ĐẠI HỌC QUỐC GIA TP.HCM
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC VÀ KỸ THUẬT MÁY TÍNH

TIỂU LUẬN VỀ TOÁN HỌC CƠ SỞ

LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ DÀNH CHO MÁY HỌC

Theory of Probability and Statistics for Machine Learning

Bộ môn: **CƠ SỞ TOÁN** (CO5263)

Người hướng dẫn khoa học: Tiến sỹ **NGUYỄN AN KHƯƠNG**
Tiến sỹ **TRẦN TUẤN ANH**

Học viên nghiên cứu: **NGUYỄN XUÂN HIỀN**

Mã học viên: 2470749

Hệ đào tạo: Cao học chính quy

Chuyên ngành: Khoa học máy tính

Tháng 06/2025

TÓM TẮT

Xác suất và Thống kê là những trụ cột toán học thiết yếu cho lĩnh vực máy học; đồng thời, đóng vai trò trung tâm trong việc mô hình hóa và xử lý các vấn đề về sự không chắc chắn vốn có trong dữ liệu và các tác vụ dự đoán. Bài tiểu luận này tổng hợp và tóm tắt các lý thuyết nền tảng của Xác suất và Thống kê nhằm cung cấp một cơ sở kiến thức vững chắc cho những người bắt đầu nghiên cứu và xây dựng mô hình trong lĩnh vực máy học.

Nội dung chính của tiểu luận này bắt đầu bằng việc giới thiệu tổng quan về tầm quan trọng của Xác suất và Thống kê trong các khía cạnh khác nhau của học máy, bao gồm học có giám sát, học không giám sát và học tăng cường. Tiếp đó, bài viết đi sâu vào các khái niệm xác suất cơ bản như không gian mẫu, biến cố, các tiên đề xác suất, và hàm xác suất. Khái niệm về biến ngẫu nhiên (rời rạc và liên tục), hàm mật độ xác suất được làm rõ, cùng với việc mở rộng sang đa biến ngẫu nhiên, bao gồm xác suất đồng thời, xác suất có điều kiện và định lý Bayes. Các tham số thống kê quan trọng dùng để mô tả dữ liệu và phân phối như giá trị kỳ vọng, phương sai, độ lệch chuẩn, và ma trận hiệp phương sai cũng được trình bày chi tiết. Một phần quan trọng của tiểu luận cũng đề cập đến các loại không chắc chắn trong học máy, cụ thể là không chắc chắn ngẫu nhiên và không chắc chắn nhận thức.

Lý thuyết được minh họa bằng các ví dụ cụ thể như bài toán tung đồng xu để giải thích luật số lớn và định lý giới hạn trung tâm, và ví dụ về xét nghiệm y tế để áp dụng định lý Bayes. Cuối cùng, một loạt các bài tập tình huống được đưa ra và giải quyết nhằm củng cố hiểu biết và khả năng áp dụng các khái niệm đã học vào thực tế. Thông qua việc hệ thống hóa của tiểu luận này, tác giả mong muốn trang bị cho người đọc những kiến thức Xác suất và Thống kê cơ bản cần thiết để tiếp cận các vấn đề trong máy học máy một cách hiệu quả.

Từ khoá: Xác suất, Thống kê, Máy học, Sự không chắc chắn, Biến ngẫu nhiên, Định lý Bayes, Giá trị kỳ vọng, Phương sai, Luật số lớn, Cơ sở toán.
Keywords: Probability, Statistics, Machine learning, Uncertainty, Random variables, Bayes' Rule, Expected value, Variance, Law of Large Numbers, Mathematical Foundations.

Thông tin về tác giả

Với sự hướng dẫn về học thuật của Tiến sỹ ***NGUYỄN AN KHƯƠNG*** và Tiến sỹ ***TRẦN TUẤN ANH***, bài tiểu luận này được biên soạn bởi học viên cao học ngành Khoa học và Kỹ thuật máy tính, Trường đại học Bách Khoa TP.Hồ Chí Minh với thông tin sau:

Họ và tên tác giả : ***NGUYỄN XUÂN HIỀN***.
Mã học viên : 2470749.
Thư điện tử : nxhien.sdh242@hcmut.edu.vn.
Điện thoại / Signal : +84 933 023 468.

MỤC LỤC

1. **Tổng quan về xác suất và thông kê trong nghiên cứu máy học:**.....1
2. **Bài toán cơ bản:**.....2
3. **Các phương pháp khác:**.....5
4. **Biến ngẫu nhiên:**.....6
5. **Đa biến ngẫu nhiên:** 7
6. **Ví dụ minh hoạ:** 9
7. **Các tham số trong thống kê:** 11
8. **Vấn đề về sự không chắc chắn:**.....15
9. **Bài tập nghiên cứu tình huống:**.....15

DANH MỤC BẢNG BIỂU, HÌNH ẢNH

Đồ thị 1 - Ước lượng giá trị xác suất của các sự kiện xảy ra đối với thử nghiệm tung đồng xu.

Đồ thị 2 - Phân phối xác suất hậu nghiệm của đồng xu không cân đối.

Đồ thị 3 - Phân phối xác suất hậu nghiệm của đồng xu không cân đối.

Biểu đồ 1 - Biểu đồ Venn biểu diễn mối quan hệ giữa biến cố A và B.

NỘI DUNG CHÍNH

1. Tổng quan về xác suất và thông kê trong nghiên cứu máy học:

Mặc dù số lượng phương pháp máy học, ***machine learning (ML)***, ngày càng gia tăng cùng với lượng dữ liệu ngày càng lớn dần nhưng vẫn luôn tồn tại sự không chắc chắn, ***uncertainty***. Đối với lĩnh vực học có giám sát, ***supervised learning***, chúng ta muốn dự báo mục tiêu, ***target***, chưa biết dựa vào các đặc trưng hay thuộc tính đã biết, ***features***, đã biết. Và chúng ta luôn cố gắng dự đoán giá trị có khả năng xảy ra nhất của mục tiêu hoặc giá trị có khoảng cách dự kiến nhỏ nhất so với mục tiêu. Và đôi khi chúng ta không chỉ muốn dự đoán một giá trị cụ thể mà còn muốn định lượng sự không chắc chắn. Trong lĩnh vực học không giám sát, ***unsupervised learning***, chúng ta thường quan tâm đến sự không chắc chắn. Để xác định khả năng bất thường của một tập hợp các phép đo, phương pháp máy học này giúp nhận biết khả năng ảnh hưởng của sự bất thường đến các giá trị quan sát trong một quần thể thuận lợi. Hơn nữa, trong học tăng cường, ***reinforcement learning***, chúng ta mong muốn phát triển các tác nhân hoạt động thông minh trong nhiều môi trường khác nhau. Điều này đòi hỏi chúng ta phải suy luận về ảnh hưởng của thay đổi môi trường và hệ quả kỳ vọng có thể xảy ra khi phản ứng với mỗi hành động khả thi.

Xác suất, ***probability***, là bộ môn toán học liên quan đến lý luận trong điều kiện không chắc chắn. Với một mô hình xác suất của một số quy trình, chúng ta có thể suy luận về khả năng xảy ra của nhiều sự kiện, ***events***, khác nhau. Việc sử dụng xác suất để mô tả tần suất của các sự kiện lặp lại sẽ hạn chế sự tranh cãi. Trên thực tế, các học giả theo trường phái tần suất, ***frequentist***, thường tuân thủ cách giải thích xác suất chỉ áp dụng cho các sự kiện lặp lại. Ngược lại, các học giả theo trường phái Bayes, ***Bayesian***, sử dụng ngôn ngữ xác suất rộng hơn để hợp lý hóa lý luận trong điều kiện không chắc chắn. Xác suất theo quan điểm

Bayes được có hai đặc trưng duy nhất: (i) chỉ định mức độ tin tưởng vào các sự kiện không lặp lại và (ii) tính chủ quan.

Khác với xác suất, thống kê, ***statistics***, giúp chúng ta suy luận ngược, bắt đầu bằng việc thu thập và tổ chức dữ liệu, ***data***, rồi suy ra những suy luận có thể rút ra về quá trình tạo ra dữ liệu. Tư duy thống kê giúp chúng ta phân tích một tập dữ liệu, ***dataset***, tìm kiếm các hình mẫu, ***pattern***, mà chúng ta kỳ vọng có thể đặc trưng cho một quần thể, ***population***, rộng hơn.

Bài tiểu luận này tổng hợp và tóm tắt một số lý thuyết để tạo nên một nền tảng cho khởi đầu nghiên cứu xây dựng mô hình.

2. Bài toán cơ bản:

Chúng ta có thể bắt đầu với bài toán cơ bản là thử nghiệm, ***experiment***,[1] tung một đồng xu hai mặt và lượng hoá khả năng xuất hiện mặt ngửa, head (h), so với mặt sấp, tail (t). Nếu đồng xu đồng chất hoặc cân đối, ***fair coin***, thì cả hai kết quả mặt sấp và mặt ngửa đều có khả năng như nhau. Hơn nữa, nếu tung đồng xu n lần thì chúng ta mong muốn tỷ lệ mặt ngửa xuất hiện phải khớp chính xác với tỷ lệ kỳ vọng của mặt sấp. Một cách trực quan để thấy điều này là tính chất đối xứng, ***symmetry***: với mọi kết quả đầu ra, ***outcome***, có thể có với n_h lần mặt ngửa và $n_t = (n - n_h)$ lần mặt sấp, thì có một kết quả gần như cân bằng với n_h lần mặt ngửa và n_t lần mặt sấp. Khả năng điều này xảy ra dựa vào trung bình của số lần thử nếu chúng ta mong đợi nhìn thấy số lần xuất hiện của mặt ngửa và mặt sấp đều là 1/2. Tất nhiên, nếu chúng ta tiến hành thí nghiệm này nhiều lần với n=1000000 lần tung, thì chúng ta không thể thấy tồn tại một lần thử chính xác $n_h = n_t$.

Về mặt hình thức, định lượng số học ½ được gọi là xác suất và khả năng nắm bắt được sự chắc chắn với bất kỳ lần tung xuất hiện mặt ngửa. Xác suất gán các điểm trong khoảng 0 và 1 cho các kết quả thuận lợi, được gọi là các sự kiện

[1] Phụ lục 1 - Lý thuyết bổ sung.

hoặc biến cố, ***event***[2]. Sự kiện thuận lợi là xuất hiện mặt ngửa và xác suất tương ứng được ký hiệu là P(heads). Xác suất bằng 1 chỉ ra sự chắc chắn tuyệt đối (một đồng xu có cả hai mặt đều là mặt ngửa) và xác suất bằng 0 chỉ ra sự bất khả thi (nếu cả hai mặt đều là mặt sấp). Tần suất n_h / n và n_t / n không phải là xác suất mà là thống kê. Xác suất là các đại lượng lý thuyết làm cơ sở cho quá trình tạo dữ liệu. Trong bài toán này, xác suất ½ là một thuộc tính của chính đồng xu. Ngược lại, thống kê là các đại lượng thực nghiệm được tính toán như các hàm của dữ liệu được quan sát. Và chúng ta quan tâm đến mối quan hệ chặt chẽ giữa các đại lượng xác suất và thống kê. Vì thế, chúng ta thường thiết kế các thống kê đặc biệt được gọi là các ước tính từ một tập dữ liệu cho trước, từ đó tạo ra các ước tính cho các tham số của mô hình như xác suất. Hơn nữa, khi các ước tính đó thỏa mãn một thuộc tính hay được gọi là tính nhất quán, ***consistency***, việc ước tính của chúng ta sẽ hội tụ về xác suất tương ứng. Đổi lại, những xác suất suy ra này cho biết về các đặc tính thống kê có thể có của dữ liệu trong cùng một quần thể mà chúng ta có thể bắt gặp trong tương lai.

Giả sử chúng ta thật sự tung một đồng xu mà không biết giá trị xác suất thực sự của mặt ngửa xuất hiện, P(heads). Để nghiên cứu định lượng bằng các phương pháp thống kê, chúng ta cần (i) thu thập một số dữ liệu; và (ii) thiết kế một ước lượng. Việc thu thập dữ liệu dễ dàng bằng phép thử tung đồng xu nhiều lần và ghi lại tất cả các kết quả. Về mặt hình thức, việc thu được các kết quả từ một số quá trình ngẫu nhiên cơ bản được gọi là lấy mẫu, ***sampling***. Như chúng ta dự đoán, một ước lượng tự nhiên là tỷ lệ giữa số mặt sấp quan sát được với tổng số lần tung.

Bây giờ, giả sử đồng xu thực sự cân đối với P(heads) = 0.50. Để mô phỏng phép thử tung một đồng xu, chúng ta có thể lập trình tạo số ngẫu nhiên và tạo mẫu ngẫu nhiên của một sự kiện với xác suất là 0.50.

[2] Phụ lục 1 - Lý thuyết bổ sung.

