3 (0,167)	0,1 (0,25)	0,3 (0,333)

Ước tính pmf (n = 2, 000)

	R1	R2	R3
1	0,582 (0,583)	0,496 (0,5)	0,417 (0,417)
2	0,248 (0,25)	0,261 (0,25)	0,244 (0,25)
3	0,171 (0,167)	0,245 (0,25)	0,339 (0,333)

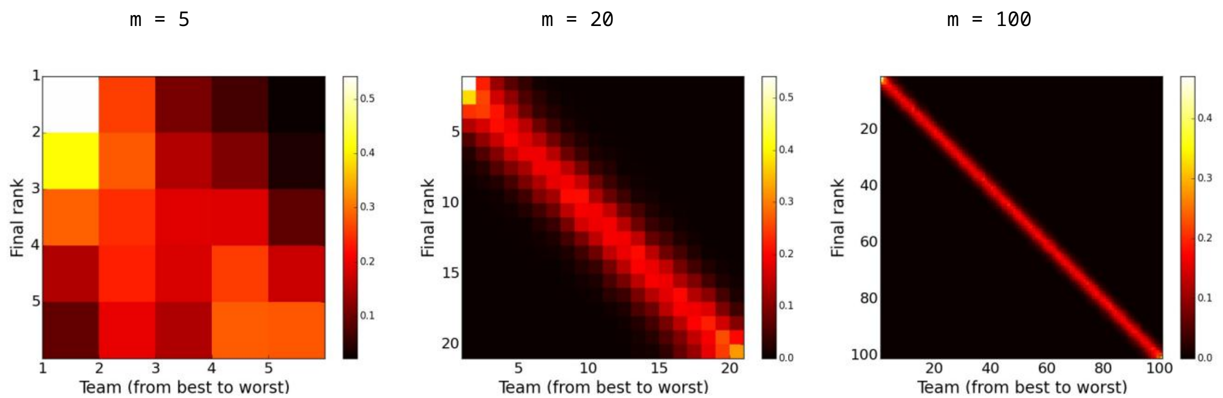
Bảng 6.2: Bảng bên trái hiển thị 10 kết quả mô phỏng của một giải đấu có ba đội (m = 3) và xếp hạng kết quả. Các bảng bên phải hiển thị pmf ước tính thu được từ mô phỏng Monte Carlo từ các kết quả mô phỏng ở bên trái (trên cùng) và từ 2.000 kết quả mô phỏng (dưới cùng). chính xác các giá trị được bao gồm trong ngoặc đơn để so sánh.



lỗi Monte Carlo

m	Lỗi trung bình
	9,28 10 3
4	12,7 10 3
5	7,95 10 3
6	7.12 10 3
7	7.12 10 3

Hình 6.5: Biểu đồ bên trái cho thấy thời gian cần thiết để có được pmf chính xác của xếp hạng cuối cùng trong Ví dụ 6.4.2 và ước tính chúng bằng phép tính gần đúng Monte Carlo sử dụng 2.000 giải đấu mô phỏng kết quả. Bảng bên phải hiển thị lỗi trung bình trên mỗi mục nhập của phép tính xấp xỉ Monte Carlo.



Hình 6.6: Ước tính pmf của các thứ hạng cuối cùng trong Ví dụ 6.4.2 sử dụng 2.000 kết quả giải đấu mô phỏng.

hiển thị thời gian chạy cần thiết để tính toán pmf chính xác và ước tính nó với Monte Cách tiếp cận Carlo cho số lượng đội khác nhau. Khi số lượng đội rất ít tính toán chính xác rất nhanh, nhưng thời gian chạy tăng theo cấp số nhân với m như mong đợi, do đó, đối với 7 đội, việc tính toán đã mất 5 phút rưỡi. Ngược lại, Monte Phép tính gần đúng của Carlo nhanh hơn đáng kể. Đối với m = 20 chỉ mất nửa giây. Hình 6.6 hiển thị pmf gần đúng của thứ hạng cuối cùng cho 5, 20 và 100 đội. Cấp bậc cao hơn có cao hơn xác suất bởi vì khi hai đội hòa nhau, họ sẽ được xếp hạng cao hơn.

Chương 7

Chuỗi Markov

Thuộc tính Markov được thỏa mãn bởi bất kỳ quá trình ngẫu nhiên nào mà tương lai độc lập với quá khứ một cách có điều kiện với hiện tại.

Định nghĩa 7.0.1 (tính chất Markov). Một quá trình ngẫu nhiên thỏa mãn tính chất Markov nếu $X(t_{i+1})$ độc lập có điều kiện với $X(t_1), \ldots, X(t_{i-1})$ cho trước $X(t_i)$ với mọi $t_1 < t_2 < \ldots < t_i < t_{i+1}$. Nếu không gian trạng thái của quá trình ngẫu nhiên là rời rạc thì với mọi $x_1, x_2, \ldots, x_{i+1}$

$$p_{X(t_{n+1})|X(t_1),X(t_2),\ldots,X(t_i)}(x_{n+1}|x_1, x_2, \ldots, x_n) = p_{X(t_{i+1})|X(t_i)}(x_{i+1}|x_i). \tag{7.1}$$

Nếu không gian trạng thái của quá trình ngẫu nhiên là liên tục (và phân bố có chung pdf),

$$f_{X(t_{i+1})|X(t_1),X(t_2),\ldots,X(t_i)}(x_{i+1}|x_1, x_2, \ldots, x_i) = f_{X(t_{i+1})|X(t_i)}(x_{i+1}|x_i). \tag{7.2}$$

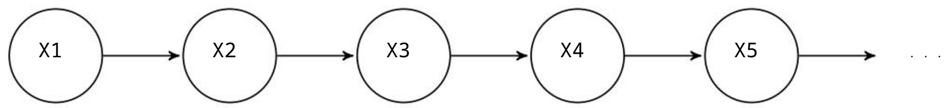
Hình 7.1 cho thấy mô hình đồ họa có hướng tương ứng với các giả định phụ thuộc hàm ý bởi thuộc tính Markov. Bất kỳ chuỗi iid nào cũng thỏa mãn tính chất Markov, vì tất cả các pmf hoặc pdf có điều kiện đều chỉ bằng các biên (trong trường hợp này sẽ không có cạnh trong đồ thị tuần hoàn có hướng của Hình 7.1). Bước đi ngẫu nhiên cũng thỏa mãn thuộc tính, vì một khi chúng ta cố định vị trí của bước đi tại một thời điểm nhất định, thì con đường mà nó đã đi trước đó không có ảnh hưởng đến các bước tiếp theo của nó.

Bổ đề 7.0.2. Bước đi ngẫu nhiên thỏa mãn tính chất Markov.

Bằng chứng. Đặt X biểu thị bước đi ngẫu nhiên được xác định trong Phần 5.6. Với điều kiện $X(j) = x_i$ với $j \leq i$ thì $X(i+1)$ bằng $x_i + S(i+1)$. Điều này không phụ thuộc vào $x_1, \ldots, x_{i-1}$, hàm ý (7.1). □

7.1 Chuỗi Markov thời gian rời rạc đồng nhất thời gian

Chuỗi Markov là một quá trình ngẫu nhiên thỏa mãn thuộc tính Markov. Ở đây, chúng tôi xem xét các chuỗi Markov thời gian rời rạc với không gian trạng thái hữu hạn, có nghĩa là quá trình chỉ có thể nhận một số lượng giá trị hữu hạn tại bất kỳ thời điểm nào. Để chỉ định chuỗi Markov, chúng ta chỉ cần xác định pmf của quá trình ngẫu nhiên tại điểm bắt đầu của nó (mà chúng ta sẽ giả sử là tại $i = 0$) và xác suất chuyển tiếp của nó. Điều này suy ra từ tính chất Markov, vì đối với bất kỳ



Hình 7.1: Mô hình đồ họa có hướng mô tả các giả định phụ thuộc hàm ý của thuộc tính Markov.

$n \geq 0$

$$p_{X(0),X(1),...,X(n)}(x_0, x_1, \ldots, x_n) := \prod_{i=0}^{n} p_{X(i)|X(0),...,X(i-1)}(x_i | x_0, \ldots, x_{i-1}) \tag{7.3}$$

$$= \prod_{i=0}^{n} p_{X(i)|X(i-1)}(x_i | x_{i-1}). \tag{7.4}$$

Nếu các xác suất chuyển đổi này giống nhau ở mọi bước thời gian (nghĩa là chúng không đổi và không phụ thuộc vào i), thì chuỗi Markov được cho là đồng nhất về thời gian. Trong trường hợp này, chúng ta có thể lưu trữ xác suất của mỗi chuyển đổi có thể có trong ma trận $s \times s$ T_X, trong đó s là số

Những trạng thái.

$$T_{X\,jk} := p_{X(i+1)|X(i)}(x_j | x_k). \tag{7.5}$$

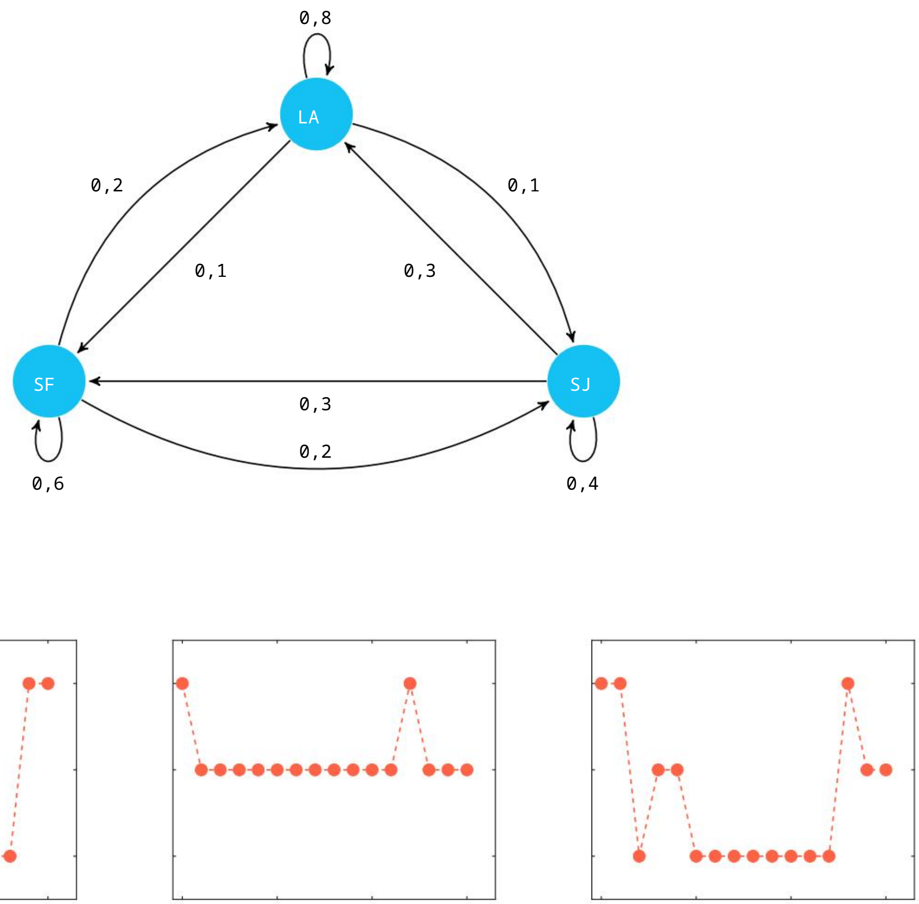
Trong chương này, chúng tôi tập trung vào các chuỗi Markov trạng thái hữu hạn đồng nhất theo thời gian. Khả năng thăm dò chuyển đổi của các chuỗi này có thể được hình dung bằng cách sử dụng sơ đồ trạng thái, sơ đồ này cho thấy từng trạng thái và xác suất của mọi chuyển đổi có thể xảy ra. Xem Hình 7.2 bên dưới để biết ví dụ. Không nên nhầm lẫn sơ đồ trạng thái với đồ thị tuần hoàn có hướng (DAG) biểu thị cấu trúc phụ thuộc của mô hình, được minh họa trong Hình 7.1. Trong sơ đồ trạng thái, mỗi nút tương ứng với một trạng thái và các cạnh để xác suất chuyển đổi giữa các trạng thái, trong khi DAG chỉ biểu thị cấu trúc phụ thuộc của quá trình ngẫu nhiên theo thời gian và thường giống nhau đối với tất cả các chuỗi Markov.

Để đơn giản hóa ký hiệu, chúng tôi xác định một vectơ s chiều được gọi là vectơ trạng thái, mà $p_{X(i)}$ chứa pmf cận biên của chuỗi Markov tại mỗi thời điểm i,

$$p_{X(i)} := \begin{bmatrix} p_{X(i)}(x_1) \\ p_{X(i)}(x_2) \\ \cdots \\ p_{X(i)}(x_s) \end{bmatrix}. \tag{7.6}$$

Mỗi mục trong vectơ trạng thái chứa xác suất chuỗi Markov ở trạng thái cụ thể đó tại thời điểm i. Nó không phải là giá trị của chuỗi Markov, mà là một biến ngẫu nhiên.

Không gian trạng thái ban đầu $p_{X(0)}$ và ma trận chuyển tiếp T_X đủ để xác định hoàn toàn chuỗi Markov trạng thái hữu hạn đồng nhất thời gian. Thật vậy, chúng ta có thể tính toán phân phối chung của in cho bất kỳ $n \geq 1$ nào từ $p_{X(0)}$ và T_X tại bất kỳ n thời điểm $i_1, i_2, \ldots,$ giảm biến trong bất kỳ bằng cách áp dụng (7.4) và xâu chuỗi thời điểm nào mà chúng tôi không quan tâm. Chúng tôi minh họa điều này trong ví dụ sau.



Hình 7.2: Biểu đồ trạng thái của chuỗi Markov được mô tả trong Ví dụ (7.1.1) (trên cùng). Mỗi mũi tên cho thấy xác suất chuyển đổi giữa hai trạng thái. Dưới đây chúng tôi chỉ ra ba hiện thực của Markov xích.

Ví dụ 7.1.1 (Cho thuê ô tô). Một công ty cho thuê xe hơi thuê bạn để mô hình vị trí của họ ô tô. Công ty hoạt động ở Los Angeles, San Francisco và San Jose. Khách hàng thường xuyên lấy một chiếc ô tô trong một thành phố và trả nó ở một thành phố khác. Sẽ rất hữu ích cho công ty nếu có thể tính toán khả năng một chiếc ô tô sẽ dừng lại ở một thành phố nhất định. Bạn quyết định lập mô hình vị trí của ô tô dưới dạng chuỗi Markov, trong đó mỗi bước thời gian tương ứng với một khách hàng mới lấy xe. Công ty phân bổ xe mới đồng đều giữa ba thành phố. quá trình chuyển đổi xác suất, thu được từ dữ liệu trong quá khứ, được đưa ra bởi

San Francisco	Los Angeles	San Jose	
0,6	0,1	0,3	San Francisco
0,2	0,8	0,3	Los Angeles
0,2	0,1	0,4	San Jose

Rõ ràng, xác suất để một khách hàng di chuyển ô tô từ San Francisco đến LA là 0,2,

Machine Translated by Google

xác suất chiếc xe ở lại San Francisco là 0,6, v.v.

Vectơ trạng thái ban đầu và ma trận chuyển tiếp của chuỗi Markov là

$$p_{X(0)} := \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \quad T_X := \begin{bmatrix} 0,6 & 0,1 & 0,3 \\ 0,2 & 0,8 & 0,3 \\ 0,2 & 0,1 & 0,4 \end{bmatrix}. \tag{7.7}$$

Bang 1 được giao cho San Francisco, bang 2 cho Los Angeles và bang 3 cho San Jose. Hình 7.2 cho thấy sơ đồ trạng thái của chuỗi Markov. Hình 7.2 cho thấy một số hiện thực của chuỗi Markov.

Công ty muốn tìm xác suất để chiếc xe xuất phát ở San Francisco, nhưng lại ở San Jose ngay sau khách hàng thứ hai. Điều này được đưa ra bởi

$$p_{X(0),X(2)}(1,3) = \sum_{tôi=1}^{3} p_{X(0),X(1),X(2)}(1,i,3) \tag{7.8}$$

$$= \sum_{tôi=1}^{3} p_{X(0)}(1)\, p_{X(1)|X(0)}(i|1)\, p_{X(2)|X(1)}(3|i) \tag{7.9}$$

$$= p_{X(0)}{}_1 \sum_{tôi=1}^{3} T_{X\,i1}\, T_{X\,3i} \tag{7.10}$$

$$= \frac{0,6 \cdot 0,2 + 0,2 \cdot 0,1 + 0,2 \cdot 0,4}{3} \approx 7,33\ 10^{-2}. \tag{7.11}$$

Xác suất là 7,33%.

Bổ đề sau cung cấp một biểu thức đơn giản cho vectơ trạng thái tại thời điểm i $p_{X(i)}$ về của TX và vectơ trạng thái trước đó.

Bổ đề 7.1.2 (Vectơ trạng thái và ma trận chuyển tiếp). Đối với chuỗi Markov X với ma trận chuyển tiếp TX

$$p_{X(i)} = T_X\, p_{X(i-1)}. \tag{7.12}$$

Nếu chuỗi Markov bắt đầu tại thời điểm 0 thì

$$p_{X(i)} = T_X^{i}\, p_{X(0)}, \tag{7.13}$$

tôi ở đâu T_X^{i} biểu thị phép nhân i lần với ma trận TX .

Bằng chứng. Bằng chứng trực tiếp từ các định nghĩa,

$$p_{X(i)} := \begin{bmatrix} p_{X(i)}(x_1) \\ p_{X(i)}(x_2) \\ \cdots \\ p_{X(i)}(x_s) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{s} p_{X(i-1)}(x_j)\, p_{X(i)|X(i-1)}(x_1|x_j) \\ \sum_{j=1}^{s} p_{X(i-1)}(x_j)\, p_{X(i)|X(i-1)}(x_2|x_j) \\ \cdots \\ \sum_{j=1}^{s} p_{X(i-1)}(x_j)\, p_{X(i)|X(i-1)}(x_s|x_j) \end{bmatrix} \tag{7.14}$$

$$= \begin{bmatrix} p_{X(i)|X(i-1)}(x_1|x_1) & p_{X(i)|X(i-1)}(x_1|x_2) & \cdots & p_{X(i)|X(i-1)}(x_1|x_s) \\ p_{X(i)|X(i-1)}(x_2|x_1) & p_{X(i)|X(i-1)}(x_2|x_2) & \cdots & p_{X(i)|X(i-1)}(x_2|x_s) \\ & \cdots & & \\ p_{X(i)|X(i-1)}(x_s|x_1) & p_{X(i)|X(i-1)}(x_s|x_2) & \cdots & p_{X(i)|X(i-1)}(x_s|x_s) \end{bmatrix} \begin{bmatrix} p_{X(i-1)}(x_1) \\ p_{X(i-1)}(x_2) \\ \cdots \\ p_{X(i-1)}(x_s) \end{bmatrix}$$

$$= T_X\, p_{X(i-1)} \tag{7.15}$$

Phương trình (7.13) thu được bằng cách áp dụng (7.12) i lần và có tính đến tính chất Markov. □

Ví dụ 7.1.3 (Cho thuê ô tô (tiếp theo)). Công ty muốn ước tính sự phân bố của các địa điểm ngay sau khi khách hàng thứ 5 đã sử dụng xe. Áp dụng Bổ đề 7.1.2 ta thu được

$$p_{X(5)} = T_X^5\, p_X(0) \tag{7.16}$$

$$= \begin{bmatrix} 0{,}281 \\ 0{,}534 \\ 0{,}185 \end{bmatrix}. \tag{7.17}$$

Mô hình ước tính rằng sau 5 khách hàng, hơn một nửa số xe ở Los Angeles.

7.2 Tái phát

Các trạng thái của chuỗi Markov có thể được phân loại tùy thuộc vào việc chuỗi Markov có được đảm bảo luôn quay lại với chúng hay liệu cuối cùng nó có thể ngừng truy cập các trạng thái đó hay không.

Định nghĩa 7.2.1 (Trạng thái lặp lại và nhất thời). Cho X là chuỗi Markov trạng thái hữu hạn đồng nhất thời gian. Chúng tôi xem xét một trạng thái cụ thể x. Nếu như

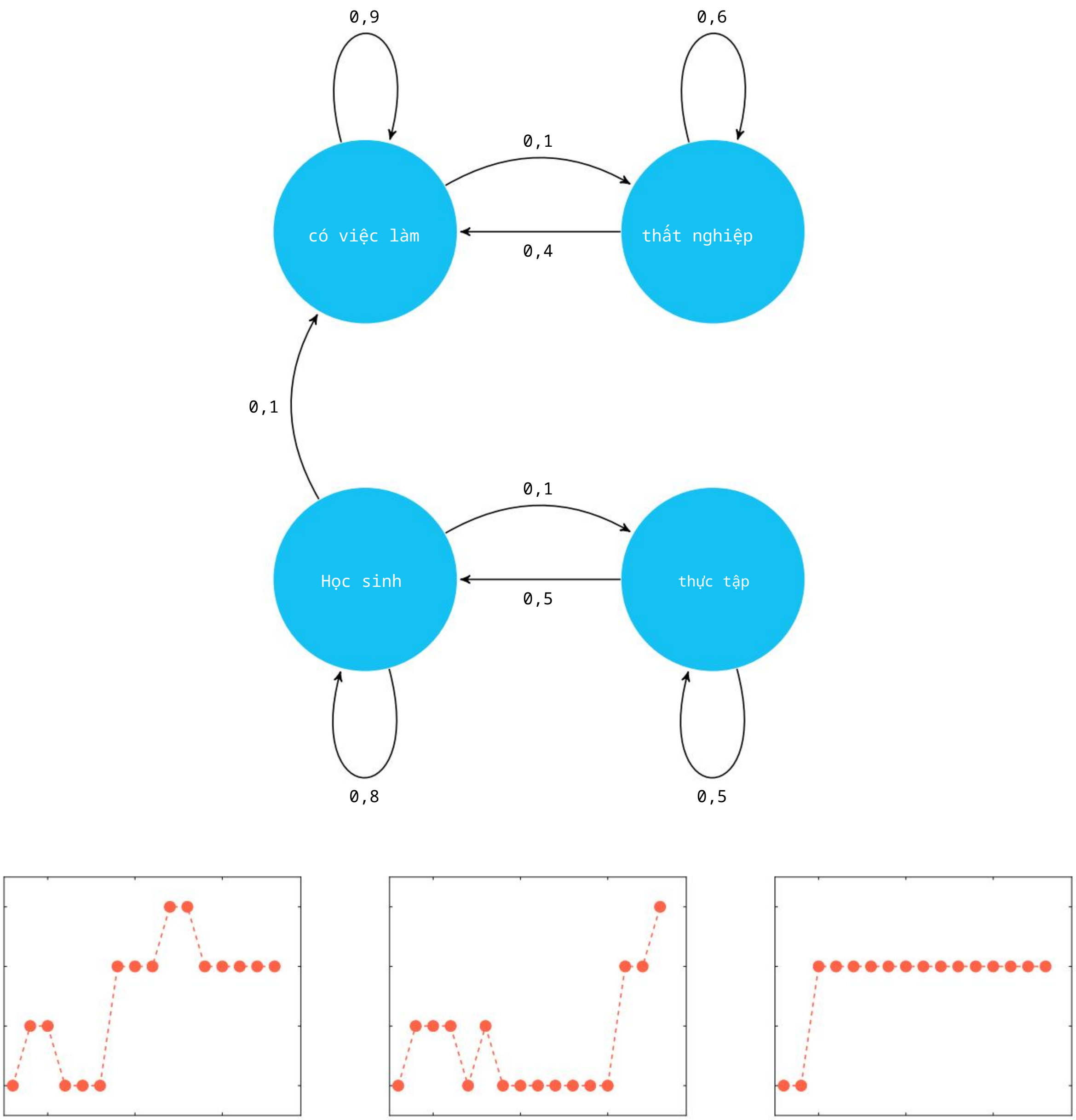
$$P\left(X(j) = s \text{ đối với } j > i \mid X(i) = s\right) = 1 \tag{7.18}$$

sau đó trạng thái được tái diễn. Nói cách khác, cho rằng chuỗi Markov ở x, xác suất để nó quay trở lại x là một. Ngược lại, nếu

$$P\left(X(j) \neq s \text{ với mọi } j > i \mid X(i) = s\right) > 0 \tag{7.19}$$

trạng thái là nhất thời. Cho rằng chuỗi Markov ở x, có xác suất khác không là nó sẽ không bao giờ quay trở lại.

Machine Translated by Google



Hình 7.3: Biểu đồ trạng thái của chuỗi Markov được mô tả trong Ví dụ (7.2.2) (trên cùng). Dưới đây chúng tôi trình bày ba nhận thức của chuỗi Markov.

Ví dụ sau minh họa sự khác biệt giữa trạng thái lặp lại và nhất thời.

Ví dụ 7.2.2 (Động lực tuyển dụng). Một nhà nghiên cứu quan tâm đến việc mô hình hóa động lực tuyển dụng của những người trẻ tuổi bằng cách sử dụng chuỗi Markov.

Cô ấy xác định rằng ở tuổi 18, một người là sinh viên với xác suất 0,9 hoặc thực tập sinh với xác suất 0,1. Sau đó, cô ấy ước tính các xác suất chuyển đổi sau:

Sinh viên	thực tập	Có việc làm	Thất nghiệp	
0,8	0,5	0	0	Học sinh
0,1	0,5	0	0	thực tập
0,1	0	0,9	0,4	có việc làm
0	0	0,1	0,6	thất nghiệp

Giả định Markov rõ ràng là không hoàn toàn chính xác, một người đã là sinh viên lâu hơn có lẽ ít có khả năng tiếp tục là sinh viên hơn, nhưng những mô hình Markov như vậy dễ phù hợp hơn (chúng ta chỉ cần ước tính xác suất chuyển đổi) và thường mang lại những hiểu biết hữu ích.

Vectơ trạng thái ban đầu và ma trận chuyển tiếp của chuỗi Markov là

$$pX(0) := \begin{bmatrix} 0,9 \\ 0,1 \\ 0 \\ 0 \end{bmatrix}, \quad TX := \begin{bmatrix} 0,8 & 0,5 & 0 & 0 \\ 0,1 & 0,5 & 0 & 0 \\ 0,1 & 0 & 0,9 & 0,4 \\ 0 & 0 & 0,1 & 0,6 \end{bmatrix}. \tag{7.20}$$

Hình 7.3 cho thấy sơ đồ trạng thái và một số thực hiện của chuỗi Markov.

Trạng thái 1 (sinh viên) và 2 (thực tập) là trạng thái nhất thời. Lưu ý rằng xác suất chuỗi Markov trở lại các trạng thái đó sau khi truy cập trạng thái 3 (được sử dụng) bằng không, vì vậy

$$PX(j) = 1 \text{ với mọi } j > i \mid X(i) = 1 \geq PX(i+1) = 3 \mid X(i) = 1 \tag{7.21}$$

$$= 0,1 > 0, \tag{7.22}$$

$$PX(j) = 2 \text{ với mọi } j > i \mid X(i) = 2 \geq PX(i+2) = 3 \mid X(i) = 2 \tag{7.23}$$

$$= 0,5 \cdot 0,1 > 0. \tag{7.24}$$

Ngược lại, trạng thái 3 và 4 (thất nghiệp) lại tái diễn. Ta chứng minh điều này cho trạng thái 3 (đối số cho trạng thái 4 hoàn toàn giống nhau):

$$PX(j) = 3 \text{ với mọi } j > i \mid X(i) = 3 \tag{7.25}$$

$$= PX(j) = 4 \text{ với mọi } j > i \mid X(i) = 3 \tag{7.26}$$

$$= \lim_{k \to \infty} PX(i+1) = 4 \mid X(i) = 3 \prod_{j=1}^{k} PX(i+j+1) = 4 \mid X(i+j) = 4 \tag{7.27}$$

$$= \lim_{k \to \infty} 0,1 \cdot 0,6^k \tag{7.28}$$

$$= 0. \tag{7.29}$$

Trong ví dụ này, không thể tiếp cận sinh viên và thực tập sinh của tiểu bang từ các tiểu bang có việc làm hoặc thất nghiệp. Các chuỗi Markov có khả năng chuyển đổi giữa hai trạng thái bất kỳ (ngay cả khi nó không trực tiếp) được gọi là bất khả quy.

Định nghĩa 7.2.3 (Chuỗi Markov bất khả quy). Chuỗi Markov trạng thái hữu hạn thuần nhất thời gian là bất khả quy nếu với bất kỳ trạng thái x nào, xác suất đạt tới mọi trạng thái khác y = x trong một số bước hữu hạn là khác không, nghĩa là tồn tại m ≥ 0 sao cho

$$PX(i + m) = y \mid X(i) = x > 0. \tag{7.30}$$

Người ta có thể dễ dàng kiểm tra xem chuỗi Markov trong Ví dụ 7.1.1 là bất khả quy, trong khi chuỗi Markov trong Ví dụ 7.2.2. Một kết quả quan trọng là tất cả các trạng thái trong chuỗi Markov bất khả quy đều tái phát.

Định lý 7.2.4 (Chuỗi Markov bất khả quy). Tất cả các trạng thái trong chuỗi Markov bất khả quy đều tái phát.

Bằng chứng. Trong bất kỳ chuỗi Markov trạng thái hữu hạn nào cũng phải có ít nhất một trạng thái lặp lại. Nếu tất cả các trạng thái đều tạm thời thì có một xác suất khác không là nó rời khỏi tất cả các trạng thái mãi mãi, điều này là không thể. Không mất tính tổng quát, giả sử trạng thái x là khả quy. Bây giờ chúng ta cung cấp một bản phác thảo chứng minh rằng một trạng thái tùy ý khác y cũng phải được lặp lại. Để giảm bớt ký hiệu, hãy để

$$p_{x,x} := PX (j) = x \text{ đối với } j > i \mid X(i) = x \quad , \tag{7.31}$$

$$p_{x,y} := PX (j) = y \text{ đối với một số } j > i \mid X(i) = x \quad , \tag{7.32}$$

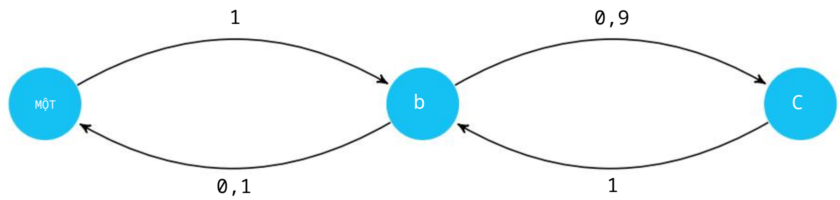
$$p_{y,x} := PX (k) = x \text{ với } j > i \mid X(i) = y . \tag{7.33}$$

Chuỗi là bất khả quy nên có một xác suất khác 0 pm > 0 để đạt được y từ x trong tối đa m bước với một số m > 0. Xác suất để chuỗi đi từ x đến y và không bao giờ quay lại x do đó ít nhất là pm (1 py,x). Tuy nhiên, x là lặp lại, vì vậy xác suất này phải bằng không! Vì pm > 0 nên điều này ngụ ý py,x = 1.

Xét sự kiện sau:

1. X đi từ y đến x.
2. X không quay lại y trong m bước sau khi đến x.
3. X cuối cùng lại đến x tại thời điểm m > m.

Xác suất của sự kiện này bằng py,x (1 pm) px,x = 1 pm (nhắc lại rằng x là lặp lại nên px,x = 1). Bây giờ hãy tưởng tượng rằng các bước 2 và 3 lặp lại k lần, tức là X không thể đi từ x đến y trong m bước k lần. Xác suất của sự kiện này là py,x (1 - pm) $^{k} p^{k}_{x,x}$ = (1 - giờ chiều)^{k}. Lấy k ∞ giá trị này bằng 0 đối với mọi m nên xác suất để X cuối cùng không trở về x phải bằng 0 (điều này có thể được thực hiện nghiêm ngặt, nhưng các chi tiết nằm ngoài phạm vi của các ghi chú này). □



Hình 7.4: Sơ đồ trạng thái của chuỗi Markov trong đó các trạng thái có giai đoạn hai.

7.3 Tính định kỳ

Một cân nhắc quan trọng khác là liệu chuỗi Markov có luôn ghé thăm một trạng thái nhất định theo các khoảng thời gian đều đặn hay không. Nếu đây là trường hợp, thì trạng thái có thời gian lớn hơn một.

Định nghĩa 7.3.1 (Chu kỳ của một trạng thái). Gọi X là chuỗi Markov trạng thái hữu hạn thuần nhất thời gian và trạng thái xa của chuỗi Markov. Khoảng thời gian m của x là số nguyên lớn nhất sao cho chỉ có thể quay về x theo số bước là bội số của m, tức là ta chỉ có thể quay về theo km bước với xác suất khác 0 trong đó k là số nguyên dương.

Hình 7.4 cho thấy một chuỗi Markov trong đó các trạng thái có chu kỳ bằng hai. Chuỗi Markov định kỳ không chứa các trạng thái có chu kỳ lớn hơn một.

Định nghĩa 7.3.2 (Chuỗi Markov định kỳ). Chuỗi Markov trạng thái hữu hạn đồng nhất thời gian X là chu kỳ nếu tất cả các trạng thái có chu kỳ bằng một.

Chuỗi Markov trong Ví dụ 7.1.1 và 7.2.2 đều có chu kỳ tuần hoàn.

7.4 Hội tụ

Trong phần này, chúng tôi nghiên cứu dưới những điều kiện nào mà chuỗi Markov X đồng nhất về thời gian ở trạng thái hữu hạn hội tụ trong phân phối. Nếu một chuỗi Markov hội tụ trong phân phối, thì vectơ $p_{X(i)}$, trạng thái của nó chứa pmf bậc nhất của X, hội tụ thành một vectơ cố định p_∞,

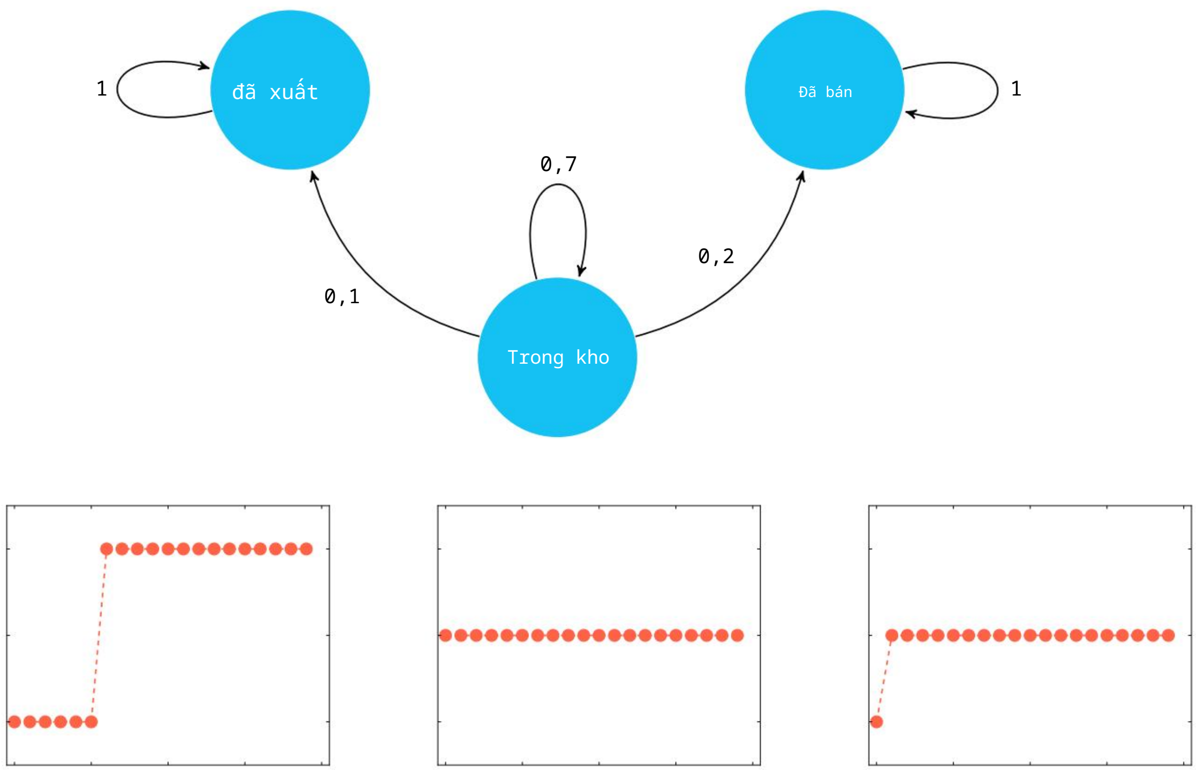
$$p_\infty := \lim_{tôi \to \infty} p_{X(i)} . \tag{7.34}$$

Trong trường hợp đó, xác suất chuỗi Markov ở mỗi trạng thái cuối cùng có xu hướng cố định (điều này không ngụ ý rằng chuỗi Markov sẽ ở trạng thái nhất định!).

Theo Bổ đề 7.1.2, chúng ta có thể biểu diễn (7.34) dưới dạng vectơ trạng thái ban đầu và quá trình chuyển đổi ma trận chuỗi Markov

$$p_\infty = \lim_{tôi \to \infty} T_X^{i}\, p_{X(0)} . \tag{7.35}$$

Việc tính toán giới hạn này một cách phân tích cho một T_X cụ thể và $p_{X(0)}$ thoạt nhìn có vẻ khó khăn. Tuy nhiên, thường có thể tận dụng phân tích riêng của ma trận chuyển tiếp (nếu nó tồn tại) để tìm p_∞. Điều này được minh họa trong ví dụ sau đây.



Hình 7.5: Biểu đồ trạng thái của chuỗi Markov được mô tả trong Ví dụ (7.4.1) (trên cùng). Dưới đây chúng tôi trình bày ba nhận thức của chuỗi Markov.

Ví dụ 7.4.1 (Điện thoại di động). Một công ty sản xuất điện thoại di động muốn lập mô hình bán hàng của một mẫu điện thoại mới mà họ vừa phát hành. Hiện 90% số máy là hàng tồn, 10% bán trong nước và chưa xuất khẩu. Dựa trên dữ liệu trong quá khứ, công ty xác định rằng mỗi ngày một chiếc điện thoại được bán với xác suất 0,2 và xuất khẩu với xác suất 0,1. Vectơ trạng thái ban đầu và ma trận chuyển tiếp của chuỗi Markov là

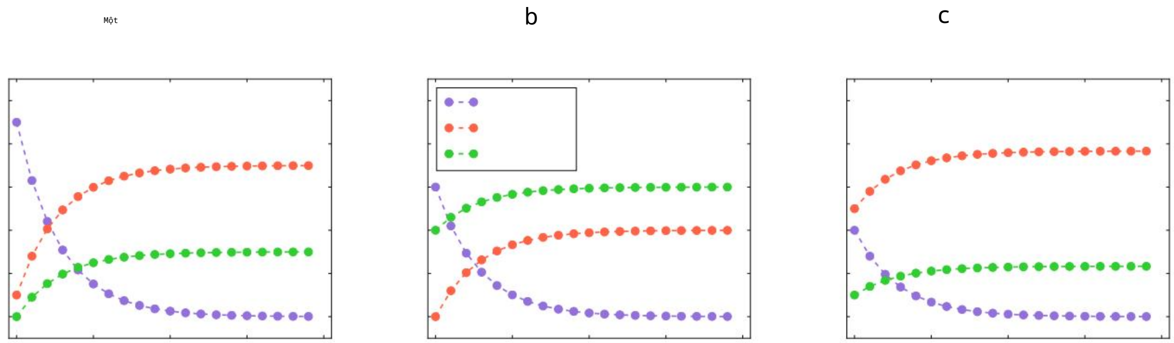
$$\text{một} := \begin{bmatrix} 0,9 \\ 0,1 \\ 0 \end{bmatrix}, \quad T_X = \begin{bmatrix} 0,7 & 0 & 0 \\ 0,2 & 1 & 0 \\ 0,1 & 0 & 1 \end{bmatrix}. \tag{7.36}$$

Chúng tôi đã sử dụng a để biểu thị vì sau này chúng tôi sẽ xem xét các vectơ trạng thái ban đầu khả dĩ khác. $p_{X(0)}$

Hình 7.6 cho thấy sơ đồ trạng thái và một số thực hiện của chuỗi Markov.

Công ty quan tâm đến số phận của mô hình mới. Đặc biệt, nó muốn tính toán phần nào điện thoại di động sẽ được xuất khẩu và phần nào sẽ được bán trong nước. Cái này

Machine Translated by Google



Hình 7.6: Sự tiến triển của vectơ trạng thái của chuỗi Markov trong Ví dụ (7.4.1) đối với các giá trị khác nhau của vectơ trạng thái ban đầu $p_X(0)$.

tương đương với tính toán

$$\lim_{i\to\infty} p_X(i) = \lim_{i\to\infty} T_X^{i}\, p_X(0) \tag{7.37}$$

$$= \lim_{i\to\infty} T_X^{i}\, a. \tag{7.38}$$

Ma trận chuyển tiếp TX có ba vectơ riêng

$$q_1 := \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad q_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad q_3 := \begin{bmatrix} 0,80 \\ 0,53 \\ 0,27 \end{bmatrix}. \tag{7.39}$$

Các giá trị riêng tương ứng là $\lambda_1 := 1$, $\lambda_2 := 1$ và $\lambda_3 := 0,7$. Chúng tôi tập hợp các vectơ riêng và giá trị riêng thành hai ma trận

$$Q := \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix}, \quad \Lambda := \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}, \tag{7.40}$$

sao cho sự phân hủy riêng của TX là

$$T_X := Q \Lambda Q^{-1}. \tag{7.41}$$

Sẽ rất hữu ích khi biểu diễn vectơ trạng thái ban đầu a theo các vectơ riêng khác nhau. Điều này đạt được bằng cách tính toán

$$Q^{-1} p_X(0) = \begin{bmatrix} 0,3 \\ 0,7 \\ 1.122 \end{bmatrix}, \tag{7.42}$$

để có thể

$$a = 0,3\, q_1 + 0,7\, q_2 + 1,122\, q_3. \tag{7.43}$$

Chúng tôi kết luận rằng

$$\lim_{i\to\infty} T_X^i a = \lim_{i\to\infty} T_X^i (0{,}3\, q_1 + 0{,}7\, q_2 + 1{,}122\, q_3) \tag{7.44}$$

$$= \lim_{i\to\infty} 0{,}3\, T_X^i q_1 + 0{,}7 T_X^i q_2 + 1{,}122 T_X^i q_3 \tag{7.45}$$

$$= \lim_{i\to\infty} 0{,}3\, \lambda_1^i q_1 + 0{,}7\, \lambda_2^i q_2 + 1{,}122\, \lambda_3^i q_3 \tag{7.46}$$

$$= \lim_{i\to\infty} 0{,}3\, q_1 + 0{,}7\, q_2 + 1{,}122\, 0{,}5^i q_3 \tag{7.47}$$

$$= 0{,}3\, q_1 + 0{,}7\, q_2 \tag{7.48}$$

$$= \begin{bmatrix} 0 \\ 0{,}7 \\ 0{,}3 \end{bmatrix}. \tag{7.49}$$

Điều này có nghĩa là cuối cùng xác suất mỗi chiếc điện thoại được bán trong nước là 0,7 và xác suất nó được xuất khẩu là 0,3. Đồ thị bên trái trong Hình 7.6 cho thấy sự tiến hóa của vectơ trạng thái. Đúng như dự đoán, cuối cùng nó hội tụ về vectơ trong phương trình (7.49).

Nói chung, do cấu trúc đặc biệt của hai vectơ riêng với giá trị riêng bằng một trong ví dụ này, chúng ta có

$$\lim_{i\to\infty} T_X^i\, p_X(0) = \begin{bmatrix} 0 \\ \left(Q^{-1} p_X(0)\right)_2 \\ \left(Q^{-1} p_X(0)\right)_1 \end{bmatrix}. \tag{7.50}$$

Điều này được minh họa trong Hình 7.6 , nơi bạn có thể thấy sự phát triển của vectơ trạng thái nếu nó được khởi tạo cho hai bản phân phối khác này:

$$b := \begin{bmatrix} 0{,}6 \\ 0 \\ 0{,}4 \end{bmatrix}, \quad Q^{-1} b = \begin{bmatrix} 0{,}6 \\ 0{,}4 \\ 0{,}75 \end{bmatrix}, \tag{7.51}$$

$$c := \begin{bmatrix} 0{,}4 \\ 0{,}5 \\ 0{,}1 \end{bmatrix}, \quad Q^{-1} c = \begin{bmatrix} 0{,}23 \\ 0{,}77 \\ 0{,}50 \end{bmatrix}. \tag{7.52}$$

Ma trận chuyển tiếp của chuỗi Markov trong Ví dụ 7.4.1 có hai vectơ riêng với giá trị riêng bằng một. Nếu chúng ta đặt vectơ trạng thái ban đầu bằng một trong hai vectơ riêng này (lưu ý rằng chúng ta phải đảm bảo chuẩn hóa chúng để vectơ trạng thái chứa pmf hợp lệ) thì

$$T_X\, p_X(0) = p_X(0), \tag{7.53}$$

để có thể

$$p_X(i) = T_X^i\, p_X(0) \tag{7.54}$$

$$= p_X(0) \tag{7.55}$$

Machine Translated by Google

cho tất cả tôi. Đặc biệt,

$$\lim_{i \to \infty} p_{X(i)} = p_{X(0)}, \tag{7.56}$$

do đó X hội tụ đến một biến ngẫu nhiên với pmf $p_{X(0)}$ trong phân phối. Một phân phối thỏa mãn fies (7.56) được gọi là phân phối dừng của chuỗi Markov.

