Hanoi Center for Financial and Industrial Mathematics
Trung Tâm Toán Tài Chính và Công Nghiệp Hà Nội

NHẬP MÔN HIỆN ĐẠI

XÁC SUẤT & THỐNG KÊ

Đỗ Đức Thái và Nguyễn Tiến Dũng

Hà Nội – Toulouse, 2009

Bản thảo này: Ngày 10 tháng 11 năm 2009

Hanoi Center for Financial and Industrial Mathematics
Hanoi National University of Education & University of Toulouse

Lời giới thiệu

Xác suất và thống kê đóng vai trò rất quan trọng trong hầu hết mọi lĩnh vực của thế giới hiện đại, từ khoa học, công nghệ, đến kinh tế, chính trị, đến sức khỏe, môi trường, v.v. Ngày nay, máy tính giúp cho việc tính toán các vấn đề xác suất thống kê ngày càng trở nên dễ dàng, một khi đã có các số liệu đúng đắn và mô hình hợp lý. Thế nhưng, bản thân máy tính không biết mô hình nào là hợp lý. Đấy là vấn đề của người sử dụng: cần phải hiểu được bản chất của các khái niệm và mô hình xác suất thống kê, thì mới có thể dùng được chúng.

Mục đích của quyển sách này chính là nhằm giúp bạn đọc *hiểu* đúng bản chất của những khái niệm và phương pháp cơ bản nhất của xác suất và thống kê, và qua đó có thể áp dụng được chúng, tìm được phương pháp thích hợp cho những tình huống cụ thể. Một số điểm mà các tác giả cố gắng đưa vào trong sách này là:

- Giải thích bản chất các khái niệm một cách trực giác, dễ hiểu nhất trong chừng mực có thể, đồng thời đảm bảo độ chặt chẽ nhất định về mặt toán học.

- Cho nhiều ví dụ và bài tập về những tình huống có thật, với số liệu có thật, nhằm giúp bạn đọc cảm nhận được các ứng dụng thực tế của xác suất và thống kê.

Quyển sách này có 5 chương. Chương 1 gồm một số khái niệm cơ sở của lý thuyết xác suất. Chương này không đòi hỏi kiến thức đặc biệt gì về toán, và học sinh phổ thông cũng có thể đọc và hiểu được phần lớn. Tuy nhiên, kiến thức của Chương 1 không hoàn toàn hiển nhiên, kể cả đối với những người đã học đại học. Trong quá trình soạn thảo, các tác giả có đem một số bài tập hơi khó của Chương 1 đố các học sinh đại học và cao học ngành toán, và phần lớn họ làm sai! Các bài tập đó không phải là khó về mặt toán học (để giải chúng chỉ cần làm vài phép tính số học đơn giản), mà là khó vì chúng chứa đựng những sự tế nhị về bản chất của xác suất. Hy vọng rằng, bạn đọc sẽ thấy được những sự tế nhị đó, và tránh được các sai lầm mà nhiều người khác hay mắc phải.

Từ Chương 2 đến Chương 4 của quyển sách là lý thuyết xác suất của các biến ngẫu nhiên. Chương 2 là về các biến ngẫu nhiên nhận giá trị thực. Chương 3 là về các bộ nhiều biến ngẫu nhiên, hay còn gọi là các vector ngẫu nhiên. Chương 4 là về các định lý giới hạn, trong đó có định lý giới hạn trung tâm, được coi là định lý quan trọng nhất của lý thuyết xác suất và là hòn đá tảng của thống kê toán học. Chương 5 của quyển sách là giới thiệu về thống kê. Bạn đọc sẽ tìm thấy trong chương này những vấn đề có thể giải quyết bằng thống kê như ước lượng, kiểm định, dự báo, những nguyên tắc cơ bản nhất của thống kê, và một số phương pháp thông kê nay đã trở thành kinh điển.

Để hiểu tốt các vấn đề được bàn tới trong Chương 2 và các chương tiếp theo, bạn đọc cần có một số kiến thức chuẩn bị về giải tích toán học, như phép tính vi tích phân và khai triển Taylor-Lagrange, cộng với một ít kiến thức về đại số tuyến tính. Nếu có thêm một ít kiến thức về tôpô và giải tích hàm thì càng tốt. Trong sách có đưa ra định nghĩa và tính chất của một số khái niệm toán học cần dùng, ví dụ như tích phân Lebesgue trên không gian xác suất, biến đổi Fourier, hội tụ yếu, v.v.

Quyển sách này có thể dùng làm sách giáo khoa hay sách tham khảo cho môn xác suất thống kê ở bậc đại học hoặc cao học nhiều ngành khác nhau. Sinh viên các ngành không phải toán có thể bỏ qua các phần chứng minh các định lý tương đối phức tạp trong sách, mà chỉ cần hiểu đúng phát biểu của các định lý quan trọng nhất và cách áp dụng chúng. Các sinh viên ngành toán thì nên tìm hiểu cả cách chứng minh các định lý.