Ví dụ 1:[3] Viết một đoạn mã tạo ra các số trong khoảng [0,1] mà xác suất nằm trong bất kỳ khoảng con [a,b]⊂[0,1]. Từ đó, chúng ta có thể lấy ra 0 và 1 với xác suất 0.50 cho mỗi lần thử bằng cách kiểm tra giá trị phần thập phân trả với 0.50. Kết quả lập trình mô phỏng với n = 100:

```
Heads: 51, Tails: 49
```

Mặc dù phép thử tung đồng xu được mô phỏng là công bằng với xác suất do chúng ta đặt ra là [0.50, 0.50], nhưng số lần xuất hiện mặt ngửa và mặt sấp có thể không giống nhau, bởi vì, chúng ta chỉ rút ra một số lượng mẫu tương đối nhỏ. Nếu không tự triển khai mô phỏng và tự thấy kết quả, thì chúng ta không thể biết được phép thử có hơi không công bằng hay độ lệch khỏi xác suất 1/2 có thể chỉ là sản phẩm của mẫu quy mô nhỏ.

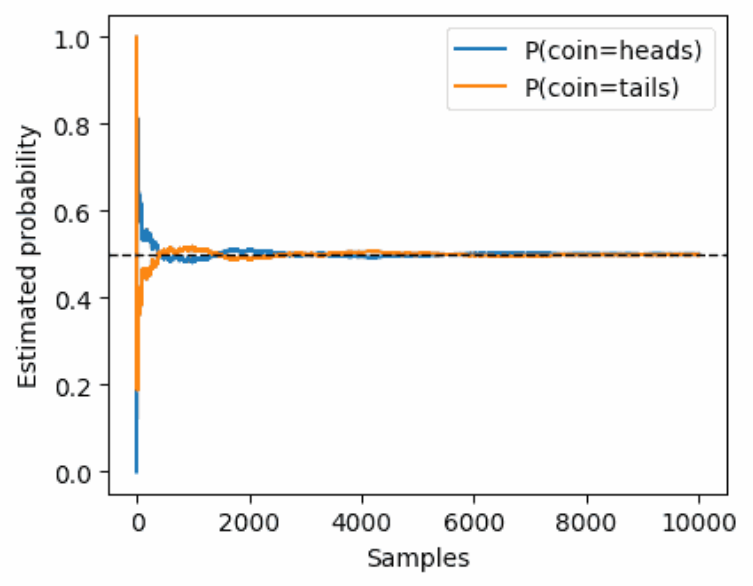
Ví dụ 2:[4] Bây giờ chúng ta thử mô phỏng phân phối đa thức, ***multinomial***, với n = 10000 lần thử. Kết quả thu được là:

```
tensor([0.4955,0.5045])
```

Tổng quát, với các giá trị trung bình của các sự kiện lặp lại (như tung đồng xu), khi số lần lặp lại tăng lên, các giá trị ước lượng được đảm bảo sẽ hội tụ thực sự về các xác suất cơ bản. Công thức toán học của hiện tượng này được gọi là luật số lớn, ***the law of large numbers***[5], và định lý giới hạn trung tâm, ***central limit theorem (Lindeberg and Lévy)***[6], cho chúng ta biết rằng trong nhiều tình huống, khi quy mô mẫu n tăng lên, các sai số sẽ giảm xuống theo tỷ lệ $(1/\sqrt{n})$.

Ví dụ 3:[7] Chúng ta tìm hiểu thêm bằng cách nghiên cứu cách ước tính tiến triển khi chúng ta tăng số lần tung từ 1 lên 10.000. Kết quả được biểu diễn dưới dạng đồ thị.

[3] Phụ lục 2 - Mã nguồn lập trình mô phỏng.
[4] Phụ lục 2 - Mã nguồn lập trình mô phỏng.
[5] Phụ lục 1 - Lý thuyết bổ sung.
[6] Phụ lục 1 - Lý thuyết bổ sung.
[7] Phụ lục 2 - Mã nguồn lập trình mô phỏng.



Đồ thị 1 - Ước lượng giá trị xác suất của các sự kiện xảy ra đối với thử nghiệm tung đồng xu.

Mỗi đường cong đặc tương ứng với một trong hai giá trị của đồng xu và xác suất ước tính đạt được sau mỗi nhóm thử nghiệm. Đường thẳng màu đen đứt nét đưa ra xác suất cơ bản thực sự. Khi chúng ta có thêm dữ liệu bằng cách tiến hành nhiều thử nghiệm hơn, các đường cong sẽ hội tụ về phía xác suất thực sự.

3. Các phương pháp khác:

Không gian mẫu, ***sample space***, hay không gian kết quả đầu ra, ***outcome space***, là tập hợp tất cả các sự kiện đơn giản, ***simple event***, hay các kết quả đầu ra duy nhất có thể xảy ra, ***distinct possible outcome***, đã được biết từ một thí nghiệm, ký hiệu là ***S***.[8,9]

Biến cố, ***events***, là một tập hợp con của không gian mẫu. Các biến cố loại trừ lẫn nhau, m***utually exclusive events***, là các biến cố không thể xảy ra đồng thời.[10]

Đối với thử nghiệm tung một đồng xu, ta có không gian mẫu ***S={heads,tails}***. Đối với thử nghiệm tung hai đồng xu, ***S={(heads, heads), (heads, tails),***

[8] Mục 4.2. "Introduction to Probability and Statistics".
[9] Mục 2.2. "A first course in probability".
[10] Mục 4.2. "Introduction to Probability and Statistics".

(tails,heads), (tails, tails)}. Trường hợp đồng xu thứ nhất xuất hiện mặt ngửa (heads) thì tương ứng với tập ***{(heads, heads), (heads, tails)}***.

Đối với thử nghiệm lắc một hột xí ngầu, ***S={1,2,3,4,5,6}***. Đặt ***A={5}*** tương ứng với biến cố xuất hiện mặt ***5*** chấm, ***B={1,3,5}*** tương ứng với biến cố xuất hiện mặt có số chấm lẻ. Trường hợp xuất hiện mặt ***5*** chấm hay ***z = 5*** thì cả biến cố ***A*** và ***B*** cùng xảy ra và $z \in A, B$. Trường hợp xuất hiện mặt ***3*** chấm hay ***z = 3*** thì biến cố ***A*** xảy ra còn ***B*** thì không và $z \in A, z \notin B$.

Hàm xác suất, ***probability function***, phản ánh các biến cố dựa vào các giá trị thực $P : A \subseteq S \rightarrow [0,1]$, ký hiệu ***P(A)***, với biến cố ***A*** thuộc không gian mẫu ***S*** cho trước với các tính chất sau:

- Xác suất bất kỳ của biến cố ***A*** là một số thực trong đoạn [0,1], $0 \leq P(A) \leq 1$.
- Xác suất của toàn bộ không gian mẫu ***S*** bằng ***1***, $P(S) = 1$.
- Với $A_1, A_2,...$ là dãy biến cố loại trừ lẫn nhau thì xác suất của bất kỳ A_i bằng tổng các xác suất riêng lẻ:

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Các tính chất trên đây chính là ba tiền đề của lý thuyết xác suất, ***axioms of probability theory***,[11] được Kolmogorov (1933) đề xuất, có thể được áp dụng để nhanh chóng rút ra một số hệ quả quan trọng. Cụ thể, phần bù, ***complement***, của một biến cố ***A***, ký hiệu là ***A'***, là một tập hợp con chứa các sự kiện trong không gian mẫu S nhưng không thuộc biến cố ***A***.[12] Nghĩa là $A \cup A' = S$, có xác suất $P(A \cup A') = 1$ và $P(A \cap A') = 0$ hay $P(A') = 1 - P(A)$.

4. Biến ngẫu nhiên:

Về mặt hình thức, các biến ngẫu nhiên, ***random variable***,[13] là các phép ánh xạ từ không gian mẫu cơ bản sang một tập hợp các giá trị. Mặc dù, biến ngẫu nhiên và không gian mẫu đều là tập hợp các kết quả đầu ra nhưng mọi giá trị được

[11] Mục 2.3. "A first course in probability".
[12] Mục 1.4. "Probability and Statistics".
[13] Phụ lục 1 - Lý thuyết bổ sung.

lấy ra bởi một biến ngẫu nhiên đều tương ứng với một tập con của không gian mẫu cơ bản. Vì vậy, sự xuất hiện của biến ngẫu nhiên X có giá trị là v, ký hiệu X = v, có xác suất là P(X = v). Biến ngẫu nhiên là một khái niệm trung tâm trong lý thuyết xác suất. Các biến ngẫu nhiên có thể là biến ngẫu nhiên rời rạc, ***Random discrete variable***, (như tung một đồng xu, lắc một hột xí ngầu) hoặc biến ngẫu nhiên liên tục, ***Random continuous variable***, (như chiều cao và cân nặng của một người được chọn mẫu ngẫu nhiên từ một quần thể). Chiều cao chính xác của một người có thể làm một số thực như 1.801392782910287192 (m), nhưng chúng ta thường quan tâm đến khả năng chiều cao của một người có nằm trong một khoảng nhất định, chẳng hạn như giữa 1,79 và 1,81 (m). Trong những trường hợp này, chúng ta làm việc với mật độ xác suất, ***probability density***. Chiều cao chính xác là 1,80 mét thì không có xác suất, nhưng có mật độ có thể khác không. Để tính xác suất được gán cho một khoảng, chúng ta phải lấy tích phân của mật độ trên khoảng đó.

5. Đa biến ngẫu nhiên:

Chúng ta có thể nhận thấy rằng hầu hết thuật toán máy học đều liên quan đến tương tác giữa nhiều biến ngẫu nhiên. Mỗi biến ngẫu nhiên sẽ biểu thị giá trị (chưa biết) của một thuộc tính khác nhau. Và khi chúng ta thực hiện lấy mẫu một cá thể từ quần thể, chúng ta quan sát thấy sự hiện thực hóa của từng biến ngẫu nhiên. Vì các giá trị do các biến ngẫu nhiên lấy ra tương ứng với các tập hợp con của không gian mẫu có thể chồng lên nhau, ***overlapping***, chồng một phần hoặc hoàn toàn không giao nhau, ***disjoint***, nên việc biết giá trị do một biến ngẫu nhiên lấy ra có thể khiến chúng ta cập nhật niềm tin về giá trị của một biến ngẫu nhiên khác có khả năng xảy ra. Chẳng hạn, nếu một bệnh nhân bước vào bệnh viện và chúng ta quan sát thấy người này gặp khó khăn khi thở và mất khứu giác, thì chúng ta tin rằng người đó có nhiều khả năng nhiễm vi rút COVID-19 hơn so với người khác không gặp khó khăn khi thở và có khứu giác hoàn toàn bình thường.

Khi làm việc với nhiều biến ngẫu nhiên, chúng ta có thể xây dựng các biến cố tương ứng với mọi tổ hợp giá trị mà các biến có thể cùng nhau lấy ra. Hàm xác suất gán các xác suất cho từng tổ hợp này được gọi là hàm xác suất kết hợp, ***joint probability function***, và chỉ trả về xác suất được gán cho phần giao nhau, ***intersection***, của các tập con tương ứng của không gian mẫu. Xác suất kết hợp được gán cho biến cố mà các biến ngẫu nhiên X và Y nhận giá trị x và y tương ứng được ký hiệu là P(X = x, Y = y), ta có bất đẳng thức:

$$P(X = x, Y = y) \leq P(X = x) \text{ và } P(X = x, Y = y) \leq P(Y = y)$$

Để cho X = x và Y = y xảy ra, thì X = x phải xảy ra và Y = y cũng phải xảy ra. Xác suất kết hợp cho chúng ta biết tất cả những điều có thể biết về các biến ngẫu nhiên này theo quan điểm xác suất và có thể được sử dụng để suy ra nhiều đại lượng khác, bao gồm cả việc khôi phục các phân phối riêng lẻ P(X) và P(Y). Để khôi phục P(X = x), chúng ta chỉ cần cộng P(X = x, Y = v) trên tất cả các giá trị v mà biến ngẫu nhiên Y có thể lấy:

$$(1)\ P(X = x) = \sum_v P(X = x, Y = v)$$

Tỷ lệ $P(X = x, Y = y) / P(X = x) \leq 1$ rất quan trọng, được gọi là xác suất có điều kiện, ***conditional probability***, được xác định bởi công thức:

$$(2)\ P(Y = y \mid X = x) = P(X = x, Y = y) / P(X = x)$$

Công thức này cho thấy mối liên hệ biến ngẫu nhiên Y = y trong điều kiện xảy ra X = x. Trường hợp biến cố B và B' không giao nhau, ta được công thức cộng xác suất:

$$(3)\ P(B \cup B' \mid A) = P(B \mid A) + P(B' \mid A)$$

Sử dụng định nghĩa xác suất có điều kiện, ta rút ra được công thức Bayes, ***Bayes's rule / formula***, như sau:

$$(4)\ P(A \mid B) = P(B \mid A)P(A) / P(B)$$

Với $P(B) = P(B|A) + P(B|A')$ ta triển khai công thức (4) thành:[14]

[14] Mục 3.3. "A first course in probability".

$$(5)\ P(A|B) = \frac{P(B|A)P(A)}{P(B|A) + P(B|A')}$$

6. Ví dụ minh hoạ:

Giả sử một bác sĩ tiến hành xét nghiệm COVID-19 cho một bệnh nhân. Xét nghiệm này khá chính xác và chỉ thất bại với xác suất 1%, tức là, bệnh nhân không nhiễm bệnh nhưng có kết quả xét nghiệm dương tính chiếm 1% trong tất cả trường hợp. Hơn nữa, xét nghiệm này không bao giờ không phát hiện ra COVID-19 nếu bệnh nhân thực sự mắc bệnh. Chúng ta đặt $D_i \in \{0,1\}$ để chỉ i lần chẩn đoán (giá trị là 0 nếu âm tính và 1 nếu dương tính) và $C \in \{0,1\}$ để biểu thị tình trạng nhiễm vi rút (giá trị 0 là không nhiễm và 1 là phơi nhiễm). Hãy tính xác suất bệnh nhân bị nhiễm vi rút COVID-19 khi làm xét nghiệm lần 1 và có kết quả dương tính, biết xác suất người bị lây nhiễm hiện tại là 0.0015.