Định nghĩa 7.4.2 (Phân bố cố định). Giả sử X là một chuỗi Markov thuần nhất thời gian ở trạng thái hữu hạn và để p_{stat} là một vectơ trạng thái chứa pmf hợp lệ trên các trạng thái có thể có của X. Nếu p_{stat} là một vectơ riêng liên kết với một giá trị riêng bằng một, do đó

$$T_X \, p_{stat} = p_{stat}, \tag{7.57}$$

thì phân phối tương ứng với p_{stat} là phân phối dừng hoặc trạng thái ổn định của X.

Xác định xem một phân phối có dừng hay không bằng cách kiểm tra xem (7.57) các giá trị giữ có thể gây khó khăn về mặt tính toán hay không nếu không gian trạng thái là rất lớn. Bây giờ chúng ta rút ra một điều kiện thay thế ngụ ý tính dừng. Trước tiên chúng ta hãy xác định khả năng đảo ngược của chuỗi Markov.

Định nghĩa 7.4.3 (Khả năng đảo ngược). Đặt X là chuỗi Markov đồng nhất thời gian ở trạng thái hữu hạn với các trạng thái s và ma trận chuyển tiếp T_X. Giả sử X(i) phân bố theo vectơ trạng thái $p \in \mathbb{R}^s$. Nếu như

$$P\left(X(i) = x_j, X(i+1) = x_k\right) = P\left(X(i) = x_k, X(i+1) = x_j\right), \text{ với mọi } 1 \le j, k \le s, \tag{7.58}$$

thì ta nói rằng X khả nghịch đối với p. Điều này tương đương với điều kiện cân bằng chi tiết

$$T_{X\,kj}\, p_j = T_{X\,jk}\, p_k, \text{ với mọi } 1 \le j, k \le s. \tag{7.59}$$

Như đã được chứng minh trong định lý sau, khả năng đảo ngược ngụ ý tính dừng, nhưng điều ngược lại không đúng. Chuỗi Markov không nhất thiết phải đảo ngược đối với phân phối cố định (và thường sẽ không như vậy). Do đó, điều kiện cân bằng chi tiết chỉ cung cấp một điều kiện đủ cho tính dừng.

Định lý 7.4.4 (Tính khả nghịch bao hàm tính dừng). Nếu chuỗi Markov thuần nhất thời gian X có thể đảo ngược đối với phân phối p_X, thì p_X là phân phối cố định của X.

Bằng chứng. Gọi p là vectơ trạng thái chứa p_X. Theo giả thiết T_X và p thỏa mãn (7.59), do đó với $1 \le j \le s$

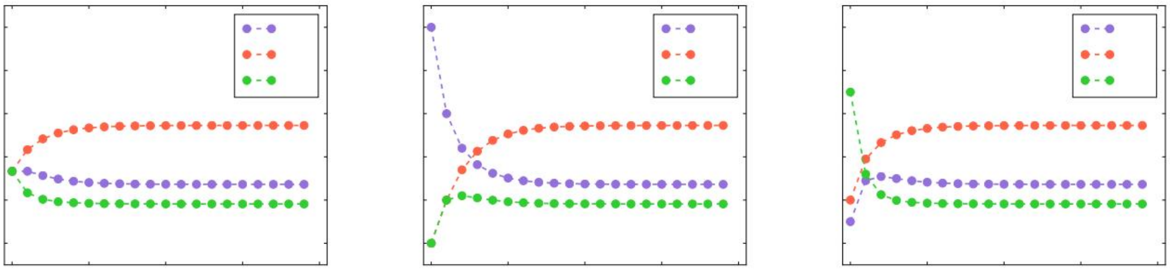
$$\left(T_X\, p\right)_j = \sum_{k=1}^{s} T_{X\,jk}\, p_k \tag{7.60}$$

$$= \sum_{k=1}^{s} T_{X\,kj}\, p_j \tag{7.61}$$

$$= p_j \sum_{k=1}^{s} T_{X\,kj} \tag{7.62}$$

$$= p_j. \tag{7.63}$$

$$p_{X(0)} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} \qquad p_{X(0)} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad p_{X(0)} = \begin{bmatrix} 0,1 \\ 0,2 \\ 0,7 \end{bmatrix}$$



Hình 7.7: Diễn biến vectơ trạng thái của chuỗi Markov trong Ví dụ (7.4.7).

Bước cuối cùng xuất phát từ thực tế là các cột của ma trận chuyển tiếp hợp lệ phải cộng thành một (chuỗi luôn phải đi đâu đó). □

Trong ví dụ 7.4.1, chuỗi Markov có hai phân bố cố định. Nó chỉ ra rằng điều này là không thể đối với các chuỗi Markov bất khả quy.

Định lý 7.4.5. Chuỗi Markov bất khả quy có một phân bố cố định duy nhất.

Bằng chứng. Điều này tuân theo định lý Perron-Frobenius, phát biểu rằng ma trận chuyển tiếp của chuỗi Markov bất khả quy có một vectơ riêng duy nhất với giá trị riêng bằng một và các phần tử không âm. □

Ngoài ra, nếu chuỗi Markov là chuỗi tuần hoàn, thì nó được đảm bảo hội tụ trong phân phối tới một biến ngẫu nhiên với phân phối dừng của nó đối với bất kỳ vectơ trạng thái ban đầu nào. Chuỗi Markov như vậy được gọi là ergodic.

Định lý 7.4.6 (Sự hội tụ của chuỗi Markov). Nếu chuỗi Markov đồng nhất thời gian rời rạc X là bất khả quy và tuần hoàn thì vectơ trạng thái của nó hội tụ với pstat phân bố cố định của X đối với bất kỳ vectơ trạng thái ban đầu nào. Điều này ngụ ý rằng X hội tụ trong phân phối thành pX(0) ngẫu nhiên. biến với pmf được đưa ra bởi pstat.

Bằng chứng của kết quả này nằm ngoài phạm vi của những ghi chú này.

Ví dụ 7.4.7 (Cho thuê ô tô (tiếp theo)). Chuỗi Markov trong ví dụ cho thuê ô tô là không thể thay đổi được và có chu kỳ . Bây giờ chúng ta sẽ kiểm tra xem nó có thực sự hội tụ trong phân phối không. Ma trận chuyển tiếp của nó có các vectơ riêng sau

$$q_1 := \begin{bmatrix} 0,273 \\ 0,545 \\ 0,182 \end{bmatrix}, \quad q_2 := \begin{bmatrix} 0,577 \\ 0,789 \\ 0,211 \end{bmatrix}, \quad q_3 := \begin{bmatrix} 0,577 \\ 0,211 \\ 0,789 \end{bmatrix}. \tag{7.64}$$

Các giá trị riêng tương ứng là $\lambda_1 := 1$, $\lambda_2 := 0,573$ và $\lambda_3 := 0,227$. Theo dự đoán của Định lý 7.4.5, chuỗi Markov có một phân phối dừng duy nhất.

Đối với bất kỳ vectơ trạng thái ban đầu nào, thành phần thẳng hàng với q_1 sẽ được bảo toàn bởi chuyển tiếp của chuỗi Markov, nhưng hai thành phần còn lại sẽ trở nên không đáng kể sau một lúc. Do đó, chuỗi hội tụ trong phân phối thành một biến ngẫu nhiên với pmf q_1 (lưu ý rằng q_1 đã được chuẩn hóa thành một pmf hợp lệ), như được tiên đoán bởi Định lý 7.4.6. Đây là minh họa trong hình 7.7. Bất kể công ty phân bổ những chiếc xe mới như thế nào, cuối cùng là 27,3% sẽ kết thúc ở San Francisco, 54,5% ở LA và 18,2% ở San Jose.

7.5 Chuỗi Markov Monte Carlo

Sự hội tụ của chuỗi Markov thành phân phối cố định rất hữu ích cho việc mô phỏng các biến ngẫu nhiên. Phương pháp Markov-chain Monte Carlo (MCMC) tạo mẫu từ mục tiêu phân phối bằng cách xây dựng chuỗi Markov sao cho phân phối cố định bằng phân phối mong muốn. Những kỹ thuật này có tầm quan trọng rất lớn trong thống kê hiện đại và trong cụ thể trong mô hình Bayesian. Trong phần này, chúng tôi mô tả một trong những MCMC phổ biến nhất phương pháp và minh họa nó bằng một ví dụ đơn giản.

Thách thức chính trong các phương pháp MCMC là thiết kế một chuỗi Markov định kỳ bất khả quy cho mà phân phối mục tiêu là cố định. Thuật toán Metropolis-Hastings sử dụng một phụ trợ Chuỗi Markov để đạt được điều này.

Thuật toán 7.5.1 (Thuật toán Metropolis-Hastings). Chúng tôi lưu trữ pmf p_X của phân phối đích trong một vectơ $p \in \mathbb{R}^s$, như vậy mà

$$p_j := p_X(x_j), \quad 1 \leq j \leq s. \tag{7.65}$$

Gọi T là ma trận chuyển tiếp của chuỗi Markov bất khả quy với cùng một không gian trạng thái $\{x_1, \ldots, x_s\}$ như p.

Khởi tạo $X(0)$ một cách ngẫu nhiên hoặc một trạng thái cố định, sau đó lặp lại các bước sau cho $i = 1, 2, 3, \ldots$.

1. Tạo biến ngẫu nhiên ứng cử viên C từ $X(i-1)$ bằng cách sử dụng ma trận chuyển tiếp T, I E

$$\text{máy tính} = k \mid X(i-1) = j = T_{kj}, \quad 1 \leq j, k \leq s. \tag{7.66}$$

2 bộ

$$X(\text{tôi}) := \begin{cases} C & \text{với xác suất } p_{acc}(X(i-1), C), \\ X(i-1) & \text{nếu không,} \end{cases} \tag{7.67}$$

trong đó xác suất chấp nhận được định nghĩa là

$$p_{acc}(j, k) := \text{tối thiểu}\left(\frac{\text{Chặc chậc}}{t_{kj}\, p_j}, 1\right) \quad 1 \leq j, k \leq s. \tag{7.68}$$

Hóa ra thuật toán này tạo ra một chuỗi Markov có thể đảo ngược đối với phân phối lãi suất, đảm bảo rằng phân phối là cố định.

Định lý 7.5.2. pmf trong p tương ứng với phân phối cố định của chuỗi Markov X thu được bằng thuật toán Metropolis-Hastings.

Machine Translated by Google

Bằng chứng. Chúng tôi chỉ ra rằng chuỗi Markov X có thể đảo ngược đối với p, nghĩa là

$$TX_{kj}\ pj = TX_{jk}\ pk, \tag{7.69}$$

đúng với mọi 1 ≤ j, k ≤ s. Điều này thiết lập kết quả theo Định lý 7.4.4. Điều kiện cân bằng chi tiết không đáng kể nếu j = k. Nếu j = k ta có

$$TX_{kj} := PX(i) = k \mid X(i\ \ 1) = j \tag{7.70}$$

$$= PX(i) = C, C = k \mid X(i\ \ 1) = j \tag{7.71}$$

$$= PX(i) = C \mid C = k, X(i\ \ 1) = j\ PC = k \mid X(i\ \ 1) = j \tag{7.72}$$

$$= pacc\ (j, k)\ Tkj \tag{7.73}$$

và bằng chính xác cùng một đối số TX_{jk} = pacc (k,j) Tjk. Chúng tôi kết luận rằng

$$TX_{kj}\ pj = pacc\ (j, k)\ Tkj\ pj \tag{7.74}$$

$$= Tkj\ pj\ \text{phút}\ \frac{\text{Chặc chặc}}{Tkj}, 1 \tag{7,75}$$

$$pj = \min\{Tjk\ pk, Tkj\ pj\} \tag{7.76}$$

$$= Tjk\ pk\ \text{tối thiểu}\ 1, \frac{tkj\ pj}{Tjk\ pk} \tag{7.77}$$

$$= pacc\ (k, j)\ Tjk\ pk \tag{7.78}$$

$$= TX_{jk}\ pk. \tag{7.79}$$

□

Ví dụ sau đây được lấy từ bài báo chuyên đề của Hastings Phương pháp lấy mẫu Monte Carlo bằng chuỗi Markov và ứng dụng của chúng.

Ví dụ 7.5.3 (Tạo biến ngẫu nhiên Poisson). Mục đích của chúng ta là tạo ra một biến ngẫu nhiên Poisson X. Lưu ý rằng chúng ta không cần biết hằng số chuẩn hóa trong Poisson miễn pmf, tương đương với e ‘ _ là chúng ta biết rằng nó tỷ lệ với

$$pX\ (x)\quad \frac{\lambda\ x}{x!} \tag{7.80}$$

Chuỗi Markov phụ trợ phải có khả năng đạt tới bất kỳ giá trị nào có thể có của X, tức là tất cả các số nguyên dương. Chúng tôi sẽ sử dụng bước đi ngẫu nhiên đã sửa đổi để thực hiện các bước lên và xuống với xác suất 1/2, nhưng không bao giờ xuống dưới 0. Ma trận chuyển tiếp của nó bằng

$$tks := \begin{cases} \frac{1}{2} & \text{nếu } j = 0 \text{ và } k = 0, \\ \frac{1}{2} & \text{nếu } k = j + \\ \frac{1}{2} & 1, \text{nếu } j > 0 \text{ và } k = j \\ & 1, 0 \text{ nếu không.} \end{cases} \tag{7.81}$$

T là đối xứng nên xác suất chấp nhận bằng tỷ lệ của pmfs:

$$p_{acc}(j, k) := \text{tối thiểu}\left(\frac{T_{jk}\, p_X(k)}{T_{kj}\, p_X(j)}, 1\right) \tag{7.82}$$

$$= \text{tối thiểu}\left(\frac{p_X(k)}{p_X(j)}, 1\right). \tag{7.83}$$

Để tính xác suất chấp nhận, chúng tôi chỉ xem xét các chuyển đổi có thể xảy ra trong bước đi ngẫu nhiên. Nếu j = 0 và k = 0

$$p_{acc}(j,k)=1. \tag{7.84}$$

Nếu k = j + 1

$$p_{acc}(j, j+1) = \text{tối thiểu}\left(\frac{\frac{\lambda^{j+1}}{(j+1)!}}{\frac{\lambda^j}{j!}}, 1\right) \tag{7.85}$$

$$= \text{tối thiểu}\left(\frac{\lambda}{j+1}, 1\right). \tag{7.86}$$

Nếu k = j 1

$$p_{acc}(j, j\ 1) = \text{tối thiểu}\left(\frac{\frac{\lambda^{j\ 1}}{(j\ 1)!}}{\frac{\lambda^j}{j!}}, 1\right) \tag{7.87}$$

$$= \text{tối thiểu}\left(\frac{j}{\lambda}, 1\right). \tag{7.88}$$

Bây giờ chúng ta giải thích các bước của phương pháp Metropolis-Hastings. Để mô phỏng bước đi ngẫu nhiên phụ trợ, chúng tôi sử dụng một chuỗi các biến ngẫu nhiên Bernoulli cho biết bước đi ngẫu nhiên đang cố gắng đi lên hay đi xuống (hoặc giữ nguyên ở mức 0). Chúng tôi khởi tạo chuỗi tại $x_0 = 0$. Sau đó, với i = 1, 2, . . ., chúng tôi

- Tạo mẫu b từ phân bố Bernoulli với tham số 1/2 và mẫu u phân bố đều trong [0, 1].

- Nếu b = 0:

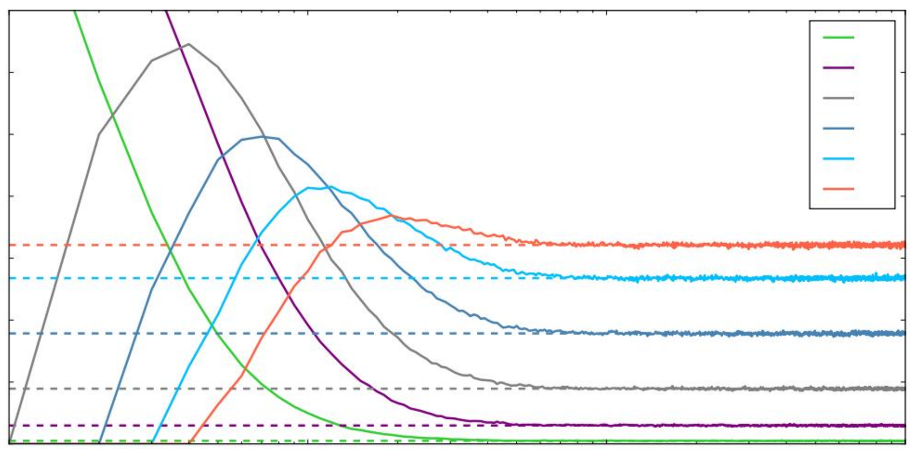
  - Nếu $x_{i\ 1} = 0$, $x_i := 0$.
  - Nếu $x_{i\ 1} > 0$:

    Nếu bạn < $\frac{x_{i\ 1}}{\lambda}$, $x_i := x_{i\ 1}\ \ 1$.

    Ngược lại $x_i := x_{i\ 1}$.

- Nếu b = 1:

  - Nếu u < $\frac{\lambda}{x_{i\ 1} + 1}$, $x_i := x_{i\ 1}+1$.
  - Ngược lại $x_i := x_{i\ 1}$.



Hình 7.8: Sự hội tụ trong phân phối của chuỗi Markov được xây dựng trong Ví dụ 7.8 cho λ := 6. Để tránh lộn xộn, chúng tôi chỉ vẽ biểu đồ phân phối theo kinh nghiệm của 6 trạng thái, được tính bằng cách chạy chuỗi Markov 104 lần.

Chuỗi Markov mà chúng tôi đã xây dựng là không thể rút gọn: có xác suất khác 0 để đi từ bất kỳ số nguyên không âm nào sang bất kỳ số nguyên không âm nào khác (mặc dù có thể mất một lúc!). Chúng ta chưa thực sự chứng minh rằng chuỗi sẽ hội tụ đến phân bố mong muốn, vì chúng ta chưa thảo luận về sự hội tụ của chuỗi Markov với không gian trạng thái vô hạn, nhưng Hình 7.8 cho thấy phương pháp này thực sự cho phép lấy mẫu từ phân bố Poisson với λ := 6.

Đối với ví dụ trong Hình 7.8, sự hội tụ gần đúng trong phân bố xảy ra sau khoảng 100 lần lặp. Đây được gọi là thời gian trộn của chuỗi Markov. Để giải thích cho vấn đề này, các phương pháp MCMC thường loại bỏ các mẫu khỏi chuỗi trong một khoảng thời gian ban đầu được gọi là thời gian lưu trữ.

Người đọc cẩn thận có thể thắc mắc về điểm sử dụng các phương pháp MCMC nếu chúng ta đã có quyền truy cập vào bản phân phối mong muốn. Thay vào đó, có vẻ đơn giản hơn nhiều nếu chỉ áp dụng phương pháp được mô tả trong Phần 2.6.1 . Tuy nhiên, phương pháp Metropolis-Hastings có thể được áp dụng cho các bản phân phối rời rạc với các hỗ trợ vô hạn và cả cho các bản phân phối liên tục (việc chứng minh điều này nằm ngoài phạm vi của các ghi chú này). Điều quan trọng, trái ngược với lấy mẫu từ chối và biến đổi nghịch đảo, Metropolis-Hastings không yêu cầu quyền truy cập vào pmf pX hoặc pdf fX của phân phối mục tiêu, mà thay vào đó là tỷ lệ pX (x) / pX (y) hoặc fX (x) /fX (y) với mọi x = y. Điều này rất hữu ích khi tính toán các phân phối có điều kiện trong các mô hình xác suất.

Hãy tưởng tượng rằng chúng ta có quyền truy cập vào phân phối cận biên của một biến ngẫu nhiên liên tục A và phân phối có điều kiện của một biến ngẫu nhiên liên tục B khác cho trước A. Tính toán điều kiện pdf

$$f_{A|B}(a|b) = \frac{f_A(a)\, f_{B|A}(b|a)}{\int_{u=-\infty}^{\infty} f_A(u)\, f_{B|A}(b|u)\, du} \tag{7.89}$$

là không cần thiết khả thi do tích phân trong mẫu số. Tuy nhiên, nếu chúng ta áp dụng Metropolis Hastings để lấy mẫu từ $f_{A|B}$ thì chúng ta không cần tính hệ số chuẩn hóa vì đối với bất kỳ $a_1 = a_2$

$$\frac{f_{A|B}(a_1|b)}{f_{A|B}(a_2|b)} = \frac{f_A(a_1)\, f_{B|A}(b|a_1)}{f_A(a_2)\, f_{B|A}(b|a_2)}. \tag{7.90}$$

Chương 8

Thống kê mô tả

Trong chương này, chúng tôi mô tả một số kỹ thuật để trực quan hóa dữ liệu, cũng như để tính toán các đại lượng tóm tắt dữ liệu một cách hiệu quả. Số lượng như vậy được gọi là thống kê mô tả. Như chúng ta sẽ thấy trong các chương sau, những số liệu thống kê này thường có thể được giải thích trong khuôn khổ xác suất, nhưng chúng cũng hữu ích khi các giả định xác suất không được đảm bảo. Vì điều này, chúng tôi trình bày chúng từ quan điểm tất định.

8.1 Biểu đồ

Chúng tôi bắt đầu bằng cách xem xét các tập dữ liệu chứa dữ liệu một chiều. Một trong những cách tự nhiên nhất để trực quan hóa dữ liệu 1D là vẽ biểu đồ của chúng. Biểu đồ thu được bằng cách chia thành nhóm dữ liệu và đếm số lượng trường hợp nằm trong mỗi ngăn. Chiều rộng của các ngăn là một tham số có thể được điều chỉnh để mang lại độ phân giải cao hơn hoặc thấp hơn. Nếu chúng ta diễn giải dữ liệu tương ứng với các mẫu từ một biến ngẫu nhiên, thì biểu đồ sẽ là hằng số gần đúng từng phần với pmf hoặc pdf của chúng.

Hình 8.1 cho thấy hai biểu đồ được tính toán từ dữ liệu nhiệt độ được thu thập tại một trạm thời tiết ở Oxford trong hơn 150 năm.[1] Mỗi điểm dữ liệu biểu thị nhiệt độ tối đa được ghi lại vào tháng Giêng hoặc tháng Tám của một năm cụ thể. Hình 8.2 thể hiện biểu đồ GDP bình quân đầu người của tất cả các quốc gia trên thế giới năm 2014 theo Liên Hợp Quốc.[2]

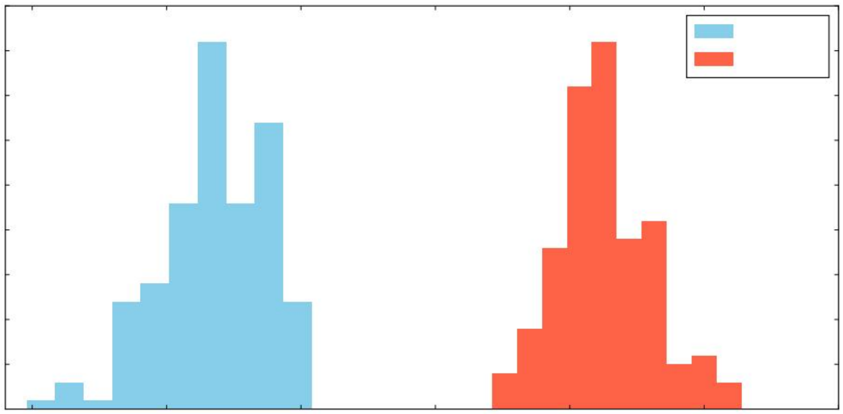
8.2 Trung bình mẫu và phương sai

Tính trung bình các phần tử trong tập dữ liệu một chiều cung cấp một bản tóm tắt dữ liệu bằng một con số, là đối chiếu xác định với giá trị trung bình của một biến ngẫu nhiên (hãy nhớ rằng chúng ta không đưa ra giả định xác suất nào trong chương này). Điều này có thể được mở rộng cho dữ liệu đa chiều bằng cách lấy trung bình trên mỗi chiều riêng biệt.

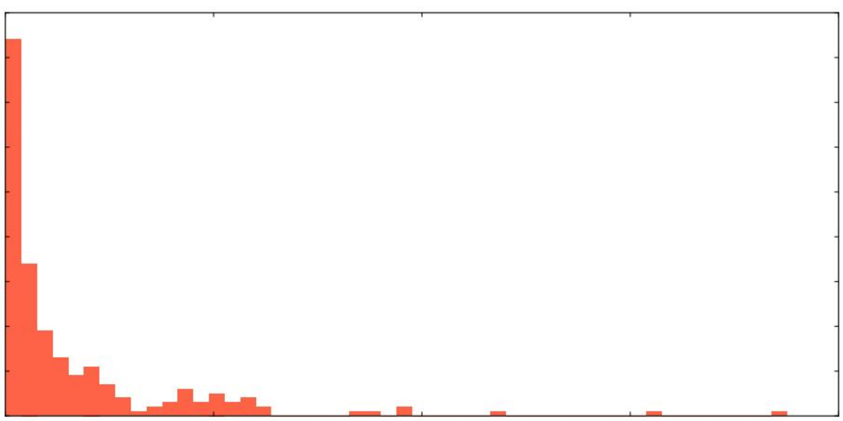
Định nghĩa 8.2.1 (Trung bình mẫu). Cho $\{x_1, x_2, . . . , x_n\}$ là tập dữ liệu có giá trị thực. Mẫu

[1] Dữ liệu có tại http://www.metoffice.gov.uk/pub/data/weather/uk/climate/stationdata/oxforddata.txt.

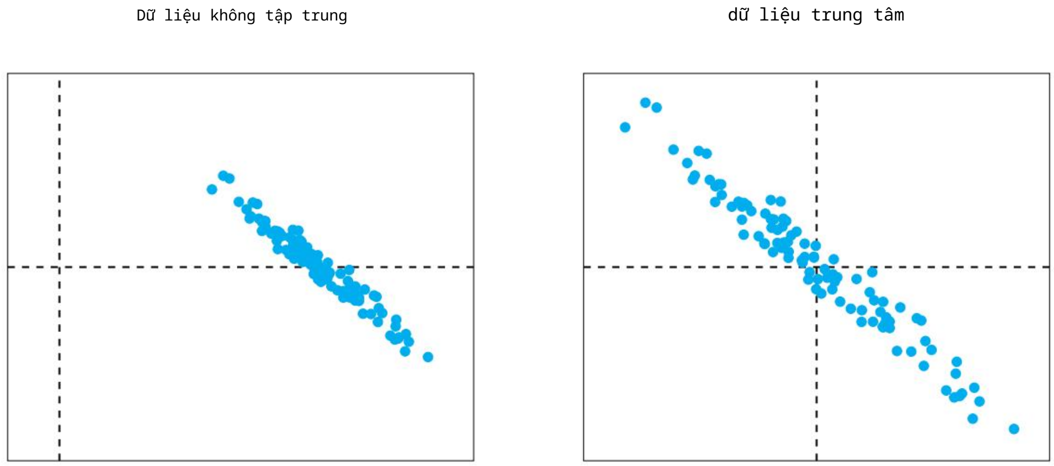
[2] Dữ liệu có tại http://unstats.un.org/unsd/snaama/selbasicFast.asp.



Hình 8.1: Biểu đồ dữ liệu nhiệt độ được lấy tại một trạm thời tiết ở Oxford trong hơn 150 năm. Mỗi điểm dữ liệu tương đương với nhiệt độ tối đa được ghi lại trong một tháng nhất định trong một năm cụ thể.



Hình 8.2: Biểu đồ GDP bình quân đầu người của các nước trên thế giới năm 2014.



Hình 8.3: Hiệu ứng căn giữa tập dữ liệu hai chiều. Các trục được mô tả bằng các đường đứt nét.

trung bình của dữ liệu được định nghĩa là

$$\mathrm{av}(x_1, x_2, \ldots, x_n) := \frac{1}{N}\sum_{\text{tôi}=1}^{N} x_i . \tag{8.1}$$

Cho $\{x_1, x_2, \ldots, x_n\}$ là một tập hợp các vectơ dữ liệu có giá trị thực d chiều. trung bình mẫu là

$$\mathrm{av}(x_1, x_2, \ldots, x_n) := \frac{1}{N}\sum_{\text{tôi}=1}^{N} x_i . \tag{8.2}$$

Giá trị trung bình mẫu của dữ liệu trong Hình 8.1 là 6,73 C vào tháng 1 và 21,3 C vào tháng 8. Giá trị trung bình mẫu của GDP bình quân đầu người trong Hình 8.2 là 16.500 USD.

Về mặt hình học, giá trị trung bình, còn được gọi là giá trị trung bình của mẫu, là trọng tâm của dữ liệu. Một bước tiền xử lý phổ biến trong phân tích dữ liệu là căn giữa một tập hợp dữ liệu bằng cách trừ đi giá trị trung bình mẫu của nó. Hình 8.3 cho thấy một ví dụ.

Thuật toán 8.2.2 (Định tâm). Cho $x_1, \ldots, x_n$ là một tập dữ liệu d-chiều. Để căn giữa tập dữ liệu, chúng tôi:

1. Tính trung bình mẫu theo Định nghĩa 8.2.1.
2. Trừ đi giá trị trung bình của mẫu từ mỗi vectơ dữ liệu. Với $1 \leq i \leq n$

$$y_i := x_i \quad \mathrm{av}(x_1, x_2, \ldots, x_n). \tag{8.3}$$

Tập dữ liệu kết quả $y_1, \ldots, y_n$ có trung bình mẫu bằng 0; nó có tâm tại gốc tọa độ.

Phương sai mẫu là giá trị trung bình của các bình phương độ lệch so với giá trị trung bình của mẫu. Về mặt hình học, nó định lượng sự thay đổi trung bình của tập dữ liệu xung quanh tâm của nó. Nó là một đối trọng xác định với phương sai của một biến ngẫu nhiên.

Định nghĩa 8.2.3 (Phương sai mẫu và độ lệch chuẩn). Cho $\{x_1, x_2, . . . , x_n\}$ là tập dữ liệu có giá trị thực. Phương sai mẫu được định nghĩa là

$$\mathrm{var}(x_1, x_2, . . . , x_n) := \frac{1}{n-1} \sum_{i=1}^{N} (x_i - \mathrm{av}(x_1, x_2, . . . , x_n))^2 \tag{8.4}$$

Độ lệch chuẩn mẫu là căn bậc hai của phương sai mẫu

$$\mathrm{std}(x_1, x_2, . . . , x_n) := \sqrt{\mathrm{var}(x_1, x_2, . . . , x_n)}. \tag{8,5}$$

Bạn có thể thắc mắc tại sao hằng số chuẩn hóa là $1/(n-1)$ thay vì $1/n$. Lý do là điều này đảm bảo rằng kỳ vọng của phương sai mẫu bằng với phương sai thực khi dữ liệu là iid (xem Bổ đề 9.2.5). Trong thực tế, không có nhiều khác biệt giữa hai chuẩn hóa.

Độ lệch chuẩn mẫu của dữ liệu nhiệt độ trong Hình 8.1 là 1,99 C vào tháng 1 và 1,73 C vào tháng 8. Độ lệch chuẩn mẫu của dữ liệu GDP trong Hình 8.2 là $25.300.

8.3 Thống kê đơn hàng

Trong một số trường hợp, một tập dữ liệu được mô tả tốt bằng giá trị trung bình và độ lệch chuẩn của nó.

Vào tháng Giêng, nhiệt độ ở Oxford vào khoảng 6,73 C hoặc giảm 2 C.

Đây là một tài khoản khá chính xác về dữ liệu nhiệt độ từ phần trước. Tuy nhiên, hãy tưởng tượng rằng ai đó mô tả tập dữ liệu GDP trong Hình 8.2 là:

Các quốc gia thường có GDP bình quân đầu người khoảng 16 500 đô la cho hoặc nhận 25 300 đô la.

Mô tả này là khá khủng khiếp. Vấn đề là hầu hết các quốc gia có GDP bình quân đầu người rất nhỏ, trong khi một số ít có GDP thực sự lớn và giá trị trung bình của mẫu và độ lệch chuẩn không thực sự truyền đạt thông tin này. Thống kê đơn hàng cung cấp một mô tả thay thế, mô tả này thường có nhiều thông tin hơn khi có các giá trị cực đoan trong dữ liệu.

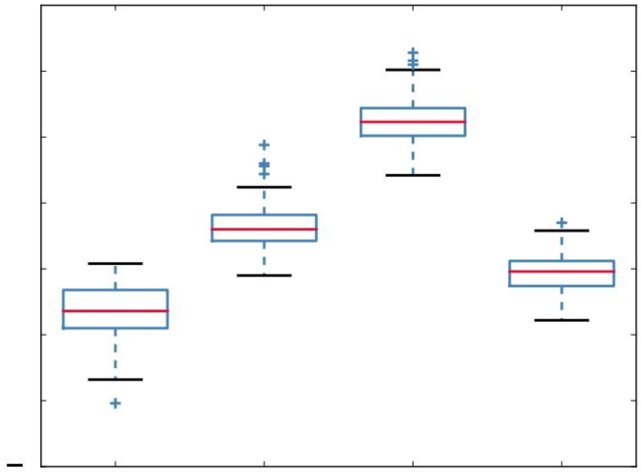
Định nghĩa 8.3.1 (Số lượng tử và phân vị). Đặt $x_{(1)} \leq x_{(2)} \leq . . . \leq x_{(n)}$ biểu thị các phần tử có thứ tự của tập dữ liệu $\{x_1, x_2, . . . , x_n\}$. Phân vị q của dữ liệu cho $0 < q < 1$ là $x_{([q(n+1)])}$, trong đó $[q(n+1)]$ là kết quả của việc làm tròn $q(n+1)$ thành số nguyên gần nhất. Phân vị 100 p được gọi là phân vị p.

Các phân vị 0,25 và 0,75 được gọi là phần tư thứ nhất và thứ ba, trong khi lượng phân vị 0,5 được gọi là trung vị mẫu. Một phần tư dữ liệu nhỏ hơn phân vị 0,25, một nửa nhỏ hơn (hoặc lớn hơn) so với trung vị và ba phần tư nhỏ hơn phân vị 0,75. Nếu n chẵn, trung vị mẫu thường được đặt thành

$$\frac{x_{(n/2)} + x_{(n/2+1)}}{2}. \tag{8.6}$$

Sự khác biệt giữa phần tư thứ ba và phần tư thứ nhất được gọi là phạm vi liên phần tư (IQR).

Hóa ra, đối với tập dữ liệu nhiệt độ trong Hình 8.1, giá trị trung bình của mẫu là 6,80 C vào tháng 1 và 21,2 C vào tháng 8, về cơ bản giống với giá trị trung bình của mẫu. IQR là



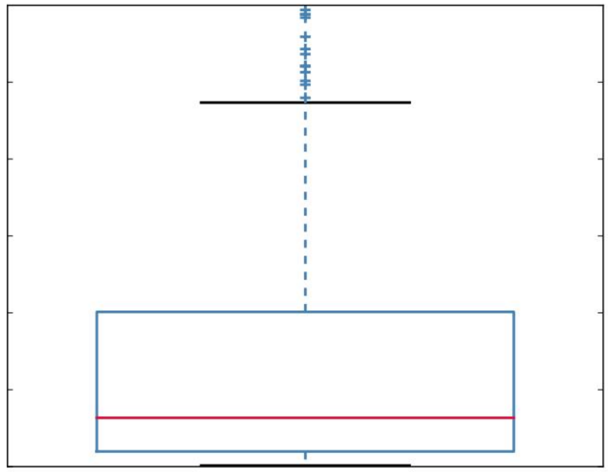
Hình 8.4: Biểu đồ hộp của bộ dữ liệu nhiệt độ Oxford được sử dụng trong Hình 8.1. Mỗi ô vuông tương ứng với nhiệt độ tối đa trong một tháng cụ thể (tháng 1, tháng 4, tháng 8 và tháng 11) trong 150 năm qua.

2,9 C vào tháng Giêng và 2,1 C vào tháng Tám. Điều này tạo ra một mức chênh lệch rất giống nhau xung quanh giá trị trung bình, giống như giá trị trung bình của mẫu. Trong ví dụ cụ thể này, dường như không có lợi thế trong việc sử dụng thống kê đơn hàng.

Đối với tập dữ liệu GDP, giá trị trung bình là $6.350. Điều này có nghĩa là một nửa số quốc gia có GDP dưới 6.350 USD. Ngược lại, 71% quốc gia có GDP bình quân đầu người thấp hơn mức trung bình của mẫu! IQR của những dữ liệu này là $18,200. Để cung cấp mô tả đầy đủ hơn về tập dữ liệu, chúng tôi có thể liệt kê một bản tóm tắt thống kê đơn hàng gồm năm số: $x_{(1)}$ nhỏ nhất, phần tư thứ nhất, trung vị mẫu, phần tư thứ ba và $x_{(n)}$ tối đa . Đối với tập dữ liệu GDP, các giá trị này lần lượt là $130, $1.960, $6.350, $20.100 và $188.000.

Chúng ta có thể hình dung số liệu thống kê thứ tự chính của một tập dữ liệu bằng cách sử dụng biểu đồ hộp, biểu đồ này hiển thị giá trị trung bình của dữ liệu được đặt trong một hộp. Phần dưới cùng và trên cùng của hộp là phần tư thứ nhất và thứ ba. Cách trực quan hóa tập dữ liệu này được đề xuất bởi nhà toán học John Tukey. Cốt truyện hộp của Tukey cũng bao gồm râu ria. Râu dưới là một đường kéo dài từ đáy hộp đến giá trị nhỏ nhất trong phạm vi 1,5 IQR của phần tư đầu tiên. Râu cao hơn kéo dài từ trên cùng của hộp đến giá trị cao nhất trong phạm vi 1,5 IQR của phần tư thứ ba. Các giá trị ngoài râu ria được coi là giá trị ngoại lệ và được vẽ riêng.

Hình 8.4 áp dụng các biểu đồ hộp để trực quan hóa bộ dữ liệu nhiệt độ được sử dụng trong Hình 8.1. Mỗi ô vuông tương ứng với nhiệt độ tối đa trong một tháng cụ thể (tháng 1, tháng 4, tháng 8 và tháng 11) trong 150 năm qua. Các ô hình hộp cho phép chúng ta nhanh chóng so sánh sự chênh lệch nhiệt độ trong các tháng khác nhau. Hình 8.5 thể hiện biểu đồ hộp của dữ liệu GDP từ Hình 8.2. Từ biểu đồ hình hộp, có thể thấy ngay rằng hầu hết các quốc gia có GDP bình quân đầu người rất nhỏ, chênh lệch giữa các quốc gia tăng đối với GDP bình quân đầu người lớn hơn và một số ít quốc gia có GDP bình quân đầu người rất lớn.



Hình 8.5: Biểu đồ hình hộp về GDP bình quân đầu người của tất cả các quốc gia trên thế giới năm 2014. Không phải tất cả các giá trị ngoại lệ đều được hiển thị.

8.4 Hiệp phương sai mẫu

Trong các phần trước, chúng ta chủ yếu xem xét các tập dữ liệu bao gồm dữ liệu một chiều (ngoại trừ khi chúng ta thảo luận về giá trị trung bình mẫu của tập dữ liệu đa chiều). Trong biệt ngữ học máy, chỉ có một tính năng cho mỗi điểm dữ liệu. Bây giờ chúng tôi nghiên cứu một kịch bản đa chiều, trong đó có một số tính năng được liên kết với từng điểm dữ liệu.

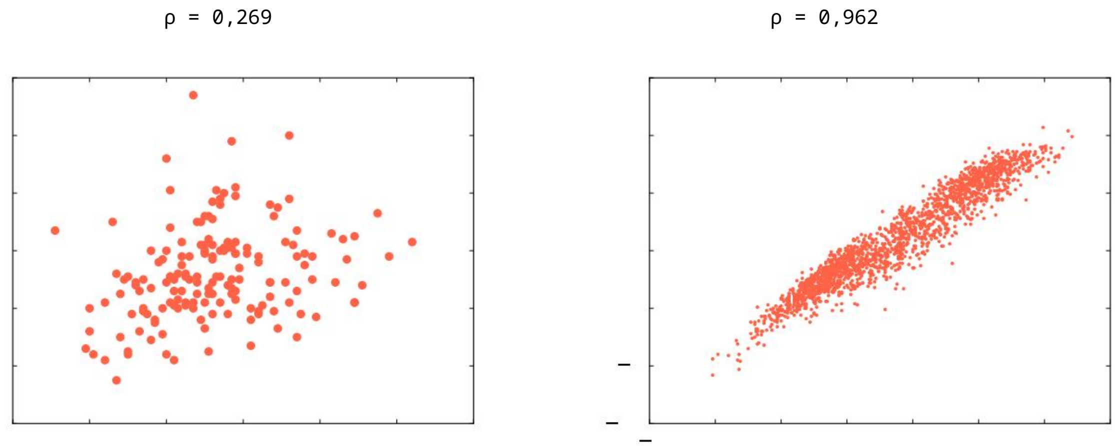
Nếu kích thước của tập dữ liệu bằng hai (tức là có hai tính năng cho mỗi điểm dữ liệu), chúng ta có thể trực quan hóa dữ liệu bằng biểu đồ phân tán, trong đó mỗi trục biểu thị một trong các tính năng. Hình 8.6 cho thấy một số đồ thị phân tán của dữ liệu nhiệt độ. Những dữ liệu này giống như trong Hình 8.1, nhưng bây giờ chúng ta đã sắp xếp chúng để tạo thành các tập dữ liệu hai chiều. Trong biểu đồ bên trái, một chiều tương ứng với nhiệt độ vào tháng 1 và chiều còn lại là nhiệt độ vào tháng 8 (có một điểm dữ liệu mỗi năm). Trong biểu đồ bên phải, một thứ nguyên biểu thị nhiệt độ tối thiểu trong một tháng cụ thể và thứ nguyên kia biểu thị nhiệt độ tối đa trong cùng tháng đó (có một điểm dữ liệu mỗi tháng).

Hiệp phương sai mẫu định lượng xem trung bình hai tính năng của tập dữ liệu hai chiều có xu hướng thay đổi theo cách giống nhau hay không, giống như hiệp phương sai định lượng biến thể chung dự kiến của hai biến ngẫu nhiên.

Định nghĩa 8.4.1 (Hiệp phương sai mẫu). Giả sử $\{(x_1, y_1),(x_2, y_2), \ldots ,(x_n, y_n)\}$ là tập dữ liệu trong đó mỗi ví dụ bao gồm phép đo của hai đối tượng địa lý khác nhau. Hiệp phương sai mẫu được định nghĩa là

$$\mathrm{cov}\left((x_1, y_1), \ldots ,(x_n, y_n)\right) := \frac{1}{n-1}\sum_{\text{tôi}=1}^{N} (x_i - \mathrm{av}(x_1, \ldots , x_n))\,(y_i - \mathrm{av}(y_1, \ldots , y_n)). \tag{8.7}$$

Để tính đến việc mỗi đối tượng riêng lẻ có thể khác nhau ở các thang đo khác nhau, bước tiền xử lý thông thường là chuẩn hóa từng đối tượng, chia cho độ lệch chuẩn mẫu của nó.



Hình 8.6: Biểu đồ phân tán nhiệt độ tháng 1 và tháng 8 (trái) và nhiệt độ cao nhất, thấp nhất tháng (phải) ở Oxford trong 150 năm qua.

Nếu chúng ta chuẩn hóa trước khi tính toán hiệp phương sai, chúng ta sẽ thu được hệ số tương quan mẫu của hai tính năng. Một trong những ưu điểm của hệ số tương quan là chúng ta không cần phải lo lắng về đơn vị đo lường các đặc điểm. Ngược lại, việc đo lường một đặc điểm đại diện cho khoảng cách tính bằng inch hoặc dặm có thể làm biến dạng nghiêm trọng hiệp phương sai, nếu chúng ta không chia tỷ lệ đặc điểm khác cho phù hợp.