Do khuôn khổ của quyển sách có hạn, nên còn rất nhiều khái niệm quan trọng của xác suất và thống kê không xuất hiện trong sách, ví dụ như quá trình ngẫu nhiên. Hy vọng rằng quyển sách này cung cấp được tương đối đầy đủ các kiến thức cơ sở, để bạn đọc có thể hiểu được các tài liệu chuyên sâu hơn về xác suất và thống kê khi cần thiết.

Để biên soạn quyển sách này, các tác giả có tham khảo nhiều sách báo liên quan đến xác suất thống kê, và có trích lại nhiều bài tập và ví dụ từ các tài liệu đó. Những sách mà các các tác giả tham khảo nhiều được liệt kê ở phần "Tài liệu tham khảo". Trong đó có những sách "nặng", có nhiều chứng minh chặt chẽ và khá nặng về toán, ví dụ như quyển "Theory of probability and random processes" của Koralev và Sinai [5], và có những sách "nhẹ", dễ đọc để có thể nắm được những ý tưởng chính, nhưng không có chứng minh, tiêu biểu như quyển "The cartoon guide to statistics" của Gonick và Smith [2].

Những bản thảo đầu tiên của quyển sách này có được một số đồng nghiệp, bạn bè và sinh viên đọc và góp ý sửa lỗi và trình bầy lại cho tốt lên. Các tác giả xin chân thành cảm ơn sự quan tâm và giúp đỡ của họ. Tất nhiên, mọi lỗi còn lại trong sách là thuộc về trách nhiệm của các tác giả.

Quyển sách này là một sản phẩm của Trung Tâm Toán Tài Chính và Công Nghiệp Hà Nội (do các tác giả thành lập vào đầu năm 2009), được viết với mục đích trước hết là để phục vụ cho nhu cầu của bản thân Trung Tâm. Các tác giả hy vọng rằng, quyển sách này sẽ có ích, không chỉ cho Trung Tâm, mà còn cho một lượng rất lớn các độc giả khác đang hoặc sẽ quan tâm về xác suất và thống kê.

Hà Nội – Toulouse, 2009

Mục lục

Chương 1

Xác suất là gì

1.1 Xác suất là gì ?

Hầu như mọi người đều biết đến khái niệm xác suất. Tuy nhiên không phải ai cũng hiểu rõ những tính chất cơ bản của nó. Ví dụ như sự phụ thuộc vào thông tin của xác suất (mỗi khi có thêm thông tin mới thì xác suất thay đổi) hay bị bỏ qua. Và có những bài toán tính toán xác suất tưởng chừng như rất đơn giản, nhưng có hơn một nửa số người đã từng học xác suất làm sai khi được hỏi, kể cả các thạc sĩ ngành toán. Bởi vậy, trong chương này, chúng ta sẽ nhấn mạnh những sự tế nhị trong xác suất, đặc biệt là với xác suất có điều kiện, mà bạn đọc cần biết đến, để tránh được những lỗi cơ bản hay gặp nhất.

Trước khi đi vào lý thuyết, có một câu đố liên quan đến xác suất sau đây dành cho bạn đọc. Giả sử có một trò chơi trên TV như sau: có 3 cánh cửa, đằng sau 1 trong 3 cánh cửa đó là 1 món quà lớn, còn sau 2 cửa còn lại không có gì. Người chơi được chọn 1 trong 3 cánh cửa, nếu chọn đúng cửa có quà thì được nhận quà. Sau khi người chơi đã chọn 1 cửa, người hướng dẫn chương trình mở một trong hai cửa còn lại ra, nhưng sẽ chỉ mở cửa không có quà. Sau đó người chơi được quyền chọn, hoặc là giữ cái cửa mình chọn ban đầu, hoặc là đổi lấy cái cửa chưa được mở còn lại. Theo bạn thì người chơi nên chọn phương án nào? Vì sao ? Hãy thử nghĩ về nó một chút trước khi tiếp tục đọc.

1.1.1 Xác suất của một sự kiện

Xác suất của một sự kiện (hay tình huống giả định) là khả năng xảy ra sự kiện (hay tình huống giả định) đó, được đánh giá dưới dạng một số thực nằm giữa 0 và 1.

Khi một sự kiện không thể xảy ra thì xác suất của nó bằng 0. Ví dụ như xác suất của sự kiện "có người sống trên sao Thổ" bằng 0.

Khi một sự kiện chắc chắn đã hoặc sẽ xảy ra thì xác suất của nó bằng 1 (hay còn viết là 100%). Ví dụ như sự kiện "tôi được sinh ra từ trong bụng mẹ" có xác suất bằng 1.