Tóm tắt dữ liệu đầu vào:

Xác suất dương tính giả: $P(D_1 = 1|C = 0) = 0.01$

Xác suất dương tính thật: $P(D_1 = 1| C = 1) = 1$

Xác suất 1 người nhiễm bệnh: $P(C = 1) = 0.0015$

Yêu cầu tính $P(C = 1|D_1 = 1)$.

Áp dụng công thức Bayes ta được:

$$P(C = 1|D_1 = 1) = P(D_1 = 1|C = 1)P(C = 1) / P(D_1 = 1)$$

Trong đó, xác suất một người có kết quả xét nghiệm dương tính:

$$P(D_1 = 1) = P(D_1 = 1,C = 1) + P(D_1 = 1,C = 0)$$

Áp dụng công thức xác suất có điều kiện:

$$P(D_1 = 1) = P(D_1 = 1|C = 1)P(C = 1) + P(D_1 = 1|C = 0)P(C = 0)$$

Thay các giá trị vào ta được kết quả:

$$P(D_1 = 1) = 1 \text{ x } 0.0015 + 0.01 \text{ x } (1 - 0.0015) = 0.011485$$

Xác suất một người nhiễm bệnh có kết quả xét nghiệm dương tính:

$$P(C = 1|D_1 = 1) = 1 \text{ x } 0.0015 / 0.011485 \approx 0.1306$$

Kết quả xét nghiệm lần 2 cũng cho kết quả dương tính. Tính xác suất một người nhiễm bệnh đều có kết quả dương tính sau 2 lần làm xét nghiệm. Biết rằng xác suất thất bại ở lần 2 là 3% và xác suất phát hiện người bị nhiễm bệnh còn 98%.

Xác suất dương tính giả lần 2: $P(D_2 = 1|C = 0) = 0.03$

Xác suất dương tính thật lần 2: $P(D_2 = 1|C = 1) = 0.98$

Xác suất dương tính thật cả 2 lần với D_1 và D_2 độc lập:

$P(D_1 = 1,D_2 = 1|C = 1) = P(D_1 = 1|C = 1)P(D_2 = 1|C = 1) = 1 \times 0.98 = 0.98$

Xác suất dương tính giả cả 2 lần với D_1 và D_2 độc lập:

$P(D_1 = 1,D_2 = 1|C = 0) = P(D_1 = 1|C = 0)P(D_2 = 1|C = 0) = 0.01 \times 0.03 = 0.0003$

Áp dụng công thức Bayes ta được:

$$P(C = 1|D_1 = 1,D_2 = 1) = P(D_1 = 1,D_2 = 1|C = 1)P(C = 1) / P(D_1 = 1,D_2 = 1)$$

Trong đó, xác suất một người có kết quả xét nghiệm hai lần đều dương tính:

$P(D_1 = 1,D_2 = 1) = P(D_1 = 1,D_2 = 1,C = 1) + P(D_1 = 1,D_2 = 1,C = 0)$

$= P(D_1 = 1,D_2 = 1|C = 1)P(C = 1) + P(D_1 = 1,D_2 = 1|C = 0)P(C=0)$

$= 0.98 \times 0.0015 + 0.0003 \times (1 - 0.0015) = 0.00176955$

Xác suất một người nhiễm bệnh có kết quả dương tính cả 2 lần xét nghiệm:

$P(C = 1|D_1 = 1,D_2 = 1) = 0.98 \times 0.0015 / 0.00176955 \approx 0.8307$

Xét nghiệm hai lần cho xác suất phát hiện người nhiễm bệnh cao hơn đáng kể. Mặc dù, xác suất riêng của lần xét nghiệm thứ hai thấp hơn nhưng vẫn cải thiện đáng kể kết quả ước tính sau cùng. Giả định quan trọng là cả hai lần xét nghiệm đều độc lập có điều kiện với nhau nên tạo khả năng ước tính chính xác hơn. Hãy lấy trường hợp bất lợi nhất khi chúng ta thực hiện cùng một xét nghiệm hai lần. Trong tình huống này, chúng ta sẽ mong đợi cùng một kết quả cả hai lần, do đó không có thêm hiểu biết nào được thu thập khi thực hiện lại cùng một bài kiểm tra. Kết quả chẩn đoán hoạt động giống như phương pháp phân loại ẩn, khả năng quyết định bệnh nhân nhiễm bệnh chỉ tăng lên khi có thêm nhiều đặc trưng hay thuộc tính, ***features***, hoặc kiểm tra kết quả đầu ra.

7. Các tham số trong thống kê:

Thông thường, việc đưa ra quyết định không chỉ xem xét các xác suất được gán cho các biến cố riêng lẻ mà còn phải tính toán xác suất kết hợp toàn bộ các biến cố để có thể cung cấp cho chúng ta hướng dẫn phù hợp. Chẳng hạn, khi các biến ngẫu nhiên liên tục lấy các giá trị vô hướng, chúng ta cần phải tính toán giá trị kỳ vọng trung bình.

Ví dụ 4: Trong đầu tư tài chính, các nhà đầu tư luôn quan tâm đến lợi nhuận kỳ vọng, ***expected return***, hay giá trị trung bình của tất cả các kết quả đầu tư có thể xảy ra và được xem xét dựa theo các xác suất xảy ra tương ứng. Giả định rằng danh mục đầu tư A thất bại với xác suất là 50%, danh mục B có thể mang lại lợi nhuận 2x (lần) với xác suất 40% và danh mục C có thể mang lại lợi nhuận 10x (lần) với xác suất 10%.

Gọi $X=\{x_1,x_2,x_3\}=\{0,2,10\}$ là đại lượng ngẫu nhiên rời rạc đại diện cho các giá trị của lợi nhuận của các danh mục đầu tư A, B, C tương ứng với xác suất xảy ra là P(X). Lợi nhuận trung bình (kỳ vọng) của toàn bộ danh mục đầu tư:

$$E[X] = \sum_{i=1}^{3} x_i . P(x_i) = 0.50 * 0 + 0.40 * 2 + 0.10 * 10 = 1.8$$

Từ ví dụ 4, ta có công thức tính giá trị kỳ vọng, ***expected value***, hay trung bình, ***mean***, tổng quát của biến ngẫu nhiên rời rạc X là:

$$(6)\ \mu = E[X] = E_{x\sim X}(x) = \sum_x xP(X = x)$$

Hoặc X là một đại lượng ngẫu nhiên liên tục:

$$(7)\ E[X] = \int_{-\infty}^{+\infty} x.p(x)dx$$

Trường hợp, tính kỳ vọng cho đại lượng có dạng hàm số f(x):

$$(7)\ E_{x\sim P}[f(x)] = \sum_x f(x)P(x)$$

$$\text{Hoặc } (8)\ E_{x\sim P}[f(x)] = \int_{-\infty}^{+\infty} f(x)p(x)dx$$

Ví dụ 5: Gọi f là hàm hữu ích của tiền, ***utility of money***, thể hiện sự hài lòng hay hạnh phúc liên quan đến lợi nhuận đầu tư. Từ dữ liệu đầu vào của ví dụ 4, chúng ta có đại lượng ngẫu nhiên đại diện cho lợi nhuận là X={0,2,10}. Gọi Y

là đại lượng ngẫu nhiên đại diện cho mức độ hài lòng và có quan hệ với X bởi hàm hữu ích f (có thể là dạng logarit), hay Y=f(X). Giả định rằng từ hàm f chúng ta có được các giá trị của $Y=\{y_1,y_2,y_3\}=\{-1,2,4\}$ tương ứng với X. Từ đó, chúng ta tính giá trị hạnh phúc kỳ vọng (trung bình) trong đầu tư là:

$$E[Y] = E[f(X)] = \sum_{i=1}^{3} f(x_i)p(x_i) = \sum_{i=1}^{3} y_i p(x_i)$$

E[Y] = 0.50*(-1) + 0.40*(2) + 0.10*4 = 0.70

Giả định rằng nhà đầu tư chỉ chấp nhận mức độ hài lòng kỳ vọng không được nhỏ hơn 1.00 nên tổng tổn thất từ đầu tư kỳ vọng, ***expected loss of utility***, tối thiểu là 1.00 - 0.70 = 0.30. Nếu thực sự đây là hàm hữu ích của nhà đầu tư thì việc gửi tiền vào ngân hàng là giải pháp tốt nhất, dù rằng, giá trị của tiền thường có xu hường dưới mức tuyến tính.

Đối với các quyết định tài chính, nhà đầu tư thường muốn đo lường mức độ rủi ro của một khoản đầu tư, không chỉ là giá trị trung bình (kỳ vọng) mà còn mức độ thay đổi các giá trị thực tế hướng đến sự thay đổi liên quan với giá trị này. Lưu ý rằng chúng ta không thể chỉ lấy trung bình (kỳ vọng) độ lệch giữa các giá trị thực tế và trung bình (kỳ vọng). Bởi vì, giá trị trung bình (kỳ vọng) của độ lệch cũng chính là chênh lệch giữa các giá trị trung bình (kỳ vọng), nghĩa là ***E[X - E[X]] = E[X] - E[E[X]] = 0***. Tuy nhiên, chúng ta có thể xem xét giá trị trung bình (kỳ vọng) của bất kỳ hàm số (không âm) của chênh lệch này. Phương sai, ***Variance***, của một biến ngẫu nhiên được tính bằng cách xem xét giá trị trung bình (kỳ vọng) của các bình phương độ lệch:

$$(9)\ \sigma^2 = \mathrm{Var}[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

Trong đó: $(X - E[X])^2 = X^2 - 2.X.E[X] + (E[X])^2$

Phương sai của đại lượng ngẫu nhiên X = f(x) (dạng hàm số) được xác định bởi công thức sau:

$$(10)\ \sigma^2 = \mathrm{Var}_{x\sim p}[f(x)] = E_{x\sim p}[f^2(x)] - (E[f(x)])^2$$

Căn bậc hai của phương sai là một đại lượng hữu ích khác được gọi là độ lệch chuẩn, ***standard deviation*** (***SD***), ta có công thức tính như sau:

$$(11)\ \sigma = SD[X] = \sqrt{\sigma^2}$$

Cả hai đại lượng độ lệch chuẩn và phương sai đều truyền tải cùng một thông tin và có thể được tính toán dựa vào giá trị của nhau, nhưng độ lệch chuẩn có đặc tính tuyệt vời là được thể hiện giống nhau với các đơn vị tượng tự như đại lượng ban đầu được đại diện bởi biến ngẫu nhiên.

Ví dụ 6: Từ Ví dụ 4, chúng ta tính được kết quả về lợi nhuận kỳ vọng hay giá trị trung bình là E[X] = 1.8. Từ đó, chúng ta có thể tính toán phương sai của khoản đầu tư, như sau:

$\sigma^2 = E[X^2] - (E[X])^2 = 0.50*0^2 + 0.4*2^2 + 0.10*10^2 - 1.8^2 = 8.36$

Và độ lệch chuẩn là: $SD[X] = \sigma = \sqrt{\sigma^2} = \sqrt{8.36} \approx 2.8914$

Với mục đích đầu tư là lợi nhuận thì đây là tập danh mục đầu tư rủi ro. Lưu ý rằng theo quy ước toán học, kỳ vọng (giá trị trung bình) và phương sai, ký hiệu là μ và σ^2, thường được sử dụng làm tham số trong phân phối chuẩn (Gaussian). Tương tự với phương pháp tiếp cận kỳ vọng và phương sai cho các biến ngẫu nhiên vô hướng, chúng ta có thể nghiên cứu tương tự đối với các biến có giá trị dạng có hướng (vectơ). Đối với tham số kỳ vọng, chúng ta có thể tính giá trị trung bình, ***mean***, của vectơ μ, như sau:

$$(12)\ \mu \stackrel{\text{def}}{=} E_{z\sim P}[Z] \text{ với } \mu_i = E_{z\sim P}[z_i]$$

Trong đó, μ_i là giá trị trung bình của biến ngẫu nhiên z_i và Z là một vectơ trong không gian n-chiều của các biến ngẫu nhiên z_i (i=1,..,n), $Z=[z_1,z_2,...,z_n]^T$. Khi đó công thức tổng quát để tính ma trận hiệp phương sai, ***covariance matrix***, ký hiệu là Σ, có dạng:

$$(13)\ \Sigma \stackrel{\text{def}}{=} Cov_{z\sim P}[Z] = E_{z\sim P}[(Z-\mu)(Z-\mu)^T]$$

Trong đó, (Z - μ) là phép tập trung dữ liệu, kết quả là một vectơ tập trung, ***zero-mean (centered) vector***. $(Z-\mu)(Z-\mu)^T$ phép nhân với vectơ bên ngoài của

vectơ trung tâm kết quả là ma trận có đường chéo chính là phương sai riêng lẻ và đường chéo còn lại là hiệp phương sai.[15] Cụ thể, các phần tử trong ma trận được tính như sau:

$$\Sigma_{ii} = \mathrm{E}[(z_i - \mu_i)^2] = \mathrm{Var}(z_i)$$

$$\Sigma_{ij} = \mathrm{E}[(z_i - \mu_i)(z_j - \mu_j)] = \mathrm{Cov}(z_i, z_j)$$

Ý nghĩa: Ma trận hiệp phương sai định lượng mức độ biến thiên cùng nhau của các thành phần trong Z:

- Giá trị đường chéo chính cho thấy độ lớn của phương sai của từng thành phần riêng lẻ.
- Giá trị ngoài đường chính chéo cho thấy mức độ ảnh hưởng của mối quan hệ tuyến tính mạnh giữa các thành phần.