Định nghĩa 8.4.2 (Hệ số tương quan mẫu). Giả sử $\{(x_1, y_1),(x_2, y_2), \ldots,(x_n, y_n)\}$ là tập dữ liệu trong đó mỗi ví dụ bao gồm hai đối tượng địa lý. Hệ số tương quan mẫu được định nghĩa là

$$\rho((x_1, y_1), \ldots,(x_n, y_n)) := \frac{\text{cov}((x_1, y_1), \ldots,(x_n, y_n))}{\text{std}(x_1, \ldots, x_n)\,\text{std}(y_1, \ldots, y_n)}. \quad (8.8)$$

Theo bất đẳng thức Cauchy-Schwarz (Định lý B.2.4), phát biểu rằng với mọi vectơ a và b

$$1 \le \frac{a^T b}{||a||_2\, ||b||_2} \le 1, \quad (8.9)$$

độ lớn của hệ số tương quan mẫu được giới hạn bởi một. Nếu nó bằng 1 hoặc -1 thì hai tập dữ liệu được căn giữa thẳng hàng. Bất đẳng thức Cauchy-Schwarz có liên quan đến bất đẳng thức Cauchy-Schwarz đối với biến ngẫu nhiên (Định lý 4.3.7), nhưng ở đây nó áp dụng cho các vectơ xác định.

Hình 8.6 được chú thích các hệ số tương quan mẫu tương ứng với hai ô.
Nhiệt độ cao nhất và thấp nhất trong cùng một tháng có mối tương quan chặt chẽ với nhau, trong khi đó nhiệt độ cao nhất tháng I và tháng VIII cùng năm chỉ tương quan với nhau rất ít.

8.5 Ma trận hiệp phương sai mẫu

8.5.1 Định nghĩa

Bây giờ chúng tôi xem xét các bộ dữ liệu đa chiều. Đặc biệt, chúng tôi quan tâm đến việc phân tích sự thay đổi trong dữ liệu. Ma trận hiệp phương sai mẫu của tập dữ liệu chứa hiệp phương sai mẫu theo cặp giữa mọi cặp đối tượng địa lý.

Định nghĩa 8.5.1 (Ma trận hiệp phương sai mẫu). Cho $\{x_1, x_2, \ldots, x_n\}$ là một tập hợp các vectơ dữ liệu có giá trị thực d chiều. Ma trận hiệp phương sai mẫu của những dữ liệu này là ma trận $d \times d$

$$\Sigma(x_1, \ldots, x_n) := \frac{1}{n-1} \sum_{\text{tôi}=1}^{N} (x_i - \operatorname{av}(x_1, \ldots, x_n))(x_i - \operatorname{av}(x_1, \ldots, x_n))^T . \tag{8.10}$$

Mục nhập (i, j) của ma trận hiệp phương sai, trong đó $1 \le i, j \le d$, được cho bởi

$$\Sigma(x_1, \ldots, x_n)_{ij} = \begin{cases} \operatorname{var}((x_1)_i, \ldots, (x_n)_i) & \text{nếu tôi} = j, \\ \operatorname{cov}((x_1)_i, (x_1)_j, \ldots, (x_n)_i, (x_n)_j) & \text{nếu tôi} = j. \end{cases} \tag{8.11}$$

Để đặc trưng cho sự biến thiên của tập dữ liệu nhiều chiều xung quanh tâm của nó, ta xét sự biến thiên của nó theo các hướng khác nhau. Biến thể trung bình của dữ liệu theo một hướng nhất định được định lượng bằng phương sai mẫu của các phép chiếu của dữ liệu theo hướng đó.
Đặt v là một vectơ định mức đơn vị được căn chỉnh theo hướng quan tâm, phương sai mẫu của tập dữ liệu theo hướng của v được cho bởi

$$\operatorname{var}(v^T x_1, \ldots, v^T x_n) = \frac{1}{n-1} \sum_{\text{tôi}=1}^{N} \left(v^T x_i - \operatorname{av}(v^T x_1, \ldots, v^T x_n)\right)^2 \tag{8.12}$$

$$= \frac{1}{n-1} \sum_{\text{tôi}=1}^{N} \left(v^T (x_i - \operatorname{av}(x_1, \ldots, x_n))\right)^2 \tag{8.13}$$

$$= v^T \left(\frac{1}{n-1} \sum_{\text{tôi}=1}^{N} (x_i - \operatorname{av}(x_1, \ldots, x_n))(x_i - \operatorname{av}(x_1, \ldots, x_n))^T \right) v$$

$$= v^T \Sigma(x_1, \ldots, x_n) v. \tag{8.14}$$

Sử dụng ma trận hiệp phương sai mẫu, chúng ta có thể biểu thị sự thay đổi theo mọi hướng! Đây là một sự tương tự xác định của thực tế là ma trận hiệp phương sai của một vectơ ngẫu nhiên mã hóa phương sai của nó theo mọi hướng.

8.5.2 Phân tích thành phần chính

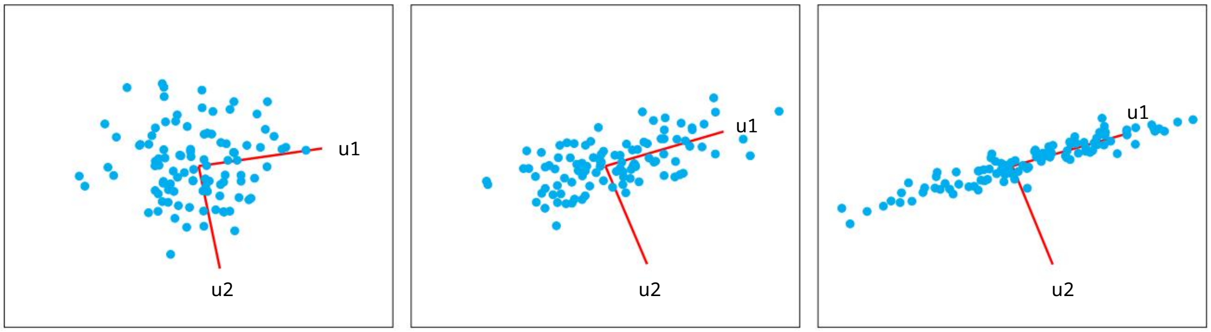
Xem xét sự phân hủy riêng của ma trận hiệp phương sai

$$\Sigma(x_1, \ldots, x_n) = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}^T . \tag{8.15}$$

Theo định nghĩa, $\Sigma(x_1, \ldots, x_n)$ đối xứng, do đó các vectơ riêng của nó $u_1, u_2, \ldots, u_n$ là trực giao. Theo phương trình (8.14) và Định lý B.7.2, các vectơ riêng và giá trị riêng đặc trưng hoàn toàn cho sự biến thiên của dữ liệu theo mọi hướng.

$\lambda 1/n = 0{,}497$ $\lambda 1/n = 0{,}967$ $\lambda 1/n = 1{,}820$

$\lambda 2/n = 0{,}476$ $\lambda 2/n = 0{,}127$ $\lambda 2/n = 0{,}021$



Hình 8.7: PCA của một tập hợp bao gồm n = 100 điểm dữ liệu hai chiều với các cấu hình khác nhau.

Định lý 8.5.2. Giả sử hiệp phương sai mẫu của một tập hợp các vectơ Σ (x1, . . . , xn) có thành phần riêng được cho bởi (8.15) trong đó các giá trị riêng được sắp xếp theo thứ tự λ1 ≥ λ2 ≥ . . . ≥ λn. Sau đó,

$$\lambda 1 = \text{tối đa}_{||v||2=1} \text{ var } v^t x1, \ . \ . \ . \ , \ v^t xn \ , \tag{8.16}$$

$$u1 = \text{đối số tối đa}_{||v||2=1} \text{ var } v^t x1, \ . \ . \ . \ , \ v^t xn \ , \tag{8.17}$$

$$\lambda k = \text{tối đa}_{||v||2=1,u \ u1,...,uk \ 1} \text{ var } v^t x1, \ . \ . \ . \ , \ v^t xn \ , \tag{8.18}$$

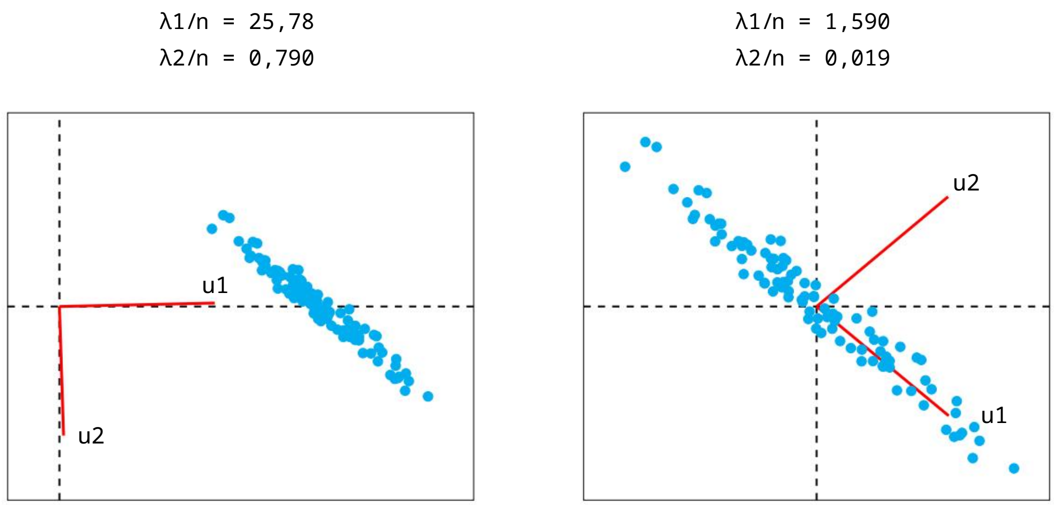
$$uk = \arg \text{tối đa}_{||v||2=1,u \ u1,...,uk \ 1} \text{ var } v^t x1, \ . \ . \ . \ , \ v^t xn \ . \tag{8.19}$$

Điều này có nghĩa là u1 là hướng biến thiên cực đại. Vectơ riêng u2 tương ứng đến giá trị riêng lớn thứ hai λ2 là hướng biến thiên cực đại trực giao đến u1. Nói chung, vectơ riêng uk tương ứng với giá trị riêng lớn thứ k λk tiết lộ hướng dao động cực đại trực giao với u1, u2, . . . , uk 1. Cuối cùng, un là hướng biến thiên nhỏ nhất.

Trong phân tích dữ liệu, các vectơ riêng của ma trận hiệp phương sai mẫu thường được gọi là hiệu trưởng hướng. Tính toán các vectơ riêng này để định lượng sự thay đổi của một tập dữ liệu theo các cách khác nhau hướng được gọi là phân tích thành phần chính (PCA). Hình 8.7 cho thấy hiệu trưởng hướng dẫn cho một số ví dụ 2D.

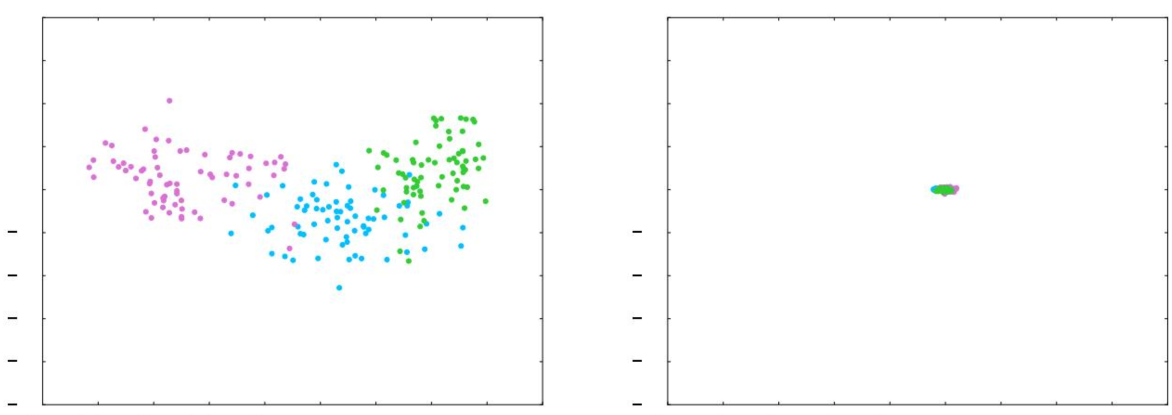
Hình 8.8 minh họa tầm quan trọng của việc định tâm trước khi áp dụng PCA. Định lý 8.5.2 vẫn đúng nếu dữ liệu không được căn giữa. Tuy nhiên, độ chuẩn của phép chiếu lên một hướng nào đó không còn phản ánh sự biến đổi của dữ liệu. Thực tế, nếu dữ liệu được tập trung xung quanh một điểm xa gốc tọa độ, hướng chính thứ nhất có xu hướng trùng với điểm đó. Cái này có ý nghĩa khi chiếu vào hướng đó sẽ thu được nhiều năng lượng hơn. Kết quả là, hiệu trưởng hướng không phản ánh hướng biến đổi tối đa trong đám mây dữ liệu. định tâm tập dữ liệu trước khi áp dụng PCA sẽ giải quyết vấn đề.

Ví dụ sau đây giải thích cách áp dụng phân tích thành phần chính để giảm kích thước. Động lực là trong nhiều trường hợp, các hướng biến đổi cao hơn có nhiều thông tin hơn về cấu trúc của tập dữ liệu.

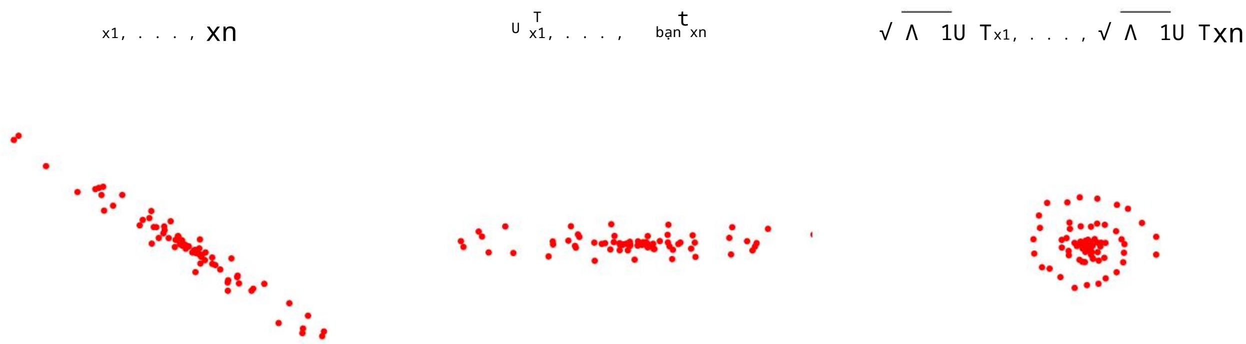


Dữ liệu không tập trung dữ liệu trung tâm

Hình 8.8: PCA áp dụng cho n = 100 điểm dữ liệu 2D. Ở bên trái, dữ liệu không được căn giữa. Kết quả là hướng chính chi phối u1 nằm theo hướng trung bình của dữ liệu và PCA không phản ánh cấu trúc thực tế. Khi chúng ta căn giữa, u1 sẽ được căn chỉnh theo hướng biến thiên cực đại.



Hình 8.9: Phép chiếu vectơ 7 chiều mô tả các hạt lúa mì khác nhau lên hai hạt đầu tiên (trái) và hai hướng chính (phải) cuối cùng của tập dữ liệu. Mỗi màu đại diện cho một loại lúa mì.



Hình 8.10: Hiệu ứng làm trắng một bộ dữ liệu. Dữ liệu gốc bị chi phối bởi một độ lệch tuyến tính (trái). Áp dụng U^T căn chỉnh các trục với các vectơ riêng của ma trận hiệp phương sai mẫu (ở giữa). Cuối cùng, $\sqrt{\Lambda}^{-1}$ cân nhắc lại dữ liệu dọc theo các trục đó sao cho chúng có cùng độ biến thiên trung bình, cho thấy cấu trúc phi tuyến tính bị che khuất bởi độ lệch tuyến tính (bên phải).

Ví dụ 8.5.3 (Giảm kích thước thông qua PCA). Chúng tôi xem xét một tập dữ liệu trong đó mỗi điểm dữ liệu tương ứng với một hạt giống có bảy đặc điểm: diện tích, chu vi, độ nén, chiều dài của hạt nhân, chiều rộng của hạt nhân, hệ số bất đối xứng và chiều dài của rãnh hạt nhân. Các hạt thuộc về ba loại lúa mì khác nhau: Kama, Rosa và Canadian.[3] Mục đích của chúng tôi là trực quan hóa dữ liệu bằng cách chiếu dữ liệu xuống hai chiều theo cách bảo tồn càng nhiều biến thể càng tốt. Điều này có thể đạt được bằng cách chiếu từng điểm lên hai chiều chính đầu tiên của tập dữ liệu.

Hình 8.9 cho thấy hình chiếu của dữ liệu lên hai hướng chính đầu tiên và hai hướng chính cuối cùng. Trong trường hợp thứ hai, hầu như không có sự thay đổi rõ rệt. Cấu trúc của dữ liệu được bảo tồn tốt hơn nhiều theo hai hướng đầu tiên, cho phép hình dung rõ ràng sự khác biệt giữa ba loại hạt. Tuy nhiên, xin lưu ý rằng phép chiếu lên các hướng chính đầu tiên chỉ đảm bảo rằng chúng ta bảo toàn được nhiều biến thể nhất có thể, nhưng nó không nhất thiết phải bảo toàn các tính năng hữu ích cho các nhiệm vụ như phân loại.

8.5.3 Làm trắng

Làm trắng là một thủ tục hữu ích để tiền xử lý dữ liệu có chứa các mẫu phi tuyến tính. Mục tiêu là loại bỏ độ lệch tuyến tính trong dữ liệu bằng cách xoay và thu gọn dữ liệu theo các hướng khác nhau để làm lộ cấu trúc phi tuyến cơ bản của nó. Điều này có thể đạt được bằng cách áp dụng một phép biến đổi tuyến tính về cơ bản là đảo ngược ma trận hiệp phương sai mẫu để kết quả không tương quan. Quá trình này được gọi là làm trắng, bởi vì các vectơ ngẫu nhiên với các mục không tương quan thường được gọi là nhiễu trắng. Nó liên quan chặt chẽ với Thuật toán 8.5.4 để tô màu các vectơ ngẫu nhiên.

Thuật toán 8.5.4 (Làm trắng). Cho $x_1, \ldots, x_n$ là một tập hợp dữ liệu d chiều, mà chúng tôi giả định là được căn giữa và có ma trận hiệp phương sai hạng đầy đủ. Để làm trắng tập dữ liệu, chúng tôi:

1. Tính toán thành phần riêng của ma trận hiệp phương sai mẫu $\Sigma(x_1, \ldots, x_n) = U \Lambda U^T$.

[3] Bạn có thể tìm thấy dữ liệu tại https://archive.ics.uci.edu/ml/datasets/seeds.

2. Đặt $y_i := \sqrt{\Lambda}^{-1} U^T x_i$, với $i = 1, \ldots, n$, ở đâu

$$\sqrt{\Lambda} := \begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{bmatrix}, \tag{8.20}$$

sao cho $\Sigma(x_1, \ldots, x_n) = U \sqrt{\Lambda} \sqrt{\Lambda} U^T$.

Tập dữ liệu được làm trắng $y_1, \ldots, y_n$ có ma trận hiệp phương sai mẫu bằng đơn vị,

$$\Sigma(y_1, \ldots, y_n) := \frac{1}{n-1} \sum_{i=1}^{n} y_i y_i^T \tag{8.21}$$

$$= \frac{1}{n-1} \sum_{i=1}^{n} \sqrt{\Lambda}^{-1} U^T x_i \left(\sqrt{\Lambda}^{-1} U^T x_i \right)^T \tag{8.22}$$

$$= \sqrt{\Lambda}^{-1} U^T \left(\frac{1}{n-1} \sum_{i=1}^{n} x_i x_i^T \right) U \sqrt{\Lambda}^{-1} \tag{8.23}$$

$$= \sqrt{\Lambda}^{-1} U^T \Sigma(x_1, \ldots, x_n) U \sqrt{\Lambda}^{-1} \tag{8.24}$$

$$= \sqrt{\Lambda}^{-1} U^T U \sqrt{\Lambda} \sqrt{\Lambda} U^T U \sqrt{\Lambda}^{-1} \tag{8.25}$$

$$= I. \tag{8.26}$$

Theo trực giác, làm trắng trước tiên xoay dữ liệu và sau đó thu nhỏ hoặc mở rộng dữ liệu sao cho biến thể trung bình giống nhau theo mọi hướng. Kết quả là các mẫu phi tuyến tính trở nên rõ ràng hơn, như được minh họa trong Hình 8.10.

Chương 9

Thống kê thường xuyên

Mục tiêu của phân tích thống kê là trích xuất thông tin từ dữ liệu bằng cách tính toán số liệu thống kê, là các chức năng xác định của dữ liệu. Trong Chương 8 , chúng tôi mô tả một số thống kê từ quan điểm xác định và hình học, mà không đưa ra bất kỳ giả định nào về quá trình tạo dữ liệu. Điều này làm cho việc đánh giá tính chính xác của thông tin thu được trở nên rất khó khăn.

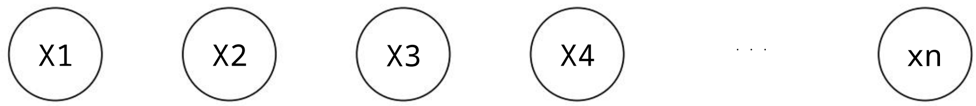
Trong chương này, chúng tôi lập mô hình quá trình thu thập dữ liệu theo xác suất. Điều này cho phép phân tích các kỹ thuật thống kê ly giải và rút ra những đảm bảo về mặt lý thuyết đối với hiệu suất của chúng. Dữ liệu được hiểu là sự thể hiện của các biến ngẫu nhiên, vectơ hoặc quy trình (tùy thuộc vào kích thước). Thông tin mà chúng tôi muốn trích xuất sau đó có thể được biểu thị dưới dạng phân phối chung của các đại lượng này. Chúng tôi coi phân phối này là không xác định nhưng cố định, theo quan điểm thường xuyên. Khung thay thế của thống kê Bayes được mô tả trong Chương 10.

9.1 Lấy mẫu phân bố đồng nhất độc lập

Trong chương này, chúng ta xem xét dữ liệu có giá trị thực một chiều, được mô hình hóa như là sự thực hiện của một chuỗi iid. Hình 9.1 mô tả mô hình đồ họa tương ứng. Đây là một giả định rất phổ biến, áp dụng cho các thử nghiệm có kiểm soát, chẳng hạn như các thử nghiệm ngẫu nhiên để thử nghiệm thuốc và thường có thể là một phép tính gần đúng trong các bối cảnh khác. Tuy nhiên, trong thực tế, điều cốt yếu là phải đánh giá xem các giả định về tính độc lập của một mô hình thực sự nắm giữ ở mức độ nào.

Ví dụ sau đây cho thấy rằng việc đo lường một đại lượng bằng cách lấy mẫu ngẫu nhiên một tập hợp con các cá thể từ một tổng thể lớn sẽ tạo ra dữ liệu thỏa mãn giả định iid, miễn là chúng ta lấy mẫu có thay thế (nếu tổng thể lớn, lấy mẫu mà không thay thế sẽ có tác động không đáng kể).

Ví dụ 9.1.1 (Lấy mẫu từ quần thể). Giả sử rằng chúng ta đang nghiên cứu một quần thể gồm m cá nhân. Chúng tôi quan tâm đến một số lượng nhất định liên quan đến mỗi người, ví dụ như mức cholesterol của họ, tiền lương của họ hoặc họ đang bỏ phiếu cho ai trong một cuộc bầu cử. Có k giá trị có thể có của đại lượng $\{z_1, z_2, . . . , z_k\}$, trong đó k có thể bằng m hoặc nhỏ hơn nhiều. Chúng tôi biểu thị bằng m_j số người có số lượng bằng z_j , $1 \leq j \leq k$. Trong trường hợp bầu cử có hai ứng cử viên, k sẽ bằng hai và m_1 và m_2 sẽ đại diện cho những người bỏ phiếu cho mỗi ứng cử viên.



Hình 9.1: Mô hình đồ thị có hướng tương ứng với dãy độc lập. Nếu trình tự cũng là phân phối đồng nhất thì $X_1, X_2, \ldots, X_n$ đều có cùng phân phối.

Chúng ta hãy giả sử rằng chúng ta chọn ngẫu nhiên n cá nhân một cách độc lập với sự thay thế, điều này có nghĩa là một cá nhân có thể được chọn nhiều lần và ghi lại giá trị của số lượng lãi. Theo các giả định này, các phép đo có thể được mô hình hóa thành một chuỗi ngẫu nhiên của các biến độc lập X. Vì xác suất chọn bất kỳ cá nhân nào là như nhau ở mọi khi chúng tôi thực hiện lựa chọn, pmf thứ tự đầu tiên của chuỗi là

$$p_{X(i)}(z_j) = P(\text{Số đo thứ i bằng } z_j) \tag{9.1}$$

$$= \frac{\text{Những người sao cho số lượng bằng } z_j}{\text{Tổng số người}} \tag{9.2}$$

$$= \frac{m_j}{\text{tôi}}, \qquad 1 \leq j \leq k, \tag{9.3}$$

cho $1 \leq i \leq n$ theo quy luật xác suất toàn phần. Chúng tôi kết luận rằng dữ liệu có thể được mô hình hóa như một thực hiện một chuỗi iid.

9.2 Sai số bình phương trung bình

Chúng tôi định nghĩa một công cụ ước tính là một hàm xác định của dữ liệu có sẵn $x_1, x_2, \ldots, x_n$ nào cung cấp một giá trị gần đúng cho một đại lượng liên quan đến phân phối tạo ra dữ liệu

$$y := h(x_1, x_2, \ldots, x_n). \tag{9.4}$$

Ví dụ, như chúng ta sẽ thấy, nếu chúng ta muốn ước tính kỳ vọng của phân phối cơ bản, một ước tính hợp lý là trung bình của dữ liệu. Vì chúng ta đang có quan điểm thường xuyên, số lượng quan tâm được mô hình hóa là tất định (ngược lại với quan điểm Bayesian sẽ mô hình hóa nó như một biến ngẫu nhiên). Đối với một tập dữ liệu cố định, công cụ ước tính là một yếu tố xác định chức năng của dữ liệu. Tuy nhiên, nếu chúng ta lập mô hình dữ liệu dưới dạng hiện thực của một chuỗi các biến ngẫu nhiên các biến, thì công cụ ước tính cũng là một hiện thực của biến ngẫu nhiên

$$Y := h(X_1, X_2, \ldots, X_n). \tag{9,5}$$

Điều này cho phép đánh giá công cụ ước tính theo xác suất (thường theo một số giả định về phân phối cơ bản). Chẳng hạn, chúng ta có thể đo lỗi do công cụ ước tính phát sinh bằng cách tính toán bình phương trung bình của sự khác biệt giữa công cụ ước tính và số lượng thực của quan tâm.

Định nghĩa 9.2.1 (Lỗi bình phương trung bình). Sai số bình phương trung bình (MSE) của công cụ ước tính Y mà xấp xỉ một đại lượng xác định $y \in \mathbb{R}$ là

$$\mathrm{MSE}(Y) := E\left((Y - y)^2\right). \tag{9.6}$$

MSE có thể được phân tách thành một thuật ngữ sai lệch và một thuật ngữ phương sai. Thuật ngữ sai lệch là sự khác biệt giữa số lượng quan tâm và giá trị mong đợi của công cụ ước tính. Thuật ngữ phương sai tương ứng với sự thay đổi của công cụ ước tính xung quanh giá trị dự kiến của nó.

Bổ đề 9.2.2 (Phân tách sai lệch phương sai). MSE của công cụ ước tính Y xấp xỉ y R thỏa mãn

$$\mathrm{MSE}(Y) = \underbrace{E\left(Y - E(Y)\right)^2}_{\text{phương sai}} + \underbrace{\left(E(Y) - y\right)^2}_{\text{Thiên kiến}}. \tag{9.7}$$

Bằng chứng. Bổ đề là hệ quả trực tiếp của tính tuyến tính của kỳ vọng. □

Nếu độ lệch bằng 0, thì công cụ ước tính bằng số lượng lãi trung bình.

Định nghĩa 9.2.3 (Ước lượng không chệch). Một ước lượng Y xấp xỉ y R là không chệch nếu độ chệch của nó bằng 0, tức là khi và chỉ khi

$$E(Y) = y. \tag{9.8}$$

Một công cụ ước tính có thể không chệch nhưng vẫn phải chịu sai số bình phương trung bình lớn do phương sai của nó.

Các bổ đề sau xác định rằng trung bình mẫu và phương sai là các ước lượng không chệch của trung bình thực và phương sai của một chuỗi iid các biến ngẫu nhiên.

Bổ đề 9.2.4 (Trung bình mẫu không chệch). Giá trị trung bình mẫu là một công cụ ước tính không chệch cho giá trị trung bình của một chuỗi iid các biến ngẫu nhiên.

Bằng chứng. Chúng tôi xem xét giá trị trung bình mẫu của chuỗi iid X với giá trị trung bình μ,

$$Y(n) := \frac{1}{N} \sum_{\text{tôi}=1}^{N} X(i). \tag{9.9}$$

Theo tuyến tính của kỳ vọng

$$EY(n) = \frac{1}{n} \sum_{i=1}^{N} EX(i) \tag{9.10}$$

$$= \mu. \tag{9.11}$$

□

Bổ đề 9.2.5 (Phương sai mẫu không chệch). Phương sai mẫu là một công cụ ước lượng không chệch của phương sai của một chuỗi iid các biến ngẫu nhiên.

Bằng chứng của kết quả này nằm trong Mục 9.7.1.

9.3 Tính nhất quán

Nếu chúng ta đang ước tính một đại lượng vô hướng, thì ước tính sẽ cải thiện khi chúng ta thu thập thêm dữ liệu. Lý tưởng nhất là ước tính sẽ hội tụ về giá trị thực trong giới hạn khi số lượng dữ liệu $n \to \infty$. Các công cụ ước tính đạt được điều này được cho là nhất quán.

Định nghĩa 9.3.1 (Tính nhất quán). Một công cụ ước tính $Y(n) := h\left(X(1), X(2), \ldots, X(n)\right)$ mà ap tiệm cận $y \in \mathbb{R}$ là nhất quán nếu nó hội tụ đến y khi $n \to \infty$ trong bình phương trung bình, với xác suất một hoặc xác suất.

Định lý sau đây cho thấy giá trị trung bình là nhất quán.

Định lý 9.3.2 (Trung bình mẫu là nhất quán). Giá trị trung bình mẫu là một công cụ ước tính nhất quán của giá trị trung bình của một chuỗi biến ngẫu nhiên iid miễn là phương sai của chuỗi được giới hạn.

Bằng chứng. Chúng tôi xem xét giá trị trung bình mẫu của chuỗi iid X với giá trị trung bình μ,

$$Y(n) := \frac{1}{N}\sum_{tôi=1}^{N} X(i). \tag{9.12}$$

Công cụ ước tính bằng với trung bình di chuyển của dữ liệu. Kết quả là nó hội tụ về μ trong bình phương trung bình (và với xác suất là một) theo định luật số lớn (Định lý 6.2.2), miễn là phương sai σ^2 của mỗi mục trong chuỗi iid bị chặn. □

Ví dụ 9.3.3 (Ước tính chiều cao trung bình). Trong ví dụ này, chúng tôi minh họa tính nhất quán của giá trị trung bình mẫu. Hãy tưởng tượng rằng chúng ta muốn ước tính chiều cao trung bình trong một quần thể. Để cụ thể, chúng tôi xem xét dân số m := 25000 người. Hình 9.2 cho thấy một biểu đồ về chiều cao của họ.[1] Như đã giải thích trong Ví dụ 9.1.1 nếu chúng ta lấy mẫu n cá nhân từ quần thể này bằng sự thay thế, thì chiều cao của họ tạo thành một chuỗi iid X. Giá trị trung bình của chuỗi này là

$$E X(i) := \sum_{j=1}^{tôi} P(\text{Người } j \text{ được chọn}) \cdot \text{chiều cao của người } j \tag{9.13}$$

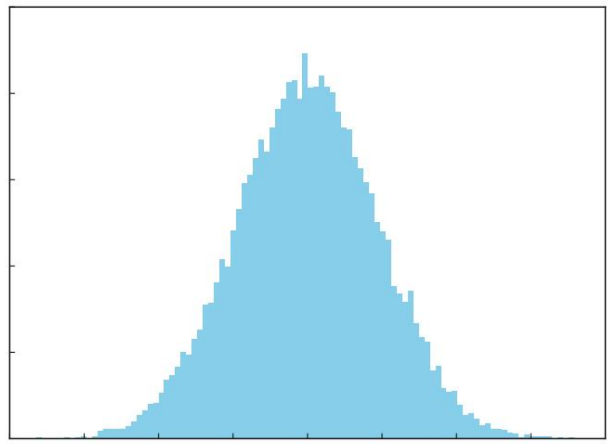
$$= \frac{1}{m}\sum_{j=1}^{tôi} h_j \tag{9.14}$$

$$= av(h_1, \ldots h_m) \tag{9.15}$$

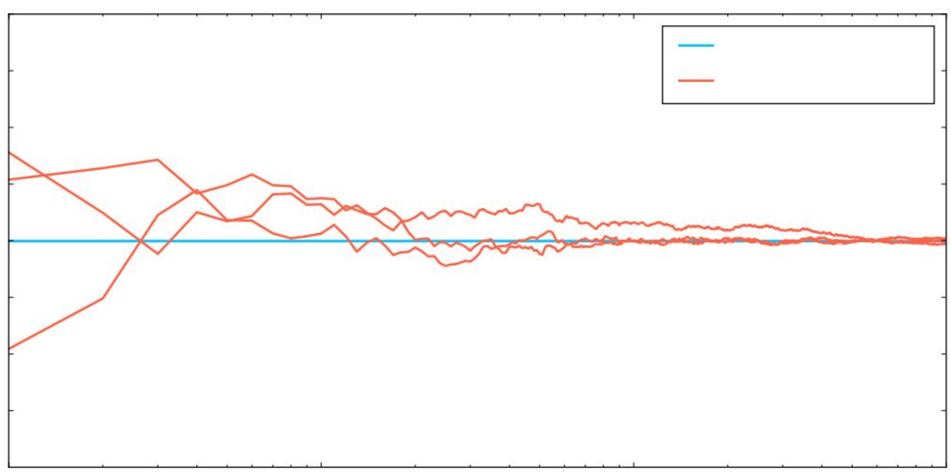
với $1 \le i \le n$, trong đó $h_1, \ldots, h_m$ là chiều cao của con người. Ngoài ra, phương sai bị giới hạn vì độ cao là hữu hạn. Theo Định lý 9.3.2, giá trị trung bình mẫu của n dữ liệu phải hội tụ thành giá trị trung bình của chuỗi iid và do đó là chiều cao trung bình trên toàn bộ tổng thể. Hình 9.3 minh họa điều này bằng số.

Nếu giá trị trung bình của phân phối cơ sở không được xác định rõ hoặc phương sai của nó không bị chặn, thì giá trị trung bình của mẫu không nhất thiết phải là một công cụ ước tính nhất quán. Điều này có liên quan đến thực tế là

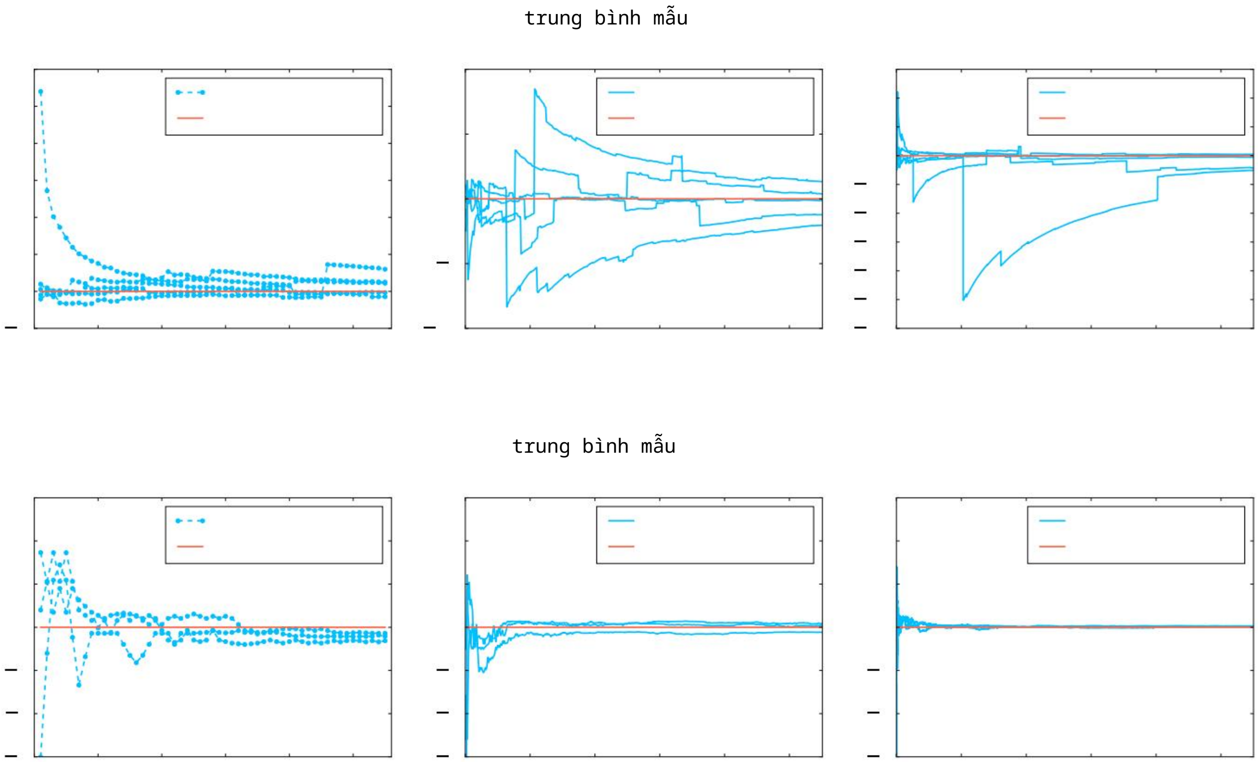
[1] Dữ liệu có tại đây: wiki.stat.ucla.edu/socr/index.php/SOCR_Data_Dinov_020108_HeightsWeights.



Hình 9.2: Biểu đồ chiều cao của một nhóm 25 000 người.



Hình 9.3: Các nhận thức khác nhau về ý nghĩa của mẫu khi các cá thể từ tổng thể trong Hình 9.2 được lấy mẫu thay thế.



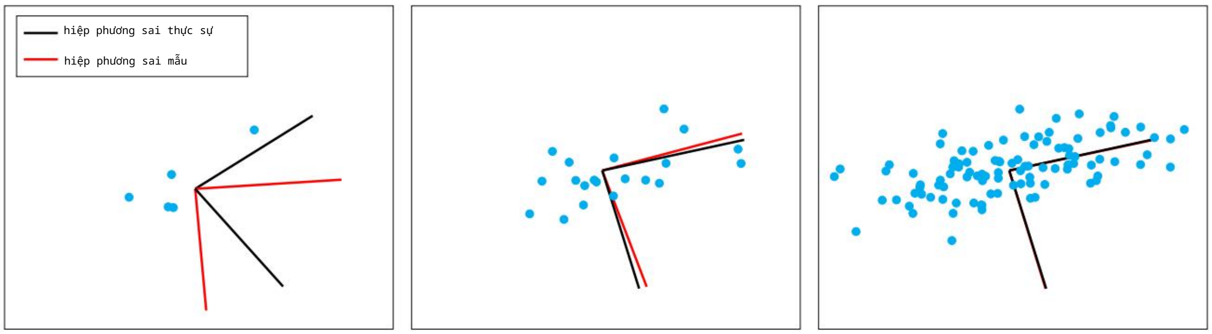
Hình 9.4: Hiện thực trung bình động của một iid dãy Cauchy (trên) so với trung bình động (dưới).

giá trị trung bình của mẫu có thể bị ảnh hưởng nghiêm trọng bởi sự hiện diện của các giá trị cực trị, như chúng ta đã thảo luận trong Phần 8.2. Ngược lại, trung vị mẫu có xu hướng mạnh hơn trong những tình huống như vậy, như đã thảo luận trong Phần 8.3. Định lý sau đây chứng minh rằng trung vị mẫu nhất quán theo giả định iid, ngay cả khi giá trị trung bình không được xác định rõ hoặc phương sai không bị chặn. Bằng chứng ở Mục 9.7.2.

Định lý 9.3.4 (Trung vị mẫu là ước lượng của trung vị). Trung vị mẫu là một công cụ ước lượng nhất quán của trung vị của một chuỗi iid các biến ngẫu nhiên.

Hình 9.4 so sánh trung bình trượt và trung vị trượt của một chuỗi iid các biến ngẫu nhiên Cauchy cho ba nhận thức khác nhau. Đường trung bình động không ổn định và không hội tụ cho dù có bao nhiêu dữ liệu, điều này không có gì đáng ngạc nhiên vì giá trị trung bình không được xác định rõ. Ngược lại, đường trung tuyến chuyển động cuối cùng sẽ hội tụ về đường trung tuyến thực như dự đoán của Định lý 9.3.4.

Phương sai mẫu và hiệp phương sai lần lượt là các công cụ ước tính nhất quán của phương sai và hiệp phương sai, theo các giả định nhất định về thời điểm cao hơn của các phân phối cơ bản. Điều này cung cấp một giải thích trực quan cho phân tích thành phần chính (xem Phần 8.5.2) với giả định rằng dữ liệu là sự thể hiện của một chuỗi iid của các vectơ ngẫu nhiên: các thành phần chính xấp xỉ với các vectơ riêng của ma trận hiệp phương sai thực (xem Phần 4.3 . 3), và do đó, các hướng của phương sai tối đa của phân phối nhiều chiều. Hình 9.5



Hình 9.5: Hướng chính của n mẫu từ phân bố Gaussian hai biến (màu đỏ) so với các vectơ riêng của ma trận hiệp phương sai của phân bố (màu đen).

minh họa điều này bằng một ví dụ số, trong đó các thành phần chính thực sự hội tụ về các vectơ riêng khi số lượng dữ liệu tăng lên.

9.4 Khoảng tin cậy

Tính nhất quán ngụ ý rằng một công cụ ước tính sẽ hoàn hảo nếu chúng ta thu được dữ liệu vô hạn, nhưng điều này tất nhiên là không thể trong thực tế. Do đó, điều quan trọng là định lượng độ chính xác của công cụ ước tính cho một số lượng dữ liệu cố định. Khoảng tin cậy cho phép thực hiện điều này từ quan điểm của người theo chủ nghĩa thường xuyên. Khoảng tin cậy có thể được hiểu là một ước tính mềm của đại lượng quan tâm xác định, đảm bảo rằng giá trị thực sẽ thuộc về khoảng với một xác suất nhất định.

Định nghĩa 9.4.1 (Khoảng tin cậy). Khoảng tin cậy $1 - \alpha$ I cho $\gamma \in \mathbb{R}$ thỏa mãn

$$P(\gamma \in I) \geq 1 - \alpha, \tag{9.16}$$

trong đó $0 < a < 1$.

Khoảng tin cậy thường có dạng $[Y - c, Y + c]$ trong đó Y là ước lượng của số lượng quan tâm và c là hằng số phụ thuộc vào số lượng dữ liệu. Định lý sau đây rút ra một khoảng tin cậy cho giá trị trung bình của một chuỗi iid. Khoảng tin cậy được căn giữa ở giá trị trung bình của mẫu.

Định lý 9.4.2 (Khoảng tin cậy cho giá trị trung bình của một dãy iid). Cho X là một dãy iid với trung bình μ và phương sai $\sigma^2 \leq b^2$ với một số $b > 0$. Với mọi $0 < \alpha < 1$

$$\text{Trong} := \left[Y_n - \frac{b}{\sqrt{\alpha n}}, Y_n + \frac{b}{\sqrt{\alpha n}} \right], \qquad Y_n := \text{av}\left(X(1), X(2), \ldots, X(n) \right), \tag{9.17}$$

là khoảng tin cậy $1 - \alpha$ cho μ.

Machine Translated by Google

Bằng chứng. Nhớ lại rằng phương sai của Yn bằng Var $\bar{X}$ của $_N$ = σ 2/n (xem phương trình (6.21) trong chứng minh Định lý 6.2.2). Chúng ta có

$$P\ \mu \quad Yn \quad \sqrt{\alpha\ n}\ \frac{b}{} , Yn + \frac{\sigma}{\sqrt{\alpha\ n}} \quad = 1 \quad P\ |Yn \quad \mu| > \sqrt{\alpha n}\ \frac{b}{} \tag{9.18}$$

$$\geq 1 \quad \frac{\alpha\ nVar\ (Yn)}{b\ 2\ \alpha\ \sigma 2} \quad \text{bởi bất đẳng thức Chebyshev (9.19)}$$

$$= 1 \frac{}{b} \tag{9.20}$$

$$2 \geq 1 \quad \alpha. \tag{9.21}$$

□

Độ rộng của khoảng được cung cấp trong định lý giảm dần theo n đối với α cố định, điều này có nghĩa là việc kết hợp nhiều dữ liệu hơn sẽ làm giảm phương sai của công cụ ước tính và do đó làm giảm độ không chắc chắn của chúng ta về nó.