Ứng dụng của ma trận hiệp phương sai cho phép chúng ta tính toán phương sai cho bất kỳ hàm tuyến tính Z bằng phép nhân ma trận đơn giản.

Ví dụ 7: Tính ma trận hiệp phương sai cho biến ngẫu nhiên hai chiều $Z=[z_1,z_2]$:

$$\Sigma = \begin{bmatrix} Var(z_1) & Cov(z_1, z_2) \\ Cov(z_2, z_1) & Var(z_2) \end{bmatrix}$$

Chúng ta có thể tăng hiệu quả và đơn giản hoá việc tính toán ma trận hiệp phương sai bằng cách áp dụng tính chất phép nhân ma trận và ma trận chuyển vị[16] thông qua một số vec tơ ánh xạ, ***projection vector***, là một vec tơ đơn vị (cố định, không ngẫu nhiên) V có cùng kích thước với Z, khi đó:

$$(14)\ V^T \Sigma V = E_{z\sim P}[V^T(Z-\mu)(Z-\mu)^T V] = Var_{z\sim P}[V^T Z]$$

Một số ứng dụng của phương pháp này:

- Phương pháp phân tích thành phần nguyên lý, ***Principal Component Analysis*** (***PCA***): thành phần nguyên lý đầu tiên là vec tơ đơn vị V để tính giá trị cực đại của phương sai được ánh xạ, $V^T\Sigma V$.

[15] Phụ lục 1 - Lý thuyết bổ sung.
[16] Phụ lục 1 - Lý thuyết bổ sung.

- Phân tích rủi ro trong tài chính, nếu như Z đại diện cho lợi nhuận của tài sản và $V^T\Sigma V$ thể hiện rủi ro danh mục đầu tư (dạng phương sai) với trọng số V.
- Phân phối chuẩn (Gaussian): Nếu $Z \sim N(\mu,\Sigma)$ thì $V^TZ \sim N(V^T\mu, V^T\Sigma V)$.

8. Vấn đề về sự không chắc chắn:

Trong máy học, một số vấn đề không chắc chắn như không chắc chắn về giá trị của nhãn từ một yếu tố đầu vào, không chắc chắn về giá trị ước tính của một tham số, hay thậm chí không chắc chắn về việc liệu dữ liệu dùng để triển khai có cùng phân phối với dữ liệu đã đào tạo. Sự không chắc chắn ngẫu nhiên, ***aleatoric uncertainty***, là sự không chắc chắn vốn có do tính ngẫu nhiên thực sự không được giải thích bởi các biến quan sát. Sự không chắc chắn về nhận thức, ***epistemic uncertainty***, là sự không chắc chắn về việc giảm giảm bớt tham số của mô hình bằng cách thu thập thêm dữ liệu.

9. Bài tập nghiên cứu tình huống:

Bài tập 1:

Hãy cho một ví dụ về việc quan sát nhiều dữ liệu hơn có thể giảm mức độ không chắc chắn về kết quả đầu ra xuống mức thấp tùy ý.

Lời giải cho bài tập 1:

Cho một ví dụ về việc giảm độ không chắc chắn về nhận thức, ***epistemic uncertainty***, bằng việc quan sát thêm dữ liệu.

1. Tình huống nghiên cứu:

Dự đoán xác suất mặt ngửa của một đồng xu không rõ tính đồng nhất hay sự cân đối.

1.1. Khởi đầu với ít dữ liệu được quan sát:

- Phép thử với một đồng xu lạ và không rõ sự cân đối.
- Thực hiện phép thử tung đồng xu 5 lần, kết quả thu được 4 lần xuất hiện mặt ngửa (head), và 1 lần xuất hiện mặt sấp (tail).
- Độ không chắc chắn cao vì:

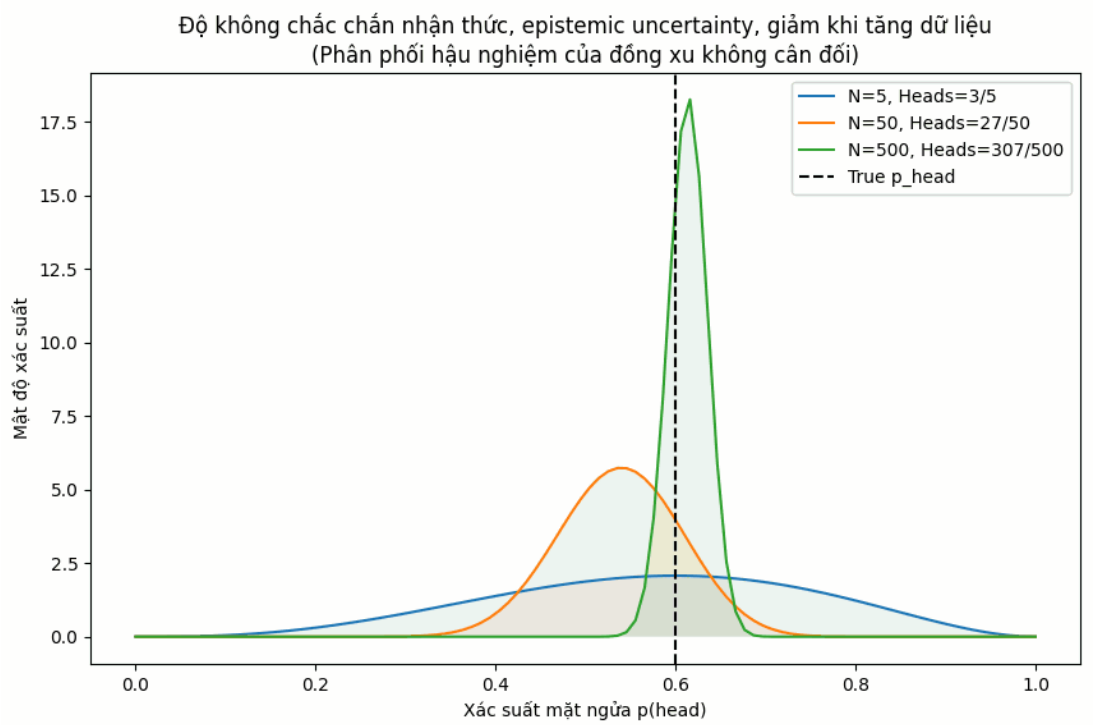
- Khả năng độ chệch của đồng xu.
- Số lần thử quá ít (5 lần) để kết luận.
- Mô hình dự đoán: Xác suất mặt ngửa = 80% ± 30% (độ tin cậy thấp).

1.2. Sau khi quan sát thêm dữ liệu:

- Thực hiện phép thử tung đồng xu 1000 lần, kết quả thu được 600 lần xuất hiện mặt ngửa và 400 lần xuất hiện mặt sấp.
- Độ không chắc chắn giảm mạnh:
  - Ước lượng xác suất mặt ngửa ≈ 60% ± 2% (độ tin cậy cao).
  - Có thể kết luận đồng xu có độ chệch về mặt ngửa với độ chắc chắn >95%.

1.3. Khi dữ liệu tiến tới vô hạn:

- Thực hiện phép thử tung đồng xu 1000000 lần, kết quả thu được 510000 lần xuất hiện mặt ngửa và 490000 lần xuất hiện mặt sấp.
- Độ không chắc chắn gần như bằng 0:
  - Xác suất hội tụ về 51% ± 0.1%.
  - Đồng xu gần như cân đối với độ lệch rất nhỏ.



Đồ thị 2 - Phân phối xác suất hậu nghiệm của đồng xu không cân đối.

2. Nhận xét:

Đây là một trường hợp ví dụ cho sự không chắc chắn về nhận thức, ***epistemic uncertainty***, bởi vì:

- Nguồn gốc của mức độ không chắc chắn là do thiếu dữ liệu ban đầu (không biết sự đồng chất hay cân đối của đồng xu.
- Phương pháp giảm thiểu sự không chắc chắn là thu thập thêm dữ liệu (càng tăng dữ liệu quan sát, ước lượng càng chính xác).
- Khác biệt với sự không chắc chắn ngẫu nhiên, ***aleatoric uncertainty***: Nếu đồng xu không đồng chất hay bị mất cân đối (nhiễu hệ thống) thì việc tăng số phép thử tung nhiều lần vẫn sẽ làm xác suất bị lệch. Trong trường hợp này, sự không chắc chắn không thể giảm bằng việc tăng dữ liệu quan sát là bản chất của sự không chắc chắn ngẫu nhiên, ***aleatoric uncertainty***.

3. Ứng dụng trong máy học:

Huấn luyện mô hình phân loại hình ảnh, ***image classification***:

- Dữ liệu huấn luyện ban đầu là 10 bức ảnh gồm cả chó và mèo, kết quả mô hình không phân biệt tốt (mức độ không chắc chắn cao).
- Sau khi huấn luyện với 10000 bức ảnh gồm cả chó và mèo, kết quả mô hình dự đoán chính xác 99%.

Lưu ý: Nếu dữ liệu có nhiễu, chẳng hạn như ảnh bị mờ, không rõ nét, độ không chắc chắn ngẫu nhiên, ***aleatoric uncertainty***, sẽ vẫn tồn tại.

4. Kết luận:

Sự không chắc chắn về nhận thức, ***epistemic uncertainty***, có thể giảm tuỳ ý nếu tăng dữ liệu quan sát với điều kiện là:

- Dữ liệu mới phải có tính đại diện cho yêu cầu của bài toán.
- Độ nhiễu trong dữ liệu, ***aleatoric***, không chiếm ưu thế hoặc đa số.

Bài tập 2:

Hãy cho ví dụ và giải thích rằng việc quan sát nhiều dữ liệu hơn sẽ chỉ làm giảm lượng bất định đến một điểm và sau đó sẽ dừng. Tìm điểm kỳ vọng có thể xảy ra này.

Lời giải cho bài tập 2:

Ví dụ minh họa về đo nhiệt độ bằng cảm biến có nhiễu.

1. Tình huống nghiên cứu:

Dùng một cảm biến nhiệt độ, ***thermometer***, để đo nhiệt độ phòng. Đặc điểm của cảm biến này:

- Nhiễu hệ thống, ***aleatoric uncertainty***, có sai số là ±0.50°C do hạn chế về kỹ thuật phần cứng của nhà sản xuất.
- Thiếu hiểu biết ban đầu, ***epistemic uncertainty***, về nhiệt độ thực tế.

2. Quá trình thu thập dữ liệu:

- Kết quả đo lần đầu với 1 mẫu thử là 25.30°C. Độ không chắc chắn cao bởi vì nhiệt độ thực có thể là 25.30 ± 2.00°C do nội hàm ẩn chứa cả độ không chắc chắn về nhận thức ngẫu nhiên, ***aleatoric uncertainty***, và độ không chắc chắn về nhận thức, ***epistemic uncertainty***.
- Kết quả đo 10 lần tiếp theo lần lượt là [25.30, 24.80, 25.60, 25.10,...] với giá trị trung bình là 25.20°C. Độ không chắc chắn về nhận thức, ***epistemic uncertainty***, giảm với ước lượng nhiệt độ thực khoảng 25.2 ± 0.3°C.
- Kết quả đo 1000 lần tiếp theo vẫn đạt giá trị trung bình xấp xỉ ~25.20°C, nhưng độ không chắc chắn không giảm thêm. Giá trị trung bình cuối cùng vẫn là 25.20 bởi sai số do nhiễu phần cứng là ±0.5°C.

3. Giải thích hiện tượng:

- Trong giai đoạn đầu, độ không chắc chắn về nhận thức, ***epistemic uncertainty***, do thiếu dữ liệu giảm nhanh khi tăng số lần đo.

- Điểm hội tụ sau khoảng 50~100 lần đo được hình thành nhưng sai số chỉ còn phụ thuộc vào độ không chắc chắn ngẫu nhiên, ***aleatoric uncertainty***, là ±0.5°C và không thể giảm thêm dù tăng số lần đo.

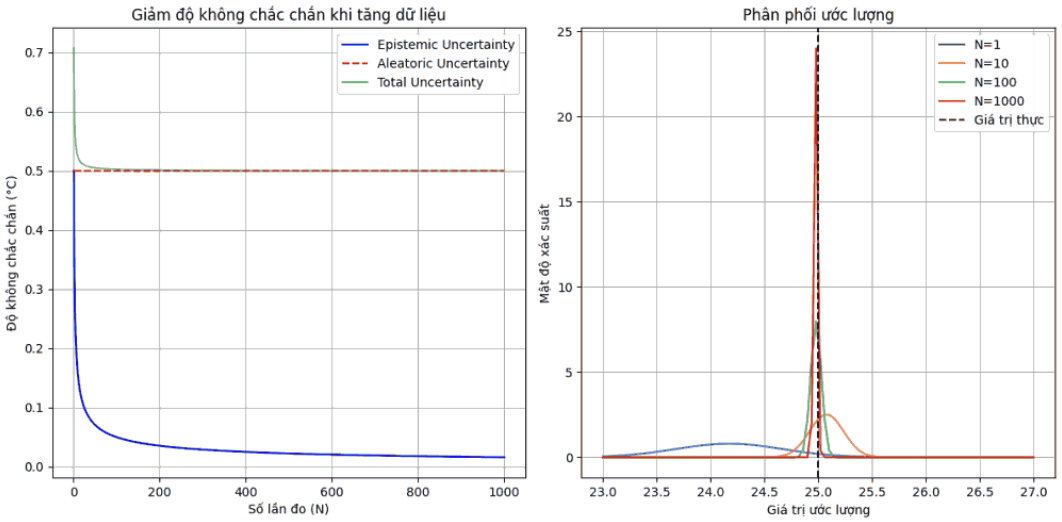
4. Phân tích nguyên nhân:

Độ không chắc chắn ngẫu nhiên, ***aleatoric uncertainty***, là nhiễu ngẫu nhiên trong hệ thống, chẳng hạn như giới hạn độ chính xác của cảm biến. Khác với độ không chắc chắn về nhận thức, ***epistemic uncertainty***, vốn giảm được bằng độ lớn dữ liệu, độ không chắc chắn ngẫu nhiên, ***aleatoric uncertainty***, là đặc tính vật lý của hệ thống.

5. Biểu diễn kết quả mô phỏngbằng đồ thị:

Kết quả tính toán bằng mô phỏng như sau:

```
N=1: Mean=24.96°C, Epistemic uncert.=0.5000°C, Total uncert.=0.7071°C
N=5: Mean=24.98°C, Epistemic uncert.=0.2236°C, Total uncert.=0.5477°C
N=10: Mean=25.06°C, Epistemic uncert.=0.1581°C, Total uncert.=0.5244°C
N=50: Mean=25.07°C, Epistemic uncert.=0.0707°C, Total uncert.=0.5050°C
N=100: Mean=25.08°C, Epistemic uncert.=0.0500°C, Total uncert.=0.5025°C
N=1000: Mean=24.99°C, Epistemic uncert.=0.0158°C, Total uncert.=0.5002°C
```



Đồ thị 3 - Phân phối ước lượng và độ không chắc chắn.

Ghi chú:

- Đường màu xanh biển là độ không chắc chắn về nhận thức, epistemic uncertainty, giảm theo hàm số $1/\sqrt{N}$.
- Đường màu đỏ là độ không chắc chắn ngẫu nhiên, aleatoric uncertainty, là hằng số 0.5°C đại diện cho giới hạn dưới.

6. Ứng dụng trong thực tế:

- Lĩnh vực y tế: xét nghiệm máu có sai số ±5% do thiết bị, nếu tăng số lần xét nghiệm lên 100 lần thì kết quả vẫn dao động trong khoảng này.
- Lĩnh vực máy học: dữ liệu hình ảnh bị mờ thì huấn luyện thêm cũng không cải thiện độ chính xác vượt ngưỡng nhất định.

7. Kết luận:

- Độ không chắc chắn về nhận thức, ***epistemic uncertainty***, được giảm bằng độ lớn dữ liệu.
- Độ không chắc chắn ngẫu nhiên, ***aleatoric uncertainty***, là hạn chế về vật lý, chỉ giảm được khi cải thiện hệ thống như dùng cảm biến có sai số thấp hơn.

Bài tập 3:

Với thử nghiệm tung đồng xu, hãy tính toán phương sai của ước tính xác suất chúng ta nhìn thấy mặt ngửa sau n lần thử.

1. Nhận xét về mối quan hệ tỷ lệ giữa phương sai với số lượng quan sát.
2. Sử dụng bất đẳng thức Chebyshev để giới hạn độ lệch chuẩn từ kỳ vọng.
3. Giải thích mối liên hệ với định lý giới hạn trung tâm (nếu có).

Lời giải cho bài tập 3:

Xét thử nghiệm tung đồng xu cân đối, ***fair coin***, với xác suất mặt ngửa (heads) là p = 0.50 và mặt sấp (tails) cũng là q = (1 - p) = 0.50. Gọi đại lượng ngẫu nhiên X là số lần xuất hiện mặt ngửa sau n lần thử và có phân phối nhị thức, X~Binomial(n, p).

1. Tính phương sai của ước tính xác suất mặt ngửa và tỷ lệ với số lượng quan sát.

Ước lượng xác suất mặt ngửa là tỉ lệ $\hat{p} = \frac{X}{n}$.

Kỳ vọng của $\hat{p}$: $\mu = E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{E[X]}{n} = \frac{np}{n} = p = 0.50$

Phương sai của $\hat{p}$: $\sigma^2 = Var[\hat{p}] = Var\left[\frac{X}{n}\right] = \frac{Var[X]}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} = \frac{0.25}{n}$

Từ kết quả tính phương sai, ta nhận thấy phương sai tỷ lệ nghịch với số lượng quan sát n hay $\sigma^2 = Var[\hat{p}] = O\left(\frac{1}{n}\right)$.

Độ lệch chuẩn, ***standard deviation*** (***SD***): $\sigma = \sqrt{Var[\hat{p}]} = \frac{0.50}{\sqrt{n}}$

2. Sử dụng bất đẳng thức Chebyshev để giới hạn độ lệch từ kỳ vọng:

Bất đẳng thức Chebyshev đưa ra đánh giá xác suất $\hat{p}$ lệch khỏi giá trị kỳ vọng $\mu = p$ với mọi $\varepsilon > 0$:

$$P(|\hat{p} - \mu| \geq \varepsilon) = P(|\hat{p} - p| \geq \varepsilon) \leq \frac{Var[\hat{p}]}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}$$

Với đồng xu cân đối:

$$P(|\hat{p} - 0.50| \geq \varepsilon) \leq \frac{0.25}{n\varepsilon^2}$$

Nhận xét: Khi tăng n thì chúng ta có thể làm giảm độ lệch giữa $\hat{p}$ so với kỳ vọng $\mu = p$ một lượng ε.

Ví dụ: $\varepsilon = 0.10$, ta được:

$$P(|\hat{p} - 0.50| \geq 0.10) \leq \frac{0.25}{n(0.10)^2} \text{ <=> } n \geq 500$$

Cần ít nhất 500 lần thử để đảm bảo $\hat{p}$ nằm trong khoảng [0.40,0.6] với độ tin cậy là 95%.

3. Mối liên hệ với định lý giới hạn trung tâm:

Định lý giới hạn trung tâm, ***central limit theorems*** (***CLT***), phát biểu rằng khi n lớn, phân phối của $\hat{p}$ sẽ tiệm cận phân phối chuẩn $\hat{p} \sim N(p, \frac{p(1-p)}{n})$.

Chuẩn hoá: $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} N(0,1)$

Ứng dụng:

- Khoảng tin cậy gần đúng: Với n lớn, xác suất để $\hat{p}$ nằm trong khoảng:

$$p \pm z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

xấp xỉ 1 - α. Ví dụ: $z_{0.025}$ = 1.96 cho khoảng 95%.

- Nếu so sánh với bất đẳng thức Chebyshev thì bất đẳng thức Chebyshev cho cận tổng quát nhưng $\hat{p}$ thường nằm trong khoảng rộng hơn trong khi CLT cho xấp xỉ chính xác hơn khi nhờ phân phối chuẩn với n lớn.
- Giải thích sự hội tụ khi n→∞ thì đồ thị tần suất của $\hat{p}$ sẽ có dạng hình chuông tập trung quanh p.

Bài tập 4:

Giả định lấy mẫu x_i từ m có phân phối xác suất với giá trị trung bình, ***mean***, bằng 0 và phương sai đơn vị. Tính trung bình của $z_m \overset{\text{def}}{=} \frac{1}{m}\sum_{i=1}^{m} x_i$ bằng cách áp dụng bất đẳng thức Chebyshev với mọi z_m độc lập (nếu có thể) và giải thích.

Lời giải cho bài tập 4:

1. Các tham số của x_i:

Kỳ vọng: $E[x_i] = 0$. Phương sai: $Var(x_i) = 1$.

2. Các tham số của z_m:

Kỳ vọng: $E(z_m) = \mathrm{E}(\frac{1}{m}\sum_{i=1}^{m} x_i) = \frac{1}{m}\sum_{i=1}^{m} E(x_i) = 0$

Phương sai: $Var(z_m) = \mathrm{Var}(\frac{1}{m}\sum_{i=1}^{m} x_i) = \frac{1}{m^2}\sum_{i=1}^{m} Var(x_i) = \frac{1}{m^2}.m = \frac{1}{m}$

3. Áp dụng bất đẳng thức Chebyshev:

Bởi vì $E(z_m)$ và $Var(z_m)$ tồn tại hữu hạn, áp dụng bất đẳng thức Chebyshev:

$$P(|z_m - E(z_m)| \geq \varepsilon) \leq \frac{Var[z_m]}{\varepsilon^2}$$

Thay kết quả đã tính toán ở phần trên ta được:

$$P(|z_m| \geq \varepsilon) \leq \frac{1}{m\varepsilon^2}$$

4. Nhận xét:

- Bất đẳng thức Chebyshev có thể áp dụng cho từng z_m riêng lẻ nếu các x_i độc lập và có cùng phân phối.
- Tuy nhiên, nếu các z_m được tính từ cùng một tập mẫu {x1,…,xn}, chẳng hạn như $\{z_2 = (x_1 + x_2) / 2, z_3 = (x_1 + x_2 + x_3) / 3,...\}$ thì các biến z_m không độc lập với nhau vì cùng chia sẻ chung các giá trị của biến x_i.

5. Khả năng áp dụng Chebyshev độc lập cho mọi z_m:

- Nếu các biến z_m được tính từ cùng một tập mẫu và có mối quan hệ phụ thuộc lẫn nhau, chẳng hạn như z_2 và z_3 đều chứa x_1,x_2; đồng thời, sự sai lệch của z_2 so với 0 có thể ảnh hưởng đến z_3.
- Bất đẳng thức Chebyshev chỉ đảm bảo ràng buộc xác suất cho từng biến ngẫu nhiên riêng lẻ, không xét đến mối tương quan giữa chúng. Nếu muốn đánh giá đồng thời nhiều z_m, cần sử dụng công cụ khác, chẳng hạn như union bound kết hợp với Chebyshev.

Bài tập 5:

Cho hai biến cố A, B có xác suất P(A) và P(B), tìm giới hạn trên và dưới của P(A∪B) và P(A∩B). Minh hoạ bằng biểu đồ Venn.

Lời giải cho bài tập 5:

1. Tìm cận trên và cận dưới của P(A ∪B):

Ứng dụng công thức cộng xác suất cơ bản, ta được:

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

Biểu đồ Venn:

- Hình tròn A và B giao nhau tại A ∩ B.
- P(A∪B) là tổng diện tích của A và B trừ phần giao.

Tìm cận dưới:

- P(A∪B) đạt giá trị nhỏ nhất khi A và B giao nhau nhiều nhất, nghĩa là A∩B lớn nhất.

- Giá trị lớn nhất của P(A∩B) là min[P(A),P(B)].

$$P(A \cup B) \geq \max[P(A), P(B)]$$

Bởi vì, nếu $A \subseteq B$ thì P(A∩B) = P(B) và ngược lại.

Tìm cận trên: P(A∪B) đạt giá trị lớn nhất khi A và B rời nhau (không giao).

$$P(A \cup B) \leq P(A) + P(B)$$

Bởi vì, nếu $A \cap B = \varnothing$ thì P(A∪B) = P(A) + P(B).

Kết luận:

$$\text{Max}[P(A),P(B)] \leq P(A \cup B) \leq P(A) + P(B)$$

2. Tìm cận trên và cận dưới của P(A ∩ B):

Từ công thức cộng xác suất cơ bản, ta được:

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

Tìm cận dưới: P(A∩B) đạt giá trị nhỏ nhất khi A và B rời nhau (không giao).

$$P(A \cap B) \geq 0$$

Bởi vì, nếu $A \cap B = \varnothing$ thì P(A∩B) = 0.

Tìm cận trên:

- P(A∩B) đạt giá trị lớn nhất khi A và B giao nhau nhiều nhất.
- Giá trị lớn nhất của P(A∩B) là min[P(A),P(B)].

$$P(A \cap B) \leq \min[P(A),P(B)]$$

Bởi vì, nếu $A \subseteq B$ thì P(A∩B) = P(A).

Kết luận:

$$0 \leq P(A \cap B) \leq \min[P(A),P(B)]$$

3. Minh họa bằng biểu đồ Venn:

- Trường hợp rời nhau: A và B không chồng lấn thì P(A∩B) = 0 nên:

$$P(A \cup B) = P(A) + P(B)$$

- Trường hợp giao nhau tối đa: A nằm hoàn toàn trong B thì P(A∩B) = P(A):

$$P(A \cup B) = P(B)$$

4. *Ứng dụng:*

Các cận này hữu ích khi không biết mối quan hệ giữa A và B, nhưng biết P(A) và P(B).