Ví dụ 9.4.3 (Những chú gấu ở Yosemite). Một nhà khoa học đang cố gắng ước tính trọng lượng trung bình của những con gấu đen ở Vườn quốc gia Yosemite. Cô quản lý để bắt 300 con gấu. Chúng tôi giả định rằng những con gấu được lấy mẫu ngẫu nhiên một cách thống nhất có thay thế (một con gấu có thể được cân nhiều lần). Theo các giả định này, trong Ví dụ 9.1.1 , chúng tôi chỉ ra rằng dữ liệu có thể được mô hình hóa dưới dạng các mẫu iid và trong Ví dụ 9.3.3 , chúng tôi chỉ ra rằng trung bình mẫu là một ước lượng nhất quán cho trung bình của toàn bộ tổng thể.

Trọng lượng trung bình của 300 con gấu bị bắt là Y := 200 lbs. Để rút ra khoảng tin cậy từ thông tin này, chúng ta cần giới hạn phương sai. Trọng lượng tối đa được ghi nhận cho một con gấu đen từng là 880 lbs. Đặt μ và σ là giá trị trung bình và phương sai (chưa biết) của ^{2} trọng số của toàn bộ tổng thể. Nếu X là trọng lượng của một con gấu được chọn ngẫu nhiên đồng đều từ toàn bộ

quần thể thì X có trung bình μ và phương sai σ^2, vì thế

$$2\sigma\ \_ = E\ X2 \quad E^2\ (X) \tag{9.22}$$

$$\leq E\ X2 \leq \tag{9.23}$$

$$8802 \text{ vì } X \leq 880. \tag{9.24}$$

Do đó, 880 là giới hạn trên của độ lệch chuẩn. Áp dụng Định lý 9.4.2,

$$Y \quad \frac{b}{\sqrt{\alpha\ n}} , Y + \frac{b}{\sqrt{\alpha n}} \quad = [\ \ 27{,}2,\ 427{,}2] \tag{9.25}$$

là khoảng tin cậy 95% cho trọng lượng trung bình của toàn bộ dân số. Khoảng không chính xác lắm vì n không lớn lắm.

Như được minh họa bởi ví dụ này, khoảng tin cậy bắt nguồn từ bất đẳng thức Chebyshev có xu hướng rất thận trọng. Một cách khác là tận dụng định lý giới hạn trung tâm (CLT). CLT đặc trưng cho sự phân bố của giá trị trung bình mẫu một cách tiệm cận, vì vậy khoảng tin cậy thu được từ nó không được đảm bảo chính xác. Tuy nhiên, CLT thường cung cấp một phép tính gần đúng rất chính xác đối với phân phối của trung bình mẫu cho n hữu hạn, như chúng tôi đã trình bày qua một số ví dụ số trong Chương 6. Để có được khoảng tin cậy cho giá trị trung bình của một chuỗi iid từ CLT như đã nêu trong Định lý 6.3.1 chúng ta cần biết phương sai thực của thứ tự, không thực tế trong thực tế. Tuy nhiên, kết quả sau đây nói rằng chúng ta có thể thay thế phương sai thực bằng phương sai mẫu. Bằng chứng nằm ngoài phạm vi của những ghi chú này.

Định lý 9.4.4 (Định lý giới hạn trung tâm với độ lệch chuẩn mẫu). Cho X là một quá trình ngẫu nhiên rời rạc iid với trung bình $\mu_X := \mu$ sao cho phương sai của nó và thời điểm thứ tư $E(X(i)^4)$ bị chặn.

$$\frac{\sqrt{n}\left(\text{av}\left(X(1), \ldots, X(n)\right) - \mu\right)}{\text{tiêu chuẩn}\ X(1), \ldots, X(n)} \tag{9.26}$$

hội tụ trong phân phối tới một biến ngẫu nhiên Gaussian tiêu chuẩn.

Nhớ lại rằng cdf của Gaussian tiêu chuẩn không có biểu thức dạng đóng. Để đơn giản hóa ký hiệu, chúng tôi biểu thị khoảng tin cậy theo hàm Q.

Định nghĩa 9.4.5 (Hàm Q). Q(x) là xác suất mà một biến ngẫu nhiên Gaussian tiêu chuẩn có thể lớn hơn x đối với x dương,

$$\text{Hỏi}(x) := \int_{u=x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du, \quad x > 0. \tag{9.27}$$

Theo tính đối xứng, nếu U là biến ngẫu nhiên Gauss tiêu chuẩn và $y < 0$

$$P(U < y) = Q(-y). \tag{9.28}$$

Hệ quả 9.4.6 (Khoảng tin cậy gần đúng cho giá trị trung bình). Cho X là một dãy iid thỏa mãn điều kiện của Định lý 9.4.4. Với mọi $0 < \alpha < 1$

$$\text{Trong} := \left[Y_n - \frac{\text{ốc}}{\sqrt{n}}\text{Hỏi}^{-1}\left(\frac{\alpha}{2}\right), Y_n + \frac{\text{ốc}}{\sqrt{n}}\text{Hỏi}^{-1}\left(\frac{\alpha}{2}\right)\right], \tag{9.29}$$

$$Y_n := \text{av}\left(X(1), X(2), \ldots, X(n)\right), \tag{9h30}$$

$$S_n := \text{std}\left(X(1), X(2), \ldots, X(n)\right), \tag{9.31}$$

là khoảng tin cậy xấp xỉ $1 - \alpha$ cho μ, nghĩa là

$$P(\mu \in I_n) \approx 1 - \alpha. \tag{9.32}$$

Bằng chứng. Theo định lý giới hạn trung tâm, khi $n \to \infty$ $\bar{X}_N$ được phân phối dưới dạng ngẫu nhiên Gaussian biến với trung bình μ và phương sai σ^2. Kết quả là

$$P(\mu \in I_n) = 1 - P\left(Y_n > \mu + \frac{\text{ốc}}{\sqrt{}}Q^{-1}\left(\frac{\alpha}{2}\right)\right) - P\left(Y_n < \mu - \frac{\text{ốc}}{\sqrt{}}Q^{-1}\left(\frac{\alpha}{2}\right)\right) \tag{9.33}$$

$$= 1 - P\left(\frac{n\sqrt{n}(Y_n - \mu)}{\text{ốc}} > Q^{-1}\left(\frac{\alpha}{2}\right)\right) - P\left(\frac{n\sqrt{n}(Y_n - \mu)}{\text{ốc}} < -Q^{-1}\left(\frac{\alpha}{2}\right)\right) \tag{9.34}$$

$$\approx 1 - 2Q\left(Q^{-1}\left(\frac{\alpha}{2}\right)\right) \quad \text{theo Định lý 9.4.4} \tag{9.35}$$

$$= 1 - \alpha. \tag{9.36}$$

□

Điều quan trọng cần nhấn mạnh là kết quả chỉ cung cấp khoảng tin cậy chính xác nếu n đủ lớn để phương sai mẫu hội tụ về phương sai thực và để CLT có hiệu lực.

Ví dụ 9.4.7 (Những chú gấu ở Yosemite (tiếp theo)). Độ lệch chuẩn mẫu của những con gấu mà nhà khoa học bắt được bằng 100 lbs. Chúng ta áp dụng Hệ quả 9.4.6 để rút ra khoảng tin cậy xấp xỉ chặt chẽ hơn khoảng tin cậy thu được khi áp dụng bất đẳng thức Chebyshev. Cho rằng Q (1,95) ≈ 0,025,

$$Y - \frac{\sigma}{\sqrt{n}} Q^{-1}\left(\frac{\alpha}{2}\right), Y + \frac{\sigma}{\sqrt{n}} Q^{-1}\left(\frac{\alpha}{2}\right) \approx [188,8,\ 211,3] \tag{9.37}$$

là khoảng tin cậy xấp xỉ 95% cho trọng lượng trung bình của quần thể gấu.

Giải thích khoảng tin cậy là hơi khó khăn. Sau khi tính toán khoảng tin cậy trong Ví dụ 9.4.7 , người ta muốn tuyên bố:

Xác suất để trọng lượng trung bình nằm trong khoảng từ 188,8 đến 211,3 lbs là 0,95.

Tuy nhiên, chúng tôi đang mô hình hóa trọng lượng trung bình dưới dạng một đại lượng xác định, vì vậy không có đại lượng ngẫu nhiên nào trong tuyên bố này! Cách giải thích đúng là nếu chúng ta lặp lại quá trình lấy mẫu dân số và tính khoảng tin cậy nhiều lần, thì giá trị thực sẽ nằm trong khoảng 95% thời gian. Điều này được minh họa trong ví dụ sau và Hình 9.6.

Ví dụ 9.4.8 (Ước tính chiều cao trung bình (tiếp theo)). Hình 9.6 cho thấy một số khoảng tin cậy 95% cho giá trị trung bình của dân số có chiều cao trong Ví dụ 9.3.3. Để tính mỗi khoảng ta chọn n cá thể rồi áp dụng Hệ quả 9.4.6. Độ rộng của các khoảng giảm khi n tăng, nhưng vì chúng đều là các khoảng tin cậy 95% nên chúng đều chứa giá trị trung bình thực với xác suất 0,95. Thật vậy, đây là trường hợp của 113 trong số 120 (94%) khoảng thời gian được vẽ.

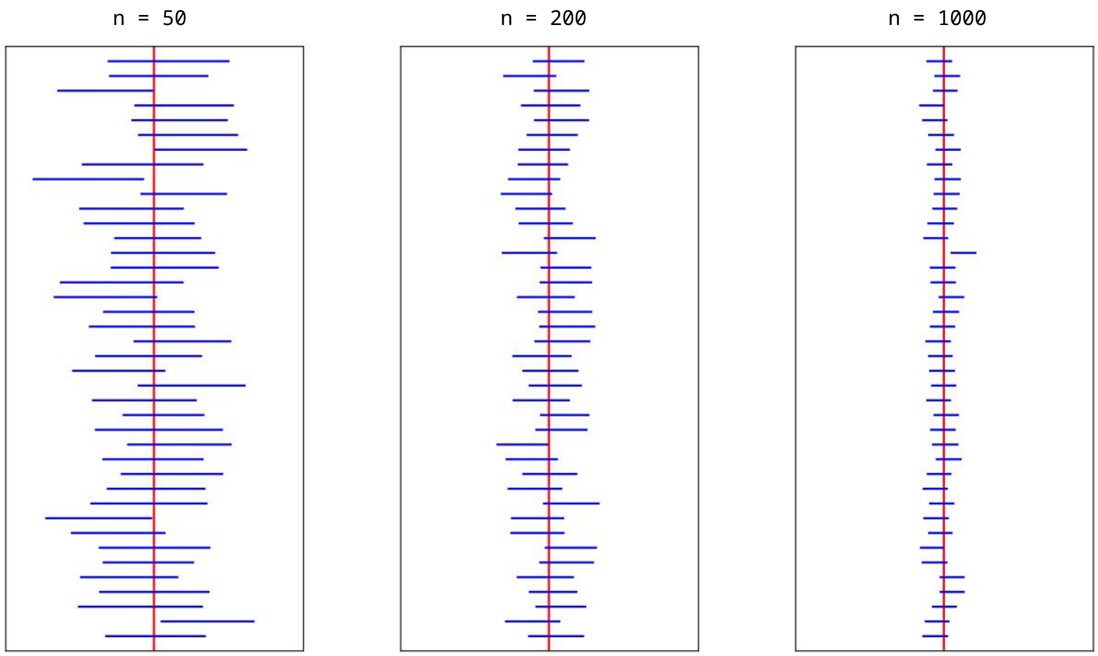
9.5 Ước lượng mô hình phi tham số

Trong phần này, chúng tôi xem xét vấn đề ước tính phân phối từ nhiều mẫu iid.
Điều này yêu cầu xấp xỉ cdf, pmf hoặc pdf của bản phân phối. Nếu chúng ta giả định rằng phân phối thuộc về một họ xác định trước, thì vấn đề sẽ giảm xuống còn việc ước lượng các tham số đặc trưng cho họ cụ thể đó, như chúng tôi giải thích chi tiết trong Phần 9.6 . Ở đây chúng tôi không đưa ra một giả định như vậy. Ước tính một phân phối trực tiếp là rất khó khăn; rõ ràng là nhiều bản phân phối khác nhau (vô hạn!) có thể đã tạo ra dữ liệu. Tuy nhiên, với đủ mẫu, thường có thể thu được các mô hình tạo ra giá trị gần đúng chính xác, miễn là giả định iid được giữ nguyên.

9.5.1 Thực nghiệm cdf

Theo giả định rằng tập dữ liệu tương ứng với các mẫu iid từ một phân phối nhất định, ước tính hợp lý cho cdf của phân phối tại một điểm x đã cho là tỷ lệ mẫu nhỏ hơn x. Điều này dẫn đến một công cụ ước lượng hằng số từng phần được gọi là cdf theo kinh nghiệm.

Machine Translated by Google



Hình 9.6: Khoảng tin cậy 95% cho giá trị trung bình của dân số có chiều cao trong Ví dụ 9.3.3.

Định nghĩa 9.5.1 (Cdf theo kinh nghiệm). cdf theo kinh nghiệm tương ứng với dữ liệu x1, . . . , xn là

$$\text{fn}\ (x) := \frac{1}{N} \sum_{\text{tôi}=1}^{N} 1_{x_i \leq x}, \tag{9.38}$$

trong đó x R.

cdf theo kinh nghiệm là một công cụ ước tính không thiên vị và nhất quán của cdf thực. Điều này được thiết lập chặt chẽ trong Định lý 9.5.2 dưới đây và được minh họa bằng thực nghiệm trong Hình 9.7. Cdf của dữ liệu chiều cao từ 25.000 người được so sánh với ba giá trị thực của cdf theo kinh nghiệm được tính từ số lượng mẫu iid khác nhau. Khi số lượng mẫu có sẵn tăng lên, phép tính gần đúng trở nên rất chính xác.

Định lý 9.5.2. Đặt X là một chuỗi iid với biên cdf FX. Với mọi x cố định R Fn (x) là một ước lượng không chệch và nhất quán của FX (x). Thật vậy, Fn (x) hội tụ bình phương trung bình tới FX (x).

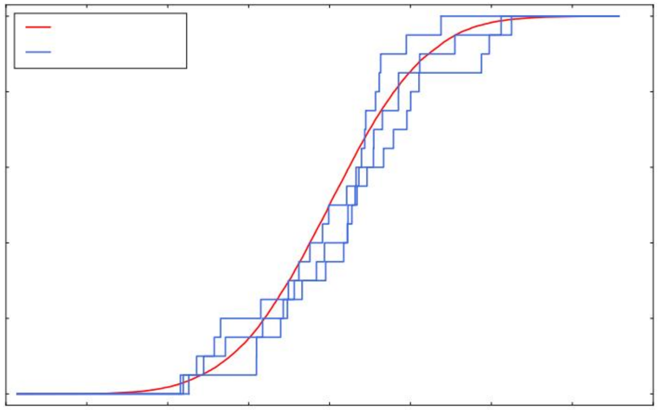
Bằng chứng. Đầu tiên, chúng tôi xác minh

$$E\ Fn\ (x) = E \frac{1}{n} \sum_{i=1}^{N} 1_{X(i) \leq x} \tag{9.39}$$

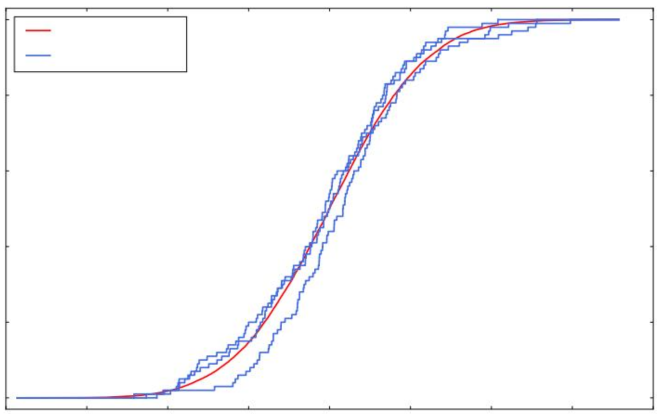
$$= \frac{1}{N} \sum_{\text{tôi}=1}^{N} P\,X\ (i) \leq x \text{ theo tuyến tính của kỳ vọng} \tag{9.40}$$

$$= \text{Ngoại hối}\ (x), \tag{9.41}$$

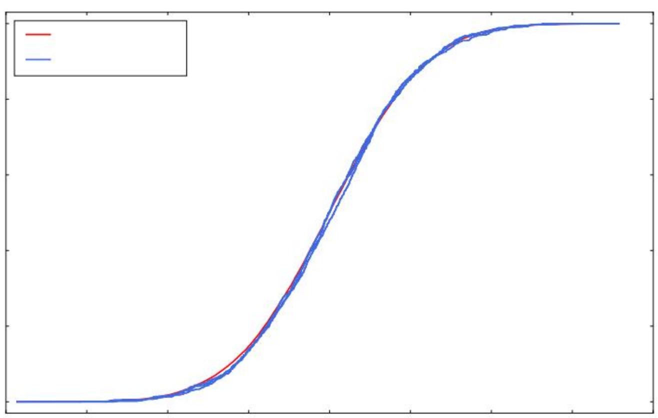
n = 10



n = 100



n = 1000



Hình 9.7: Cdf của dữ liệu độ cao trong Hình 2.13 cùng với ba lần thực hiện cdf theo kinh nghiệm được tính toán với n iid mẫu cho n = 10, 100, 1000.

vì vậy công cụ ước tính là không thiên vị. Bây giờ chúng tôi ước tính bình phương trung bình của nó

$$E\, F_N^2(x) = E\left(\frac{1}{n^2}\sum_{i=1}^{N}\sum_{j=1}^{N} 1_{X(i)\le x}\, 1_{X(j)\le x}\right) \tag{9.42}$$

$$= \frac{1}{n^2}\sum_{\text{tôi}=1}^{N} P\left(X(i) \le x\right) + \frac{1}{n^2}\sum_{i=1}^{N}\sum_{j=1, i=j}^{N} P\left(X(i) \le x, X(j) \le x\right) \tag{9.43}$$

$$= \frac{\text{Ngoại hối}(x)}{N} + \frac{1}{n^2}\sum_{i=1}^{N}\sum_{j=1, i=j}^{N} F_{X(i)}(x)\, F_{X(j)}(x) \text{ theo tính độc lập} \tag{9.44}$$

$$= \frac{\text{Ngoại hối}(x)}{N} + \frac{n-1}{N} F_X^2(x). \tag{9.45}$$

Do đó, phương sai bằng

$$\text{Biến } F_n(x) = E\, F_n(x)^2 - E^2\, F_n(x) \tag{9.46}$$

$$= \frac{\text{Ngoại hối}(x)\,(1 - \text{Ngoại hối}(x))}{N}. \tag{9.47}$$

Chúng tôi kết luận rằng

$$\lim_{n\to\infty} E\left(F_X(x) - F_n(x)\right)^2 = \lim_{n\to\infty} \operatorname{Var} F_n(x) = 0. \tag{9.48}$$

□

9.5.2 Ước tính mật độ

Ước tính pdf của một số lượng liên tục khó hơn nhiều so với ước tính cdf. Nếu chúng tôi có đủ dữ liệu, tỷ lệ mẫu nhỏ hơn một số x nhất định sẽ cung cấp ước tính tốt cho cdf tại thời điểm đó. Tuy nhiên, bất kể chúng ta có bao nhiêu dữ liệu, thì xác suất chúng ta sẽ thấy bất kỳ mẫu nào chính xác tại x là không đáng kể: một công cụ ước tính mật độ theo kinh nghiệm theo điểm sẽ bằng 0 ở hầu hết mọi nơi (ngoại trừ tại các mẫu có sẵn).

Hy vọng duy nhất của chúng tôi để tạo ra một công cụ ước tính chính xác là nếu bản pdf mà chúng tôi muốn ước tính mượt mà. Trong trường hợp đó, chúng ta có thể ước tính giá trị của nó tại một điểm x từ các mẫu được quan sát nằm ở các vị trí lân cận. Nếu có nhiều mẫu gần với x thì điều này cho thấy rằng ước tính tại x phải lớn, trong khi nếu tất cả các mẫu ở xa thì ước lượng đó sẽ nhỏ. Ước tính mật độ hạt nhân đạt được điều này bằng cách lấy trung bình các mẫu.

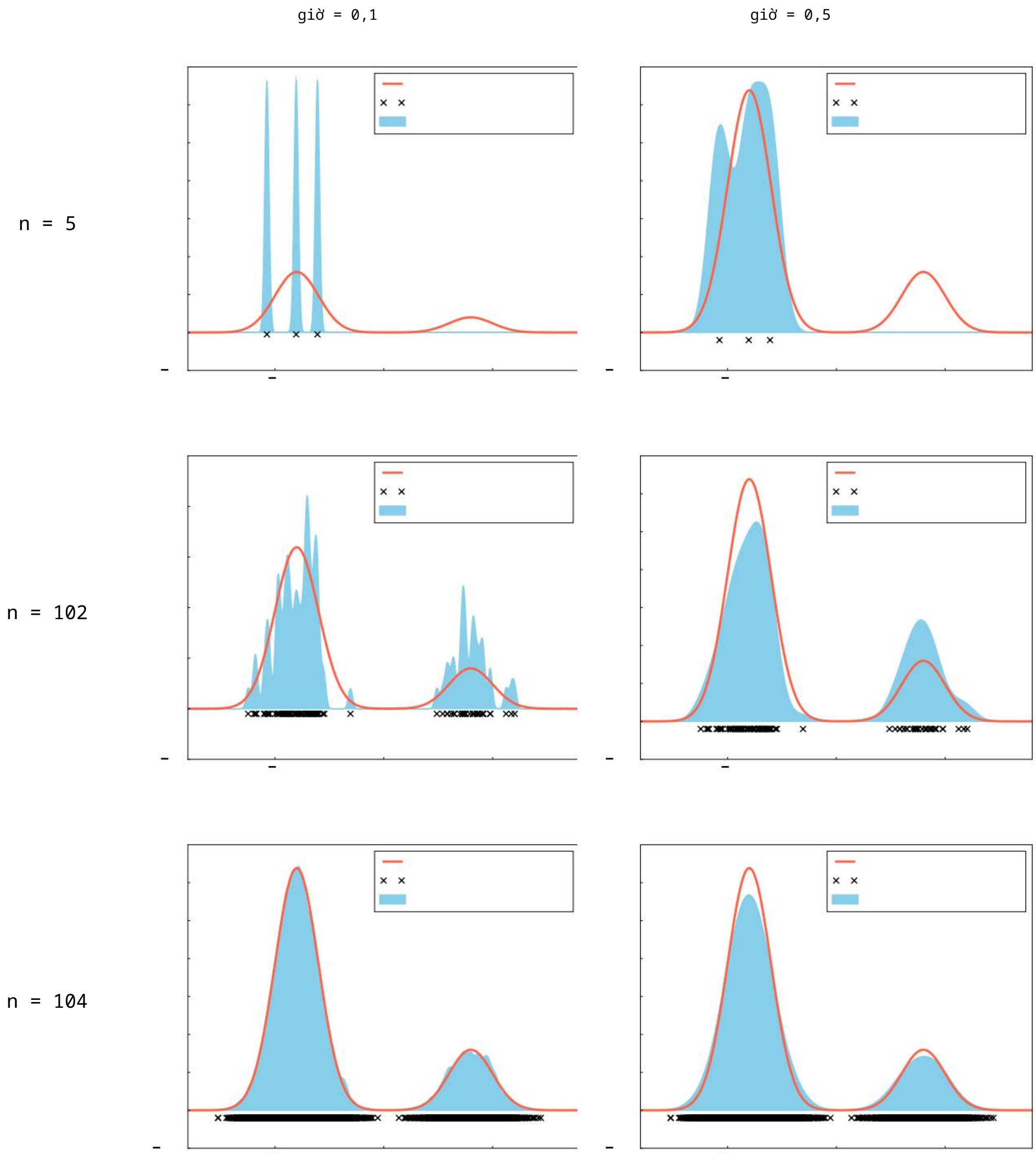
Định nghĩa 9.5.3 (Công cụ ước tính mật độ hạt nhân). Ước tính mật độ hạt nhân với băng thông h của phân phối $x_1, \ldots, x_n$ tại $x \in \mathbb{R}$ là

$$f_{h,n}(x) := \frac{1}{nh}\sum_{\text{tôi}=1}^{N} k\left(\frac{x - x_i}{h}\right), \tag{9.49}$$

trong đó k là hàm hạt nhân có tâm tại gốc tọa độ thỏa mãn

$$k(x) \ge 0 \text{ với mọi } x \in \mathbb{R}, \tag{9.50}$$

$$\int_{\mathbb{R}} k(x)\,dx = 1. \tag{9.51}$$



Hình 9.8: Ước tính mật độ hạt nhân cho hỗn hợp Gauss được mô tả trong Ví dụ 9.6.5 cho các số lượng mẫu iid và các giá trị khác nhau của băng thông hạt nhân h.

Tác dụng của hạt nhân là cân từng mẫu theo khoảng cách của chúng đến điểm mà chúng ta đang ước tính pdf x. Việc chọn một nhân hình chữ nhật mang lại ước tính mật độ theo kinh nghiệm là hằng số từng phần và gần giống như một biểu đồ (các trọng số tương ứng là hằng số hoặc bằng 0). Một giải pháp thay thế phổ biến là nhân Gaussian $k(x) = \exp\left(-x^2\right) / \sqrt{\pi}$, tạo ra một ước lượng mật độ mượt mà. Hạt nhân phải phân rã sao cho $k((x - x_i)/h)$ lớn khi mẫu x_i ở gần x và nhỏ khi ở xa. Sự phân rã này bị chi phối bởi băng thông h, được chọn trước dựa trên kỳ vọng của chúng tôi về độ mượt của pdf và lượng dữ liệu có sẵn. Nếu băng thông rất nhỏ, các mẫu riêng lẻ có ảnh hưởng lớn đến ước tính mật độ. Điều này cho phép tái tạo các hình dạng bất thường dễ dàng hơn, nhưng cũng tạo ra các dao động giả không có trong đường cong thực, đặc biệt nếu chúng ta không có nhiều mẫu. Việc tăng băng thông sẽ giúp loại bỏ những biến động như vậy và mang lại các ước tính ổn định hơn khi số lượng dữ liệu nhỏ. Tuy nhiên, nó cũng có thể làm trơn quá mức ước tính. Theo nguyên tắc thông thường, chúng ta nên giảm băng thông của nhân khi số lượng dữ liệu tăng lên.

Hình 9.8 và 9.9 minh họa ảnh hưởng của việc thay đổi băng thông h ở các tốc độ lấy mẫu khác nhau. Trong Hình 9.8 Ước tính mật độ nhân Gaussian được áp dụng để ước tính hỗn hợp Gaussian được mô tả trong Ví dụ 9.6.5. Hình 9.9 cho thấy một ví dụ trong đó kỹ thuật tương tự được sử dụng trên dữ liệu thực: mục đích là để ước tính mật độ trọng lượng của một quần thể ốc biển.[2] Toàn bộ quần thể bao gồm 4.177 cá thể. Ước tính mật độ hạt nhân được tính toán từ 200 mẫu iid cho các giá trị khác nhau của băng thông hạt nhân.

9.6 Ước lượng mô hình tham số

Trong phần trước, chúng tôi mô tả cách ước tính phân phối bằng cách ước tính trực tiếp tệp cdf hoặc pdf tạo dữ liệu. Trong phần này, chúng tôi thảo luận về một lộ trình thay thế dựa trên giả định rằng loại phân phối tạo ra dữ liệu đã được biết trước. Nếu đây là trường hợp, vấn đề tóm lại là điều chỉnh các tham số đặc trưng cho phân phối dữ liệu. Nhớ lại rằng từ quan điểm của người theo chủ nghĩa thường xuyên, phân phối thực sự là cố định, do đó, các tham số tương ứng được mô hình hóa dưới dạng đại lượng xác định (ngược lại, trong khung Bayesian, chúng được mô hình hóa dưới dạng biến ngẫu nhiên).

9.6.1 Phương pháp khoảnh khắc

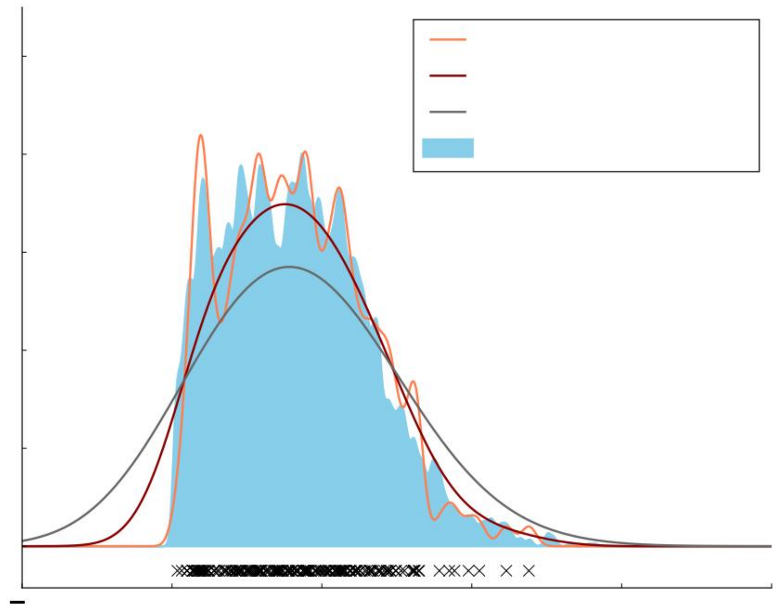
Phương pháp khoảnh khắc điều chỉnh các tham số của phân phối sao cho khoảnh khắc của phân phối trùng với khoảnh khắc mẫu của dữ liệu (tức là giá trị trung bình, bình phương trung bình hoặc phương sai, v.v.). Nếu phân phối chỉ phụ thuộc vào một tham số, thì chúng tôi sử dụng giá trị trung bình mẫu làm giá trị thay thế cho giá trị trung bình thực và tính giá trị tương ứng của tham số. Đối với một hàm mũ với tham số λ và nghĩa là μ ta có

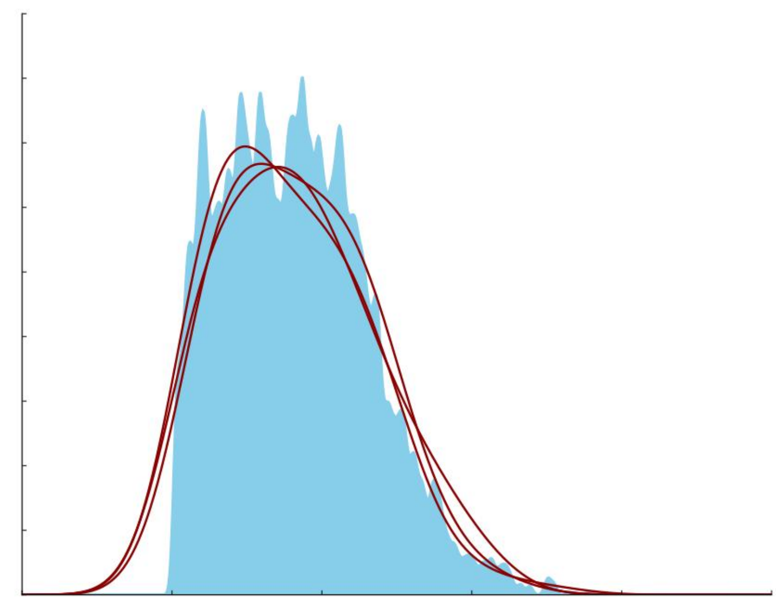
$$\mu = \frac{1}{\lambda}. \tag{9.52}$$

Giả sử rằng chúng tôi có quyền truy cập vào n mẫu iid $x_1, \ldots, x_n$ từ phân phối hàm mũ, ước tính phương pháp thời điểm của λ bằng

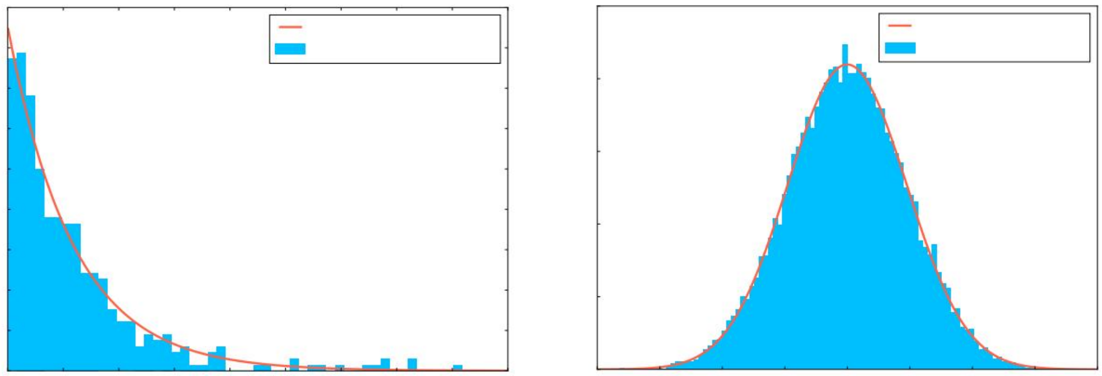
$$\lambda_{MM} := \frac{1}{\text{av}(x_1, \ldots, x_n)}. \tag{9.53}$$

[2] Dữ liệu có sẵn tại archive.ics.uci.edu/ml/datasets/Abalone





Hình 9.9: Ước tính mật độ nhân cho trọng lượng của một quần thể bào ngư, một loài ốc biển. Trong biểu đồ trên, mật độ được ước tính từ 200 mẫu iid sử dụng nhân Gaussian với ba băng thông khác nhau. Chữ thập màu đen đại diện cho các mẫu riêng lẻ được hiển thị bên dưới. Trong biểu đồ bên dưới, chúng ta thấy kết quả của việc lặp lại quy trình ba lần bằng cách sử dụng băng thông cố định bằng 0,25.



Hình 9.10: Phân phối hàm mũ phù hợp với dữ liệu bao gồm thời gian đến giữa các cuộc gọi tại một trung tâm cuộc gọi ở Israel (trái). Phân phối Gaussian phù hợp với dữ liệu chiều cao (phải).

Biểu đồ bên phải của Hình 9.10 cho thấy kết quả của việc khớp số mũ với dữ liệu trung tâm cuộc gọi trong Hình 2.11. Tương tự, để khớp Gaussian bằng cách sử dụng phương pháp khoảnh khắc, chúng ta đặt giá trị trung bình bằng giá trị trung bình mẫu của nó và phương sai bằng phương sai mẫu, như được minh họa bằng đồ thị bên phải của Hình 9.10 sử dụng dữ liệu từ Hình 2.13 .

9.6.2 Khả năng tối đa

Phương pháp phổ biến nhất để học các mô hình tham số là khả năng phù hợp tối đa. Hàm khả năng là pmf hoặc pdf chung của dữ liệu, được hiểu là một hàm của các tham số chưa biết. Chi tiết hơn, chúng ta hãy biểu thị dữ liệu bằng $x_1, \ldots, x_n$ và giả sử rằng chúng là các thực thể của tập hợp các biến ngẫu nhiên rời rạc $X_1, \ldots, X_n$ có một pmf chung phụ thuộc vào một vectơ tham số θ. Để nhấn mạnh rằng liên kết pmf phụ thuộc vào θ chúng ta ký hiệu nó là $p_\theta := p_{X_1,...,X_n}$. PMf này được đánh giá tại dữ liệu quan sát được

$$p_\theta (x_1, \ldots, x_n) \tag{9.54}$$

là hàm khả năng, khi chúng ta hiểu nó là hàm của θ. Đối với các biến ngẫu nhiên liên tục, thay vào đó, chúng tôi sử dụng bản pdf chung của dữ liệu.

Định nghĩa 9.6.1 (Hàm khả năng). Cho một nhận thức $x_1, \ldots, x_n$ của một tập hợp các biến dom chạy rời rạc $X_1, \ldots, X_n$ với khớp pmf p_θ, trong đó θ $R^{tôi}$ là một vectơ tham số, hàm khả năng là

$$L_{x_1,...,x_n} \theta := p_\theta (x_1, \ldots, x_n). \tag{9.55}$$

Nếu các biến ngẫu nhiên liên tục với pdf f_θ, trong đó θ R^m, hàm khả năng là

$$L_{x_1,...,x_n} \theta := f_\theta (x_1, \ldots, x_n). \tag{9.56}$$

Hàm khả năng log bằng logarit của hàm khả năng log $L_{x_1,...,x_n} \theta$.

Khi dữ liệu được mô hình hóa dưới dạng các mẫu iid, khả năng xảy ra sẽ tính thành tích của pmf hoặc pdf cận biên, do đó, khả năng xảy ra của nhật ký có thể được phân tích thành một tổng.

Trong trường hợp phân phối rời rạc, đối với θ cố định, khả năng xảy ra là xác suất mà $X_1, \ldots, X_n$ bằng số liệu quan sát được. Nếu chúng ta không biết θ, thì nên chọn một giá trị cho θ sao cho xác suất này càng cao càng tốt, nghĩa là để tối đa hóa khả năng xảy ra. Đối với các bản phân phối liên tục, chúng tôi áp dụng nguyên tắc tương tự cho bản pdf chung của dữ liệu.

Định nghĩa 9.6.2 (Công cụ ước tính khả năng tối đa). Công cụ ước tính khả năng (ML) tối đa cho vectơ tham số $\theta \quad R^{tôi}$ là

$$\theta_{ML}(x_1, \ldots, x_n) := \arg\max_{\theta} L_{x_1,\ldots,x_n}\theta \tag{9.57}$$

$$= \arg \text{cực đại}_{\theta} \log L_{x_1,\ldots,x_n}\theta . \tag{9.58}$$

Cực đại của hàm khả năng và của hàm khả năng log là ở cùng một vị trí vì logarit là đơn điệu.

Trong những điều kiện nhất định, người ta có thể chỉ ra rằng công cụ ước tính khả năng cực đại là nhất quán: nó hội tụ xác suất thành tham số thực khi số lượng dữ liệu tăng lên. Người ta cũng có thể chỉ ra rằng phân phối của nó hội tụ với phân phối của biến ngẫu nhiên Gaussian (hoặc vectơ), giống như phân phối của giá trị trung bình mẫu. Những kết quả này nằm ngoài phạm vi của khóa học. Tuy nhiên, hãy nhớ rằng chúng chỉ giữ nếu dữ liệu thực sự được tạo bởi loại phân phối mà chúng tôi đang xem xét.

Bây giờ chúng tôi chỉ ra cách rút ra khả năng tối đa cho phân phối Bernoulli và Gaussian. Các công cụ ước tính kết quả cho các tham số giống như các công cụ ước tính phương pháp thời điểm (ngoại trừ một chút khác biệt trong ước tính của tham số phương sai Gaussian).

Ví dụ 9.6.3 (Công cụ ước lượng ML của tham số phân phối Bernoulli). Chúng tôi lập mô hình một tập hợp xn liệu $x_1, \ldots,$ dưới dạng các mẫu iid từ phân phối Bernoulli với tham số θ (trong trường hợp này là dữ chỉ có một tham số). Hàm khả năng bằng

$$L_{x_1,\ldots,x_n}(\theta) = p_\theta(x_1, \ldots, x_n) \tag{9.59}$$

$$= \prod_{tôi=1} \left(1_{x_i=1}\theta + 1_{x_i=0}(1 \quad \theta)\right) \tag{9.60}$$

$$= \theta^{n_1}(1 \quad \theta)^{n_0} \tag{9.61}$$

và hàm log-likelihood để

$$\log L_{x_1,\ldots,x_n}(\theta) = n_1 \log\theta + n_0 \log(1 \quad \theta), \tag{9.62}$$

trong đó n_1 là số mẫu bằng một và n_0 là số mẫu bằng không.
Công cụ ước tính ML của tham số θ là

$$\theta_{ML} = \arg\log \text{tối}_{\theta} \text{ đa } L_{x_1,\ldots,x_n}(\theta) \tag{9.63}$$

$$= \arg \text{cực đại}_{\theta} \; n_1 \log\theta + n_0 \log(1 \quad \theta). \tag{9.64}$$

Chúng tôi tính toán đạo hàm và đạo hàm cấp hai của hàm khả năng log, d log

$$\frac{L_{x_1,...,x_n}(\theta)}{d\theta} = \frac{n_1}{\theta} - \frac{n_0}{1-\theta}, \tag{9.65}$$

$$\frac{d^2 \log L_{x_1,...,x_n}(\theta)}{d\theta^2} = -\frac{n_1}{\theta^2} - \frac{n_0}{(1-\theta)^2} < 0. \tag{9.66}$$

Hàm lõm, vì đạo hàm cấp hai âm. Do đó, mức tối đa là tại điểm mà đạo hàm đầu tiên bằng 0, cụ thể là

$$\theta_{ML} = \frac{n_1}{n_0 + n_1}. \tag{9.67}$$

Ước lượng bằng tỷ lệ các mẫu bằng một.

Ví dụ 9.6.4 (Công cụ ước tính ML của các tham số của phân bố Gaussian). Cho $x_1, x_2, ...$ là dữ liệu mà chúng tôi muốn lập mô hình dưới dạng các mẫu iid từ phân phối Gauss với giá trị trung bình μ và độ lệch chuẩn σ. Hàm khả năng bằng

$$L_{x_1,...,x_n}(\mu,\sigma) = f_{\mu,\sigma}(x_1,...,x_n) \tag{9.68}$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \tag{9.69}$$

và hàm log-likelihood để

$$\log L_{x_1,...,x_n}(\mu,\sigma) = -\frac{n\log(2\pi)}{2} - n\log\sigma - \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{2\sigma^2}. \tag{9.70}$$

Công cụ ước tính ML của các tham số μ và σ là

$$\{\mu_{ML},\sigma_{ML}\} = \arg\max_{\{\mu,\sigma\}} \log L_{x_1,...,x_n}(\mu,\sigma) \tag{9.71}$$

$$= \arg\max_{\{\mu,\sigma\}} -n\log\sigma - \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{2\sigma^2}. \tag{9.72}$$

Chúng tôi tính toán đạo hàm riêng của hàm log-likelihood,

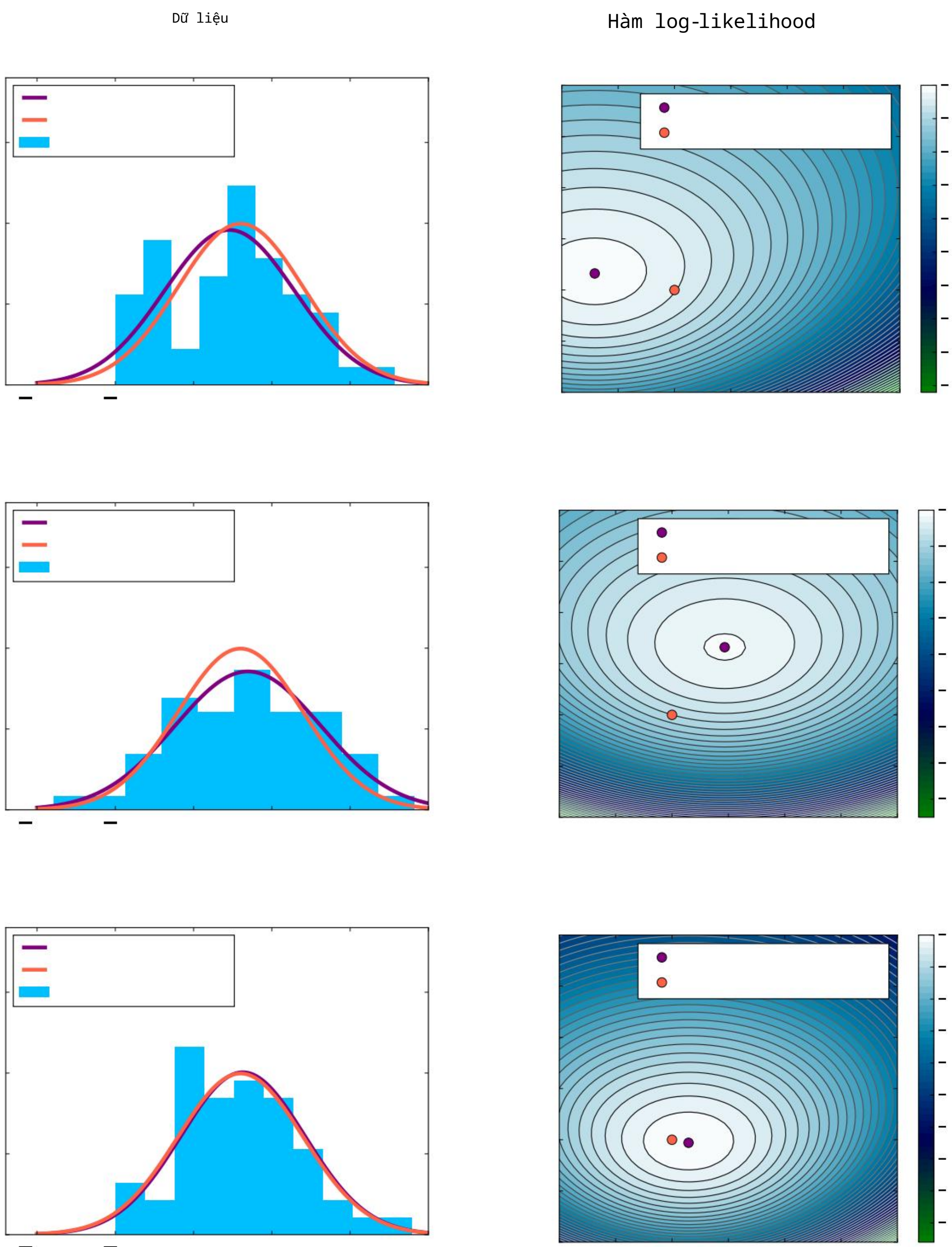
$$\frac{\partial \log L_{x_1,...,x_n}(\mu,\sigma)}{\partial\mu} = \sum_{i=1}^{N}\frac{x_i-\mu}{\sigma^2}, \tag{9.73}$$

$$\frac{\partial \log L_{x_1,...,x_n}(\mu,\sigma)}{\partial\sigma} = -\frac{N}{\sigma} + \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{\sigma^3}. \tag{9.74}$$

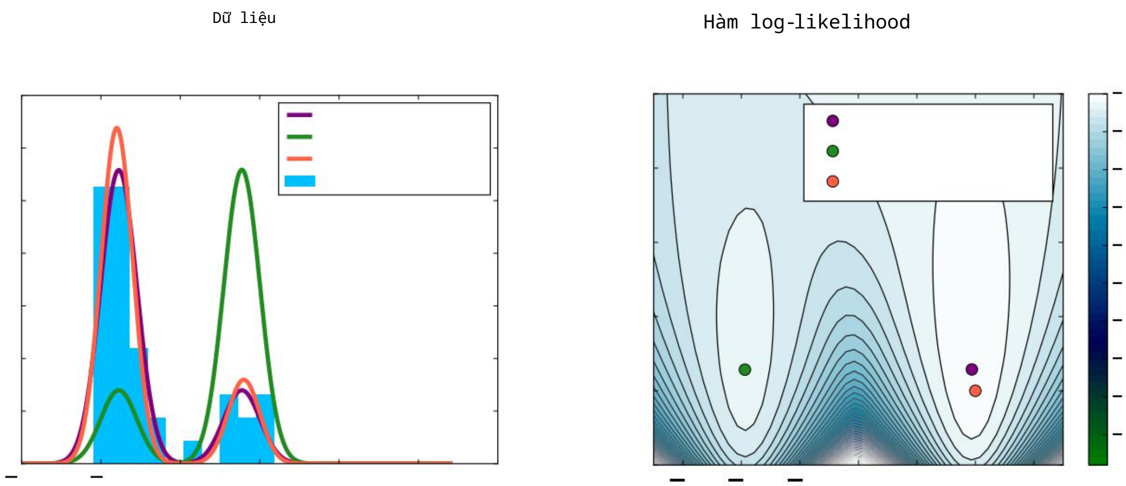
Hàm mà chúng ta đang cố cực đại hóa hoàn toàn lõm trong $\{\mu,\sigma\}$. Để chứng minh điều này, chúng ta sẽ phải chỉ ra rằng Hessian của hàm là xác định dương. Chúng tôi bỏ qua các tính toán cho thấy rằng đây là trường hợp. Đặt các đạo hàm riêng bằng 0, chúng ta thu được

$$\mu_{ML} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{9,75}$$

$$\sigma^2_{ML} = \frac{1}{n}\sum_{i=1}^{N}(x_i-\mu_{ML})^2. \tag{9.76}$$



Hình 9.11: Cột bên trái hiển thị biểu đồ của 50 mẫu iid từ phân phối Gaussian, cùng nhau với bản pdf của bản phân phối ban đầu, cũng như ước tính khả năng xảy ra tối đa. Cột bên phải hiển thị hàm log-likelihood tương ứng với dữ liệu và vị trí cực đại của nó và của điểm tương ứng với các tham số thực.



Hình 9.12: Hình ảnh bên trái hiển thị biểu đồ của 40 mẫu iid từ hỗn hợp Gaussian được xác định trong Ví dụ 9.6.5, cùng với bản pdf của phân phối ban đầu. Hình ảnh bên phải hiển thị hàm khả năng ghi nhật ký tương ứng với dữ liệu, có mức tối đa cục bộ khác với mức tối đa toàn cầu. Các ước tính mật độ tương ứng với hai cực đại được hiển thị bên trái.

Công cụ ước tính cho giá trị trung bình chỉ là giá trị trung bình mẫu. Công cụ ước tính cho phương sai là phương sai mẫu được thay đổi tỷ lệ.

Hình 9.11 hiển thị hàm log-likelihood tương ứng với 50 mẫu iid từ phân phối Gaussian với μ := 3 và σ := 4. Nó cũng cho thấy giá trị gần đúng với pdf thực thu được theo khả năng tối đa. Trong Ví dụ 9.6.3 và 9.6.4 , hàm log-likelihood hoàn toàn lõm.
Điều này có nghĩa là hàm có giá trị cực đại duy nhất có thể được định vị bằng cách đặt độ dốc thành 0. Khi điều này tạo ra các phương trình phi tuyến tính không thể giải trực tiếp, chúng ta có thể tận dụng các phương pháp tối ưu hóa, chẳng hạn như tăng dần độ dốc sẽ hội tụ đến mức tối đa. Tuy nhiên, hàm log-likelihood không phải lúc nào cũng lõm. Như được minh họa bằng ví dụ sau, trong những trường hợp như vậy, nó có thể có nhiều cực đại cục bộ, điều này có thể khiến việc tính toán ước lượng khả năng cực đại trở nên khó khăn.

Ví dụ 9.6.5 (Hàm log-likelihood của hỗn hợp Gaussian). Đặt X là hỗn hợp Gaussian được định nghĩa là

$$X := \begin{cases} G1 \text{ với xác suất } \frac{1}{5}, \\ G2 \text{ với xác suất } \frac{4}{5}, \end{cases} \tag{9.77}$$

trong đó G1 là một biến ngẫu nhiên Gaussian với giá trị trung bình μ và phương ^{2}, trong khi G2 cũng sai σ Gaussian với các tham số trung bình ^{2}là . Chúng tôi đã tham số hóa hỗn hợp chỉ với hai μ và phương sai σ để chúng ta có thể hình dung khả năng log theo hai chiều. Cho x1, x2, . . . là dữ liệu được mô hình hóa dưới dạng các mẫu iid từ X. Hàm khả năng bằng

$$L_{x_1,...,x_n}(\mu, \sigma) = f_{\mu,\sigma}(x_1, \ldots, x_n) \tag{9.78}$$

$$= \prod_{i=1}^{N} \frac{1}{5\sqrt{2\pi}\sigma} e^{-\frac{(x_i+\mu)^2}{2\sigma^2}} + \frac{4}{5\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \tag{9.79}$$

và hàm log-likelihood để

$$\log L_{x_1,...,x_n}(\mu, \sigma) = \sum_{i=1}^{N} \log \frac{1}{5\sqrt{2\pi}\sigma} e^{-\frac{(x_i+\mu)^2}{2\sigma^2}} + \frac{4}{5\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}. \tag{9.80}$$

Hình 9.12 cho thấy hàm log-likelihood cho 40 mẫu iid của phân bố khi $\mu := 4$ và $\sigma := 1$. Hàm này có cực đại cục bộ khác với cực đại toàn cục. Điều này có nghĩa là nếu chúng ta sử dụng phương pháp đi lên cục bộ để tìm công cụ ước tính ML, chúng ta có thể không tìm thấy mức tối đa toàn cầu mà thay vào đó vẫn bị kẹt ở mức tối đa cục bộ. Ước tính tương ứng với cực đại cục bộ (hiển thị bên trái) có cùng phương sai với cực đại toàn cầu nhưng μ gần bằng 4 thay vì 4. Mặc dù ước tính không khớp với dữ liệu lắm nhưng nó tối ưu cục bộ, nhỏ sự thay đổi của μ và σ mang lại sự phù hợp tồi tệ hơn (xét về khả năng xảy ra).

Để kết thúc phần này, chúng tôi mô tả một thuật toán học máy dành cho học có giám sát dựa trên việc khớp tham số bằng cách sử dụng ước tính ML.

Ví dụ 9.6.6 (Phân tích biệt thức bậc hai). Phân tích biệt thức bậc hai là một thuật toán cho việc học có giám sát. Đầu vào của thuật toán là hai bộ dữ liệu huấn luyện, bao gồm bn thuộc hai lớp khác nhau (ing của các thức có thể dễ dàng mở rộng để xử lý nhiều lớp hơn). Mục vectơ d chiều $a_1, \ldots, a_n$ và $b_1, \ldots,$ phương tiêu là phân loại các phiên bản mới dựa trên cấu trúc của dữ liệu.

Để thực hiện phân tích phân biệt bậc hai, trước tiên, chúng tôi điều chỉnh phân phối Gaussian d chiều cho dữ liệu của từng lớp bằng cách sử dụng công cụ ước tính ML cho ma trận trung bình và hiệp phương sai, tương ứng với ma trận trung bình và hiệp phương sai mẫu của dữ liệu đào tạo (lên đến một chút thay đổi tỷ lệ của hiệp phương sai mẫu). Chi tiết hơn, $a_1, \ldots, a_n$ được sử dụng để ước tính giá trị trung bình μ_a và ma trận hiệp phương sai Σ_a, trong khi $b_1, \ldots, b_n$ được sử dụng để ước lượng μ_b và Σ_b,

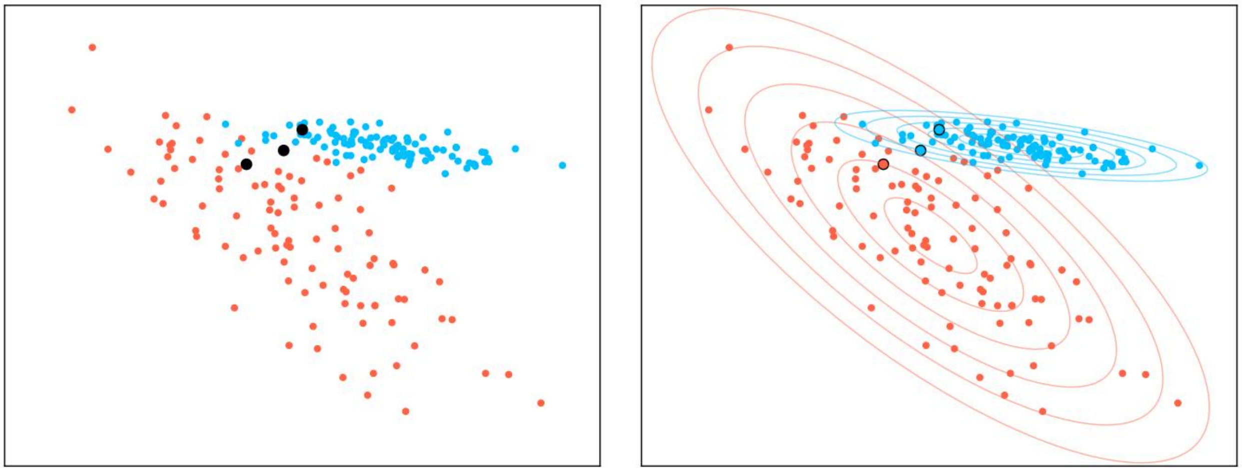
$$\{\mu_a, \Sigma_a\} := \arg\max_{\mu,\Sigma} L_{a_1,...,a_n}(\mu, \Sigma), \tag{9.81}$$

$$\{\mu_b, \Sigma_b\} := \arg\max_{\mu,\Sigma} L_{b_1,...,b_n}(\mu, \Sigma). \tag{9.82}$$

Sau đó, đối với mỗi ví dụ mới x, giá trị của hàm mật độ tại ví dụ cho cả hai lớp được đánh giá. Nếu như

$$f_{\mu_a,\Sigma_a}(x) > f_{\mu_b,\Sigma_b}(x) \tag{9.83}$$

thì x được khai báo thuộc lớp thứ nhất, ngược lại nó được khai báo thuộc lớp thứ hai. Hình 9.13 cho thấy kết quả của việc áp dụng phương pháp cho dữ liệu được mô phỏng bằng cách sử dụng hai phân phối Gaussian.



Hình 9.13: Phân tích phân biệt bậc hai áp dụng cho dữ liệu từ hai lớp khác nhau (trái). Dữ liệu tương ứng với hai lớp khác nhau có màu cam và xanh dương. Ba ví dụ mới được tô màu đen. Hai Gaussian hai biến phù hợp với dữ liệu. Các đường đồng mức của chúng được thể hiện bằng màu tương ứng của từng lớp bên phải. Các bản phân phối này được sử dụng để phân loại các ví dụ mới, được tô màu theo lớp ước tính của chúng.

9.7 Bằng chứng

9.7.1 Chứng minh bổ đề 9.2.5

Chúng tôi xem xét phương sai mẫu của một chuỗi iid X với trung bình μ và phương sai σ^2,

$$Y(n) := \frac{1}{n-1}\sum_{i=1}^{n}\left(X(i) - \frac{1}{n}\sum_{j=1}^{n} X(j)\right)^2 \tag{9.84}$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(X(i) - \frac{1}{n}\sum_{j=1}^{n} X(j)\right)^2 \qquad (9,85)$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(X(i)^2 + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n} X(j)X(k) - \frac{2}{n}\sum_{j=1}^{n} X(i)X(j)\right) \tag{9.86}$$

Để đơn giản hóa ký hiệu, chúng tôi biểu thị bình phương trung bình $EX(i)^2 = \mu^{2}\, 2 + \sigma$ bởi ξ. Chúng ta có

$$EY(n) = \frac{1}{n\ \ 1} \sum_{tôi=1}^{N} EX(i)^2 + \frac{1}{n2} \sum_{j=1}^{N} EX(j)^2 + \frac{1}{n2} \sum_{j=1}^{N} \sum_{k=1, k=j}^{N} EX(j)X(k) \tag{9.87}$$

$$- \frac{2}{N} EX(i)^2 - \frac{2}{n} \sum_{j=1, j=i}^{N} EX(i)X(j) \tag{9.88}$$

$$= \frac{1}{n\ \ 1} \sum_{tôi=1}^{N} \xi\ \xi + \frac{n}{+ n2} \frac{n(n\ \ 1)\mu^2}{n2} - \frac{2\xi}{N} - \frac{2(n\ \ 1)\mu^2}{N} \tag{9.89}$$

$$= \frac{1}{n\ \ 1} \sum_{tôi=1}^{N} \frac{n\ \ 1}{N} \xi \quad \mu^2 \tag{9,90}$$

$$2 = \sigma. \tag{9.91}$$

9.7.2 Chứng minh Định lý 9.3.4

Ta ký hiệu trung vị mẫu là Y(n). Mục đích của chúng ta là chứng minh rằng với mọi > 0

$$\lim_{n \to \infty} P\,Y(n)\ \ y \geq = 0. \tag{9.92}$$

Chúng tôi sẽ chứng minh rằng

$$\lim_{n \to \infty} P\,Y(n) \geq y + = 0. \tag{9.93}$$

Lập luận tương tự cho phép thiết lập

$$\lim_{n \to \infty} P\,Y(n) \leq y \quad = 0. \tag{9.94}$$

Nếu chúng ta đặt bộ X (1), . . . , X(n) , thì Y(n) bằng phần tử thứ (n + 1)/2 nếu n lẻ và trung bình cộng của phần tử thứ n/2 và (n/2 + 1) nếu n chẵn. Biến cố Y(n) ≥ y + do đó ngụ ý rằng ít nhất (n + 1)/2 phần tử lớn hơn y + .

Với mỗi cá nhân X(i), xác suất để X(i) > y + là

$$p := 1 \quad F_{X(i)}(y +) = 1/2 \tag{9,95}$$

trong đó chúng tôi giả định rằng > 0. Nếu không phải như vậy thì cdf của chuỗi iid không đổi tại y và trung vị không được xác định rõ. Số biến ngẫu nhiên trong tập X(1), . . . , X(n) lớn hơn y + được phân phối dưới dạng biến ngẫu nhiên nhị thức Bn với tham số n và P. Kết quả là, chúng ta có

$$PY(n) \geq y + \leq P \quad \frac{n+}{2} \text{1 mẫu trở lên lớn hơn hoặc bằng y +} \tag{9.96}$$

$$= P\, Bn \geq \frac{n+1}{2} \tag{9.97}$$

$$= P\, Bn \quad np \geq \frac{n+1}{2} - np \tag{9,98}$$

$$1 \leq P\, |Bn \quad np| \geq n + 2 \; - \tag{9,99}$$

$$\leq \frac{\text{Biến } (Bn)}{n + \frac{1}{2}^{2}} \quad \text{bằng bất đẳng thức Chebyshev} \tag{9.100}$$

$$= \frac{np\,(1 - p)}{n2 + \frac{1}{2n}^{2}} \tag{9.101}$$

$$= \frac{p\,(1 \quad p)}{1n + 2n \; -^{2}}, \tag{9.102}$$

hội tụ về 0 khi n ∞. Điều này thiết lập (9,93).

Chương 10

Thống kê Bayes

Trong mô hình người theo chủ nghĩa thường xuyên, chúng tôi lập mô hình dữ liệu dưới dạng thực hiện từ một phân phối cố định. Đặc biệt, nếu mô hình là tham số, thì các tham số là đại lượng xác định. Ngược lại, trong mô hình tham số Bayes, các tham số được mô hình hóa dưới dạng các biến ngẫu nhiên. Mục tiêu là có sự linh hoạt để định lượng trước sự không chắc chắn của chúng tôi về phân phối cơ bản, chẳng hạn như để tích hợp thông tin có sẵn trước đó về dữ liệu.

10.1 Các mô hình tham số Bayes

Trong phần này, chúng tôi mô tả cách khớp một mô hình tham số với một tập dữ liệu trong khuôn khổ Bayesian. Như trong Phần 9.6, chúng tôi giả định rằng dữ liệu được tạo bằng cách lấy mẫu từ các phân phối đã biết với các tham số chưa biết. Sự khác biệt quan trọng là chúng tôi mô hình hóa các tham số là ngẫu nhiên thay vì xác định. Điều này đòi hỏi phải chọn phân phối trước của chúng trước khi khớp dữ liệu, điều này cho phép định lượng trước sự không chắc chắn của chúng tôi về giá trị của các tham số. Một mô hình tham số Bayes được chỉ định bởi:

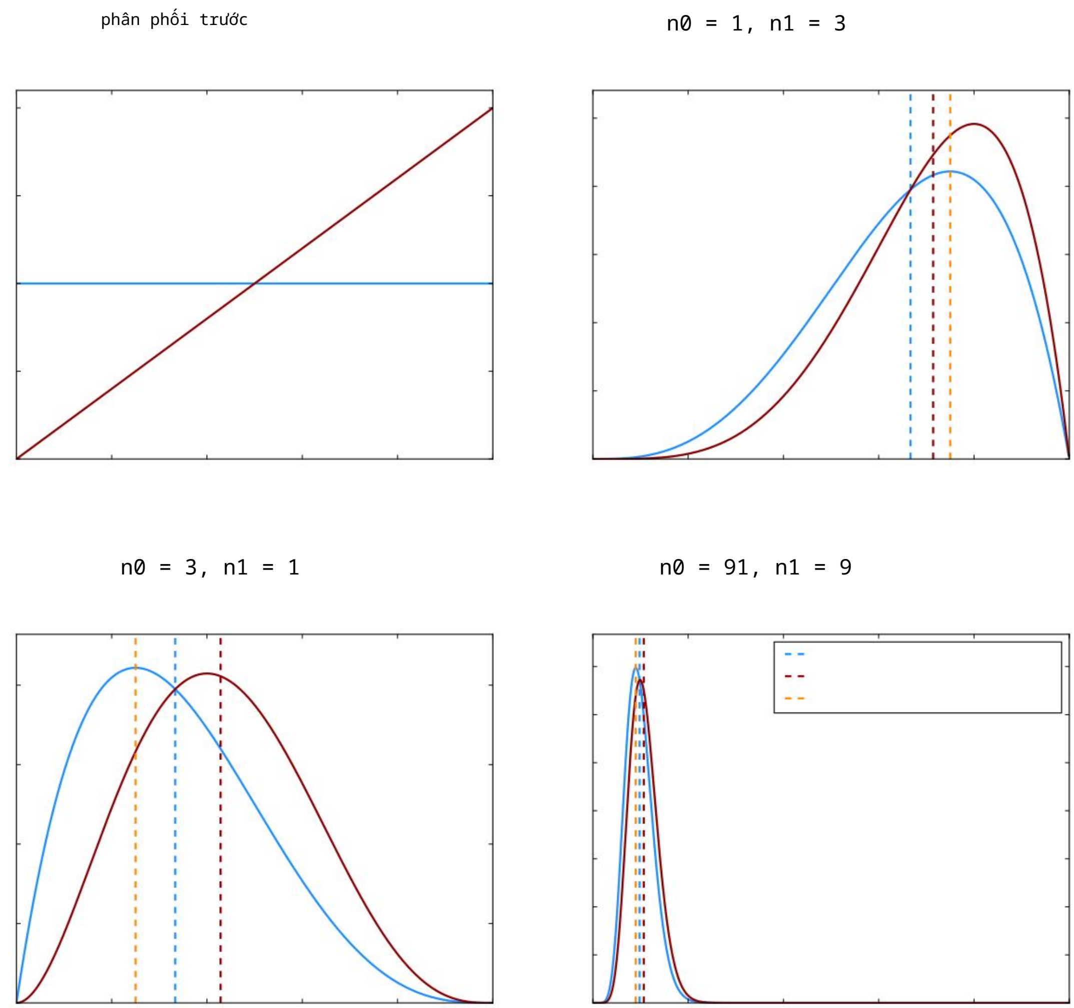
1. Phân phối trước là phân phối của Θ, mã hóa sự không chắc chắn của chúng ta về mô hình trước khi xem dữ liệu.

2. Khả năng xảy ra là phân phối có điều kiện của X cho trước Θ, xác định mức độ phụ thuộc của dữ liệu vào các tham số. Trái ngược với khung phổ biến, khả năng không được hiểu là hàm xác định của các tham số.

Mục tiêu của chúng ta khi học một mô hình Bayes là tính toán phân phối hậu nghiệm của các tham số Θ đã cho X . Đánh giá phân phối sau này khi thực hiện x cho phép cập nhật sự không chắc chắn của chúng tôi về Θ bằng cách sử dụng dữ liệu.

Ví dụ sau phù hợp với mô hình Bayesian với các mẫu iid từ biến ngẫu nhiên Bernoulli.

Ví dụ 10.1.1 (Phân phối Bernoulli). Đặt x là một vectơ dữ liệu mà chúng tôi muốn mô hình hóa dưới dạng các mẫu iid từ phân phối Bernoulli. Vì chúng tôi đang sử dụng phương pháp Bayes nên chúng tôi chọn phân phối trước cho tham số của Bernoulli. Chúng ta sẽ xem xét hai công cụ ước lượng Bayesian khác nhau Θ1 và Θ2:



Hình 10.1: Phân phối trước của 01 (màu xanh lam) và 02 (màu đỏ đậm) trong Ví dụ 10.1.1 được hiển thị trong biểu đồ trên cùng bên trái. Phần còn lại của biểu đồ hiển thị các phân phối sau tương ứng cho các tập dữ liệu khác nhau.

1. 01 đại diện cho một ước lượng thận trọng về mặt thông tin trước đó. Chúng tôi chỉ định một pdf thống nhất cho tham số. Bất kỳ giá trị nào trong khoảng thời gian đơn vị đều có cùng mật độ xác suất:

$$f01\ (\theta) = \begin{cases} 1 \text{ cho } 0 \leq \theta \leq 1, \\ 0 \text{ khác.} \end{cases} \tag{10.1}$$

2. 02 là một công cụ ước tính giả định rằng tham số gần với 1 hơn là 0. Chẳng hạn, chúng ta có thể sử dụng nó để nắm bắt nghi ngờ rằng đồng xu thiên về mặt ngửa. Chúng tôi chọn một bản pdf bị lệch tăng tuyến tính từ 0 đến 1,

$$f02\ (\theta) = \begin{cases} 2\ \theta \text{ với } 0 \leq \theta \leq 1, \\ 0 \text{ khác.} \end{cases} \tag{10.2}$$

Theo giả định iid, khả năng, chỉ là pmf có điều kiện của dữ liệu được cung cấp tham số của Bernoulli, bằng

$$p_{X|\Theta}(x|\theta) = \theta^{n_1}(1-\theta)^{n_0}, \tag{10.3}$$

trong đó n_1 là số đơn vị trong dữ liệu và n_0 là số không (xem Ví dụ 9.6.3).
Do đó, các pdf sau của hai công cụ ước tính bằng

$$f_{\Theta_1|X}(\theta|x) = \frac{f_{\Theta_1}(\theta)\, p_{X|\Theta_1}(x|\theta)}{p_X(x)} \tag{10.4}$$

$$= \frac{f_{\Theta_1}(\theta)\, p_{X|\Theta_1}(x|\theta)}{\int_u f_{\Theta_1}(u)\, p_{X|\Theta_1}(x|u)\, du} \tag{10,5}$$

$$= \frac{\theta^{n_1}(1-\theta)^{n_0}}{\int_u u^{n_1}(1-u)^{n_0}\, du} \tag{10.6}$$

$$= \frac{\theta^{n_1}(1-\theta)^{n_0}}{\beta(n_1+1, n_0+1)}, \tag{10.7}$$

$$f_{\Theta_2|X}(\theta|x) = \frac{f_{\Theta_2}(\theta)\, p_{X|\Theta_2}(x|\theta)}{p_X(x)} \tag{10.8}$$

$$= \frac{\theta^{n_1+1}(1-\theta)^{n_0}}{\int_u u^{n_1+1}(1-u)^{n_0}\, du} \tag{10.9}$$

$$= \frac{\theta^{n_1+1}(1-\theta)^{n_0}}{\beta(n_1+2, n_0+1)}, \tag{10.10}$$

(10.11)

Ở đâu

$$\beta(a, b) := \int_u u^{a-1}(1-u)^{b-1}\, du \tag{10.12}$$

là một hàm đặc biệt gọi là hàm beta hay tích phân Euler loại một, được lập thành bảng.

Hình 10.1 cho thấy đồ thị của phân phối sau đối với các giá trị khác nhau của n_1 và n_0. Nó cũng cho thấy công cụ ước lượng khả năng xảy ra cực đại của tham số, chỉ là $n_1/(n_0+n_1)$ (xem Ví dụ 9.6.3). Đối với một số lần lật nhỏ, pdf sau của 02 bị lệch sang phải so với 01, phản ánh niềm tin trước đó rằng tham số gần với 1 hơn. Tuy nhiên, đối với một số lượng lớn các lần lật, cả hai mật độ sau đều rất gần nhau.

10.2 Liên hợp trước

Cả hai bản phân phối sau trong Ví dụ 10.1.1 đều là bản phân phối beta (xem Định nghĩa 2.3.12) và bản phân phối trước cũng vậy. Phần trước đồng nhất của 01 là beta với các tham số a = 1 và b = 1, trong khi phần trước sai lệch của 02 là phân phối beta với các tham số a = 2 và b = 1. Vì phần trước và phần sau thuộc cùng một họ, nên tính toán sau tương đương với việc chỉ cập nhật các tham số. Khi phần trước và phần sau được đảm bảo thuộc về cùng một họ phân phối cho một khả năng cụ thể, thì các phân phối được gọi là phần trước liên hợp.

Định nghĩa 10.2.1 (Liên hợp linh mục). Một họ phân phối liên hợp cho một khả năng nhất định thỏa mãn tính chất sau: nếu cái trước thuộc về họ, thì cái sau cũng thuộc về họ.

Phân phối beta là linh mục liên hợp khi khả năng là nhị thức.

Định lý 10.2.2 (Phân phối beta liên hợp với khả năng nhị thức). Nếu phân phối trước của θ là phân phối beta với tham số a và b và khả năng dữ liệu X cho θ là nhị thức với tham số n và x, thì phân phối sau của θ cho X là phân phối beta với tham số $x + a$ và $n - x + b$.

Bằng chứng.

$$f_{\theta \mid X}(\theta \mid x) = \frac{f_{\theta}(\theta)\, p_{X \mid \theta}(x \mid \theta)}{p_X(x)} \tag{10.13}$$

$$= \frac{p_{X \mid \theta}(x \mid \theta)\, f_{\theta}}{\int_u (u)\, p_{X \mid \theta}(x \mid u)\, du} \tag{10.14}$$

$$= \frac{\theta^{a-1}(1-\theta)^{b-1}\binom{n}{x}\theta^{x}(1-\theta)^{n-x}}{\int_u u^{a-1}(1-u)^{b-1}\binom{n}{x}u^{x}(1-u)^{n-x}\, du} \tag{10.15}$$

$$= \frac{\theta^{x+a-1}(1-\theta)^{n-x+b-1}}{\int_u u^{x+a-1}(1-u)^{n-x+b-1}\, du} \tag{16.10}$$

$$= f_{\beta}(\theta; x + a, n - x + b). \tag{10.17}$$

□

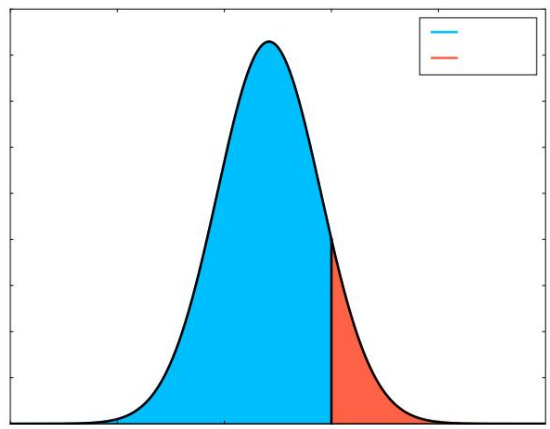
Lưu ý rằng các hậu nghiệm thu được trong Ví dụ 10.1.1 theo ngay lập tức từ định lý.

Ví dụ 10.2.3 (Thăm dò ở New Mexico). Trong một cuộc thăm dò ở New Mexico cho cuộc bầu cử Hoa Kỳ năm 2016, 429 người tham gia, 227 người dự định bỏ phiếu cho Clinton và 202 người cho Trump (dữ liệu là từ một cuộc thăm dò thực tế[1] , nhưng để đơn giản, chúng tôi đang bỏ qua các ứng cử viên khác và những người chưa quyết định) . Mục đích của chúng tôi là sử dụng khung Bayesian để dự đoán kết quả của cuộc bầu cử ở New Mexico bằng cách sử dụng những dữ liệu này.

Chúng tôi lập mô hình tỷ lệ người bỏ phiếu cho Trump như một biến ngẫu nhiên θ. Chúng tôi giả định rằng n người trong cuộc thăm dò được chọn ngẫu nhiên thống nhất với sự thay thế từ dân số, do đó, với $\theta = \theta$ số cử tri Trump là một nhị thức với các tham số n và θ. Chúng tôi không có bất kỳ thông tin bổ sung nào về giá trị có thể có của θ, vì vậy chúng tôi cho rằng đó là phân phối beta đồng nhất hoặc tương đương với các tham số $a := 1$ và $b := 1$.

Theo Định lý 10.2.2, phân phối sau của θ với dữ liệu mà chúng ta quan sát được là phân phối beta với các tham số $a := 203$ và $b := 228$, được mô tả trong Hình 10.2. Xác suất tương ứng mà $\theta \geq 0,5$ là 11,4%, đây là ước tính của chúng tôi về xác suất Trump thắng trong Mexico mới.

[1]Kết quả cuộc thăm dò được lấy từ https://www.abqjournal.com/883092/clinton-still-ahead-in-new-mexico.html



Hình 10.2: Phân phối hậu thế của tỷ lệ cử tri Trump ở New Mexico dựa trên dữ liệu thăm dò trong Ví dụ 10.2.3.

10.3 Công cụ ước lượng Bayes

Phương pháp Bayesian để học các mô hình xác suất mang lại toàn bộ phân phối sau của các tham số quan tâm. Trong phần này, chúng tôi mô tả hai lựa chọn thay thế để có được một ước tính duy nhất của các tham số từ phân phối sau.

10.3.1 Ước tính sai số bình phương trung bình tối thiểu

Giá trị trung bình của phân phối sau là kỳ vọng có điều kiện của các tham số được cung cấp dữ liệu. Việc chọn giá trị trung bình sau làm công cụ ước tính cho các tham số θ có cơ sở lý thuyết vững chắc: đảm bảo đạt được sai số bình phương trung bình tối thiểu (MSE) trong số tất cả các công cụ ước tính có thể. Tất nhiên, điều này chỉ đúng nếu tất cả các giả định đều đúng, tức là các tham số được tạo theo thông số trước và dữ liệu sau đó được tạo theo khả năng xảy ra, điều này có thể không đúng với trường hợp của dữ liệu thực.

Định lý 10.3.1 (Trung bình sau tối thiểu hóa MSE). Giá trị trung bình sau là ước tính sai số trung bình bình phương tối thiểu (MMSE) tối thiểu của tham số θ với dữ liệu X . Để chính xác hơn, chúng ta hãy định nghĩa

$$\theta_{MMSE}(x) := E\left(\theta \mid X = x\right). \quad (10.18)$$

Đối với bất kỳ ước lượng tùy ý $\theta_{other}(x)$,

$$E\left(\theta_{khác}(X) - \theta\right)^2 \geq E\left(\theta_{MMSE}(X) - \theta\right)^2. \quad (10.19)$$

Bằng chứng. Chúng tôi bắt đầu bằng cách tính toán MSE của công cụ ước tính tùy ý dựa trên X = x trong số hạng của kỳ vọng có điều kiện của θ cho trước X,

$$E\left[\left(\theta_{khác}(X) - \theta\right)^2 \mid X = x\right] \quad (10.20)$$

$$= E\left[\left(\theta_{other}(X) - \theta_{MMSE}(X) + \theta_{MMSE}(X) - \theta\right)^2 \mid X = x\right] \quad (21.10)$$

$$= (\theta_{khác}(x) - \theta_{MMSE}(x))^2 + E\left[\left(\theta_{MMSE}(X) - \theta\right)^2 \mid X = x\right] \quad (10.22)$$

$$+ 2\,(\theta_{khác}(x) - \theta_{MMSE}(x))\left(E\left[\theta_{MMSE}(x) - E[\theta \mid X = x]\right]\right)$$

$$= (\theta_{khác}(x) - \theta_{MMSE}(x))^2 + E\left[\left(\theta_{MMSE}(X) - \theta\right)^2 \mid X = x\right]. \quad (10.23)$$

Bằng kỳ vọng lặp đi lặp lại,

$$E\left[\left(\theta_{khác}(X) - \theta\right)^2\right] = E\left[E\left[\left(\theta_{khác}(X) - \theta\right)^2 \mid X\right]\right] \quad (10.24)$$

$$= E\left[\left(\theta_{khác}(X) - \theta_{MMSE}(X)\right)^2\right] + E\left[E\left[\left(\theta_{MMSE}(X) - \theta\right)^2 \mid X\right]\right]$$

$$= E\left[\left(\theta_{khác}(X) - \theta_{MMSE}(X)\right)^2\right] + E\left[\left(\theta_{MMSE}(X) - \theta\right)^2\right] \quad (10.25)$$

$$\geq E\left[\left(\theta_{MMSE}(X) - \theta\right)^2\right], \quad (26.10)$$

vì kỳ vọng của một đại lượng không âm là không âm. □

Ví dụ 10.3.2 (Phân phối Bernoulli (tiếp theo)). Để có được ước tính điểm cho tham số trong Ví dụ 10.1.1, chúng tôi tính toán phương tiện sau:

$$E(\theta_1 \mid X = x) = \int_0^1 \theta f_{\theta_1|X}(\theta|x)\, d\theta \quad (27.10)$$

$$= \frac{\int_0^1 \theta^{n_1+1}(1-\theta)^{n_0}\, d\theta}{\beta(n_1 + 1, n_0 + 1)} \quad (10.28)$$

$$= \frac{\beta(n_1 + 2, n_0 + 1)}{\beta(n_1 + 1, n_0 + 1)}, \quad (29.10)$$

$$E(\theta_2 \mid X = x) = \int_0^1 \theta f_{\theta_2|X}(\theta|x)\, d\theta \quad (10.30)$$

$$= \frac{\beta(n_1 + 3, n_0 + 1)}{\beta(n_1 + 2, n_0 + 1)}. \quad (31.10)$$

Hình 10.1 cho thấy trung bình sau đối với các giá trị khác nhau của n_0 và n_1.

10.3.2 Ước lượng cực đại hậu nghiệm

Một thay thế cho giá trị trung bình sau là chế độ sau, là mức tối đa của pdf hoặc pmf của phân phối sau.

Định nghĩa 10.3.3 (Ước tính cực đại-a-posteriori). Ước lượng cực đại-a-posteriori (MAP) của tham số θ dữ liệu đã cho x được mô hình hóa dưới dạng hiện thực của vectơ ngẫu nhiên X là

$$\theta_{MAP}(x) := \arg\max_{\theta} p_{\theta \mid X}(\theta \mid x) \tag{10.32}$$

nếu θ được mô hình hóa như một biến ngẫu nhiên rời rạc và

$$\theta_{MAP}(x) := \arg\max_{\theta} f_{\theta \mid X}(\theta \mid x) \tag{10.33}$$

nếu nó được mô hình hóa như một biến ngẫu nhiên liên tục.

Trong Hình 10.1, công cụ ước tính ML của θ là chế độ (giá trị lớn nhất) của phân phối sau khi phân phối trước đồng nhất. Đây không phải là sự trùng hợp ngẫu nhiên, theo thống nhất trước đó, ước tính MAP và ML giống nhau.

Bổ đề 10.3.4. Công cụ ước tính khả năng tối đa của tham số θ là chế độ (giá trị tối đa) của pdf của phân phối sau được cung cấp dữ liệu X nếu phân phối trước của nó là đồng nhất.

Bằng chứng. Chúng tôi chứng minh kết quả khi mô hình cho dữ liệu và các tham số là liên tục, nếu bất kỳ hoặc cả hai đều rời rạc thì bằng chứng là giống hệt nhau (trong trường hợp đó, công cụ ước tính ML là chế độ của pmf sau). Nếu phân phối trước của các tham số là đồng nhất, thì $f_\theta(\theta)$ là hằng số đối với bất kỳ θ nào, nghĩa là

$$\text{đối số tối đa}_{\theta} f_{\theta \mid X}(\theta \mid x) = \arg\max_{\theta} \frac{f_\theta(\theta) f_{X \mid \theta}(x \mid \theta)}{\int_u f_\theta(u) f_{X \mid \theta}(x \mid u)\, du} \tag{10.34}$$

$$= \arg \text{cực đại}_{\theta} f_{X \mid \theta}(x \mid \theta) \quad \text{(các số hạng còn lại không phụ thuộc vào θ)}$$

$$= \arg \text{cực đại}_{\theta} \mathcal{L}_x(\theta). \tag{10.35}$$

□

Lưu ý rằng các ưu tiên thống nhất chỉ được xác định rõ trong các trường hợp tham số bị giới hạn trong một tập hợp giới hạn.

Bây giờ chúng tôi mô tả một tình huống trong đó công cụ ước tính MAP là tối ưu. Nếu tham số θ chỉ có thể nhận một tập hợp các giá trị rời rạc, thì công cụ ước tính MAP sẽ giảm thiểu khả năng đưa ra lựa chọn sai.

Định lý 10.3.5 (Bộ ước lượng MAP giảm thiểu xác suất lỗi). Đặt θ là một vectơ ngẫu nhiên rời rạc và đặt X là một vectơ ngẫu nhiên mô hình hóa dữ liệu. Chúng tôi xác định

$$\theta_{MAP}(x) := \arg\max_{\theta} p_{\theta \mid X}(\theta \mid X = x). \tag{10.36}$$

Đối với bất kỳ ước lượng tùy ý $\theta_{other}(x)$,

$$P(\theta_{khác}(X) = \theta) \geq P(\theta_{MAP}(X) = \theta). \tag{10.37}$$

Nói cách khác, công cụ ước tính MAP giảm thiểu xác suất lỗi.

Bằng chứng. Chúng ta giả sử rằng X là một vectơ ngẫu nhiên liên tục, nhưng cũng áp dụng lập luận tương tự nếu nó rời rạc. Chúng ta có

$$P\,\theta = \theta_{khác}(X) = \int_x f_X(x)\, P\,\theta = \theta_{other}(x) \mid X = x\, dx \quad (10.38)$$

$$= \int_x f_X(x)\, p_{\theta \mid X}(\theta_{khác}(x) \mid x)\, dx \quad (10.39)$$

$$\leq \int_x f_X(x)\, p_{\theta \mid X}(\theta_{MAP}(x) \mid x)\, dx \quad (10.40)$$

$$= P\,\theta = \theta_{MAP}(X), \quad (10.41)$$

trong đó (10,40) xuất phát từ định nghĩa của công cụ ước tính MAP là chế độ sau. □

Ví dụ 10.3.6 (Gửi bit). Chúng tôi xem xét một mô hình rất đơn giản cho một kênh liên lạc trong đó chúng tôi muốn gửi tín hiệu θ bao gồm một bit đơn lẻ. Kiến thức trước đây của chúng tôi chỉ ra rằng tín hiệu bằng một với xác suất 1/4.

$$p(1) = \frac{1}{4}, \quad p_\theta(0) = \frac{3}{4}. \quad (10.42)$$

Do có nhiễu trong kênh, chúng tôi gửi tín hiệu n lần. Tại cơ quan tiếp nhận, chúng tôi quan sát

$$X_i = \theta + Z_i, \quad 1 \leq i \leq n, \quad (10.43)$$

trong đó Z chứa n iid biến ngẫu nhiên Gaussian chuẩn. Mô hình hóa nhiễu loạn như Gaus sian là một lựa chọn phổ biến trong truyền thông. Nó được chứng minh bằng định lý giới hạn trung tâm, với giả định rằng tiếng ồn là sự kết hợp của nhiều hiệu ứng nhỏ gần như độc lập.

Bây giờ chúng ta sẽ tính toán và so sánh các công cụ ước tính ML và MAP của θ dựa trên các quan sát. Xác suất bằng

$$L_x(\theta) = \prod_{tôi=1}^{N} f_{X_i \mid \theta}(x_i \mid \theta) \quad (10.44)$$

$$= \prod_{tôi=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - \theta)^2}{2}}. \quad (10.45)$$

Việc xử lý hàm log-likelihood sẽ dễ dàng hơn,

$$\log L_x(\theta) = -\sum_{tôi=1}^{N} \frac{(x_i - \theta)^2}{2} - \frac{N}{2}\log 2\pi. \quad (10.46)$$

Vì θ chỉ nhận hai giá trị nên ta có thể so sánh trực tiếp. Chúng ta sẽ chọn $\theta_{ML}(x) = 1$ nếu

$$\log L_x(1) = -\sum_{tôi=1}^{N} \frac{x_i^2 - 2x_i + 1}{2} - \frac{N}{2}\log 2\pi \quad (10.47)$$

$$\geq -\sum_{tôi=1}^{n} \frac{x_i^2}{2} - \frac{N}{2}\log 2\pi \quad (10.48)$$

$$= \log L_x(0). \quad (10.49)$$

Machine Translated by Google

tương đương,

$$\theta_{ML}(x) = \begin{cases} 1 & \text{nếu } \frac{1}{N}\sum_{tôi=1}^{N} x_i > \frac{1}{2}, \\ 0 & \text{khác.} \end{cases} \quad (10.50)$$

Quy tắc rất có ý nghĩa: nếu giá trị trung bình mẫu của dữ liệu gần với 1 hơn 0 thì ước tính của chúng ta bằng 1. Theo luật xác suất tổng, xác suất sai sót của công cụ ước tính này bằng

$$P\left(\theta \neq \theta_{ML}(X)\right) = P\left(\theta \neq \theta_{ML}(X) \mid \theta = 0\right) P(\theta = 0) + P\left(\theta \neq \theta_{ML}(X) \mid \theta = 1\right) P(\theta = 1)$$
$$= P\left(\frac{1}{n}\sum_{i=1}^{N} X_i > \frac{1}{2} \mid \theta = 0\right) P(\theta = 0) + P\left(\frac{1}{N}\sum_{tôi=1}^{N} X_i < \frac{1}{2} \mid \theta = 1\right) P(\theta = 1)$$
$$= Q\left(\sqrt{n}/2\right), \quad (10.51)$$

trong đó đẳng thức cuối cùng xuất phát từ thực tế là nếu chúng ta đặt điều kiện trên $\theta = \theta$ thì trung bình thực nghiệm là Gaussian với phương sai σ^2/n và trung bình θ (xem phần chứng minh của Định lý 6.2.2).