5. *Ví dụ cụ thể:*

Nếu P(A) = 0.60 và P(B) = 0.40 thì:

$$0.60 \leq P(A \cup B) \leq 1.00$$

$$0.00 \leq P(A \cap B) \leq 0.40$$

Biểu đồ 1 - Biểu đồ Venn biểu diễn mối quan hệ giữa biến cố A và B.

Bài tập 6:

Giả sử chúng ta có một chuỗi Markov gồm các biến ngẫu nhiên A, B, C với B chỉ phụ thuộc vào A và C chỉ phụ thuộc vào B, hãy đơn giản hóa xác suất kết hợp P(A,B,C).

Lời giải cho bài tập 6:

1. Phân tích mối quan hệ phụ thuộc:

Theo đề bài, các biến ngẫu nhiên có mối quan hệ phụ thuộc như sau:

- B chỉ phụ thuộc vào A: P(B|A,C) = P(B|A)
- C chỉ phụ thuộc vào B: P(C|A,B) = P(C|B

Đây chính là tính chất Markov của chuỗi $A \rightarrow B \rightarrow C$, nghĩa là, tương lai C chỉ phụ thuộc vào hiện tại B, không phụ thuộc vào quá khứ A.

2. Đơn giản hóa P(A, B,C):

Xác suất đồng thời có thể phân tích thành:

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

Theo tính chất Markov, P(C|A,B) = P(C|B), nên:

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

Kết quả:

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

3. Minh họa bằng sơ đồ:

Sơ đồ phụ thuộc: $A \rightarrow B \rightarrow C$

Giải thích:

- A ảnh hưởng đến B, B ảnh hưởng đến C.
- C không phụ thuộc trực tiếp vào A khi biết B.

4. Ví dụ minh hoạ:

Giả sử:

- A: Thời tiết hôm nay (Nắng/Mưa).
- B: Số lượng người đi công viên, phụ thuộc vào thời tiết.

- C: Doanh thu bán kem, phụ thuộc vào số lượng người đi công viên.

Xác suất đồng thời:

P(Nắng,Đông người,Doanh thu cao) = P(Nắng).P(Đông người|Nắng).P(Doanh thu cao|Đông người)

5. Kiểm tra tính chất Markov:

Nếu C phụ thuộc cả vào A và B, chuỗi không còn là Markov. Trong trường hợp này, C chỉ phụ thuộc B, nên: P(C|A,B) = P(C|B)

6. Kết luận:

Công thức tổng quát cho chuỗi Markov $A \rightarrow B \rightarrow C$:

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

7. Ứng dụng:

- Mô hình hóa các hệ thống có phụ thuộc bậc thang như tín hiệu truyền qua nhiều bước....
- Đơn giản hóa tính toán xác suất trong mạng Bayesian.

8. Ví dụ mô phỏng:

Để kiểm tra, có thể mô phỏng bằng Python với các giá trị xác suất cụ thể:

Cho P(A) = 0.60, P(B|A) = 0.70, P(C|B) = 0.80 ta tính được P(A,B,C) = 0.336.

Bài tập 7:

Từ ví dụ minh hoạ tại mục 6, giả định rằng các kết quả đầu ra của 2 lần xét nghiệm không độc lập. Đồng thời, tỷ lệ dương tính giả 0.10 và âm tính giả là 0.01. Nghĩa là P(D = 1|C = 0) = 0.1 và P(D = 0|C = 1) = 0.01. Giả định trường hợp nhiễm bệnh C = 1, kết quả xét nghiệm chỉ độc lập có điều kiện.

1. Tính bảng xác suất chung cho D_1 và D_2, với C = 0 dựa trên thông tin đã có cho đến nay.
2. Tính xác suất bệnh nhân bị bệnh (C = 1) sau khi một xét nghiệm cho kết quả dương tính. Giả định xác suất cơ sở P(C = 1) = 0.0015.

3. Tính xác suất bệnh nhân bị bệnh (C = 1) sau khi cả hai xét nghiệm đều có kết quả dương tính.

Lời giải cho bài tập 7:

Tổng hợp giả thiết:

Xác suất tiên nghiệm người đó bị nhiễm bệnh: $P(C = 1) = 0.0015$

Xác suất người đó không bị nhiễm bệnh: $P(C = 0) = 1 - 0.0015 = 0.9985$

Đặc tính của mỗi xét nghiệm (D1 và D2 có cùng đặc tính):

- Tỷ lệ dương tính giả: $P(D = 1|C = 0) = 0.10$
- Tỷ lệ âm tính giả: $P(D = 0|C = 1) = 0.01$
- Từ tỷ lệ âm tính giả, suy ra tỷ lệ dương tính thật:

$$P(D = 1|C = 1) = 1 - P(D = 0|C = 1) = 1 - 0.01 = 0.99$$

Kết quả của hai lần xét nghiệm (D_1,D_2) được giả định là không độc lập một cách vô điều kiện.

Khi một người thực sự nhiễm bệnh (C = 1), kết quả của hai xét nghiệm là độc lập có điều kiện: $P(D_1,D_2|C = 1) = P(D_1|C = 1)P(D_2|C = 1)$

Đối với trường hợp một người không nhiễm bệnh (C = 0), để giải quyết phần 1 của bài toán, chúng ta sẽ giả định rằng kết quả của hai xét nghiệm cũng độc lập có điều kiện: $P(D1,D2|C = 0) = P(D1|C = 0)P(D2|C = 0)$

Giả định này là cần thiết vì không có thông tin nào khác về sự phụ thuộc có điều kiện khi C = 0 nào được cung cấp; đồng thời, nhất quán với việc các đặc tính xét nghiệm riêng lẻ được áp dụng cho cả hai xét nghiệm.

1. Lập bảng xác suất chung cho D_1 và D_2 khi C = 0:

Với giả định D_1 và D_2 độc lập có điều kiện khi C = 0, chúng ta tính được:

$P(D_1 = 1|C = 0) = 0.10 \Longrightarrow P(D_1 = 0|C = 0) = 1 - 0.10 = 0.90$

$P(D_2 = 1|C = 0) = 0.10 \Longrightarrow P(D_2 = 0|C = 0) = 1 - 0.10 = 0.90$

Các xác suất kết hợp có điều kiện là:

$P(D_1 = 0,D_2 = 0|C = 0) = P(D_1 = 0|C = 0)*P(D_2 = 0|C = 0) = 0.90*0.90 = 0.81$

$P(D_1 = 0,D_2 = 1|C = 0) = P(D_1 = 0|C = 0)*P(D_2 = 1|C = 0) = 0.90*0.10 = 0.09$

$P(D_1 = 1,D_2 = 0|C = 0) = P(D_1 = 1|C = 0)*P(D_2 = 0|C = 0) = 0.10*0.90 = 0.09$

$P(D_1 = 1,D_2 = 1|C = 0) = P(D_1 = 1|C = 0)*P(D_2 = 1|C = 0) = 0.10*0.10 = 0.01$

Từ các kết quả tính toán trên chúng ta lập bảng xác suất chung $P(D_1,D_2|C = 0)$:

	$D_2 = 0$	***$D_2 = 1$***	***Tổng (D_1)***
$D_1 = 0$	0.81	0.09	***0.90***
$D_1 = 1$	0.09	0.01	***0.10***
Tổng (D_2)	***0.90***	***0.10***	***1.00***

2. Xác suất bệnh nhân bị bệnh (C = 1) sau khi một xét nghiệm cho kết quả dương tính.

Sử dụng công thức Bayes:

$P(C = 1|D_1 = 1) = P(D_1 = 1)*P(D_1 = 1|C = 1)*P(C = 1)$

Trước hết, tính $P(D_1 = 1)$ (xác suất biên của một xét nghiệm dương tính):

$P(D_1 = 1) = P(D_1 = 1|C = 1)*P(C = 1) + P(D_1 = 1|C = 0)*P(C = 0)$

$P(D_1 = 1) = (0.99*0.0015) + (0.10*0.9985) = 0.101335$

Bây giờ, tính $P(C = 1|D_1 = 1)$:

$P(C = 1|D_1 = 1) = 0.101335*0.99*0.0015 \approx 0.01465$

Vậy, xác suất bệnh nhân bị bệnh sau khi một xét nghiệm cho kết quả dương tính là khoảng 1.465%.

3. Xác suất bệnh nhân bị bệnh (C = 1) sau khi cả hai xét nghiệm đều có kết quả dương tính.

Sử dụng công thức Bayes:

$P(C = 1|D_1 = 1,D_2 = 1) = P(D_1 = 1,D_2 = 1)*P(D_1 = 1,D_2 = 1|C = 1)*P(C = 1)$

Tính các thành phần cần thiết:

Do D_1, D_2 độc lập có điều kiện khi C = 1 nên:

$P(D_1 = 1,D_2 = 1|C = 1) = P(D_1 = 1|C = 1)*P(D_2 = 1|C = 1) = 0.99*0.99 = 0.9801$

Từ các giả định độc lập có điều kiện khi C = 0: $P(D_1 = 1,D_2 = 1|C = 0) = 0.01$

Xác suất biên của cả hai xét nghiệm đều dương tính:

$P(D_1=1,D_2=1) = P(D_1=1,D_2=1|C=1)*P(C=1) + P(D_1=1,D_2=1|C=0)*P(C=0)$

$P(D_1 = 1,D_2 = 1) = (0.9801*0.0015) + (0.01*0.9985) = 0.01145515$

Bây giờ, tính $P(C = 1|D_1 = 1,D_2 = 1)$:

$P(C = 1|D_1 = 1,D_2 = 1) = 0.01145515*0.9801*0.0015 \approx 0.12833$

Vậy, xác suất bệnh nhân bị bệnh sau khi cả hai xét nghiệm đều cho kết quả dương tính là khoảng 12.833%.

Bài tập 8:

Với vai trò là một nhà quản lý tài sản cho một ngân hàng đầu tư, hãy ra quyết định đầu tư vào các chứng khoán s_i. Danh mục đầu tư với các tỷ trọng α_i cho mỗi cổ phiếu phải có tổng bằng 1. Các cổ phiếu có lợi nhuận trung bình là $\mu=E_{s\sim P}[s]$ và ma trận hiệp phương sai $\Sigma=Cov_{s\sim P}[s]$.

1. Tính lợi nhuận kỳ vọng cho một danh mục đầu tư α đã cho trước.
2. Hãy lý giải sự lựa chọn đầu tư để tối đa hoá lợi nhuận của danh mục.
3. Tính phương sai của danh mục.
4. Xây dựng bài toán tối ưu hóa để tối đa hóa lợi nhuận trong khi vẫn giữ phương sai ở mức giới hạn trên. Ứng dụng phương pháp giải phương trình bậc hai.

Lời giải cho bài tập 8:

Giả định rằng danh mục đầu tư gồm hai loại cổ phiếu với các tham số đầu vào là $\alpha=[0.60,0.40]$, $\mu=[0.10,0.20]$ và $\Sigma = \begin{bmatrix} 0.04 & 0.02 \\ 0.02 & 0.09 \end{bmatrix}$

1. Tính lợi nhuận kỳ vọng của danh mục α:

Lợi nhuận kỳ vọng của danh mục là tổng trọng số của lợi nhuận từng cổ phiếu được xác định bởi công thức:

$$\mu = E[R_\alpha] = \sum_i \alpha_i \mu_i = \alpha^T \mu = 0.60 * 0.10 + 0.40 * 0.20 = 0.14$$

2. *Tối ưu hóa để tối đa hóa lợi nhuận:*

Chiến lược: Đầu tư toàn bộ vốn vào cổ phiếu có lợi nhuận trung bình cao nhất:

$$\alpha_i = \begin{cases} 1 \; nếu \; \mu_i = \max(\mu) \\ 0 \quad\quad ngược \; lại \end{cases}$$

Lưu ý: Chiến lược này bỏ qua rủi ro (phương sai) nên có thể dẫn đến thiệt hại (lỗ) lớn nếu cổ phiếu được chọn biến động (giảm) mạnh về giá.

3. *Tính phương sai (rủi ro) của danh mục:*

Phương sai danh mục phụ thuộc vào hiệp phương sai giữa các cổ phiếu:

$$\sigma^2 = Var[R_\alpha] = \sum_{i,j} \alpha_i \alpha_j \sum_{ij} \quad = \alpha^T \Sigma \alpha$$

=> Var(R_α) = 0.60^2 * 0.04 + 0.40^2 * 0.09 + 2 * 0.60 * 0.40 * 0.02 = 0.0384

4. *Bài toán tối ưu Markowitz (Nobel 1990):*

Tối đa hóa lợi nhuận với ràng buộc rủi ro tối đa:

Hàm mục tiêu:

$$\mu = E[\alpha_i] = \alpha^T \mu \rightarrow \max \text{ (cực đại)}$$

Các ràng buộc:

- Rủi ro tối đa chấp nhận được: $\alpha^T \Sigma \alpha \leq \max(\sigma^2)$
- Tỷ trọng danh mục: $\sum_i \alpha_i = 1$
- Không được bán khống: $\alpha_i \geq 0$

Mô phỏng bài toán: ta thu được kết quả:

```
Trọng số tối ưu: [0.37716097 0.62283903]
```

5. *Kết luận:*

- Lợi nhuận kỳ vọng: $\alpha^T \mu$.
- Tối đa hóa lợi nhuận: Đầu tư 100% vào cổ phiếu có μ cao nhất (rủi ro cao).
- Phương sai danh mục: $\alpha^T \Sigma \alpha$
- Tối ưu Markowitz: Cân bằng giữa lợi nhuận và rủi ro bằng các ràng buộc phương sai, cần giải đẳng thức bậc hai.