Để tính toán ước tính MAP, chúng ta phải tìm giá trị lớn nhất của pdf sau của θ với dữ liệu được quan sát. Tương tự, chúng ta tìm giá trị lớn nhất của logarit của nó (điều này tương đương vì logarit là một hàm đơn điệu),

$$\log p_{\theta|X}(\theta|x) = \log \frac{\prod_{i=1}^{N} f_{X_i|\theta}(x_i|\theta)\, p_\theta(\theta)}{f_X(x)} \quad (10.52)$$
$$= \sum_{tôi=1}^{N} \log f_{X_i|\theta}(x_i|\theta)\, p_\theta(\theta) - \log f_X(x) \quad (10.53)$$
$$= -\sum_{tôi=1}^{N} \frac{x_i^2 - 2x_i\theta + \theta}{2} - \frac{N}{2}\log 2\pi + \log p_\theta(\theta) - \log f_X(x). \quad (10.54)$$

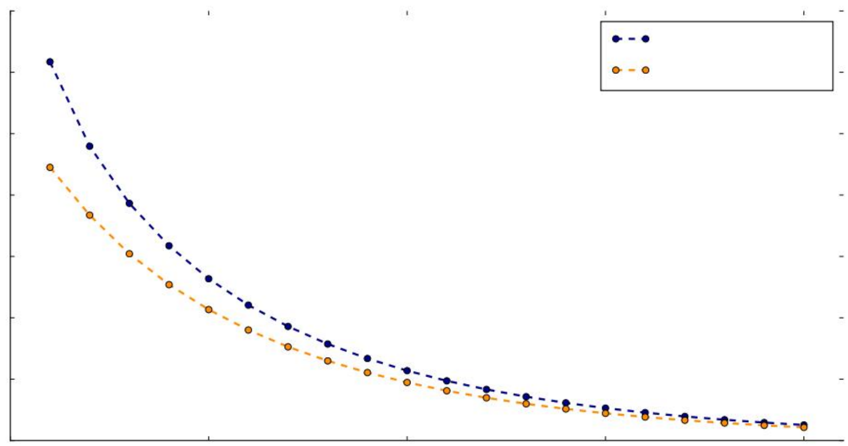
Chúng ta so sánh giá trị của hàm này với hai giá trị có thể có của θ: 0 và 1. Chúng ta chọn $\theta_{MAP}(x) = 1$ nếu

$$\log p_{\theta|X}(1|x) + \log f_X(x) = -\sum_{tôi=1}^{N} \frac{x_i^2 - 2x_i + 1}{2} - \frac{N}{2}\log 2\pi - \log 4 \quad (10.55)$$
$$\geq -\sum_{tôi=1}^{n} \frac{x_i^2}{2} - \frac{N}{2}\log 2\pi - \log 4 + \log 3 \quad (10.56)$$
$$= \log p_{\theta|X}(0|x) + \log f_X(x). \quad (10.57)$$

tương đương,

$$\theta_{MAP}(x) = \begin{cases} 1 & \text{nếu } \frac{1}{N}\sum_{i=1}^{N} x_i > \frac{1}{2} + \frac{\log 3}{n}, \\ 0 & \text{ngược lại.} \end{cases} \quad (10.58)$$

Ước tính MAP thay đổi ngưỡng đối với ước tính ML để tính đến việc θ có xu hướng bằng 0 hơn. Tuy nhiên, số hạng hiệu chỉnh có xu hướng bằng 0 khi chúng tôi thu thập thêm bằng chứng, vì vậy nếu có nhiều dữ liệu thì hai công cụ ước tính sẽ rất giống nhau.



Hình 10.3: Xác suất lỗi của các công cụ ước tính ML và MAP trong Ví dụ 10.3.6 đối với các giá trị khác nhau của n.

Xác suất lỗi của công cụ ước tính MAP bằng

$$P\left(\theta = \theta_{MAP}(X)\right) = P\left(\theta = \theta_{MAP}(X) \mid \theta = 0\right) P(\theta = 0) + P\left(\theta = \theta_{MAP}(X) \mid \theta = 1\right) P(\theta = 1)$$

$$= P\left(\frac{1}{n}\sum_{i=1}^{N} X_i > \frac{1}{2} + \frac{\log 3}{N} \,\middle|\, \theta = 0\right) P(\theta = 0) \tag{10.59}$$

$$+ P\left(\frac{1}{N}\sum_{i=1}^{N} X_i < \frac{1}{2} + \frac{\log 3}{N} \,\middle|\, \theta = 1\right) P(\theta = 1)$$

$$= \frac{3}{4} Q\left(\sqrt{n}/2 + \frac{\log 3}{\sqrt{n}}\right) + \frac{1}{4} Q\left(\sqrt{n}/2 - \frac{\log 3}{\sqrt{n}}\right). \tag{10.60}$$

Chúng tôi so sánh xác suất lỗi của các công cụ ước tính ML và MAP trong Hình 10.3. Ước tính MAP dẫn đến hiệu suất tốt hơn, nhưng sự khác biệt trở nên nhỏ khi n tăng lên.

chương 11

kiểm định giả thuyết

Trong một nghiên cứu y học, chúng tôi quan sát thấy rằng 10% phụ nữ và 12,5% nam giới mắc bệnh tim. Nếu có 20 người trong nghiên cứu, có lẽ chúng ta sẽ do dự khi tuyên bố rằng phụ nữ ít bị bệnh tim hơn nam giới; rất có thể kết quả xảy ra một cách tình cờ. Tuy nhiên, nếu có 20.000 người tham gia nghiên cứu, thì có vẻ như chúng ta đang quan sát một hiện tượng có thật. Thử nghiệm giả thuyết làm cho trực giác này trở nên chính xác; đó là một khuôn khổ cho phép chúng tôi quyết định xem các mẫu mà chúng tôi quan sát được trong dữ liệu của mình có khả năng là kết quả của các biến động ngẫu nhiên hay không.

11.1 Khung kiểm tra giả thuyết

Mục đích của thử nghiệm giả thuyết là để đánh giá một phỏng đoán được xác định trước. Trong ví dụ trên, đây có thể là bệnh tim phổ biến ở nam giới hơn nữ giới. Giả thuyết cho rằng phỏng đoán của chúng ta là sai được gọi là giả thuyết vô hiệu, ký hiệu là H_0. Trong ví dụ của chúng tôi, giả thuyết vô hiệu sẽ là bệnh tim ít nhất cũng phổ biến ở nam giới như ở nữ giới. Nếu giả thuyết không đúng, thì bất kỳ mẫu nào chúng tôi đang phát hiện trong dữ liệu của mình dường như hỗ trợ cho phỏng đoán của chúng tôi đều chỉ là một sự may rủi. Tình cờ có rất nhiều đàn ông mắc bệnh tim (hoặc phụ nữ không mắc bệnh) trong nghiên cứu. Ngược lại, giả thuyết theo đó phỏng đoán của chúng ta là đúng được gọi là giả thuyết thay thế, ký hiệu là H_1. Trong chương này, chúng ta có quan điểm thường xuyên: các giả thuyết hoặc đúng hoặc không, chúng không được mô hình hóa theo xác suất.

Thử nghiệm là một thủ tục để xác định xem chúng ta có nên bác bỏ giả thuyết không dựa trên dữ liệu hay không. Bác bỏ giả thuyết không có nghĩa là chúng ta cho rằng điều đó khó có thể xảy ra, đó là bằng chứng ủng hộ giả thuyết thay thế. Nếu chúng tôi không bác bỏ giả thuyết không, điều này không có nghĩa là chúng tôi cho rằng nó có khả năng xảy ra, chỉ là chúng tôi không có đủ thông tin để loại bỏ nó. Hầu hết các bài kiểm tra đưa ra quyết định bằng cách tạo ngưỡng cho một thống kê kiểm tra, đây là một chức năng ánh xạ dữ liệu (tứclà một vectơ trong R^n) thành một số duy nhất. Kiểm định bác bỏ giả thuyết không nếu thống kê kiểm định thuộc vùng bác bỏ R. Ví dụ, chúng ta có thể có

$$R := \{t \mid t \geq \eta\}, \tag{11.1}$$

trong đó t là thống kê kiểm tra được tính toán từ dữ liệu và η là ngưỡng được xác định trước. Trong trường hợp này, chúng ta chỉ bác bỏ giả thuyết không nếu t lớn hơn η.

Như thể hiện trong Bảng 11.1, có hai lỗi mà chúng ta có thể mắc phải. Lỗi loại I là sai: phỏng đoán của chúng tôi là sai, nhưng chúng tôi bác bỏ giả thuyết không. Lỗi loại II là

Bác bỏ H0?

	KHÔNG	Đúng
H0 là đúng		Lỗi loại I
H1 đúng là	lỗi Loại II	

Bảng 11.1: Lỗi loại I và II.

phủ định sai: phỏng đoán của chúng tôi đúng, nhưng chúng tôi không bác bỏ giả thuyết không. Trong thử nghiệm giả thuyết, ưu tiên của chúng tôi là kiểm soát lỗi Loại I. Khi bạn đọc trong một nghiên cứu rằng kết quả có ý nghĩa thống kê ở mức 0,05, điều này có nghĩa là xác suất phạm phải lỗi Loại I bị giới hạn bởi 5%.

Định nghĩa 11.1.1 (Mức ý nghĩa và kích thước). Kích thước của một bài kiểm tra là xác suất mắc lỗi Loại I. Mức ý nghĩa của một thử nghiệm là giới hạn trên của kích thước.

Việc bác bỏ giả thuyết không có ý nghĩa định lượng về mức độ mà dữ liệu không tương thích với giả thuyết không. Giá trị p là một chức năng của dữ liệu đóng vai trò này.

Định nghĩa 11.1.2 (giá trị p). Giá trị p là mức ý nghĩa nhỏ nhất mà tại đó chúng ta sẽ bác bỏ giả thuyết không đối với dữ liệu chúng ta quan sát được.

Đối với mức ý nghĩa cố định, nên chọn phép thử giảm thiểu xác suất mắc lỗi Loại II. Tương tự, chúng tôi muốn tối đa hóa xác suất bác bỏ giả thuyết không khi nó không đúng. Xác suất này được gọi là sức mạnh của bài kiểm tra.

Định nghĩa 11.1.3 (Công suất). Sức mạnh của một bài kiểm tra là xác suất bác bỏ giả thuyết không nếu nó không đúng.

Lưu ý rằng để mô tả sức mạnh của một thử nghiệm, chúng ta cần biết phân phối dữ liệu theo giả thuyết thay thế, điều này thường không thực tế (nhớ lại rằng giả thuyết thay thế chỉ là phần bổ sung của giả thuyết không và do đó bao gồm nhiều khả năng khác nhau) .

Quy trình chuẩn để áp dụng thử nghiệm giả thuyết trong khoa học ứng dụng như sau:

1. Chọn một phỏng đoán.

2. Xác định giả thuyết vô hiệu tương ứng.

3. Chọn một bài kiểm tra.

4. Thu thập dữ liệu.

5. Tính toán thống kê kiểm tra từ dữ liệu.

6. Tính toán giá trị p và bác bỏ giả thuyết không nếu nó nằm dưới giới hạn được xác định trước (thường là 1% hoặc 5%).

Ví dụ 11.1.4 (Ly hợp). Chúng tôi muốn kiểm tra phỏng đoán rằng một cầu thủ nào đó ở NBA đang chơi ly hợp, tức là anh ta ghi được nhiều điểm hơn vào cuối trận đấu so với phần còn lại của trận đấu. Giả thuyết vô hiệu là không có sự khác biệt trong hiệu suất của anh ta. Thống kê kiểm tra t mà chúng tôi chọn là liệu anh ấy kiếm được nhiều hay ít điểm mỗi phút trong quý cuối cùng so với phần còn lại của trò chơi

$$t(x) = \sum_{tôi=1}^{N} 1_{x_i>0}, \tag{11.2}$$

trong đó x_i là hiệu số giữa số điểm mỗi phút anh ta ghi được trong hiệp thứ 4 và trong phần còn lại của hiệp thứ i với điều kiện $1 \leq i \leq n$.

Miền bác bỏ của phép thử có dạng

$$R := \{t(x) \mid t(x) \geq \eta\}, \tag{11.3}$$

cho một ngưỡng cố định η. Theo giả thuyết không, xác suất ghi được nhiều điểm hơn mỗi phút trong hiệp 4 là 1/2 (để đơn giản, chúng tôi bỏ qua khả năng anh ta ghi được số điểm như nhau), vì vậy chúng tôi có thể lập mô hình thống kê kiểm tra theo giả thuyết không là biến ngẫu nhiên nhị thức với tham số n và 1/2. Nếu η là một số nguyên trong khoảng từ 0 đến n, thì xác suất để thống kê kiểm định nằm trong vùng bác bỏ nếu giả thuyết không đúng là

$$P(T_0 \geq \eta) = 2 \frac{1}{n} \sum_{k=\eta}^{N} \binom{N}{k}. \tag{11.4}$$

Vì vậy, kích thước của bài kiểm tra là $\frac{1}{2^n} \sum_{k=\eta}^{N} \binom{N}{k}$. Bảng 11.2 cho thấy giá trị này cho tất cả các giá trị có thể có của η. Nếu chúng ta muốn mức ý nghĩa là 1% hoặc 5% thì chúng ta cần đặt ngưỡng tương ứng là η = 16 hoặc η = 15.

Chúng tôi thu thập dữ liệu từ 20 trò chơi x và tính toán giá trị của thống kê kiểm tra t(x) (lưu ý rằng chúng tôi sử dụng chữ cái viết thường vì đó là một nhận thức cụ thể), kết quả là 14 (anh ấy ghi được nhiều điểm hơn mỗi phút trong quý thứ tư trong 14 trò chơi). Điều này là không đủ để bác bỏ giả thuyết không đối với mức xác định trước của chúng tôi là 1% hoặc 5%. Do đó kết quả không có ý nghĩa thống kê.

Trong mọi trường hợp, chúng tôi tính toán giá trị p, là mức nhỏ nhất mà tại đó kết quả sẽ có ý nghĩa. Từ bảng nó bằng 0,058. Lưu ý rằng trong khuôn khổ người theo chủ nghĩa thường xuyên, chúng ta không thể giải thích đây là xác suất mà giả thuyết không đúng (tức là người chơi không tốt hơn trong hiệp thứ tư) bởi vì giả thuyết này không phải là ngẫu nhiên, nó có thể đúng hoặc không.
Kết quả của chúng tôi hầu như rất quan trọng và mặc dù chúng tôi không có đủ bằng chứng để hỗ trợ phỏng đoán của mình, nhưng có vẻ hợp lý khi người chơi thể hiện tốt hơn trong quý IV.

11.2 Thử nghiệm tham số

Trong phần này, chúng tôi thảo luận về thử nghiệm giả thuyết với giả định rằng dữ liệu của chúng tôi được lấy mẫu từ một phân phối đã biết với các tham số chưa biết. Một lần nữa chúng tôi có một quan điểm thường xuyên,

η	1	2	3	4	5	6	7	số 8	9	10
$P(T_0 \geq \eta)$	1.000	1.000	1.000	0.999	0.994	0.979	0.942	0.868	0.748	0.588
η	11	12	13	14	15	16	17	18	19	20
$P(T_0 \geq \eta)$	0,412	0,252	0,132	0,058	0,021	0,006	0,001	0,000	0,000	0,000

Bảng 11.2: Xác suất mắc lỗi Loại I tùy thuộc vào giá trị của ngưỡng trong Ví dụ 11.1.4. Các giá trị được làm tròn đến ba chữ số thập phân.

như thường được thực hiện trong hầu hết các nghiên cứu về khoa học ứng dụng. Do đó, tham số là xác định và các giả thuyết cũng vậy: giả thuyết không là đúng hay không, không có như vậy điều như xác suất mà giả thuyết không nắm giữ.

Để đơn giản hóa giải thích, chúng tôi giả sử rằng phân phối xác suất chỉ phụ thuộc vào một tham số mà chúng ta biểu thị bằng θ. P_θ là thước đo xác suất của không gian xác suất của chúng ta nếu θ là giá trị của tham số. X là một vectơ ngẫu nhiên được phân phối theo P_θ. dữ liệu thực tế mà chúng tôi quan sát, mà chúng tôi biểu thị bằng x được coi là một nhận thức từ vectơ ngẫu nhiên này.

Giả sử rằng giả thuyết vô hiệu là $\theta = \theta_0$. Trong trường hợp đó, kích thước của một bài kiểm tra với thống kê kiểm tra T và vùng bác bỏ R bằng

$$\alpha = P_{\theta_0}\left(T(X) \quad R\right). \tag{11.5}$$

Với miền bác bỏ có dạng (11.1) ta có

$$\alpha := P_{\theta_0}\left(T(X) \geq \eta\right). \tag{11.6}$$

Nếu việc thực hiện thống kê kiểm định là $T(x_1, \ldots, x_n)$ thì mức ý nghĩa tại đó chúng ta sẽ bác bỏ H_0 sẽ là

$$p = P_{\theta_0}\left(T(X) \geq T(x)\right), \tag{11.7}$$

đó là giá trị p nếu chúng ta quan sát x. Do đó, giá trị p có thể được hiểu là xác suất quan sát một kết quả cực đoan hơn những gì chúng ta quan sát được trong dữ liệu nếu giả thuyết không giữ.

Giả thuyết có dạng $\theta = \theta_0$ được gọi là giả thuyết đơn giản. Nếu một giả thuyết có dạng $\theta \quad S$ với tập S nào đó thì giả thuyết là hợp số. Đối với một giả thuyết null tổng hợp $\theta \quad H_0$ chúng tôi xác định lại kích thước và giá trị p theo cách sau,

$$\alpha = \text{hỗ trợ}_{\theta \;\; H_0} P_\theta\left(T(X) \geq \eta\right), \tag{11.8}$$

$$p = \text{hỗ trợ}_{\theta \;\; H_0} P_\theta\left(T(X) \geq T(x)\right). \tag{11.9}$$

Để mô tả sức mạnh của thử nghiệm cho một mức ý nghĩa nhất định, chúng tôi tính toán chức năng nguồn.

Định nghĩa 11.2.1 (Hàm công suất). Đặt P_θ là độ đo xác suất được tham số hóa bởi θ và đặt R là vùng bác bỏ cho phép kiểm tra dựa trên thống kê kiểm tra T(x). Hàm công suất của bài kiểm tra được định nghĩa là

$$\beta(\theta) := P_\theta\left(T(X) \in R\right) \tag{11.10}$$

Lý tưởng nhất là chúng ta muốn $\beta(\theta) \approx 0$ cho $\theta \in H_0$ và $\beta(\theta) \approx 1$ cho $\theta \in H_1$.

Ví dụ 11.2.2 (Tung đồng xu). Chúng tôi quan tâm đến việc kiểm tra xem một đồng xu có thiên về mặt ngửa hay không. Giả thuyết vô hiệu là với mỗi lần tung đồng xu, xác suất xuất hiện mặt ngửa là $\theta \leq 1/2$.
Do đó, giả thuyết thay thế là $\theta > 1/2$. Chúng ta hãy xem xét một thống kê kiểm tra bằng số lượng mặt ngửa được quan sát trong một chuỗi n lần lật,

$$T(x) = \sum_{i=1}^{n} 1_{x_i=1}, \tag{11.11}$$

trong đó x_i là một nếu lần tung đồng xu thứ i là mặt ngửa và bằng 0 nếu ngược lại. Vùng đào thải tự nhiên là

$$T(x) \geq \eta. \tag{11.12}$$

Đặc biệt, chúng tôi xem xét hai ngưỡng có thể

1. $\eta = n$, tức là chúng ta chỉ bác bỏ giả thuyết vô hiệu nếu tất cả các lần tung đồng xu đều là mặt ngửa,
2. $\eta = 3n/5$, tức là chúng ta bác bỏ giả thuyết không nếu ít nhất ba phần năm số lần tung đồng xu là cái đầu.

Chúng ta nên sử dụng phép thử nào nếu số lần tung đồng xu là 5, 50 hoặc 100? Các kiểm định có mức ý nghĩa 5% không? Sức mạnh của các bài kiểm tra cho các giá trị này của n là gì?

Để trả lời những câu hỏi này, chúng ta tính hàm lũy thừa của phép thử cho cả hai phương án. Nếu $\eta = n$,

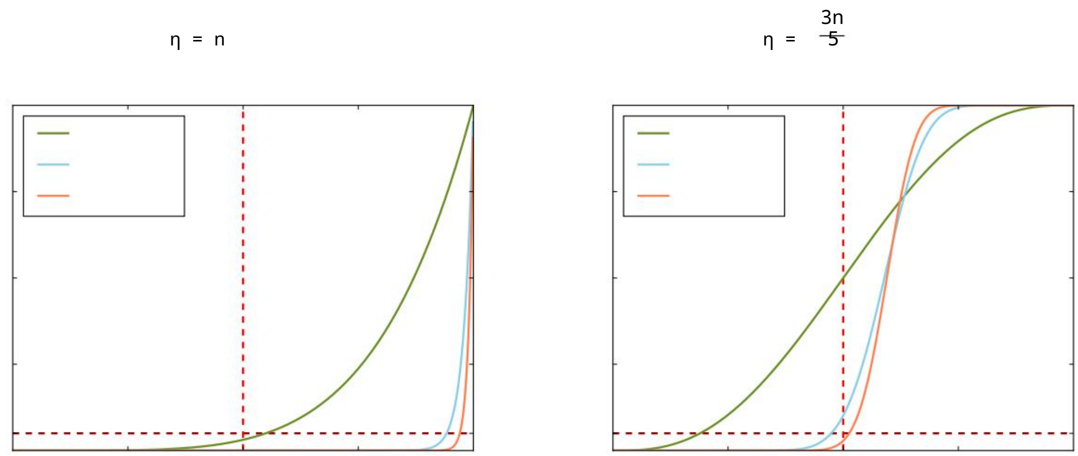
$$\beta_1(\theta) = P_\theta\left(T(X) \in R\right) \tag{13.11}$$

$$= \theta^n. \tag{14.11}$$

Nếu $\eta = 3n/5$,

$$\beta_2(\theta) = \sum_{k=3n/5}^{n} \binom{n}{k} \theta^k (1-\theta)^{n-k}. \tag{15.11}$$

Hình 11.1 cho thấy hai hàm công suất. Nếu $\eta = n$ thì kiểm định có mức ý nghĩa 5% đối với ba giá trị của n. Tuy nhiên, công suất rất thấp, đặc biệt đối với n lớn. Điều này có ý nghĩa: ngay cả khi đồng xu khá thiên vị thì xác suất xuất hiện n mặt ngửa là cực kỳ thấp. Nếu $\eta = 3n/5$, thì với n = 5, phép thử có mức ý nghĩa trên 5%, vì ngay cả khi đồng xu không bị sai lệch thì xác suất quan sát được 3 mặt ngửa trong số 5 lần tung là khá cao. Tuy nhiên, đối với n lớn, thử nghiệm có công suất cao hơn nhiều so với tùy chọn đầu tiên. Nếu độ lệch của đồng xu trên 0,7, chúng tôi sẽ từ chối giả thuyết không với xác suất cao.



Hình 11.1: Các chức năng nguồn cho các thử nghiệm được mô tả trong Ví dụ 11.2.2.

Một phương pháp có hệ thống để xây dựng các bài kiểm tra theo các giả định tham số là ngưỡng tỷ lệ giữa khả năng xảy ra của dữ liệu theo giả thuyết không và khả năng xảy ra của dữ liệu theo giả thuyết không. giả thuyết thay thế. Nếu tỷ lệ này cao, dữ liệu tương thích với giả thuyết không, vì vậy nó không nên bị từ chối.

Định nghĩa 11.2.3 (Kiểm tra tỷ lệ khả năng xảy ra). Đặt $L_x(\theta)$ biểu thị hàm khả năng tương ứng với một vectơ dữ liệu x. H_0 và H_1 lần lượt là các tập tương ứng với các giả thuyết hy không và thay thế. Tỷ lệ xác suất là

$$\Lambda(x) := \frac{\sup_{\theta \in H_0} L_x(\theta)}{\sup_{\theta \in H_1} L_x(\theta)}. \tag{16.11}$$

Phép thử tỷ lệ khả năng có vùng bác bỏ có dạng $\{\Lambda(x) \leq \eta\}$, đối với ngưỡng không đổi η.

Ví dụ 11.2.4 (Gaussian với phương sai đã biết). Hãy tưởng tượng rằng bạn có một số dữ liệu được mô hình hóa tốt như iid Gaussian với phương sai σ đã biết. Ý nghĩa là không rõ và chúng tôi là quan tâm đến việc thiết lập rằng nó không bằng một giá trị nhất định μ_0. tương ứng là gì kiểm tra tỷ lệ khả năng và nên đặt ngưỡng như thế nào để chúng ta có mức ý nghĩa α?

Đầu tiên, từ Ví dụ 9.6.4, giá trị trung bình của mẫu đạt được giá trị lớn nhất của hàm khả năng của một Gaussian

$$av(x) := \arg \text{tối đa}_{\mu} L_x(\mu, \sigma) \tag{17.11}$$

với mọi giá trị của σ. Sử dụng kết quả này, chúng ta có

$$\Lambda(x) = \frac{\sup_{\mu \in H_0} L_x(\mu)}{\sup_{\mu \in H_1} L_x(\mu)} \tag{18.11}$$

$$= \frac{L_x(\mu_0)}{L_x(av(x))}. \tag{19.11}$$

Cắm các biểu thức cho khả năng chúng ta có được,

$$\Lambda(x) = \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - av(x))^2 - (x_i - \mu_0)^2\right) \quad (11.20)$$

$$= \text{kinh nghiệm}\left(-\frac{1}{2\sigma^2}\left(-2\,av(x)\sum_{i=1}^{n}x_i + n\,av(x)^2 - 2\mu_0\sum_{i=1}^{n}x_i + n\mu_0^2\right)\right) \quad (21.11)$$

$$= \text{điểm kinh nghiệm}\left(-\frac{n(av(x) - \mu_0)^2}{2\sigma^2}\right). \quad (22.11)$$

Lấy logarit, bài kiểm tra có dạng

$$\frac{|av(x) - \mu_0|}{\sigma} \geq \sqrt{\frac{2\log\eta}{N}}. \quad (23.11)$$

Giá trị trung bình mẫu của n biến ngẫu nhiên Gaussian độc lập với giá trị trung bình μ_0 và phương sai σ^2 là Gaussian với giá trị trung bình μ_0 và phương sai σ^2/n, ngụ ý

$$\alpha = P_{\mu_0}\left(\left|\frac{av(X) - \mu_0}{\sigma/\sqrt{n}}\right| \geq \sqrt{2\log\eta}\right) \quad (24.11)$$

$$= 2Q\left(\sqrt{2\log\eta}\right). \quad (25.11)$$

Nếu chúng tôi cố định kích thước mong muốn α thì thử nghiệm sẽ trở thành

$$\frac{|av(x) - \mu_0|}{\sigma} \geq \frac{Q^{-1}(\alpha/2)}{\sqrt{n}}. \quad (26.11)$$

Một lập luận thúc đẩy để sử dụng kiểm tra tỷ lệ khả năng là nếu các giả thuyết không và giả thuyết thay thế là đơn giản, thì nó là tối ưu về mặt sức mạnh.

Bổ đề 11.2.5 (Bổ đề Neyman-Pearson). Nếu cả giả thuyết không và giả thuyết thay thế đều đơn giản, tức là tham số θ chỉ có thể có hai giá trị θ0 và θ1, thì phép thử tỷ lệ hợp lý có công suất cao nhất trong số tất cả các phép thử có kích thước cố định.

Bằng chứng. Nhớ lại rằng sức mạnh là xác suất bác bỏ giả thuyết không nếu nó không đúng. Nếu chúng ta biểu thị vùng từ chối của phép thử tỷ lệ khả năng bằng RLR thì sức mạnh của nó là

$$P_{\theta_1}\left(X \in R_{LR}\right). \quad (27.11)$$

Giả sử rằng chúng ta có một phép thử khác với vùng bác bỏ R. Công suất của nó bằng

$$P_{\theta_1}\left(X \in R\right). \quad (28.11)$$

Để chứng minh rằng sức mạnh của kiểm tra tỷ lệ khả năng lớn hơn, chúng ta chỉ cần thiết lập rằng

$$P_{\theta_1}\left(X \in R^c \cap R_{LR}\right) \geq P_{\theta_1}\left(X \in R^c_{LR} \cap R\right). \quad (29.11)$$

Machine Translated by Google

Chúng ta hãy giả sử rằng dữ liệu là các biến ngẫu nhiên liên tục (đối số cho các biến ngẫu nhiên rời rạc trên thực tế là giống nhau) và pdf khi giả thuyết không và giả thuyết thay thế được giữ lần lượt là $f_{\theta 0}$ và $f_{\theta 1}$. Theo định nghĩa vùng bác bỏ của phép thử tỷ lệ hợp lý, nếu Λ (x) RLR

$$f_{\theta 1}(x) \geq \frac{f_{\theta 0}(x)}{\eta}, \tag{11.30}$$

ngược lại nếu Λ(x) R^c_{LR}

$$f_{\theta 1}(x) \leq \frac{f_{\theta 0}(x)}{\eta}. \tag{31.11}$$

Nếu cả hai phép thử đều có kích thước α thì

$$P_{\theta 0}\ X\ \ R = \alpha = P_{\theta 0}\ X\ \ R_{LR}\ . \tag{32.11}$$

và do đó

$$P_{\theta 0}\ X\ \ R^c \cap R_{LR} = P_{\theta 0}\ X\ \ R_{LR}\ \ P_{\theta 0}\ X\ \ R \cap R_{LR} \tag{33.11}$$

$$= P_{\theta 0}\ X\ \ R\ \ P_{\theta 0}\ X\ \ R \cap R_{LR} \tag{11.34}$$

$$= P_{\theta 0}\ X\ \ R \cap R^c_{LR}\ . \tag{11.35}$$

Bây giờ ta chứng minh rằng (11.29) đúng,

$$P_{\theta 1}\ X\ \ R^c \cap R_{LR} = \int_{x\ \ R^c \cap R_{LR}} f_{\theta 1}(x)\, dx \tag{36.11}$$

$$\frac{1}{\eta} f_{\theta 0}(x)dx \text{ theo (11.30) } \geq \eta\ x\ \ R^c \cap R_{LR}\ 1 \tag{37.11}$$

$$= \frac{1}{\eta} P_{\theta 0}\ X\ \ R^c \cap R_{LR} \tag{11.38}$$

$$= \frac{1}{\eta} P_{\theta 0}\ X\ \ R \cap R^c_{LR} \quad \text{bởi (11.35)} \tag{39.11}$$

$$= \frac{1}{\eta} \int_{x\ \ R \cap R^c_{LR}} f_{\theta 0}(x)\, dx \tag{11.40}$$

$$\geq \int_{x\ \ R \cap R^c_{LR}} f_{\theta 1}(x)\, dx \text{ theo (11.31)} \tag{11.41}$$

$$= P_{\theta 1}\ X\ \ R \cap R^c_{LR}\ . \tag{11.42}$$

□

11.3 Kiểm định phi tham số: Kiểm định hoán vị

Trong các tình huống thực tế, chúng tôi có thể không thiết kế được một mô hình tham số phù hợp với dữ liệu của mình. Kiểm định phi tham số là kiểm định giả thuyết không giả định rằng dữ liệu tuân theo bất kỳ phân phối với một hình thức được xác định trước. Trong phần này, chúng tôi mô tả phép thử hoán vị, một phép thử phi tham số có thể được sử dụng để so sánh hai tập dữ liệu xA và xB nhằm đánh giá các phỏng đoán có dạng xA được lấy mẫu từ phân phối có giá trị trung bình cao hơn xB hoặc xB được lấy mẫu từ một phân phối có phương sai cao hơn xA. Giả thuyết không là hai bộ dữ liệu thực sự được lấy mẫu từ cùng một phân phối.

Thống kê kiểm tra trong kiểm tra hoán vị là sự khác biệt giữa các giá trị của thống kê kiểm tra quan tâm t được đánh giá trên hai bộ dữ liệu

$$t_{diff}(x) := t(x_A) \quad t(x_B), \tag{11.43}$$

trong đó x là tất cả dữ liệu được hợp nhất với nhau. Mục tiêu của chúng tôi là kiểm tra xem t(xA) có lớn hơn t(xB) ở một mức ý nghĩa nhất định hay không. Vùng bác bỏ tương ứng có dạng R := {t | t ≥ η}. Vấn đề là cố định ngưỡng như thế nào để kiểm định có mức ý nghĩa mong muốn.

Hãy tưởng tượng rằng chúng ta hoán vị ngẫu nhiên các nhãn A và B trong tập dữ liệu đã hợp nhất x. Do đó, một số dữ liệu được gắn nhãn là A sẽ được gắn nhãn là B và ngược lại. Nếu chúng ta tính toán lại tdiff (x) thì rõ ràng chúng ta sẽ thu được một giá trị khác. Tuy nhiên, phân phối của biến ngẫu nhiên tdiff(X) theo giả thuyết dữ liệu được lấy mẫu từ cùng một phân phối không thay đổi. Thật vậy, giả thuyết không ngụ ý rằng phân phối của bất kỳ hàm nào của XXX chỉ phụ thuộc vào lớp được gán cho mỗi biến là bất biến đối với permu 1, 2, . . . , ga. Chính thức hơn, chuỗi ngẫu nhiên có thể hoán đổi đối với các chức năng như vậy.

Xem xét giá trị của tdiff cho tất cả các hoán vị có thể có của các nhãn: tdiff,1, tdiff,2, . . . tdiff,n! . Nếu giả thuyết không đúng, thì sẽ rất ngạc nhiên khi thấy rằng tdiff (x) lớn hơn hầu hết các tdiff,i. Thực tế, theo giả thuyết không, biến ngẫu nhiên tdiff(X) được phân phối đều trong tập {tdiff,1, tdiff,2, . . . tdiff,n!}, sao cho

$$P\left(t_{diff}(X) \geq \eta\right) = n! \frac{1}{} \sum_{tôi=1}^{N!} 1_{t_{diff,i} \geq \eta}. \tag{11.44}$$

Đây chính xác là kích thước của bài kiểm tra. Do đó, chúng ta có thể tính giá trị p của thống kê quan sát được tdiff (x) là

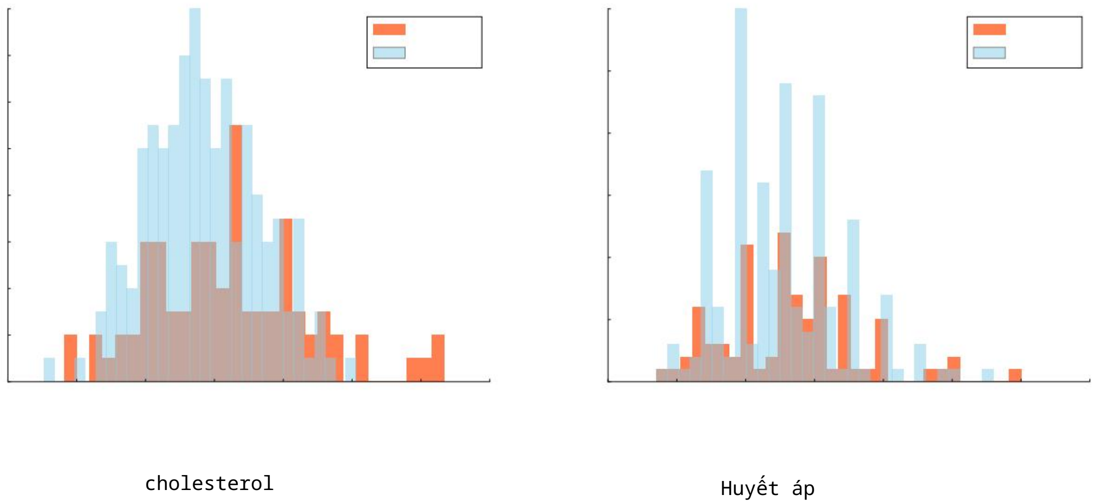
$$p = P\left(t_{diff}(X) \geq t_{diff}(x)\right) \tag{11.45}$$

$$= \frac{1}{N!} \sum_{tôi=1}^{N!} 1_{t_{diff,i} \geq t_{diff}(x)}. \tag{11.46}$$

Nói một cách dễ hiểu, giá trị p là một phần của các hoán vị mang lại một thống kê thử nghiệm cực đoan hơn so với thống kê mà chúng ta quan sát được. Thật không may, thường rất khó tính toán (11,46) một cách chính xác. Ngay cả đối với các tập dữ liệu có kích thước vừa phải, số lượng hoán vị có thể có thường là quá lớn (ví dụ: 40! > 8 1047) để có thể tính toán được. Trong những trường hợp như vậy, giá trị p có thể được xấp xỉ bằng cách lấy mẫu một số lượng lớn các hoán vị và thực hiện phép tính xấp xỉ Monte Carlo của (11,46) với giá trị trung bình của nó.

Trước khi xem xét một ví dụ, chúng ta hãy xem lại các bước cần tuân theo khi áp dụng phép thử hoán vị.

1. Chọn một phỏng đoán về xA và xB khác nhau như thế nào.



Hình 11.2: Biểu đồ cholesterol và huyết áp của nam và nữ trong Ví dụ 11.3.1.

2. Chọn một thống kê kiểm định tdiff.

3. Tính tdiff (x).

4. Hoán vị các nhãn m lần và tính các giá trị tương ứng của tdiff: tdiff,1, tdiff,2, . . .tdiff,m.

5. Tính giá trị p gần đúng

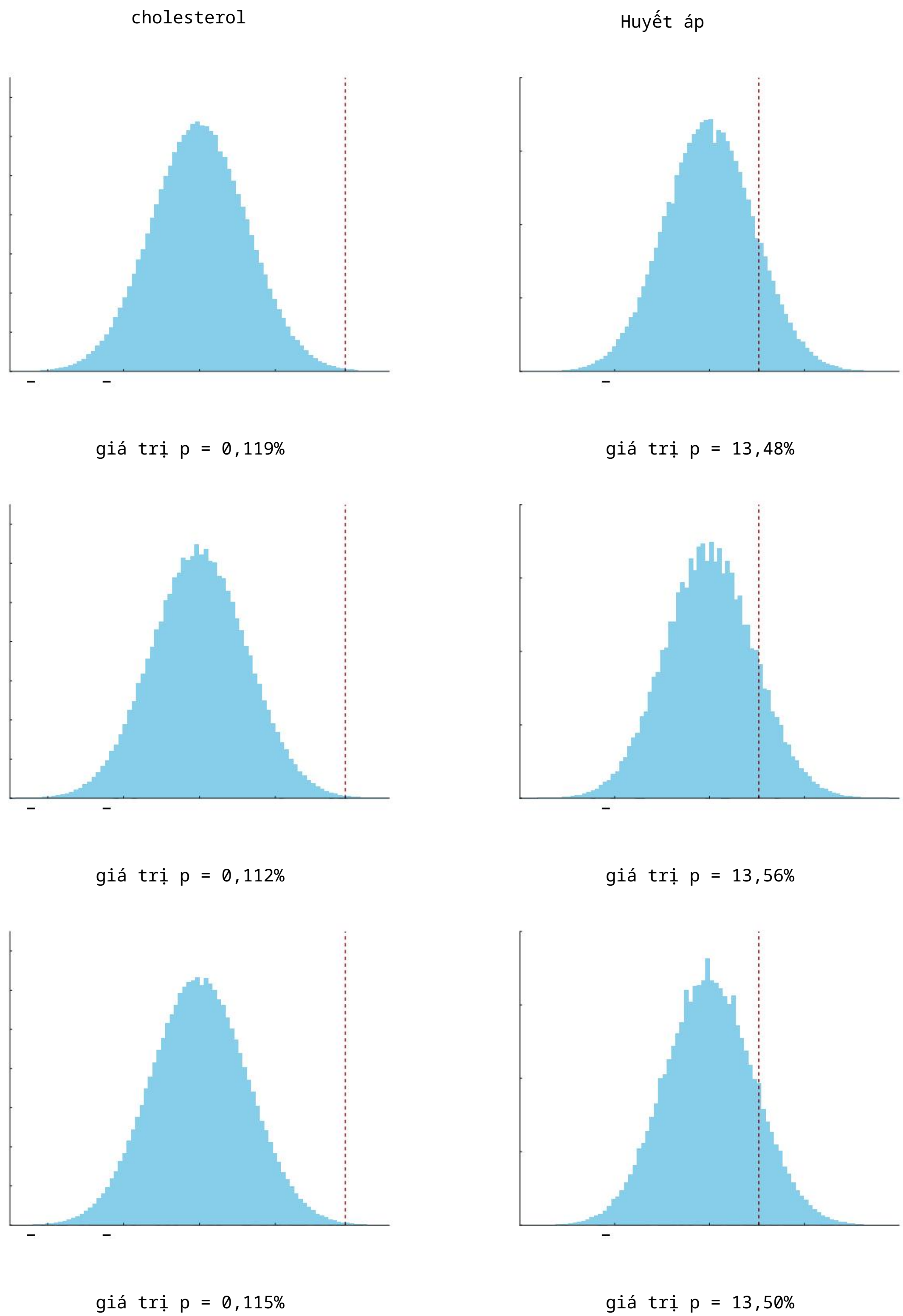
$$p = P\ \text{tdiff}(X\ ) \geq \text{tdiff}\ (x) \tag{11.47}$$

$$= \frac{1}{\text{tôi}} \sum_{\text{tôi}=1}^{\text{tôi}} 1_{\text{tdiff},i \geq \text{tdiff}(x)} \tag{11.48}$$

và bác bỏ giả thuyết không nếu nó nằm dưới giới hạn được xác định trước (thường là 1% hoặc 5%).

Ví dụ 11.3.1 (Cholesterol và huyết áp). Một nhà khoa học muốn xác định xem nam giới có cholesterol và huyết áp cao hơn hay không. Cô thu thập dữ liệu từ 86 nam giới và 182 phụ nữ. Hình 11.2 cho thấy biểu đồ cholesterol và huyết áp của nam giới và nữ giới. Từ các biểu đồ, có vẻ như nam giới có mức cholesterol và huyết áp cao hơn. Giá trị trung bình của mẫu đối với cholesterol là 261,3 mg/dl ở nam giới và 242,0 mg/dl ở nữ giới. Giá trị trung bình của mẫu đối với huyết áp là 133,2 mmHg ở nam giới và 130,6 mmHg ở nữ giới.
phụ nữ.

Để định lượng xem những khác biệt này có đáng kể hay không, chúng tôi tính toán phân phối hoán vị mẫu của sự khác biệt giữa các phương tiện mẫu bằng cách sử dụng 106 hoán vị. Để đảm bảo rằng kết quả ổn định, chúng tôi lặp lại quy trình ba lần. Kết quả được thể hiện trong hình 11.3. Đối với cholesterol, giá trị p là khoảng 0,1%, vì vậy chúng tôi có bằng chứng rất mạnh mẽ chống lại giả thuyết không. Ngược lại, giá trị p của huyết áp là 13% nên kết quả không thuyết phục lắm, chúng ta không thể bác bỏ khả năng sự khác biệt chỉ là do dao động ngẫu nhiên.



Hình 11.3: Phân phối gần đúng theo giả thuyết không về sự khác biệt giữa các mẫu phương tiện cholesterol và huyết áp ở nam giới và phụ nữ. Giá trị quan sát được của thống kê kiểm định là được đánh dấu bằng một đường đứt nét.

11.4 Thử nghiệm nhiều lần

Trong một số ứng dụng, người ta thường tiến hành nhiều kiểm định giả thuyết đồng thời. Ví dụ, trong bộ gen tính toán, một nhà nghiên cứu có thể quan tâm đến việc kiểm tra xem có bất kỳ gen nào trong một nhóm vài nghìn gen có liên quan đến một bệnh nhất định hay không. Nếu chúng ta áp dụng thử nghiệm giả thuyết với kích thước α trong cài đặt này, thì xác suất nhận được kết quả dương tính giả đối với một gen cụ thể là α. Bây giờ, giả sử rằng chúng ta kiểm tra n gen và gen sự kiện i là dương tính giả, $1 \leq i \leq n$ đều độc lập lẫn nhau. Xác suất để có ít nhất một lần dương tính giả là

$$P\,(\text{ít nhất một dương tính giả}) = 1 - P\,(\text{không có dương tính giả}) \tag{11.49}$$

$$= 1 - (1 - \alpha)^N. \tag{11.50}$$

Với $\alpha = 0{,}01$ và $n = 500$, xác suất này bằng 0,99! Nếu chúng tôi muốn kiểm soát xác suất mắc lỗi Loại I, chúng tôi phải tính đến việc chúng tôi đang thực hiện nhiều thử nghiệm cùng một lúc. Một quy trình phổ biến để làm điều này là phương pháp Bonferroni.

Định nghĩa 11.4.1 (Phương pháp Bonferroni). Cho n phép thử giả thuyết, hãy tính các giá trị p tương ứng $p_1, \ldots, p_n$. Đối với mức ý nghĩa cố định, α bác bỏ giả thuyết không thứ i nếu

$$p_i \leq \frac{\alpha}{N}. \tag{11.51}$$

Bổ đề sau đây cho thấy rằng phương pháp này đảm bảo rằng mức ý nghĩa mong muốn được giữ đồng thời cho tất cả các kiểm định.

Bổ đề 11.4.2. Nếu chúng ta áp dụng phương pháp của Bonferroni, xác suất mắc lỗi Loại I bị giới hạn bởi α.

Bằng chứng. Kết quả theo sau trực tiếp từ ràng buộc hợp, kiểm soát xác suất hợp của các sự kiện với tổng các xác suất riêng lẻ của chúng.

Định lý 11.4.3 (Liên kết hợp). Cho $(\Omega, \mathcal{F}, P)$ là một không gian xác suất và $S_1, S_2, \ldots$ một tập hợp các sự kiện trong $\mathcal{F}$. Sau đó

$$P\left(\cup_i S_i\right) \leq \sum_i P\,(S_i). \tag{11.52}$$

Bằng chứng. Hãy xác định các tập hợp:

$$\tilde{S}_i = S_i \cap \cap_{j=1}^{i-1} S_j^c. \tag{11.53}$$

$\cup_i \tilde{S}_i$. Dễ dàng chứng minh bằng quy nạp rằng $\cup_{j=1}^n S_j =$ đặt $\cup_{j=1}^n \tilde{S}_j$ với n bất kỳ, vì vậy $\cup_i S_i =$ $\tilde{S}_1, \tilde{S}_2, \ldots$ được rời rạc bởi xây dựng, vì vậy

$$P\left(\cup_i S_i\right) = P\left(\cup_i \tilde{S}_i\right) = \sum_i P\,\tilde{S}_i \quad \text{theo Tiên đề 2 trong Định nghĩa 1.1.4} \tag{11.54}$$

$$\leq \sum_i P\,(S_i) \quad \text{vì } \tilde{S}_i \subseteq S_i. \tag{11.55}$$

□

Áp dụng giới hạn,

$$P\,(\text{Lỗi loại I}) = P\,(\cup_{i=1}^{n} \text{Lỗi loại I đối với bài kiểm tra i}) \tag{11.56}$$

$$\leq \sum_{\text{tôi}=1}^{N} P\,(\text{Lỗi loại I đối với bài kiểm tra i}) \quad \text{bởi liên minh ràng buộc} \tag{11.57}$$

$$= n \cdot \frac{\alpha}{N} = \alpha. \tag{11.58}$$

□

Ví dụ 11.4.4 (Ly hợp (tiếp theo)). Nếu chúng ta áp dụng phép thử trong Ví dụ 11.1.4 cho 10 người chơi, xác suất mà một trong số họ dường như ăn khớp chỉ do ngẫu nhiên sẽ tăng lên đáng kể. Để kiểm soát điều này, theo phương pháp của Bonferroni, chúng tôi phải chia giá trị p của các bài kiểm tra riêng lẻ cho 10. Do đó, để duy trì mức ý nghĩa 0,05, chúng tôi sẽ yêu cầu mỗi người chơi ghi được nhiều điểm hơn mỗi phút trong quý cuối cùng của 17 năm 20 trò chơi thay vì 15 (xem Bảng 11.2) để bác bỏ giả thuyết không.

Chương 12

hồi quy tuyến tính

Trong thống kê, hồi quy là bài toán mô tả mối quan hệ giữa một đại lượng quan tâm y nhất định, được gọi là phản ứng hoặc biến phụ thuộc, với một số biến quan sát $x_1, x_2, \ldots, x_p$, được gọi là đồng biến, đặc trưng hoặc biến độc lập. Ví dụ: câu trả lời có thể là giá của một ngôi nhà và các đồng biến có thể tương ứng với phần mở rộng, số phòng, năm xây dựng, v.v. Một mô hình hồi quy sẽ mô tả giá nhà bị ảnh hưởng như thế nào bởi tất cả các yếu tố này.

Chính thức hơn, giả định chính trong các mô hình hồi quy là bộ dự đoán được tạo theo một hàm h được áp dụng cho các tính năng và sau đó bị nhiễu bởi một số nhiễu z không xác định, thường là phép cộng,

$$y = h(x) + z. \tag{12.1}$$

Mục đích là học h từ n ví dụ về phản hồi và các đặc điểm tương ứng của chúng

$$\left(y^{(1)}, x^{(1)}\right), \left(y^{(2)}, x^{(2)}\right), \ldots, \left(y^{(n)}, x^{(n)}\right). \tag{12.2}$$

Trong chương này chúng ta tập trung vào trường hợp h là một hàm tuyến tính.

12.1 Mô hình tuyến tính

Nếu hàm hồi quy h trong mô hình có dạng 12.1 là tuyến tính, thì phản hồi được mô hình hóa dưới dạng tổ hợp tuyến tính của các yếu tố dự báo:

$$y^{(i)} = x^{(i)\,T} \beta + z^{(i)}, \quad 1 \le i \le n, \tag{12.3}$$

ở đâu $z^{(i)}$ là một mục của vectơ nhiễu chưa biết. Hàm được tham số hóa bởi một vectơ trọng số $\beta \in \mathbb{R}^p$. Tất cả những gì chúng ta cần để mô hình tuyến tính phù hợp với dữ liệu là ước tính các trọng số này.

Biểu diễn hệ tuyến tính (12.3) ở dạng ma trận mang lại biểu diễn sau của mô hình hồi quy tuyến tính

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \cdots \\ y^{(N)} \end{bmatrix} = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_p^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_p^{(2)} \\ \cdots & \cdots & \cdots & \cdots \\ x_1^{(n)} & x_2^{(n)} & \cdots & x_p^{(n)} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} z^{(1)} \\ z^{(2)} \\ \cdots \\ z^{(N)} \end{bmatrix}. \tag{12.4}$$

tương đương,

$$y = X\beta + z, \quad (12,5)$$

trong đó X là ma trận ×p chứa các đặc trưng, y chứa phản hồi và z R ^{N} đa i diê n tiếng ồn.

Ví dụ 12.1.1 (Mô hình tuyến tính cho GDP). Ta xét bài toán xây dựng mô hình tuyến tính để dự đoán tổng sản phẩm quốc nội (GDP) của một tiểu bang ở Hoa Kỳ từ dân số và tỷ lệ thất nghiệp. Chúng tôi có sẵn các dữ liệu sau:

	GDP (triệu USD)	dân số	thất nghiệp tỷ lệ (%)
Bắc Dakota	52 089	757 952	2.4
Alabama	204 861	4 863 300	3,8
Mississippi	107 680	2 988 726	5.2
Arkansas	120 689	2 988 248	3,5
Kansas	153 258	2 907 289	3,8
Gruzia	525 360	10 310 371	4,5
Iowa	178 766	3 134 693	3.2
phia Tây Virginia	73 374	1 831 102	5.1
Kentucky	197 043	4 436 974	5.2
Tennessee	???	6 651 194	3.0

Trong ví dụ này, GDP là câu trả lời, còn dân số và tỷ lệ thất nghiệp là các tính năng. Mục tiêu của chúng tôi là khớp một mô hình tuyến tính với dữ liệu để chúng tôi có thể dự đoán GDP của Tennessee, sử dụng mô hình tuyến tính. Chúng tôi bắt đầu bằng cách định tâm và chuẩn hóa dữ liệu. trung bình của phản hồi và của các tính năng là

$$av(y) = 179\ 236, \qquad av(X) = 3\ 802\ 073\ 4.1\ . \quad (12.6)$$

Độ lệch chuẩn thực nghiệm là

$$\text{tiêu chuẩn}(y) = 396\ 701, \qquad \text{tiêu chuẩn}(X) = 7\ 720\ 656\ 2,80\ . \quad (12.7)$$

Chúng tôi trừ trung bình và chia cho độ lệch chuẩn để cả phản hồi và các tính năng được tập trung và trên cùng một tỷ lệ,

$$y = \begin{bmatrix} 0,321 \\ 0,065 \\ 0,180 \\ 0,148 \\ 0,065 \\ 0,872 \\ 0,001 \\ 0,267 \\ 0,045 \end{bmatrix}, \quad X = \begin{bmatrix} 0,394 & 0,600 \\ 0,137 & 0,099 \\ 0,105 & 0,401 \\ 0,105 & 0,207 \\ 0,116 & 0,099 \\ 0,843 & 0,151 \\ 0,086 & 0,314 \\ 0,255 & 0,366 \\ 0,082 & 0,401 \end{bmatrix}. \tag{12.8}$$

Để có được ước tính cho GDP của Tennessee, chúng tôi điều chỉnh mô hình

$$y \approx X\beta, \tag{12.9}$$

thay đổi tỷ lệ theo độ lệch chuẩn (12,7) và gần đây hơn bằng cách sử dụng mức trung bình (12,6). ước tính cuối cùng là

$$y_{Ten} = av(y) + std(y)\, x_{Ten\ định\ mức}, \beta \tag{12.10}$$

xTen ở đâu $_{định\ mức}$ được căn giữa bằng cách sử dụng av(X) và được chuẩn hóa bằng cách sử dụng std(X).

12.2 Ước lượng bình phương nhỏ nhất

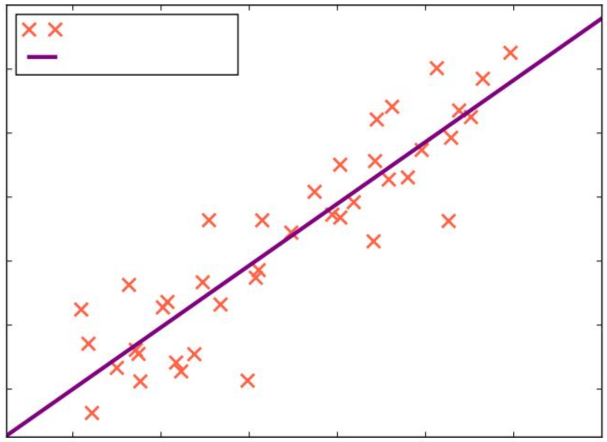
Để hiệu chỉnh mô hình hồi quy tuyến tính, chúng ta cần ước tính vectơ trọng số sao cho nó phù hợp với dữ liệu. Chúng ta có thể đánh giá mức độ phù hợp cho một lựa chọn cụ thể của $\beta \in \mathbb{R}^p$ bằng cách sử dụng tổng bình phương của sai số,

$$\sum_{tôi=1}^{N} \left(y^{(i)} - x^{(i)\,T}\beta \right)^2 = \left\| y - X\beta \right\|_2^2. \tag{12.11}$$

Ước lượng bình phương nhỏ nhất β_{LS} là véc tơ trọng số làm cực tiểu hóa hàm chi phí này,

$$\beta_{LS} := \arg\min_{\beta} \left\| y - X\beta \right\|_2. \tag{12.12}$$

Hàm chi phí bình phương nhỏ nhất thuận tiện từ góc nhìn tính toán, vì nó lồi và có thể được cực tiểu hóa một cách hiệu quả (thực tế, như chúng ta sẽ thấy trong giây lát, nó có nghiệm dạng đóng). Ngoài ra, nó có các diễn giải xác suất và hình học trực quan. Hình 12.1 cho thấy mô hình tuyến tính đã học bằng cách sử dụng bình phương tối thiểu trong một ví dụ đơn giản chỉ có một đối tượng địa lý (p = 1) và 40 ví dụ (n = 40).



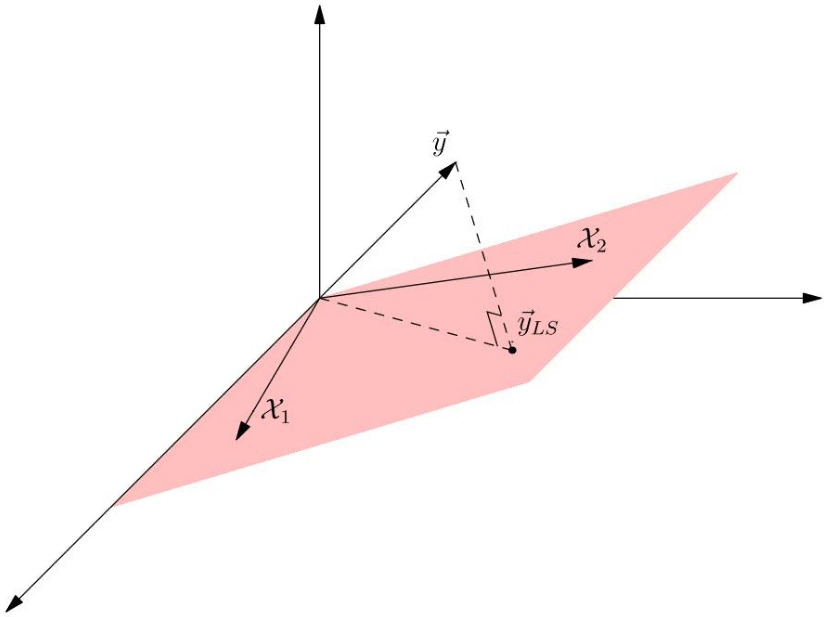
Hình 12.1: Mô hình tuyến tính được học thông qua khớp bình phương nhỏ nhất cho một ví dụ đơn giản khi chỉ có một tính năng (p = 1) và 40 ví dụ (n = 40).

Ví dụ 12.2.1 (Mô hình tuyến tính cho GDP (tiếp theo)). Ước lượng bình phương nhỏ nhất cho các hệ số hồi quy trong mô hình GDP tuyến tính bằng

$$\beta_{LS} = \begin{matrix} 1.019 \\ 0,111 \end{matrix} . \quad (13.12)$$

GDP dường như tỷ lệ thuận với dân số và tỷ lệ nghịch với tỷ lệ thất nghiệp. Bây giờ chúng ta so sánh sự phù hợp được cung cấp bởi mô hình tuyến tính với dữ liệu gốc, như cũng như dự đoán về GDP của Tennessee:

	Ước tính GDP	
Bắc Dakota	52 089	46 241
Alabama	204 861	239 165
Mississippi	107 680	119 005
Arkansas	120 689	145 712
Kansas	153 258	136 756
Gruzia	525 360	513 343
Iowa	178 766	158 097
Tây Virginia	73 374	59 969
Kentucky	197 043	194 829
Tennessee	328 770	345 352



Hình 12.2: Minh họa Hệ quả 12.2.3. Giải pháp bình phương nhỏ nhất là phép chiếu dữ liệu lên không gian con được kéo dài bởi các cột của X, kí hiệu là X_1 và X_2.

12.2.1 Diễn giải hình học

Định lý sau, được chứng minh trong Mục 12.2.2, cho thấy bài toán bình phương nhỏ nhất có nghiệm dạng đóng.

Định lý 12.2.2 (Giải pháp bình phương nhỏ nhất). Với $p \geq n$, nếu X là hạng đầy đủ thì nghiệm của bài toán bình phương nhỏ nhất (12.12) là

$$\beta_{LS} := \left(X^T X\right)^{-1} X^T y. \tag{12.14}$$

Hệ quả của kết quả này cung cấp một diễn giải hình học cho ước lượng bình phương nhỏ nhất của y: nó thu được bằng cách chiếu phản hồi lên không gian cột của ma trận được tạo bởi các yếu tố dự đoán.

Hệ quả 12.2.3. Với $p \geq n$, nếu X là hạng đầy đủ thì $X\beta_{LS}$ là hình chiếu của y lên không gian cột của X.

Chúng tôi cung cấp một bằng chứng chính thức trong Mục 12.5.2 của phụ lục, nhưng kết quả rất trực quan. Bất kỳ vectơ nào có dạng X đều nằm trong khoảng của các cột của X. Theo định nghĩa, ước lượng bình phương nhỏ nhất là vectơ gần y nhất có thể được biểu diễn theo cách này, do đó, nó là hình chiếu của y lên không gian cột của X. Điều này được minh họa trong Hình 12.2.

12.2.2 Giải thích xác suất

Nếu chúng ta lập mô hình nhiễu trong (12.5) như là một nhận thức từ một vectơ ngẫu nhiên Z có các mục nhập là các biến ngẫu nhiên Gaussian độc lập với trung bình bằng 0 và một phương sai σ^2 nhất định thì chúng ta có thể giải thích ước tính bình phương nhỏ nhất là ước tính khả năng tối đa. Theo giả định đó, dữ liệu là sự hiện thực hóa của vectơ ngẫu nhiên

$$Y := X\beta + Z, \tag{15.12}$$

đó là một vectơ ngẫu nhiên Gaussian iid với trung bình $X\beta$ và ma trận hiệp phương sai $\sigma^2 I$. Khớp pdf của Y bằng

$$f_Y(a) = \prod_{tôi=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \text{điểm kinh nghiệm} \left(-\frac{1}{2\sigma^2}\left(a_i - X\beta\right)^2\right) \tag{12.16}$$

$$= \frac{1}{\sqrt{(2\pi)^n}\sigma^n} \text{điểm kinh nghiệm} \left(-\frac{1}{2\sigma^2}\left\|a - X\beta\right\|_2^2\right). \tag{12.17}$$

Khả năng xảy ra là hàm mật độ xác suất của Y được đánh giá tại dữ liệu y được quan sát và được hiểu là hàm của vectơ trọng số β,

$$L_y(\beta) = \frac{1}{\sqrt{(2\pi)^N}} \text{điểm kinh nghiệm} \left(-\frac{1}{2}\left\|y - X\beta\right\|_2^2\right). \tag{18.12}$$

Để tìm ước tính ML, chúng tôi tối đa hóa khả năng đăng nhập. Ta kết luận rằng nó được cho bởi nghiệm của bài toán bình phương nhỏ nhất, vì

$$\beta_{ML} = \arg \text{tối đa}_{\beta} L_y(\beta) \tag{19.12}$$

$$= \arg\max_{\beta} \log L_y(\beta) \tag{12.20}$$

$$= \arg\min_{\beta} \left\|y - X\beta\right\|_2^2 \tag{21.12}$$

$$= \beta_{LS}. \tag{12.22}$$

12.3 Trang bị thừa

Hãy tưởng tượng rằng một người bạn nói với bạn:

Tôi đã tìm ra một cách hay để dự đoán nhiệt độ ở New York: Đó chỉ là sự kết hợp tuyến tính của nhiệt độ ở mọi tiểu bang khác. Tôi điều chỉnh mô hình dựa trên dữ liệu từ tháng rưỡi trước và nó thật hoàn hảo!

Bạn của bạn không nói dối, nhưng vấn đề là cô ấy đang sử dụng một số điểm dữ liệu để khớp với mô hình tuyến tính gần bằng với số lượng tham số. Nếu $n \leq p$ chúng ta có thể tìm được β sao cho $y = X\beta$ chính xác, ngay cả khi y và X không liên quan gì đến nhau! Điều này được gọi là trang bị quá mức và thường xảy ra do sử dụng một mô hình quá linh hoạt đối với số lượng dữ liệu có sẵn.

Để đánh giá xem một mô hình có bị quá khớp hay không, chúng tôi tách dữ liệu thành tập huấn luyện và tập kiểm tra. Tập huấn luyện được sử dụng để điều chỉnh mô hình và tập kiểm tra được sử dụng để đánh giá lỗi. Một mô hình overfit tập huấn luyện sẽ có lỗi rất thấp khi được đánh giá trên các ví dụ huấn luyện, nhưng sẽ không tổng quát hóa tốt cho các ví dụ kiểm tra.

Hình 12.3 thể hiện kết quả đánh giá sai số huấn luyện và sai số kiểm tra của một mô hình tuyến tính với p = 50 tham số được lắp từ n ví dụ huấn luyện. Dữ liệu huấn luyện và kiểm tra được tạo bằng cách cố định một vectơ trọng số β và sau đó tính toán

$$\text{ytrain} = \text{Xtrain}\ \beta + \text{ztrain}, \tag{12.23}$$

$$\text{ytest} = \text{Xtest}\ \beta\ , \tag{12.24}$$

trong đó các mục của Xtrain, Xtest, ztrain và β được lấy mẫu ngẫu nhiên một cách độc lập từ phân phối Gauss với giá trị trung bình bằng 0 và phương sai đơn vị. Các lỗi đào tạo và kiểm tra được định nghĩa là

$$\text{hệ thống lỗi} = \frac{\text{Xtrain}\ \beta\ \text{LS} \quad \text{ytrain}\quad {}^{2}}{||\text{ytrain}||2}, \tag{12.25}$$

$$\text{kiểm tra lỗi} = \frac{\text{Xtest}\ \beta\ \text{LS} \quad \text{ytest}\quad {}^{2}}{||\text{ytest}||2}. \tag{26.12}$$

Lưu ý rằng ngay cả β thực cũng không đạt được sai số huấn luyện bằng 0 do có nhiễu, nhưng sai số kiểm tra thực sự bằng 0 nếu chúng ta có thể ước tính chính xác β .

Lỗi đào tạo của mô hình tuyến tính tăng theo n. Điều này có ý nghĩa vì mô hình phải phù hợp với nhiều dữ liệu hơn bằng cách sử dụng cùng một số tham số. Khi n gần bằng p := 50, mô hình được trang bị tốt hơn nhiều so với mô hình thực trong việc sao chép dữ liệu huấn luyện (lỗi của mô hình thực được hiển thị bằng màu xanh lá cây). Đây là một dấu hiệu của việc trang bị quá mức: mô hình đang thích nghi với tiếng ồn và không học được cấu trúc tuyến tính thực sự. Thật vậy, trong chế độ đó, lỗi kiểm tra là cực kỳ cao. Ở mức n lớn hơn, lỗi đào tạo tăng lên đến mức mà mô hình tuyến tính thực sự đạt được và lỗi kiểm tra giảm xuống, cho thấy rằng chúng ta đang học mô hình cơ bản.

12.4 Sự nóng lên toàn cầu

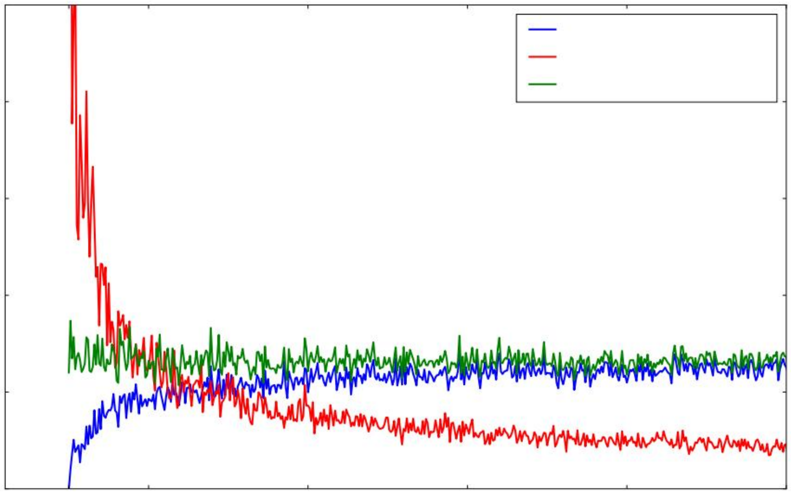
Trong phần này, chúng tôi mô tả một ứng dụng hồi quy tuyến tính cho dữ liệu khí hậu. Cụ thể, chúng tôi phân tích dữ liệu nhiệt độ được lấy tại một trạm thời tiết ở Oxford trong hơn 150 năm.[1] Mục tiêu của chúng tôi không phải là thực hiện dự đoán, mà là xác định xem nhiệt độ đã tăng hay giảm trong 150 năm qua ở Oxford.

Để tách nhiệt độ thành các thành phần khác nhau giải thích cho các hiệu ứng theo mùa, chúng tôi sử dụng một mô hình tuyến tính đơn giản với ba yếu tố dự báo và một giao điểm chặn.

$$\text{yt} \approx \beta\ 0 + \beta\ 1 \cos \frac{2\pi t}{12} + \beta\ 2\ \text{tội lỗi}\ \frac{2\pi t}{12} + \beta\ 3\ t \tag{27.12}$$

trong đó 1 ≤ t ≤ n biểu thị thời gian tính bằng tháng (n bằng 12 nhân 150). ma trận tương ứng

1Dữ liệu có tại http://www.metoffice.gov.uk/pub/data/weather/uk/climate/stationdata/oxforddata.txt.



Hình 12.3: Sai số chuẩn 2 tương đối trong ước tính phản hồi đạt được bằng cách sử dụng hồi quy bình phương nhỏ nhất cho các giá trị khác nhau của n (số lượng dữ liệu huấn luyện). Lỗi đào tạo được vẽ bằng màu xanh lam, trong khi lỗi kiểm tra được vẽ bằng màu đỏ. Đường màu xanh lá cây cho biết lỗi đào tạo của mô hình thực được sử dụng để tạo dữ liệu.

của dự đoán là

$$X := \begin{matrix} 1 & \cos \frac{2\pi t_1}{12} & \text{tội lỗi } \frac{2\pi t_1}{12} & t_1 \\ 1 & \cos \frac{2\pi t_2}{12} & \sin \frac{2\pi t_2}{12} & t_2 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & \cos \frac{2\pi t_n}{12} & \text{tội lỗi } \frac{2\pi t_n}{12} & t_n \end{matrix} . \qquad (28.12)$$

Hệ số chặn β 0 đại diện cho nhiệt độ trung bình, β 1 và β 2 tính đến các dao động định kỳ hàng năm và β 3 là xu hướng tổng thể. Nếu β 3 dương thì mô hình chỉ ra rằng nhiệt độ đang tăng, nếu nó âm thì nó chỉ ra rằng nhiệt độ đang giảm.

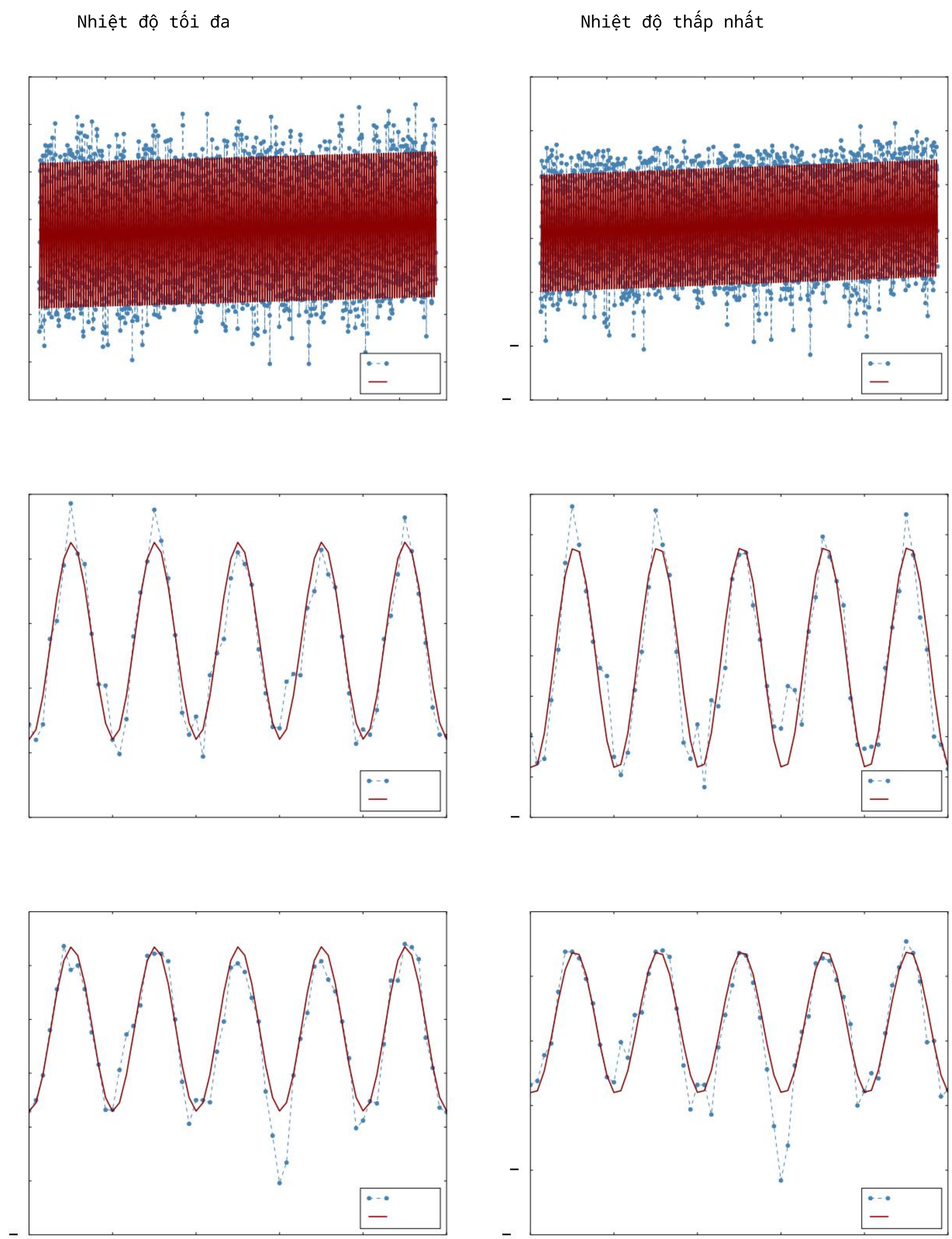
Kết quả điều chỉnh mô hình tuyến tính được thể hiện trên Hình 12.4 và 12.5. Mô hình phù hợp chỉ ra rằng cả nhiệt độ tối đa và tối thiểu đều có xu hướng tăng khoảng 0,8 độ C (khoảng 1,4 độ F).

12.5 Bằng chứng

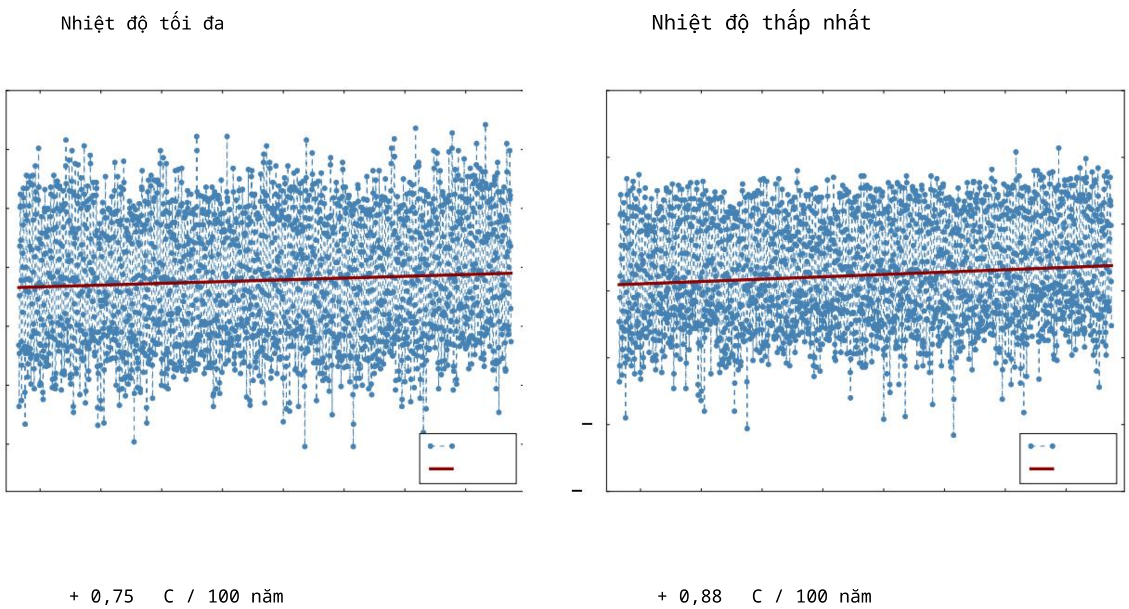
12.5.1 Chứng minh Mệnh đề 12.2.2

Đặt $X = U\Sigma V^T$ là phân tách giá trị kỳ dị (SVD) của X . Theo các điều kiện của định lý, $X^t X^{1} X^t y = V \Sigma U^t$. Chúng tôi bắt đầu bằng cách tách y thành hai thành phần

$$y = UU^T y + I \quad UU^T \quad y \qquad (29.12)$$



Hình 12.4: Dữ liệu nhiệt độ cùng với mô hình tuyến tính được mô tả bởi (12.27) cho cả cực đại và nhiệt độ tối thiểu.



Hình 12.5: Xu hướng nhiệt độ thu được bằng cách điều chỉnh mô hình được mô tả bởi (12.27) cho cả nhiệt độ tối đa và tối thiểu.

trong đó UUT y là hình chiếu của y lên không gian cột của X . Lưu ý rằng I UUT y trực giao với không gian cột của X và do đó trực giao với cả UUT y và X β đối với bất kỳ β nào. Theo định lý Pythagoras

$$y \quad X\,\beta \;{}^{2}_{2} = \text{Tôi} \quad UUT \quad y \;{}^{2}_{2} + UUT\ y \quad X\,\beta \;{}^{2}_{2}. \tag{12.30}$$

Giá trị tối thiểu của hàm chi phí này có thể đạt được bằng cách tối ưu hóa trên beta là $||yX \quad ||_2$. Điều này có thể đạt được bằng cách giải hệ phương trình

$$UUT\ y = X\,\beta = U\Sigma\ VT\ \beta. \tag{31.12}$$

Vì U TU = I vì p ≥ n, nhân cả hai vế của đẳng thức ta được hệ thức tương đương

$$\text{bạn}^{t}\ y = \Sigma\ VT\ \beta. \tag{12.32}$$

Vì X là hạng đầy đủ, Σ và V vuông góc và khả nghịch (và theo định nghĩa của SVD V $^{1} = V^{t}$), vì thế

$$\beta_{LS} = V\ \Sigma\ U\ \ T\ y \tag{12.33}$$

là nghiệm duy nhất của hệ và do đó cũng là nghiệm của bài toán bình phương bé nhất.

12.5.2 Chứng minh Hệ quả 12.2.3

Đặt $X = U\Sigma V^t$ là phân tách giá trị kỳ dị của X . Vì X là số hạng đầy đủ và $p \geq n$ ta có $U^TU = I$, $V^tV = I$ và Σ là ma trận vuông khả nghịch, ngụ ý

$$X\beta_{LS} = X\left(X^tX\right)^{-1}X^Ty \quad (12.34)$$

$$= U\Sigma V^t\left(V\Sigma U^TU\Sigma V^t\right)^{-1}V\Sigma U^Ty \quad (12.35)$$

$$= UU^Ty. \quad (12.36)$$

Phụ lục A

lý thuyết tập hợp

Chương này cung cấp một đánh giá về các khái niệm cơ bản trong lý thuyết tập hợp.

A.1 Định nghĩa cơ bản

Một tập hợp là một tập hợp các đối tượng. Tập hợp chứa mọi đối tượng có thể mà chúng ta xem xét trong một tình huống nhất định được gọi là vũ trụ và thường được ký hiệu là Ω. Nếu một đối tượng x trong Ω thuộc tập S, ta nói rằng x là phần tử của S và viết x S. Nếu x không phải là phần tử của S thì ta viết x / S. Tập rỗng, thường được ký hiệu là , là một tập hợp sao cho x / với mọi x Ω (tức là nó không có phần tử nào). Nếu tất cả các phần tử trong tập hợp B cũng thuộc tập hợp A thì B là tập hợp con của A, ký hiệu là B A. Ngoài ra, nếu có ít nhất một phần tử của A không thuộc B thì B là tập con thực sự của A, ký hiệu là B A.

Các phần tử của một tập hợp có thể là các đối tượng tùy ý và đặc biệt chúng có thể là chính các tập hợp đó. Đây là trường hợp cho tập lũy thừa của một tập hợp, được xác định trong phần tiếp theo.

Một cách hữu ích để xác định một tập hợp là thông qua một tuyên bố liên quan đến các phần tử của nó. Gọi S là tập hợp các phần tử sao cho một câu lệnh s(x) nào đó đúng, để xác định S ta viết

$$S := \{x \mid s(x)\} . \tag{A.1}$$

Ví dụ: A := {x | 1 < x < 3} là tập hợp tất cả các phần tử lớn hơn 1 và nhỏ hơn 3. Hãy để chúng tôi xác định một số tập hợp quan trọng và tập hợp hoạt động bằng cách sử dụng ký hiệu này.

A.2 Các thao tác cơ bản

Định nghĩa A.2.1 (Tập hợp các hoạt động).

- Phần bù S c của tập hợp S chứa tất cả các phần tử không thuộc S.

$$S^c := \{x \mid x / \ S\} . \tag{A.2}$$

- Hợp của hai tập hợp A và B chứa các đối tượng thuộc A hoặc B.

$$A \quad B := \{x \mid x \quad A \text{ hoặc } x \quad B\} . \tag{A.3}$$

Machine Translated by Google

Điều này có thể được khái quát hóa thành một chuỗi các bộ $A_1, A_2, \ldots$

$$\text{Một}_N := \{x \mid x \quad A_n \text{ với } n\}, \tag{A.4}$$

trong đó chuỗi có thể là vô hạn.

- Giao của hai tập hợp A và B chứa các đối tượng thuộc A và B.

$$A \cap B := \{x \mid x \quad A \text{ và } x \quad B\}. \tag{A.5}$$

Một lần nữa, điều này có thể được khái quát hóa thành một chuỗi,

$$\text{Một}_N := \{x \mid x \quad A_n \text{ với mọi } n\}. \tag{A.6}$$

- Hiệu của hai tập hợp A và B chứa các phần tử thuộc A mà không thuộc B.

$$A/B := \{x \mid x \quad A \text{ và } x \,/\, B\}. \tag{A.7}$$

- Bộ nguồn 2^S của một tập hợp S là tập hợp tất cả các tập hợp con có thể có của S, bao gồm và S.

$$^{S2} – := S \mid S \quad S. \tag{A.8}$$

- Tích cartesian của hai tập hợp S1 và S2 là tập hợp tất cả các cặp phần tử có thứ tự trong bộ

$$S_1 \times S_2 := \{(x_1, x_2) \mid x_1 \quad S_1, x_2 \quad S_2\}. \tag{A.9}$$

Một ví dụ là $R^2 = R \times R$, tập hợp tất cả các cặp số thực có thể có.

Hai tập hợp bằng nhau nếu chúng có cùng phần tử, nghĩa là A = B khi và chỉ khi A B và B A. Thật dễ dàng để xác minh rằng (Ac) được gọi $^c = A$, S Ω = Ω, S ∩ Ω = S hoặc các đẳng thức sau là định luật De Morgan.

Định lý A.2.2 (Định luật De Morgan). Với hai tập hợp A và B bất kì

$$(A \quad B)^c = \text{Một}^c \cap B^c, \tag{A.10}$$

$$(A \cap B)^c = \text{Một}^c \quad B^c. \tag{A.11}$$

Bằng chứng. Hãy để chúng tôi chứng minh danh tính đầu tiên; bằng chứng của điều thứ hai gần như giống hệt nhau.

Đầu tiên ta chứng minh rằng $(A \quad B)^c \quad A^c \cap B^c$. Một cách tiêu chuẩn để chứng minh sự bao hàm của một tập hợp trong một tập hợp khác là chỉ ra rằng nếu một phần tử thuộc tập hợp đầu tiên thì nó cũng phải thuộc tập hợp thứ hai. Bất kỳ phần tử x nào trong $(A \quad B)^c$ (nếu tập rỗng thì phép bao hàm không đáng kể, vì S đối với tập S bất kỳ) đều thuộc Ac ; nếu không nó sẽ thuộc về A và do đó thuộc về A B. Tương tự, x cũng thuộc về Bc . Ta kết luận rằng x thuộc Ac ∩ Bc Để hoàn thành chứng , điều đó chứng tỏ sự bao hàm. minh ta thiết lập Ac ∩ Bc (A B) nên x / A B và do ^{c}. Nếu x Ac ∩ Bc, thì x / A và x / B, đó x $(A \quad B)^c$. □

Phụ lục B

Đại số tuyến tính

Chương này cung cấp một đánh giá về các khái niệm cơ bản trong đại số tuyến tính.

B.1 Không gian vectơ

Chắc chắn bạn đã quen thuộc với các vectơ trong R^2 hoặc R^3, I E

$$x = \begin{bmatrix} 2.2 \\ 3 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix}. \tag{B.1}$$

Theo quan điểm của đại số, vectơ là đối tượng tổng quát hơn nhiều. Chúng là các phần tử của các tập hợp được gọi là không gian vectơ thỏa mãn định nghĩa sau.

Định nghĩa B.1.1 (Không gian vectơ). Một không gian vectơ bao gồm một tập hợp V và hai phép toán + và · thỏa mãn các điều kiện sau.

1. Với mọi cặp phần tử $x, y \in V$ thì vectơ tổng $x + y$ thuộc V.
2. Với $x \in V$ bất kỳ và bội số vô hướng $\alpha \in R$ bất kỳ $\alpha \cdot x \in V$.
3. Tồn tại một vectơ không hoặc gốc 0 sao cho $x + 0 = x$ với $x \in V$ bất kỳ.
4. Với mọi $x \in V$, tồn tại một phép cộng nghịch đảo y sao cho $x + y = 0$, thường được ký hiệu là $-x$.
5. Tổng vectơ có tính chất giao hoán và kết hợp, nghĩa là với mọi $x, y \in V$

$$x + y = y + x, \quad (x + y) + z = x + (y + z). \tag{B.2}$$

6. Phép nhân vô hướng có tính chất kết hợp, với mọi $\alpha, \beta \in R$ và $x \in V$

$$\alpha(\beta \cdot x) = (\alpha \beta) \cdot x. \tag{B.3}$$

7. Tổng vô hướng và vectơ đều có tính chất phân phối, nghĩa là với mọi $\alpha, \beta \in R$ và $x, y \in V$

$$(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x, \quad \alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y. \tag{B 4}$$

Không gian con của không gian vectơ V là một tập con của V mà chính nó cũng là một không gian vectơ.

Từ nay về sau, để dễ ký hiệu ta bỏ qua ký hiệu tích vô hướng ·, viết $\alpha \cdot x$ là αx.

Ghi chú B.1.2 (Định nghĩa tổng quát hơn). Chúng ta có thể định nghĩa các không gian vectơ trên một trường tùy ý, thay vì R, chẳng hạn như các số phức C. Chúng ta tham khảo bất kỳ văn bản đại số tuyến tính nào để biết thêm chi tiết.

Chúng ta có thể dễ dàng kiểm tra rằng R^N là một không gian vectơ hợp lệ cùng với phép cộng vectơ thông thường và tích vectơ-vô hướng. Trong trường hợp này, vectơ không là vectơ toàn không $0\ 0\ 0\ ...^T$. Khi nghĩ về không gian vectơ, nên có R^2 hoặc R^3 trong đầu để có được trực giác, nhưng điều quan trọng cần lưu ý là chúng ta có thể xác định các tập hợp vectơ trên nhiều đối tượng khác, chẳng hạn như chuỗi vô hạn, đa thức, hàm và thậm chí cả các biến ngẫu nhiên như trong ví dụ sau.

Định nghĩa của không gian vectơ đảm bảo rằng mọi tổ hợp tuyến tính của các vectơ trong không gian vectơ V, thu được bằng cách cộng các vectơ sau khi nhân với các hệ số vô hướng, đều thuộc về V.
Cho trước một tập hợp các vectơ, một câu hỏi tự nhiên được đặt ra là liệu chúng có thể được biểu diễn dưới dạng tổ hợp tuyến tính của nhau hay không, nghĩa là liệu chúng phụ thuộc tuyến tính hay độc lập.

Định nghĩa B.1.3 (Phụ thuộc/độc lập tuyến tính). Một bộ gồm m véc tơ $x_1, x_2, ..., x_m$ phụ thuộc sớm lin nếu tồn tại m hệ số vô hướng $\alpha_1, \alpha_2, ..., \alpha_m$ không phải tất cả đều bằng 0 và sao cho

$$\sum_{tôi = 1}^{tôi} \alpha_i x_i = 0 . \quad (B.5)$$

Mặt khác, các vectơ độc lập tuyến tính.

Tương tự, ít nhất một vectơ trong tập phụ thuộc tuyến tính có thể được biểu diễn dưới dạng tổ hợp tuyến tính của phần còn lại, trong khi điều này không đúng với các tập độc lập tuyến tính.

Hãy để chúng tôi kiểm tra sự tương đương. Phương trình (B.5) đúng với $\alpha_j = 0$ với một số j khi và chỉ khi

$$x_j = \frac{1}{\alpha_j} \sum_{i \ \{1,...,m\}/\{j\}} a_i x_i . \quad (B.6)$$

Chúng tôi xác định khoảng của một tập hợp các vectơ $\{x_1, ..., x_m\}$ là tập hợp tất cả các tổ hợp tuyến tính có thể có của các vectơ:

$$\text{span}(x_1, ..., x_m) := y \mid y = \sum_{tôi = 1}^{tôi} \alpha_i x_i \text{ cho một số } \alpha_1, \alpha_2, ..., \alpha_m \quad R . \quad (B.7)$$

Điều này hóa ra là một không gian vectơ.

Bổ đề B.1.4. Khoảng của bất kỳ tập hợp vectơ $x_1, ..., x_m$ thuộc một không gian vectơ V là một không gian con của V .