6. Ứng dụng thực tế:

- Các quỹ đầu tư cần xây dựng danh mục dựa trên rủi ro kỳ vọng.
- Tránh tập trung vào một cổ phiếu để giảm rủi ro hệ thống (đa dạng hoá danh mục đầu tư).

PHỤ LỤC 1

LÝ THUYẾT BỔ SUNG

1. Phép thử hay thí nghiệm:[17]

1.1. Định nghĩa:

Trong lý thuyết xác suất, một thí nghiệm, ***experiment***, là một quy trình hay một phép đo thu được một quan sát hoặc thu được một kết quả đầu ra, ***outcome***, không được dự đoán chắc chắn, ***certainty***.

1.2. Ví dụ minh hoạ:

Quan sát, ***observation***, hoặc phép đo, ***measurement***, được tạo ra bởi một thí nghiệm có thể hoặc không thể được định lượng bằng giá trị số học. Sau đây là một số ví dụ về các thí nghiệm:

- Ghi lại điểm kiểm tra;
- Đo lượng mưa hàng ngày;
- Phỏng vấn người dân để lấy ý kiến về một sắc lệnh;
- Kiểm tra sản phẩm được sản xuất ra để xác định xem đó có phải là sản phẩm lỗi hay sản phẩm chấp nhận được;
- Tung đồng xu và quan sát mặt xuất hiện.

2. Sự kiện (đơn giản):[18]

2.1. Định nghĩa:

Khi thực hiện một thí nghiệm, điều chúng ta quan sát được là một kết quả được gọi là sự kiện đơn giản, ***simple event***, ký hiệu là E_i.

2.2. Ví dụ minh hoạ:

Gọi E_i là sự kiện đơn giản xuất hiện mặt thứ i khi tung hột xí ngầu gồm 6 mặt, i=(1,...,6). Ta có: E_1 là sự kiện xuất hiện 1 mặt có 1 chấm,..., E_6 là sự kiện xuất hiện 1 mặt có 6 chấm.

---

[17] Mục 4.2. "Introduction to Probability and Statistics".
[18] Mục 4.2. "Introduction to Probability and Statistics".

3. Biến cố:[19]

3.1. Định nghĩa:

Biến cố, ***event***, là một tập hợp các sự kiện đơn giản.

3.2. Ví dụ minh hoạ:

Từ ví dụ 2.2 ta có các biến cố sau:

- A={E_1,E_3,E_5} là biến cố quan sát có một số lẻ.
- B={E_1,E_2,E_3} là biến cố quan sát có một số nhỏ hơn 4.
- C={E_2,E_4,E_6} là biến cố quan sát có một số chẵn.

4. Biến ngẫu nhiên:[20]

4.1. Định nghĩa:

Biến ngẫu nhiên, ***Random variable***, X có giá trị được giả định là thu được tương ứng với kết quả của một phép thử, là một cơ hội hoặc sự kiện ngẫu nhiên.

4.2. Ví dụ minh hoạ:

- X = Số lỗi trên một mẫu hàng nội thất được chọn ngẫu nhiên.
- X = Điểm số SAT cho một đơn ứng tuyển được chọn ngẫu nhiên.
- X = Số lượng cuộc gọi điện thoại được nhận bởi đường dây nóng trong khoảng thời gian được chọn ngẫu nhiên.

4.3. Phân loại:[21,22]

Một biến ngẫu nhiên rời rạc, ***Random discreted variable***, chỉ có thể nhận một giá trị (hay không gian của chính nó) là hữu hạn hoặc đếm được.

Một biến ngẫu nhiên liên tục, ***Random continuous variable***, có thể nhận vô số giá trị tương ứng với các điểm trên một khoảng tuyến tính.

5. Luật số lớn:[23]

5.1. Bất đẳng thức Markov và Chebyshev:

---

[19] Mục 4.2. "Introduction to Probability and Statistics".
[20] Mục 4.8. "Introduction to Probability and Statistics".
[21] Mục 1.2. "Introduction to Probability and Statistics".
[22] Mục 1.6.1. "Introduction to Mathematical Statictics".
[23] Mục 6.2. "Probability and Statistics".

Giả định rằng X là một biến ngẫu nhiên có xác suất $P(X \geq 0) = 1$. Tồn tại kỳ vọng E(X) và phương sai Var(X). Khi đó, với mọi số thực $\varepsilon > 0$ ta được:

5.1.1. Bất đẳng thức Markov:

$$P(X \geq \varepsilon) \leq E[X] / \varepsilon$$

5.1.2. Bất đẳng thức Chebyshev:

$$P(|X - E(X)| \geq \varepsilon) \leq Var(X) / \varepsilon^2$$

5.2. Luật số lớn:

5.2.1. Định nghĩa sự hội tụ trong xác suất:

Giả sử ta có dãy các biến ngẫu nhiên X_1, X_2,... hội tụ đến một số b cho trước, nếu phân phối xác suất của X_n ngày càng tập trung quanh b khi $n \to \infty$ (sự hội tụ trong xác suất, ***Convergence in Probability***). Với mọi $\varepsilon > 0$, ta được:

$$\lim_{n\to\infty} P(|X_n - b| < \varepsilon) = 1$$

Ký hiệu: $X_n \xrightarrow{p} b$

5.2.2. Định lý về Luật số lớn:

Giả sử X_1,..., X_n có dạng một mẫu ngẫu nhiên từ một phân phối có giá trị trung bình, ***Mean***, ký hiệu là μ, thì trung bình mẫu:

$$\bar{X}_n \xrightarrow{p} \mu$$

5.2.3. Định lý về Hàm liên tục của biến ngẫu nhiên:

Nếu $X_n \xrightarrow{p} b$ và g(x) là một hàm số liên tục khi x = b, thì $g(X_n) \xrightarrow{p} g(b)$

6. Định lý giới hạn trung tâm:[24]

6.1. Nội hàm của Định lý giới hạn trung tâm:

Nếu biến ngẫu nhiên X_1,...,X_n có dạng mẫu ngẫu nhiên có kích thước n từ một phân phối cho trước với giá trị trung bình (kỳ vọng), ***expected value***, là μ và phương sai, ***Variance***, là σ^2 ($0 < \sigma^2 < \infty$), thì với mỗi giá trị cố định x ta được:

[24] Mục 6.3. "Probability and Statistics".

$$\lim_{n\to\infty} P\left[\frac{(\bar{X}_n - \mu)\sqrt{n}}{\sigma} \leq x\right] = \Phi(x)$$

Trong đó, Φ(x) là hàm phân phối tích lũy, ***Cumulative distribution function*** (***CDF***), của phân phối chuẩn tắc, ***standard normal distribution***.

6.2. Tác động của Định lý giới hạn trung tâm.

Định lý giới hạn trung tâm đưa ra lời giải thích hợp lý cho thực tế là phân phối xác suất của nhiều biến ngẫu nhiên được nghiên cứu trong các thí nghiệm vật lý là gần đúng với phân phối chuẩn. Về tổng quát, định lý giới hạn trung tâm chỉ ra rằng phân phối xác suất của tổng nhiều biến ngẫu nhiên có thể gần đúng với phân phối chuẩn, mặc dù phân phối của từng biến ngẫu nhiên trong tổng khác với phân phối chuẩn.

7. Phân phối xác suất của hai biến ngẫu nhiên:[25]

7.1. Định nghĩa véc tơ ngẫu nhiên:

Cho một phép thử ngẫu nhiên với không gian mẫu S và hai biến ngẫu nhiên X_1, X_2, với mỗi phần tử $s \in S$ sao cho $X_1(s) = x_1$ và $X_2(s) = x_2$ khi đó (X_1,X_2) là một véc tơ ngẫu nghiên, ***random vertor***. Không gian của (X_1,X_2) là một tập:

$$D=\{(x_1, x_2) : x_1 = X_1(s), x_2 = X_2(s), s \in S\}$$

7.2. Phân phối xác suất:

Với hàm phân phối xác suất tích luỹ (CDF) có dạng:

$$F_{X_1,X_2}(x_1, x_2) = P[\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\}] \ \forall (x_1,x_2) \in R^2$$

Với các tập có dạng $(a_1,b_1]\times(a_2,b_2]$ thì CDF trở thành hàm hàm phân phối xác suất tích luỹ chung, ***joint cumulative distribution function***, của (X_1,X_2) có dạng:

$$P[a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2] = F_{X_1,X_2}(b_1, b_2) - F_{X_1,X_2}(a_1, b_2) - F_{X_1,X_2}(a_2, b_1) + F_{X_1,X_2}(a_1, a_2)$$

[25] Mục 2.1. "Introduction to Mathematical Statistics".

Trường hợp (X_1,X_2) là véc tơ ngẫu nhiên rời rạc, ***discrete random vector***, thì X_1, X_2 là biến ngẫu nhiên rời rạc và không gian S là hữu hạn hay đếm được. Và hàm xác suất chung, ***joint probability mass function***, được xác định:

$$p_{x_1,x_1}(x_1,x_2) = P[X_1 = x_1, X_2 = x_2] \ \forall (x_1,x_2) \in S$$

7.3. Tham số:

Kỳ vọng: giả định rằng $Y = g(X_1,X_2)$ với $g:R^2 \rightarrow R$:

Trường hợp (X_1,X_2) là véc tơ ngẫu nhiên rời rạc, luôn tồn tại:

$$E[Y] = \sum_{x_1} \sum_{x_2} g(x_1,x_2) p_{X_1,X_2}(x_1,x_2)$$

Trường hợp (X_1,X_2) là véc tơ ngẫu nhiên liên tục, luôn tồn tại:

$$E[Y] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x_1,x_2) f_{X_1,X_2}(x_1,x_2) d_{x_1} d_{x_2}$$

7.4. Hàm tạo sinh động lượng:

Đặt $X = (X_1,X_2)'$ là véc tơ ngẫu nhiên. Nếu tồn tại $E(e^{t_1X_1+t_2X_2})$ với $|t_1|<h_1$ và $|t_2|<h_2$ ($h_1,h_2>0$), ta gọi $M_{X_1,X_2}(t_1,t_2)$ là Hàm tạo sinh động lượng, ***moment generating function***, của X. Đặt $t = (t_1,t_2)'$ ta được:

$$M_{X_1,X_2}(t) = E[e^{t'X}]$$

Kỳ vọng:

$$E[X] = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix}$$

7.5. Các định lý liên quan:

Với (X_1,X_2) là véc tơ ngẫu nhiên, $Y_1 = g_1(X_1,X_2)$ và $Y_2 = g_2(X_1,X_2)$, $k_1,k_2 \in R$, ta được:

$$E(k_1Y_1 + k_2Y_2) = k_1E(Y_1) + k_2E(Y_2)$$

Với (X_1,X_2) là véc tơ ngẫu nhiên và $Var(X_2)$ là hữu hạn, ta được:

$$E[E(X_2|X_1)] = E(X_2)$$

$$Var[E(X_2|X_1)] \leq Var(X_2)$$

7.6. Hiệp phương sai và hệ số tương quan:

Cho hai biến ngẫu nhiên X và Y có hàm phân phối xác suất chung là f(x,y). Nếu u(x,y) là hàm theo x và y thì E[u(X,Y)] được xác định tồn tại.