Bằng chứng. Khoảng là tập con của V do Điều kiện 1 và 2 trong Định nghĩa B.1.1. Bây giờ chúng ta chỉ ra rằng nó là một không gian vectơ. Điều kiện 5, 6 và 7 trong Định nghĩa B.1.1 đúng vì V là một không gian vectơ. Ta kiểm tra các Điều kiện 1, 2, 3 và 4 bằng cách chứng minh rằng với hai phần tử tùy ý của khoảng

$$y1 = \sum_{tôi=1}^{tôi} \alpha i\ xi\ ,\ y2 = \sum_{tôi=1}^{tôi} \beta i\ xi\ ,\ \alpha 1, \ldots, \alpha m, \beta 1, \ldots, \beta m \quad R, \tag{B.8}$$

γ1 y1 + γ2 y2 cũng thuộc nhịp. Điều này giữ bởi vì

$$\gamma 1\ y1 + \gamma 2\ y2 = \sum_{tôi=1}^{tôi} (\gamma 1\ \alpha i + \gamma 2\ \beta i)\ xi \quad , \tag{B.9}$$

vì vậy γ1 y1 + γ2 y2 nằm trong khoảng (x1, . . . , xm). Bây giờ để chứng minh Điều kiện 1, chúng ta đặt γ1 = γ2 = 1, cho Điều kiện 2 γ2 = 0, cho Điều kiện 3 γ1 = γ2 = 0 và cho Điều kiện 4 γ1 = 1, γ2 = 0. □

Khi làm việc với một không gian vectơ, sẽ hữu ích khi xem xét tập hợp các vectơ có lực lượng nhỏ nhất bao trùm không gian. Đây được gọi là một cơ sở của không gian vectơ.

Định nghĩa B.1.5 (Cơ sở). Cơ sở của không gian vectơ V là tập các vectơ độc lập {x1, . . . , xm} sao cho

$$V = \text{span}\ (x1, \ldots, xm). \tag{B.10}$$

Một tính chất quan trọng của tất cả các cơ sở trong không gian vectơ là chúng có cùng lực lượng.

Định lý B.1.6. Nếu một không gian vectơ V có một cơ sở với lực lượng hữu hạn thì mọi cơ sở của V đều chứa cùng một số vectơ.

Định lý này, được chứng minh trong Mục B.8.1, cho phép chúng ta xác định chiều của một vectơ không gian.

Định nghĩa B.1.7 (Thứ nguyên). Số chiều dim (V) của một không gian vectơ V là lực lượng của bất kỳ cơ sở nào của nó, hoặc tương đương với số lượng nhỏ nhất các vectơ độc lập tuyến tính trên V.

Định nghĩa này trùng với khái niệm hình học thông thường về kích thước trong R^2 3 và R : một dòng có kích thước 1, trong khi một mặt phẳng có kích thước 2 (miễn là chúng chứa gốc tọa độ). Lưu ý rằng tồn tại không gian vectơ vô hạn chiều, chẳng hạn như các hàm giá trị thực liên tục được xác định trên [0, 1].

Không gian vectơ mà chúng ta sử dụng để mô hình hóa một bài toán nào đó thường được gọi là không gian xung quanh và thứ nguyên của nó là thứ nguyên xung quanh. Trong trường hợp R^n, kích thước xung quanh là n.

Bổ đề B.1.8 (Thứ nguyên của R^n). chiều của R^N là n.

Bằng chứng. Xét tập hợp các vectơ e1, . . . , en R^n xác định bởi

$$e1 = \begin{bmatrix} 1 \\ 0 \\ \cdots \\ 0 \end{bmatrix}, \quad e2 = \begin{bmatrix} 0 \\ 1 \\ \cdots \\ 0 \end{bmatrix}, \quad \ldots, \quad vi = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 1 \end{bmatrix}. \tag{B.11}$$

Machine Translated by Google

Người ta có thể dễ dàng kiểm tra xem tập hợp này có phải là cơ sở hay không. Trên thực tế nó là cơ sở chuẩn của R^N . □

B.2 Tích trong và chuẩn

Cho đến nay, các hoạt động duy nhất chúng tôi đã xem xét là cộng và nhân với một vô hướng. Trong phần này, chúng tôi giới thiệu phép toán thứ ba, tích trong giữa hai vectơ.

Định nghĩa B.2.1 (Tích trong). Tích trong trên không gian vectơ V là phép toán ·, · ánh xạ các cặp vectơ vào R và thỏa mãn các điều kiện sau.

- Nó đối xứng với mọi x, y V

$$x, y = y, x. \tag{B.12}$$

- Nó là tuyến tính, nghĩa là với mọi α R và mọi x, y, z V

$$\alpha x, y = \alpha y, x, \tag{B.13}$$

$$x + y, z = x, z + y, z. \tag{B.14}$$

- Nửa xác định dương: x, x không âm với mọi x V và nếu x, x = 0 thì x = 0.

Không gian vectơ có tích trong được gọi là không gian tích trong. Một trường hợp quan trọng của tích trong là tích vô hướng giữa hai vectơ x, y R N BẰNG

$$x \cdot y := x [i] y [i] , \tag{B.15}$$

trong đó x[i] là mục nhập thứ i của x. Trong phần này, chúng tôi sử dụng xi để biểu thị một vectơ, nhưng trong một số phần khác của ghi chú, nó cũng có thể biểu thị một mục của một vectơ x; điều này sẽ rõ ràng từ bối cảnh. Dễ dàng kiểm tra xem tích vô hướng có phải là tích trong hợp lệ hay không. R n có tích vô hướng thường được gọi là không gian Euclide có chiều n.

Chuẩn của một vectơ là một tổng quát của khái niệm độ dài.

Định nghĩa B.2.2 (Định mức). Cho V là một không gian vectơ, một chuẩn là một hàm ||·|| từ V đến R thỏa mãn các điều kiện sau.

- Nó đồng nhất. Với mọi α R và x V

$$||\alpha x|| = |\alpha| \, ||x|| . \tag{B.16}$$

- Thỏa mãn bất đẳng thức tam giác

$$||x + y|| \leq ||x|| + ||y|| . \tag{B.17}$$

  Đặc biệt, nó không âm (đặt y = x).

- ||x|| = 0 ngụ ý rằng x là vectơ không 0.

Không gian vectơ có chuẩn được gọi là không gian chuẩn. Khoảng cách trong một không gian định chuẩn có thể được đo bằng cách sử dụng định mức hiệu giữa các vectơ.

Định nghĩa B.2.3 (Khoảng cách). Khoảng cách giữa hai vectơ x và y trong một không gian chuẩn tắc với chuẩn ||·|| là

$$d(x, y) := ||x \quad y|| . \tag{B.18}$$

Không gian sản phẩm bên trong là không gian được định mức bởi vì chúng ta có thể xác định một chỉ tiêu hợp lệ bằng cách sử dụng sản phẩm bên trong. Định mức gây ra bởi một tích bên trong có được bằng cách lấy căn bậc hai của tích bên trong của vectơ với chính nó,

$$||x||_{\cdot,\cdot} := \sqrt{x, x} . \tag{B.19}$$

Chuẩn do tích bên trong sinh ra rõ ràng là đồng nhất bởi tính tuyến tính và tính đối xứng của tích bên trong. $||x||_{\cdot,\cdot} = 0$ hàm ý $x = 0$ vì tích bên trong là nửa xác định dương.

Chúng ta chỉ cần chứng minh rằng bất đẳng thức tam giác đúng để đảm bảo rằng tích bên trong là một chuẩn hợp lệ. Điều này suy ra từ một bất đẳng thức cổ điển trong đại số tuyến tính, được chứng minh trong Phần B.8.2.

Định lý B.2.4 (Bất đẳng thức Cauchy-Schwarz). Đối với bất kỳ hai vectơ x và y trong một sản phẩm bên trong không gian

$$|x, y| \leq ||x||_{\cdot,\cdot} \; ||y||_{\cdot,\cdot} . \tag{B.20}$$

Giả sử $||x||_{\cdot,\cdot} = 0$,

$$\cdot,\cdot \; x, y = \quad ||x||_{\cdot,\cdot} \; ||y||_{\cdot,\cdot} \qquad y = \quad || \frac{||y||}{x||_{\cdot,\cdot} \; ||} x, \tag{B.21}$$

$$x, y = ||x||_{\cdot,\cdot} \; ||y||_{\cdot,\cdot} \qquad y = ||x||_{\cdot,\cdot} \frac{y||_{\cdot,\cdot}}{} x. \tag{B.22}$$

Hệ quả B.2.5. Định mức gây ra bởi một sản phẩm bên trong thỏa mãn bất đẳng thức tam giác.

Bằng chứng.

$$||x + y||^2_{\cdot,\cdot} = ||x||^2_{\cdot,\cdot} + ||y||^2_{\cdot,\cdot} + 2\; x, y \tag{B.23}$$

$$\leq ||x||^2_{\cdot,\cdot} + ||y||^2_{\cdot,\cdot} + 2\; ||x||_{\cdot,\cdot} \; ||y||_{\cdot,\cdot} \text{ theo bất đẳng thức Cauchy-Schwarz}$$

$$= ||x||_{\cdot,\cdot} + ||y||_{\cdot,\cdot}{}^2 . \tag{B.24}$$

□

Định mức Euclide hoặc 2 là định mức được tạo ra bởi tích vô hướng trong R^n,

$$||x||_2 := \sqrt{x \cdot x} = \sqrt{\sum_{tôi = 1}^{N} x[i]^2} . \tag{B.25}$$

trong trường hợp của R^2 3 hoặc R nó là cái mà chúng ta thường coi là độ dài của vectơ.

B.3 Tính trực giao

Một khái niệm quan trọng trong đại số tuyến tính là tính trực giao.

Định nghĩa B.3.1 (Tính trực giao). Hai vectơ x và y trực giao với nhau nếu

$$x, y = 0. \tag{B.26}$$

Một vectơ x trực giao với tập hợp S, nếu

$$x, s = 0, \text{ với mọi } s \quad S. \tag{B.27}$$

Hai tập hợp S1, S2 là trực giao nếu với mọi x S1, y S2

$$x, y = 0. \tag{B.28}$$

Phần bù trực giao của một không gian con S là

$$S \quad := \{x \mid x, y = 0 \text{ với mọi } y \quad S\} . \tag{B.29}$$

Khoảng cách giữa các vectơ trực giao được đo theo định mức gây ra bởi bên trong sản phẩm dễ tính toán.

Định lý B.3.2 (Định lý Pitago). Nếu x và y là các vectơ trực giao

$$||x + y||^2 = ||x||^2 + ||y||^2 . \tag{B.30}$$

Bằng chứng. Theo tuyến tính của sản phẩm bên trong

$$||x + y||^2 = ||x||^2 + ||y||^2 + 2 x, y \tag{B.31}$$

$$= ||x||^2 + ||y||^2 . \tag{B.32}$$

□

Nếu chúng ta muốn chứng minh rằng một vectơ trực giao với một không gian con nhất định, thì chỉ cần chứng minh rằng nó trực giao với mọi vectơ trong một cơ sở của không gian con.

Bổ đề B.3.3. Cho x là một vectơ và S là một không gian con của chiều n. Nếu với mọi cơ sở b1, b2, . . . , tỷ của S,

$$x, b_i = 0, \quad 1 \leq i \leq n, \tag{B.33}$$

thì x trực giao với S.

Bằng chứng. Bất kỳ vectơ v S nào cũng có thể được biểu diễn dưới dạng $v = \sum_{i=1}^{N} \alpha_i b_i$ cho α1, . . . , αn R, từ (B.33)

$$x, v = x, \sum_{tôi=1}^{N} \alpha_i b_i = \sum_{tôi=1}^{N} \alpha_i \; x, b_i = 0. \tag{B.34}$$

□

Bây giờ chúng tôi giới thiệu các cơ sở trực giao.

Định nghĩa B.3.4 (Cơ sở trực giao). Cơ sở của các vectơ trực giao với nhau có chuẩn bằng nhau với một được gọi là một cơ sở trực giao.

Rất dễ tìm các hệ số của một véc tơ trong một cơ sở trực giao: chúng ta chỉ cần tính các tích vô hướng với các véc tơ cơ sở.

Bổ đề B.3.5 (Các hệ số trong cơ sở trực giao). Nếu $\{u_1, \ldots, u_n\}$ là một cơ sở trực giao của không gian vectơ V, với mọi vectơ $x \in V$

$$x = \sum_{tôi=1}^{N} \langle x, u_i \rangle u_i . \tag{B.35}$$

Bằng chứng. Vì $\{u_1, \ldots, u_n\}$ là một cơ sở,

$$x = \sum_{tôi=1}^{tôi} \alpha_i u_i \text{ cho một số } \alpha_1, \alpha_2, \ldots, \alpha_m \in R . \tag{B.36}$$

Ngay lập tức,

$$\langle x, u_i \rangle = \left\langle \sum_{tôi=1}^{tôi} u_i u_i , \right\rangle = \sum_{tôi=1}^{tôi} \alpha_i \langle u_i, u_i \rangle = \alpha_i \tag{B.37}$$

bởi vì $\langle u_i, u_i \rangle = 1$ và $\langle u_i, u_j \rangle = 0$ với $i \neq j$. □

Đối với bất kỳ không gian con nào của phương pháp R^N chúng ta có thể có được một cơ sở trực chuẩn bằng cách áp dụng Gram-Schmidt thành một tập hợp các vectơ độc lập tuyến tính bao trùm không gian con.

Thuật toán B.3.6 (Gram-Schmidt). Xét tập hợp các vectơ độc lập tuyến tính $x_1, \ldots, x_m$ trong R^N. Để có được một cơ sở trực giao của khoảng của các vectơ này, chúng tôi:

1. Đặt $u_1 := x_1 / ||x_1||_2$.

2. Với $i = 1, \ldots, m$, tính toán

$$v_i := x_i - \sum_{j=1}^{i-1} \langle u_j, x_i \rangle u_j . \tag{B.38}$$

và đặt $u_i := v_i / ||v_i||_2$.

Không khó để chứng minh rằng tập hợp các vectơ $u_1, \ldots, u_m$ là cơ sở trực giao cho khoảng $x_1, \ldots, x_m$. Điều này đặc biệt ngụ ý rằng chúng ta luôn có thể giả sử rằng một không gian con có một cơ sở trực giao.

Định lý B.3.7. Mọi không gian vectơ hữu hạn chiều đều có một cơ sở trực giao.

Bằng chứng. Để thấy rằng phương pháp Gram-Schmidt tạo ra một cơ sở trực giao cho span của các vectơ đầu vào, chúng ta có thể kiểm tra span $(x_1, \ldots, x_i)$ = span $(u_1, \ldots u_i)$ và $u_1, \ldots, u_i$ là tập hợp các vectơ trực giao. □

B.4 Phép chiếu

Hình chiếu của vectơ x lên không gian con S là vectơ trong S gần x nhất. Để xác định điều này một cách chặt chẽ, chúng tôi bắt đầu bằng cách giới thiệu khái niệm tổng trực tiếp. Nếu hai không gian con rời nhau, nghĩa là điểm chung duy nhất của chúng là gốc tọa độ, thì một vectơ có thể viết dưới dạng tổng của một vectơ từ mỗi không gian con được gọi là thuộc tổng trực tiếp của chúng.

Định nghĩa B.4.1 (Tổng trực tiếp). Cho V là một không gian vectơ. Với mọi không gian con S_1, S_2 V sao cho

$$S_1 \cap S_2 = \{0\} \tag{B.39}$$

tổng trực tiếp được định nghĩa là

$$S_1 \quad S_2 := \{x \mid x = s_1 + s_2 \; s_1 \quad S_1, s_2 \quad S_2\} . \tag{B.40}$$

Biểu diễn của một vectơ trong tổng trực tiếp của hai không gian con là duy nhất.

Bổ đề B.4.2. Mọi vectơ x S1 S2 có một biểu diễn duy nhất

$$x = s_1 + s_2 \; s_1 \quad S_1, s_2 \quad S_2. \tag{B.41}$$

Bằng chứng. Nếu x S1 S2 thì theo định nghĩa tồn tại s_1 S1, s_2 S2 sao cho $x = s_1 + s_2$. Giả sử $x = v_1 + v_2$, v_1 S1, v_2 S2, khi đó s_1 $v_1 = s_2$ v_2. Điều này ngụ ý rằng s_1 v_1 và s_2 v_2 nằm trong S1 và cả trong S2. Tuy nhiên $S_1 \cap S_2 = \{0\}$ nên ta kết luận $s_1 = v_1$ và $s_2 = v_2$. □

Bây giờ chúng ta có thể định nghĩa phép chiếu của một vectơ x lên một không gian con S bằng cách tách vectơ đó thành một thành phần thuộc S và một thành phần khác thuộc phần bù trực giao của nó.

Định nghĩa B.4.3 (Phép chiếu trực giao). Cho V là một không gian vectơ. Hình chiếu trực giao của một vectơ x V lên một không gian con S V là một vectơ ký hiệu là $P_S\, x$ sao cho x $P_S\, x$ S .

Định lý B.4.4 (Tính chất của các phép chiếu trực giao). Cho V là một không gian vectơ. Mỗi vectơ x V có một phép chiếu trực giao duy nhất $P_S\, x$ lên bất kỳ không gian con S V có số chiều hữu hạn.
Đặc biệt x có thể được biểu thị dưới dạng

$$x = P_S\, x + P_S \quad x. \tag{B.42}$$

Với mọi vectơ s S

$$x, s = P_S\, x, s. \tag{B.43}$$

Đối với mọi cơ sở trực giao $b_1, \ldots, b_m$ của S,

$$P_S\, x = \sum_{tôi = 1}^{tôi} x, b_i \, b_i . \tag{B.44}$$

Bằng chứng. Gọi số chiều của S là m. Vì m hữu hạn nên tồn tại một cơ sở trực giao S: bb Xét vectơ 1 , · m: · ,

$$p := \sum_{tôi=1}^{tôi} \langle x, b_{tôi} \rangle b_{tôi} . \tag{B.45}$$

Hóa ra x p trực giao với mọi vectơ trong cơ sở. Với $1 \leq j \leq m$,

$$\langle x - p, b_j \rangle = \langle x - \sum_{tôi=1}^{tôi} \langle x, b_{tôi} \rangle bb_{tôi}, j \rangle \tag{B.46}$$

$$= \langle x, b_j \rangle - \sum_{tôi=1}^{tôi} \langle x, b_{tôi} \rangle \langle bb_{tôi}, j \rangle \tag{B.47}$$

$$= \langle x, b_j \rangle - \langle x, b_j \rangle = 0, \tag{B.48}$$

nên x p S và p là một phép chiếu trực giao. Vì $S \cap S = \{0\}$ vectơ khác x1 [1] không thể có hai S, x1 S sao cho $x = x_1 + x_2$ nên phép chiếu trực giao là duy nhất. sao cho x o = p thuộc S và .

Lưu ý rằng o := x p là một vectơ trong S do đó thuộc S và lập (B.42).
Điều này ngụ ý rằng o là hình chiếu trực giao của x lên S

Phương trình (B.43) suy ngay từ tính trực giao của bất kỳ vectơ s S và PS x.

Phương trình (B.44) suy ra từ (B.43). □

Việc tính toán chuẩn của phép chiếu của một vectơ lên một không gian con sẽ dễ dàng nếu chúng ta tiếp cận được một cơ sở trực chuẩn (miễn là chuẩn được tạo ra bởi tích bên trong).

Bổ đề B.4.5 (Định mức của phép chiếu). Chuẩn của phép chiếu của một vectơ tùy ý x V lên một không gian con S V có chiều d có thể được viết là

$$||P_S\, x||_{\langle\cdot,\cdot\rangle} = \sqrt{\sum_{tôi}^{đ} \langle b_i, x \rangle^2} \tag{B.49}$$

cho mọi cơ sở trực giao b1, . . . , bd của S .

Bằng chứng. Theo (B.44)

$$||P_S\, x||^2_{\langle\cdot,\cdot\rangle} = \langle P_S x, P_S x \rangle \tag{B.50}$$

$$= \left\langle \sum_{tôi}^{đ} \langle b_i, x \rangle b_i, \sum_{j}^{đ} \langle b_j, x \rangle b_j \right\rangle \tag{B.51}$$

$$= \sum_{tôi}^{đ} \sum_{j}^{đ} \langle b_i, x \rangle \langle b_j, x \rangle \langle b_i, b_j \rangle \tag{B.52}$$

$$= \sum_{tôi}^{đ} \langle b_i, x \rangle^2 . \tag{B.53}$$

□

1Đối với bất kỳ vectơ v nào thuộc cả S và S v, $v = ||v||_2^2 = 0$, nghĩa là $v = 0$.

Machine Translated by Google

Ví dụ B.4.6 (Phép chiếu lên không gian con một chiều). Để tính toán hình chiếu của một vectơ x trên một không gian con một chiều được mở rộng bởi một vectơ v, chúng ta sử dụng thực tế là

$v/ \|v\|_{\cdot,\cdot}$ là một cơ sở cho span (v) (nó là một tập hợp chứa một vectơ đơn vị kéo dài không gian con) và áp dụng (B.44) để thu được

$$P_{\text{span}(v)}\, x = \frac{v, x}{\|v\|^2_{\cdot,\cdot}} v. \tag{B.54}$$

Cuối cùng, ta chứng minh rằng hình chiếu của vectơ x lên không gian con S thực sự là vectơ trong S gần x nhất trong khoảng cách do định mức tích bên trong gây ra.

Định lý B.4.7 (Hình chiếu trực giao gần nhất). Phép chiếu trực giao của một vectơ x lên một không gian con S thuộc cùng một không gian tích trong là vectơ gần x nhất thuộc S xét về chuẩn do tích bên trong gây ra. Chính thức hơn, PS x là lời giải bài toán tối ưu

$$\text{giảm thiểu}_{\text{bạn}} \quad \|x \quad u\|_{\cdot,\cdot} \tag{B.55}$$

$$\text{chịu} \quad u \quad S \;. \tag{B.56}$$

Bằng chứng. Lấy bất kỳ điểm s S sao cho s = PS x

$$\|x \quad s\|^2_{\cdot,\cdot} = \|x \quad PS\ x + PS\ x \quad s\|^2_{\cdot,\cdot} \tag{B.57}$$

$$= \|x \quad PS\ x\|^2_{\cdot,\cdot} + \|PS\ x \quad s\|^2_{\cdot,\cdot} \tag{B.58}$$

$$> \|x \quad PS\ x\|^2_{\cdot,\cdot} \quad \text{vì } s = PS\ x, \tag{B.59}$$

trong đó (B.58) suy ra từ định lý Pythagore vì vì PS x := x PS x thuộc về $S_{\perp}$ và PS x s đến S. □

B.5 Ma trận

Một ma trận là một mảng hình chữ nhật của các số. Ta ký hiệu không gian vectơ của ma trận m × n là $r^{m\times n}$. Ta ký hiệu hàng thứ i của ma trận A là Ai: , cột thứ j bởi A:j và mục nhập (i, j) của Aij . Chuyển vị của một ma trận thu được bằng cách chuyển đổi các hàng và cột của nó.

Định nghĩa B.5.1 (Chuyển cung). Chuyển vị AT của ma trận A R $^{m\times n}$ là một ma trận trong Một $R^{m\times n}$

$$\text{MỘT}^{t}_{ij} = A_{ji}. \tag{B.60}$$

Một ma trận đối xứng là một ma trận bằng với chuyển vị của nó.

Ma trận ánh xạ các vectơ tới các vectơ khác thông qua một phép toán tuyến tính được gọi là tích ma trận-vector.

Machine Translated by Google

Định nghĩa B.5.2 (Tích ma trận-vectơ). Tích của ma trận A R x R $^{m\times n}$ và một véc tơ ^{N} là một vectơ Ax R ^{N}, như vậy mà

$$(\text{Cây rìu})_i = \sum_{j=1}^{N} A_{ij}x[j] \tag{B.61}$$

$$= \langle A_{i:}, x \rangle, \tag{B.62}$$

tức là mục thứ i của Ax là tích vô hướng giữa hàng thứ i của A và x.

tương đương,

$$\text{rìu} = \sum_{j=1}^{N} A_{:j}x[j], \tag{B.63}$$

tức là Ax là sự kết hợp tuyến tính của các cột của A được đánh trọng số bởi các mục trong x.

Người ta có thể dễ dàng kiểm tra xem chuyển vị của tích của hai ma trận A và B có bằng các phép chuyển vị nhân theo thứ tự nghịch đảo hay không,

$$(AB)^t = B^t \text{MỘT}^t. \tag{B.64}$$

Chúng ta có thể biểu diễn tích vô hướng giữa hai vectơ x và y dưới dạng

$$\langle x, y \rangle = x^T y = y^T x. \tag{B.65}$$

Ma trận đơn vị là một ma trận ánh xạ bất kỳ vectơ nào tới chính nó.

Định nghĩa B.5.3 (Ma trận định danh). Ma trận đơn vị trong R $^{n\times n}$ là

$$\text{tôi} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{B.66}$$

Rõ ràng, với mọi x R ^{N} ix = x.

Định nghĩa B.5.4 (Nhân ma trận). Tích của hai ma trận A RR n×p là ma trận AB R m×p $^{m\times n}$ và B , như vậy mà

$$(AB)_{ij} = \sum_{k=1}^{N} A_{ik}B_{kj} = \langle \text{Ái}:, B_{:,j} \rangle, \tag{B.67}$$

tức là mục (i,j) của AB là tích giữa hàng thứ i của A và cột thứ j của B.

Tương tự, cột thứ j của AB là kết quả của phép nhân A với cột thứ j của B

$$AB = \sum_{k=1}^{N} A_{ik}B_{kj} = \langle \text{Ái}:, B_{:,j} \rangle, \tag{B.68}$$

và hàng thứ i của AB là kết quả của phép nhân hàng thứ i của A và B.

Ma trận vuông có thể có nghịch đảo. Nếu đúng như vậy, thì ma trận nghịch đảo là một ma trận đảo ngược hiệu ứng của ma trận của bất kỳ vectơ nào.

Định nghĩa B.5.5 (Nghịch đảo ma trận). Ma trận nghịch đảo của ma trận vuông $A \in R^{n\times n}$ là một ma trận $A^{-1} \in R^{n\times n}$ như vậy mà

$$AA^{-1} = A^{-1}A = I . \quad (B.69)$$

Bổ đề B.5.6. Ma trận nghịch đảo là duy nhất.

Bằng chứng. Giả sử có một ma trận M khác sao cho AM = I, khi đó

$$M = A^{-1}AM \text{ bởi (B.69)} \quad (B.70)$$

$$= A^{-1} . \quad (B.71)$$

□

Một loại ma trận quan trọng là ma trận trực giao.

Định nghĩa B.5.7 (Ma trận trực giao). Một ma trận trực giao là một ma trận vuông sao cho nghịch đảo của nó bằng chuyển vị của nó,

$$U^TU = UU^T = I \quad (B.72)$$

Theo định nghĩa, các cột $U_{:1}, U_{:2}, \ldots, U_{:n}$ của bất kỳ ma trận trực giao nào đều có chuẩn đơn vị và trực giao với nhau, vì vậy chúng tạo thành một cơ sở trực chuẩn (thay vào đó, hơi khó hiểu là ma trận trực giao không được gọi là ma trận trực giao). Chúng ta có thể hiểu việc áp dụng $U^T x$ là tính toán các hệ số biểu diễn thành một véc tơ của nó trong cơ sở được hình thành bởi các cột của U.
Áp dụng U cho $U^t x$ phục hồi x bằng cách chia tỷ lệ từng vectơ cơ sở với hệ số tương ứng:

$$x = UU^T x = \sum_{i=1}^{N} \langle U_{:i}, x\rangle U_{:i} . \quad (B.73)$$

Áp dụng một ma trận trực giao cho một vectơ không ảnh hưởng đến chuẩn của nó, nó chỉ xoay vectơ.

Bổ đề B.5.8 (Các ma trận trực giao bảo toàn chuẩn). Đối với ma trận trực giao bất kỳ $U \in R^{n\times n}$ và bất kỳ vectơ $x \in R^N$,

$$||Ux||_2 = ||x||_2 . \quad (B.74)$$

Bằng chứng. Theo định nghĩa của một ma trận trực giao

$$||Ux||_2^2 = x^TU^TUx = x^T \quad (B.75)$$

$$x = || \quad (B.76)$$

$$x||_2^2 . \quad (B.77)$$

□

B.6 Sự phân hủy bản địa

Một vectơ riêng v của ma trận A thỏa mãn

$$av = \lambda v \tag{B.78}$$

đối với một vô hướng λ là giá trị riêng tương ứng. Ngay cả khi A là số thực, thì các vectơ riêng và giá trị riêng của nó có thể phức tạp.

Bổ đề B.6.1 (Phân tích bản địa). Nếu một ma trận vuông A R có n $^{n\times n}$ độc lập tuyến tính với các các vectơ riêng giá trị riêng λ1, . . . , λn nó có thể được biểu diễn dưới dạng ma trận Q, v1, . . . , có các cột là các vectơ riêng và một ma trận đường chéo chứa các giá trị riêng,

$$một = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}^{1} \tag{B.79}$$

$$= Q\Lambda Q^{1} \tag{B.80}$$

Bằng chứng.

$$AQ = \begin{bmatrix} Av_1 & Av_2 & \cdots & Av_n \end{bmatrix} \tag{B.81}$$

$$= \begin{bmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_2 v_n \end{bmatrix} \tag{B.82}$$

$$= Q\Lambda. \tag{B.83}$$

Nếu các cột của một ma trận vuông đều độc lập tuyến tính, thì ma trận này có một nghịch đảo, do đó, việc nhân biểu thức với Q 1 ở cả hai vế sẽ hoàn thành chứng minh. □

Bổ đề B.6.2. Không phải tất cả các ma trận đều có phân tích riêng

Bằng chứng. Xét ví dụ ma trận

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{B.84}$$

Giả sử λ có một giá trị riêng khác 0 tương ứng với một véc tơ riêng với các mục v[1] và v[2], khi đó

$$\begin{bmatrix} v[2] \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v[1] \\ v[2] \end{bmatrix} = \begin{bmatrix} \lambda v[2] \\ \lambda v[2] \end{bmatrix}, \tag{B.85}$$

ngụ ý rằng v[2] = 0 và do đó v[1] = 0, vì chúng ta đã giả sử rằng λ = 0. Điều này ngụ ý rằng ma trận không có các giá trị riêng khác 0 liên kết với các vectơ riêng khác 0. □

Một ứng dụng thú vị của phép phân tích riêng là tính toán các tích ma trận kế tiếp nhau rất nhanh. Giả sử rằng chúng ta muốn tính toán

$$AA \cdots Ax = A^k x, \tag{B.86}$$

tức là ta muốn áp A lên xk lần. Ak không thể được tính toán bằng cách lấy sức mạnh của các mục của nó (hãy thử một ví dụ đơn giản để thuyết phục chính bạn). Tuy nhiên, nếu A có một phân tích riêng,

$$A^k = Q\Lambda Q^{-1}Q\Lambda Q^{-1} \cdots Q\Lambda Q^{-1} \tag{B.87}$$

$$= Q\Lambda^k Q^{-1} \tag{B.88}$$

$$= Q \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \lambda_n^k \end{bmatrix} Q^{-1}, \tag{B.89}$$

sử dụng thực tế là đối với các ma trận đường chéo, việc áp dụng ma trận nhiều lần tương đương với việc lấy sức mạnh của các mục đường chéo. Điều này cho phép tính k tích ma trận chỉ bằng 3 tích ma trận và lấy lũy thừa của n số.

Từ trường trung học hoặc đại số đại học, bạn có thể nhớ cách tính toán các vectơ riêng bằng cách sử dụng định thức. Trong thực tế, đây thường không phải là một lựa chọn khả thi do các vấn đề về độ ổn định. Một kỹ thuật phổ biến để tính toán các vectơ riêng dựa trên hiểu biết sâu sắc sau đây. Cho $A \in \mathbb{R}^{n\times n}$ là ma trận có phân tích riêng $Q\Lambda Q^{-1}$ và x là vectơ tùy ý trong $\mathbb{R}^n$. Kể từ khi cột của Q độc lập tuyến tính, chúng tạo cơ sở cho $\mathbb{R}^n$, vì vậy chúng ta có thể đại diện cho x như

$$x = \sum_{i=1}^{n} \alpha_i Q_{:i}, \quad \alpha_i \in \mathbb{R},\ 1 \le i \le n. \tag{B.90}$$

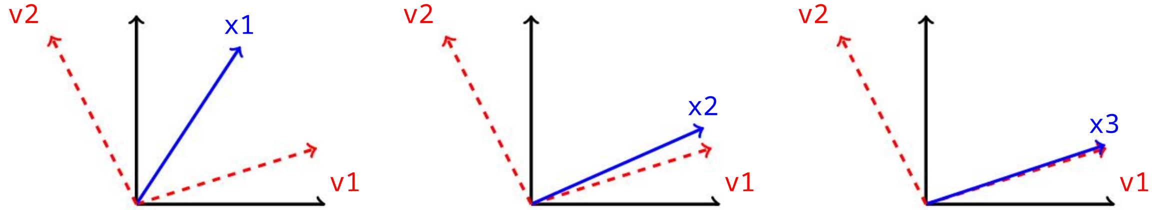
Bây giờ chúng ta hãy áp dụng A cho xk lần,

$$A^k x = \sum_{i=1}^{n} \alpha_i A^k Q_{:i} \tag{B.91}$$

$$= \sum_{i=1}^{n} \alpha_i \lambda_i^k Q_{:i}. \tag{B.92}$$

Nếu chúng ta giả định rằng các vectơ riêng được sắp xếp theo độ lớn của chúng và độ lớn của một trong số chúng lớn hơn các vectơ còn lại, thì $|\lambda_1| > |\lambda_2| \ge \ldots$, và $\alpha_1 \neq 0$ (xảy ra với xác suất cao nếu chúng ta rút ngẫu nhiên một x) thì khi k càng lớn thì số hạng $\alpha_1\lambda_1^k Q_{:1}$ chiếm ưu thế. Số hạng sẽ tăng vọt hoặc có xu hướng bằng 0 trừ khi chúng ta chuẩn hóa mọi lần trước khi áp dụng A. Việc thêm bước chuẩn hóa vào quy trình này dẫn đến phương pháp lũy thừa hoặc phép lặp lũy thừa, một thuật toán có tầm quan trọng lớn trong đại số tuyến tính số.

Thuật toán B.6.3 (Phương pháp lũy thừa).
Input: Một ma trận A .
Đầu ra: Ước tính của vectơ riêng của A tương ứng với giá trị riêng lớn nhất.



Hình B.1: Minh họa ba lần lặp đầu tiên của phương pháp lũy thừa đối với ma trận có các vectơ riêng v1 và v2, có các giá trị riêng tương ứng là $\lambda 1$ = 1,05 và $\lambda 2$ = 0,1661.

Khởi tạo: Đặt $x1 := x/ ||x||2$, trong đó các mục nhập của x được rút ngẫu nhiên.

Với tôi = 1, . . . , k, tính toán

$$xi := \frac{\text{trục} \quad 1}{||Axi \quad 1||2} . \tag{B.93}$$

Hình B.1 minh họa phương pháp lũy thừa trên một ví dụ đơn giản, trong đó ma trận bằng

$$\text{một} = \begin{matrix} 0{,}930 & 0{,}388 \\ 0{,}237 & 0{,}286 \end{matrix} \tag{B.94}$$

Sự hội tụ đến vector riêng tương ứng với giá trị riêng có độ lớn lớn nhất diễn ra rất nhanh.

B.7 Phân tích riêng của ma trận đối xứng

Các ma trận đối xứng thực luôn có một phân tích riêng. Ngoài ra, các giá trị riêng của chúng là thực và các vectơ riêng của chúng đều trực giao.

Định lý B.7.1 (Định lý phổ cho ma trận đối xứng thực). Nếu A R thì nó $^{n\times n}$ là đối xứng, có phân tích riêng dạng

$$\text{một} = \begin{matrix} u1 & u2 & \cdots & un \end{matrix} \begin{matrix} \lambda 1 & 0 & \cdots & 0 \\ 0 & \lambda 2 & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & \lambda n \end{matrix} \begin{matrix} u1 & u2 & \cdots & un \end{matrix}^{t} , \tag{B.95}$$

trong đó các giá trị riêng u1, u2, . . . , un là số thực và các vectơ riêng u1, u2, . . . , un là số thực và trực giao.

Bằng chứng. Việc chứng minh rằng mọi ma trận đối xứng thực đều có n véc tơ riêng nằm ngoài phạm vi của các ghi chú này. Theo giả định rằng đây là trường hợp, chúng tôi bắt đầu bằng cách chứng minh rằng các giá trị riêng là có thật. Xét một giá trị riêng tùy ý λi và véc tơ riêng chuẩn hóa tương ứng vi chúng ta có ,

$$_{vi} Avi = \lambda v \text{ tôi } vi = \lambda, \tag{B.96}$$

$$^{\text{tôi}}_{-} Avi = (Avi) \quad vi = (\lambda vi) \quad = \lambda v^{-} \quad i\ vi = \lambda. \tag{B.97}$$

Điều này ngụ ý rằng λ là thực vì $\lambda = \bar{\lambda}$, vì vậy chúng ta có thể hạn chế các vectơ riêng là thực (vì giá trị riêng là thực, cả phần thực và phần ảo của vectơ riêng đều là các vectơ riêng của chúng và ít nhất một trong số chúng phải khác không). Nếu một số vectơ riêng độc lập tuyến tính có cùng giá trị riêng, thì một cơ sở trực chuẩn của khoảng của chúng cũng sẽ bao gồm các vectơ riêng của ma trận. Tất cả những gì còn lại để chứng minh là các vectơ riêng tương ứng với các giá trị riêng khác nhau là trực giao. Giả sử vi và vj là các vectơ riêng tương ứng với các giá trị riêng khác nhau $\lambda_i = \lambda_j$, khi đó

$$u_i^T u_j = \frac{1}{\lambda_i} (A u_i)^T u_j \tag{B.98}$$

$$= \frac{1}{\lambda_i} u_i^T A^T u_j \tag{B.99}$$

$$= \frac{\lambda_j}{\lambda_i} u_i^T u_j . \tag{B.100}$$

$$= \tag{B.101}$$

Điều này chỉ có thể xảy ra nếu $u_i^T u_j = 0$. □

Các giá trị riêng của ma trận đối xứng xác định giá trị của dạng bậc hai:

$$q(x) := x^T A x = \sum_{i=1}^{N} \lambda_i \left(x^T\right)^2 \tag{B.102}$$

Nếu chúng ta sắp xếp các giá trị riêng $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ thì giá trị riêng thứ nhất là giá trị lớn nhất mà hàm bậc hai đạt được nếu đầu vào của nó có chuẩn đơn vị 2 , giá trị riêng thứ hai là giá trị lớn nhất mà dạng bậc hai đạt được nếu chúng ta giới hạn đối số của nó là chuẩn hóa và trực giao với vectơ riêng thứ nhất, và sớm.

Định lý B.7.2. Với mọi ma trận đối xứng A R n với các vectơ riêng chuẩn hóa $u_1, u_2, \ldots,$ với các giá trị riêng bỏ tương ứng $\lambda_1, \lambda_2, \ldots, \lambda_n$

$$\lambda_1 = \max_{||u||_2=1} u^T A u, \tag{B.103}$$

$$u_1 = \arg\max_{||u||_2=1} u^T A u, \tag{B.104}$$

$$\lambda_k = \max_{||u||_2=1, u \perp u_1,\ldots,u_{k-1}} u^T A u, \tag{B.105}$$

$$u_k = \arg\max_{||u||_2=1, u \perp u_1,\ldots,u_{k-1}} u^T A u. \tag{B.106}$$

Bằng chứng. Các vectơ riêng là một cơ sở trực chuẩn (chúng trực giao lẫn nhau và chúng ta giả sử rằng chúng đã được chuẩn hóa), vì vậy chúng ta có thể biểu diễn bất kỳ vectơ chuẩn đơn vị hk nào trực giao với $u_1, \ldots, u_{k-1}$ dưới dạng

$$k = \sum_{i=k}^{n} a_i u_i \tag{B.107}$$

Ở đâu

$$\text{HK}\frac{2}{2} = \sum_{\text{tôi} = k}^{\text{tôi}} \alpha_{\text{tôi}}^{2} = 1, \qquad \text{(B.108)}$$

bởi bổ đề B.4.5. Lưu ý rằng h1 chỉ là một vectơ định mức đơn vị tùy ý.

Bây giờ chúng ta sẽ chỉ ra rằng giá trị của dạng bậc hai khi đầu vào chuẩn hóa bị hạn chế là trực giao với u1, . . . , uk 1 không thể lớn hơn λk,

$$h_k^t àk = \sum_{\text{tôi} = 1}^{N} \text{tôi} \left(\sum_{j=k}^{\text{tôi}} a_j u_i^T u_j \right)^2 \quad \text{bởi (B.102) và (B.107)} \qquad \text{(B.109)}$$

$$= \sum_{\text{tôi} = 1}^{N} \lambda i \alpha^2 \quad \text{vì } u1, . . . , um \text{ là một cơ sở trực giao} \qquad \text{(B.110)}$$

$$\leq \lambda k \sum_{\text{tôi} = k}^{\text{tôi}} \alpha_{\text{tôi}}^{2} \quad \text{bởi vì } \lambda k \geq \lambda k+1 \geq . . . \geq \lambda m \qquad \text{(B.111)}$$

$$= \lambda k, \text{ bởi (B.108).} \qquad \text{(B.112)}$$

Điều này thiết lập (B.103) và (B.105). Để chứng minh (B.104) và (B.106) ta chỉ cần chứng minh uk đạt cực đại

$$u_k^T auk = \sum_{\text{tôi} = 1}^{N} \lambda i \left(u_i^T uk\right)^2 \qquad \text{(B.113)}$$

$$= \lambda k. \qquad \text{(B.114)}$$

□

B.8 Bằng chứng

B.8.1 Chứng minh Định lý B.1.6

Ta chứng minh khẳng định bằng mâu thuẫn. Giả sử rằng chúng ta có hai cơ sở {x1, . . . , xm} và {y1, . . . , yn} sao cho m < n (hoặc tập hợp thứ hai có lực lượng vô hạn). Chứng minh xuất phát từ việc áp dụng m lần bổ đề sau (đặt r = 0, 1, ... , m 1) để chứng minh rằng {y1, . . . , ym} kéo dài V và do đó {y1, . . . , yn} phải phụ thuộc tuyến tính.

Bổ đề B.8.1. Theo giả thiết của định lý, nếu {y1, y2, . . . , năm, xr+1, . . . , xm} kéo dài V sau đó {y1, . . . , năm+1, xr+2, . . . , xm} cũng kéo dài V (có thể sau khi sắp xếp lại các chỉ số r+1, . . . , m) cho r = 0, 1, . . . , m 1.

Bằng chứng. Vì {y1, y2, . . . , năm, xr+1, . . . , xm} kéo dài V

$$\text{năm+1} = \sum_{\text{tôi} = 1}^{r} i\, yi + \sum_{i=r+1}^{\text{tôi}} yi\, xi\,,\ \beta 1, . . . , \beta r,\ \gamma r+1, . . . , \gamma m \quad R, \qquad \text{(B.115)}$$

Machine Translated by Google

trong đó ít nhất một trong số yj khác 0, như {y1, . . . , yn} độc lập tuyến tính theo giả thiết. Không làm mất tính tổng quát (đây là nơi chúng ta có thể cần sắp xếp lại các chỉ số), chúng ta giả sử rằng yr+1 = 0, do đó

$$x_{r+1} = \frac{1}{y_{r+1}} \sum_{i=1}^{r} \beta_i y_i - \sum_{i=r+2}^{m} y_i x_i \,. \tag{B.116}$$

Điều này ngụ ý rằng bất kỳ vectơ nào trong khoảng {y1, y2, . . . , năm, xr+1, . . . , xm}, tức là trong V, có thể được biểu diễn dưới dạng tổ hợp tuyến tính của các vectơ trong {y1, . . . , năm+1, xr+2, . . . , xm}, hoàn thành bằng chứng. □

B.8.2 Chứng minh Định lý B.2.4

Nếu ||x||·,· = 0 thì x = 0 vì tích bên trong là nửa xác định dương, hàm ý x, y = 0 và do đó (B.20) có đẳng thức. Điều này cũng đúng nếu ||y||·,· = 0.

Bây giờ giả sử rằng ||x||·,· = 0 và ||y||·,· = 0. Bằng tính nửa xác định của tích trong,

$$0 \leq \left\| \, ||y||_{\cdot,\cdot}\, x + ||x||_{\cdot,\cdot}\, y \right\|^2 = 2\, ||x||^2_{\cdot,\cdot}\, ||y||^2_{\cdot,\cdot} + 2\, ||x||_{\cdot,\cdot}\, ||y||_{\cdot,\cdot}\, x, y, \tag{B.117}$$

$$0 \leq \left\| \, ||y||_{\cdot,\cdot}\, x \quad ||x||_{\cdot,\cdot}\, y \right\|^2 = 2\, ||x||^2_{\cdot,\cdot}\, ||y||^2_{\cdot,\cdot} \quad 2\, ||x||_{\cdot,\cdot}\, ||y||_{\cdot,\cdot}\, x, y. \tag{B.118}$$

Những bất bình đẳng này thiết lập (B.20).

Hãy chứng minh (B.21) bằng cách chứng minh cả hai hàm ý.

( ) Giả sử x, y = ||x||·,· ||y||·,·. Khi đó (B.117) bằng không, vì vậy ||y||·,· x = ||x||·,· y vì tích bên trong là nửa xác định dương. ( ) Giả sử ||y||·,· x = ||x||·,· y. Khi đó người ta có thể dễ dàng kiểm tra xem (B.117) có bằng 0 hay không, nghĩa là x, y = ||x||·,· ||y||·,·.

Chứng minh của (B.22) là đồng nhất (dùng (B.118) thay cho (B.117)).