Giả định các giá trị trung bình của X, Y là μ_1, μ_2 luôn tồn tại, có được từ hàm u(x,y) và phương sai của X, Y là σ_1^2, σ_2^2 tồn tại từ việc đặt u(x,y) bằng $(x - \mu_1)^2$ và $(x - \mu_2)^2$. Ta có kỳ vọng toán:

$$E[(X - \mu_1)(Y - \mu_2)] = E(XY) - \mu_1\mu_2$$

Gọi Corr(X,Y) là hệ số tương quan, ***Correlation coefficient***, và Cov(X,Y) là hiệp phương sai, ***Covariance***, ta được:

$$Corr(X,Y) = \frac{E[(X-\mu_1)(Y-\mu_2)]}{\sigma_1\sigma_2} = \frac{Cov(X,Y)}{\sigma_1\sigma_2}$$

Từ hai phương trình chúng ta thu được:

$$E(XY) = \mu_1\mu_2 + Corr(X,Y)\sigma_1\sigma_2 = \mu_1\mu_2 + Cov(X,Y)$$

Định lý: Nếu E(Y|X) là tuyến tính theo X thì

$$E(Y|X) = \mu_2 + Corr(X,Y)\frac{\sigma_2}{\sigma_1}(X - \mu_1)$$

$$E(var(Y|X)) = \sigma_2^2[1 - [Corr(X,Y)]^2]$$

8. Một số tính chất về ma trận:

8.1. Tính chất của phép nhân ma trận:

- Tính chất kết hợp: $(AB)C = A(BC)$
- Tính chất phân phối đối với phép cộng:

$$A(B + C) = AB + AC \text{ hoặc } (A + B)C = AC + BC$$

- Tính chất nhân với hằng số: $\alpha(AB) = (\alpha A)B = A(\alpha B)$
- Phép nhân ma trận đơn vị: $AI = IA = A$

8.2. Tính chất ma trận chuyển vị:

- $(A^T)^T = A$
- $(A + B)^T = A^T + B^T$
- $(\alpha A)^T = \alpha A^T$
- $(AB)^T = B^T A^T$

8.3. Tính chất ma trận nghịch đảo:

- $(A^{-1})^{-1} = A$

- $(A^{-1}B^{-1}) = (B^{-1})(A^{-1})$
- $XA = B \Rightarrow X = BA^{-1}$
- $AX = B \Rightarrow X = A^{-1}.B$
- $A.X.B = C \Rightarrow X = A^{-1}.C.B^{-1}$

PHỤ LỤC 2

MÃ NGUỒN LẬP TRÌNH MÔ PHỎNG

(Minh hoạ cho ví dụ bằng ngôn ngữ lập trình Python)

```
# Ví dụ 1:
# Tải thư viện
import random
# Đặt số phép thử
num_tosses = 100
# Tạo thử ngẫu kết quả số lần hiện mặt ngửa (head)
results = [random.choice(['Heads', 'Tails']) for _ in
range(num_tosses)]
heads = results.count('Heads')
# Tính kết quả số lần hiện mặt sấp (tail)
tails = num_tosses - heads
# Xuất kết quả ra màn hình.
print(f"Heads: {heads}, Tails: {tails}")
```

```
# Ví dụ 2:
# Tải thư viện
import torch
# Đặt số phép thử hay cỡ mẫu
num_samples = 10000
# Xác suất phép thử công bằng
fair_probs = torch.tensor([0.5, 0.5])
# Tạo 1 mẫu với 10000 lần thử từ phân phối đa thức
counts = torch.distributions.Multinomial(total_count=10000,
probs=fair_probs).sample()
# Tính tỷ lệ
probabilities = counts / 10000
print(probabilities)
```

```
# Ví dụ 3:
# Tải thư viện
import torch
import matplotlib.pyplot as plt
# Đặt số phép thử hay cỡ mẫu
num_samples = 10000
# Xác suất phép thử công bằng
fair_probs = torch.tensor([0.5, 0.5])
# Lấy mẫu từ phân phối Multinomial
counts = torch.multinomial(fair_probs, num_samples,
replacement=True)
counts = torch.stack([(counts == i).cumsum(0) for i in
range(2)], dim=1).float()
# Ước lượng xác suất
estimates = counts / counts.sum(dim=1, keepdim=True)
estimates = estimates.numpy()
# Biểu diễn bằng đồ thị
plt.figure(figsize=(4.5, 3.5))
plt.plot(estimates[:, 0], label="P(coin=heads)")
plt.plot(estimates[:, 1], label="P(coin=tails)")
plt.axhline(y=0.5, color='black', linestyle='dashed',
linewidth=1)
plt.xlabel('Samples')
plt.ylabel('Estimated probability')
plt.legend()
plt.show()
```

PHỤ LỤC 3

MÔ PHỎNG LỜI GIẢI CHO BÀI TẬP TÌNH HUỐNG

(Minh hoạ bằng ngôn ngữ lập trình Python)

```
# Bài tập 1
# Tải thư viện
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import beta
# Giả lập đồng xu mất cân đối thật sự
true_p_head = 0.6
np.random.seed(0)
# Thí nghiệm: Tung đồng xu N lần và cập nhật độ không chắc
chắn
def experiment(N_trials):
    data = np.random.binomial(1, true_p_head, N_trials)
    heads = sum(data)
    tails = N_trials - heads
    # Bayesian suy luận với ưu tiên là Beta(1,1) ~ Đồng nhất
(Uniform)
    posterior = beta(1 + heads, 1 + tails)
    # Vẽ phân phối hậu nghiệm (posterior)
    x = np.linspace(0, 1, 100)
    plt.plot(x, posterior.pdf(x), label=f'N={N_trials},
Heads={heads}/{N_trials}')
    plt.fill_between(x, posterior.pdf(x), alpha=0.1)
# Chạy thí nghiệm với số lần tung khác nhau
plt.figure(figsize=(10, 6))
# Rất ít dữ liệu -> Độ không chắc chắn cao
experiment(N_trials=5)
# Độ không chắc chắn giảm
experiment(N_trials=50)
```

```
# Hội tụ về xác suất thực
experiment(N_trials=500)
# Biểu diễn bằng đồ thị
plt.axvline(true_p_head, color='black', linestyle='--',
label='True p_head')
plt.title('Độ không chắc chắn nhận thức, epistemic
uncertainty, giảm khi tăng dữ liệu\n(Phân phối hậu nghiệm
của đồng xu không cân đối)')
plt.xlabel('Xác suất mặt ngửa p(head)')
plt.ylabel('Mật độ xác suất')
plt.legend()
plt.show()
```

```
# Bài tập 2
# Tải thư viện
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm
# Thiết lập tham số
np.random.seed(42)
# Giá trị đo lường thực (nhiệt độ)
true_value = 25.0
# Nhiễu hệ thống không thể loại bỏ (aleatoric)
aleatoric_noise_std = 0.5
# Hàm mô phỏng quá trình đo lường
def simulate_measurement(N_observations):
    # Tạo dữ liệu: giá trị thực + nhiễu ngẫu nhiên
    measurements = true_value + np.random.normal(0,
aleatoric_noise_std, N_observations)
    # Tính trung bình và độ không chắc chắn
    mean_estimate = np.mean(measurements)
    # Giảm theo căn bậc 2 của N
```

```
    epistemic_uncertainty = aleatoric_noise_std /
np.sqrt(N_observations)
    # Tổng độ không chắc chắn = epistemic + aleatoric (không
đổi)
    total_uncertainty = np.sqrt(epistemic_uncertainty**2 +
aleatoric_noise_std**2)
    return mean_estimate, epistemic_uncertainty,
total_uncertainty
# Chạy mô phỏng với số lần đo từ 1 đến 1000
N_range = np.arange(1, 1001)
results = [simulate_measurement(N) for N in N_range]
means, epistemic_uncerts, total_uncerts = zip(*results)
# Vẽ đồ thị
plt.figure(figsize=(12, 6))
# Đồ thị 1: Độ không chắc chắn theo số lần đo
plt.subplot(1, 2, 1)
plt.plot(N_range, epistemic_uncerts, label='Epistemic
Uncertainty', color='blue')
plt.plot(N_range, [aleatoric_noise_std]*len(N_range), 'r--',
label='Aleatoric Uncertainty')
plt.plot(N_range, total_uncerts, 'g-', label='Total
Uncertainty', alpha=0.5)
plt.xlabel('Số lần đo (N)')
plt.ylabel('Độ không chắc chắn (°C)')
plt.title('Giảm độ không chắc chắn khi tăng dữ liệu')
plt.legend()
plt.grid(True)
# Đồ thị 2: Phân phối ước lượng ở các giai đoạn
plt.subplot(1, 2, 2)
for N in [1, 10, 100, 1000]:
    mean, epistemic_uncert, _ = simulate_measurement(N)
    x = np.linspace(true_value-2, true_value+2, 100)
```

```
    y = norm.pdf(x, mean, epistemic_uncert)
    plt.plot(x, y, label=f'N={N}')
plt.axvline(true_value, color='black', linestyle='--',
label='Giá trị thực')
plt.xlabel('Giá trị ước lượng')
plt.ylabel('Mật độ xác suất')
plt.title('Phân phối ước lượng')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
# In kết quả tại các mốc quan trọng
for N in [1, 5, 10, 50, 100, 1000]:
    mean, epistemic, total = simulate_measurement(N)
    print(f'N={N}: Mean={mean:.2f}°C, Epistemic
uncert.={epistemic:.4f}°C, Total uncert.={total:.4f}°C')
```

```
# Bài tập 5
# Tải thư viện
import matplotlib.pyplot as plt
from matplotlib_venn import venn2
# Xác suất của biến cố A và B
P_A = 0.60
P_B = 0.40
# Trường hợp 1: A và B rời nhau (không giao nhau)
venn2(subsets=(P_A, P_B, 0), set_labels=('A', 'B'))
plt.title(f'Trường hợp rời nhau\n$P(A \cap B) = 0$\n$P(A
\cup B) = {P_A + P_B}$')
plt.show()
# Trường hợp 2: A và B giao nhau tối đa (A nằm trong B)
venn2(subsets=(0, P_B - P_A, P_A), set_labels=('A', 'B'))
```

```
plt.title(f'Trường hợp giao tối đa\n$P(A \cap B) =
{P_A}$\n$P(A \cup B) = {P_B}$')
plt.show()
# Trường hợp 3: A và B giao nhau một phần (tuỳ chọn)
P_AnB = 0.2  # Giả sử P(A ∩ B) = 0.2
venn2(subsets=(P_A - P_AnB, P_B - P_AnB, P_AnB),
set_labels=('A', 'B'))
plt.title(f'Trường hợp giao một phần\n$P(A \cap B) =
{P_AnB}$\n$P(A \cup B) = {P_A + P_B - P_AnB}$')
plt.show()
```

```
# Bài tập 6:
# Giả sử:
P_A = 0.6
P_B_given_A = 0.7
P_C_given_B = 0.8
# Tính P(A,B,C)
joint_prob = P_A * P_B_given_A * P_C_given_B
print(f"P(A, B, C) = {joint_prob:.3f}")
```

```
# Bài tập 8
# Tải thư viện
import numpy as np
import cvxpy as cp
# Dữ liệu mẫu
mu = np.array([0.1, 0.2])
Sigma = np.array([[0.04, 0.02], [0.02, 0.09]])
sigma_max = 0.05
# Biến tối ưu
alpha = cp.Variable(2)
objective = cp.Maximize(mu.T @ alpha)
constraints = [
```

```
    cp.quad_form(alpha, Sigma) <= sigma_max,
    cp.sum(alpha) == 1,
    alpha >= 0
]
prob = cp.Problem(objective, constraints)
prob.solve()
print("Trọng số tối ưu:", alpha.value)
```

PHỤ LỤC 4

NHẬT KÝ CẬP NHẬT TÀI LIỆU

Ngày phát thành: 2025-06-03. Tổng số trang gồm bìa là 56.

DANH MỤC THUẬT NGỮ CHUYÊN NGÀNH

Axioms of probability theory	Tiền đề của lý thuyết xác suất.
Bayes's rule / formula	Công thức Bayes.
Central limit theorems - CLT	Định lý giới hạn trung tâm.
Chebyshev's inequality	Bất đẳng thức Chebyshev.
Complement	Phần bù.
Conditional probability	Xác suất có điều kiện.
Correlation coefficient	Hệ số tương quan.
Consistency	Tính nhất quán.
Covariance	Hiệp phương sai.
Cumulative distribution function - CDF	Hàm phân phối tích lũy.
Data	Dữ liệu.
Dataset	Tập dữ liệu.
Disjoint	Không giao nhau.
Distinct possible outcome	Kết quả đầu ra duy nhất có thể xảy ra.
Expected value / Mean	Giá trị kỳ vọng / trung bình.
Event	Sự kiện, biến cố
Experiment	Thí nghiệm, thử nghiệm.
Feature	Đặc trưng hay thuộc tính.
Frequentist	Học giả theo trường phái tần suất.
Intersection	Phần giao nhau.
Joint probability function	Hàm xác suất kết hợp.
Laws of large numbers	Luật số lớn.

Machine learning - ML	Máy học.
Measurement	Phương pháp (phép) đo.
Multinomial distribution	Phân phối đa thức.
Normal distribution	Phân phối chuẩn.
Observation	Quan sát.
Outcome	Kết quả đầu ra.
Overlap	Chồng lên nhau.
Pattern	Hình mẫu.
Population	Quần thể.
Posterior probabilities	Xác suất hậu nghiệm.
Prior probabilities	Xác suất tiên nghiệm.
Probability	Xác suất.
Probability density	Mật độ xác suất.
Random discrete variable	Biến ngẫu nhiên rời rạc.
Random continuous variable	Biến ngẫu nhiên liên tục.
Random variable	Biến ngẫu nhiên.
Reinforcement learning	Học tăng cường.
Sample space	Không gian mẫu.
Simple event	Sự kiện đơn giản.
Statistics	Thống kê.
Supervised learning	Học có giám sát.
Standard deviation - SD	Độ lệch chuẩn.
Standard normal distribution	Phân phối chuẩn tắc.
Uncertainty	Sự không chắc chắn.

Unsupervised learning	Học không có giám sát.
Variance	Phương sai.
Venn diagram	Biểu đồ Venn.

DANH MỤC CÁC NGUỒN THAM KHẢO

1. Danh mục sách tham khảo:

[1] Morris H. DeGroot, Mark J. Schervish. 2012. “Probability and Statistics”. 4th edition.

[2] Robert V. Hogg, Joseph W. McKean, Allen T. Craig. 2013. “Introduction to Mathematical Statistics”. 7th edition.

[3] Sheldon Ross. 2020. “A first course in probability”. 10th edition.

[4] William Mendenhall, Robert J. Beaver, Barbara M. Beaver. 2009. “Introduction to Probability and Statistics”. 13th edition.

2. Mã nguồn:

Liên kết Github:

https://github.com/hiennx2k4/HCMUT_MathematicalFoundations.git