Khi một sự kiện có thể xảy ra và cũng có thể không xảy ra, và chúng ta không biết nó có chắn chắn xảy ra hay không, thì chúng ta có thể coi xác suất của nó lớn hơn 0 và nhỏ hơn 1. Sự kiện nào được coi là càng dễ xảy ra thì có xác suất càng lớn (càng gần 1), và ngược lại nếu càng khó xảy ra thì xác suất càng nhỏ (càng gần 0). Ví dụ tôi mua một vé xổ số. Tôi không biết nó sẽ trúng giải hay không, có thể có mà cũng có thể không. Nếu như cứ 100 vé xổ số chỉ có 1 vé trúng giải, thì tôi sẽ coi xác suất trúng giải của vé của tôi là 1%. Con số 1% ở đây chính là tần suất, hay tỷ lệ trúng giải của các vé xổ số: nó bằng số các vé trúng giải chia cho tổng số các vé.

Không những chỉ các sự kiện trong tương lai, mà cả các sự kiện trong quá khứ, mà chúng ta thiếu thông tin để có thể biết chắc là chúng đã thực sự xảy ra hay không, thì chúng ta vẫn có thể gán cho các sự kiện đó một xác suất nào đó, ứng với độ tin tưởng của chúng ta về việc sự kiện đó đã thực sự xảy ra hay không. Ví dụ như, nữ hoàng Cleopatra của Ai Cập có tự tử bằng cách để cho rắn độc cắn không ? Đấy là một giả thuyết, mà theo các nhà sử học thì có nhiều khả năng xảy ra, nhưng không chắc chắn.

1.1.2 Ba tiên đề về sự nhất quán của xác suất

Tiên đề 1. Như đã viết phía trên, nếu A là một sự kiện (giả định) và ký hiệu $P(A)$ là **xác suất của** A thì

$$0 \leq P(A) \leq 1 \tag{1.1}$$

Tiên đề 2. Nếu A là một sự kiện, và ký hiệu $\overline{A}$ là sự kiện *phủ định của* A thì

$$P(A) + P(\overline{A}) = 1 \tag{1.2}$$

Ý nghĩa triết học của tiên đề 2 tương đối hiển nhiên: Trong hai sự kiện "A" và "phủ định của A" có 1 và chỉ 1 sự kiện xảy ra. Nếu "A" càng có nhiều khả năng xả ra thì "phủ định của A" càng có ít khả năng xảy ra, và ngược lại.

Ví dụ 1.1. Một học sinh đi thi vào một trường đại học. Nếu xác suất thi đỗ là 80% thì xác suất thi trượt là 20% (= 100% - 80%), chứ không thể là 30%, vì nếu xác suất thi đỗ là 80% và xác suất thi trượt là 30% thì không nhất quán.

Ví dụ 1.2. Tôi tung một đồng tiền, khi nó rơi xuống thì có thể hiện mặt sấp hoặc mặt ngửa. Tổng xác suất của hai sự kiện "mặt sấp" và "mặt ngửa" bằng 1. Nếu tôi không có lý do đặc biệt gì để nghĩ rằng mặt nào dễ hiện lên hơn mặt nào, thì tôi coi rằng hai mặt có xác suất hiện lên bằng nhau. Khi đó sự kiện "mặt ngửa" có xác suất bằng sự kiện "mặt sấp" và bằng 1/2.

Tiên đề 3. Với hai sự kiện A và B, ta sẽ ký hiệu sự kiện "cả A và B đều xảy ra" bằng $A \cap B$ và sự kiện "ít nhất một trong hai sự kiện A hoặc B xảy ra" bằng $A \cup B$. Khi đó nếu hai sự kiện A và B không thể cùng xảy ra, thì xác suất của sự kiện "xảy ra A hoặc B" bằng tổng các xác suất của A và của B:

$$P(A \cap B) = 0 \Rightarrow P(A \cup B) = P(A) + P(B) \tag{1.3}$$

Ví dụ 1.3. Một học sinh được cho điểm một bài kiểm tra. Có thể được 7 điểm, có thể được 8 điểm, hoặc có thể được điểm khác, nhưng không thể vừa được 7 điểm vừa được 8 điểm. Bởi vậy $P((7d) \cup (8d)) = P(7d) + P(8d)$

Tiên đề 3 có thể phát biểu một cách tổng quát hơn như sau:

Tiên đề 3'. Nếu X và Y là hai sự kiện bất kỳ thì

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \tag{1.4}$$

Bài tập 1.1. Chứng minh rằng tiên đề 3 tương đương với tiên đề 3'.

1.1.3 Xác suất phụ thuộc vào những gì ?

Xác suất của một sự kiện không nhất thiết phải là một hằng số, mà nó có thể thay đổi, phụ thuộc vào nhiều yếu tố. (Từ *sự kiện* ở đây hiểu theo nghĩa thông thường, chứ không phải theo nghĩa "một tập hợp trong một không gian xác suất với 1 độ đo xác suất đã cố định" trong mô hình toán học)

Xác suất thay đổi theo thời gian. Ví dụ, ông Obama được bầu làm tổng thống Mỹ vào tháng 11/2008. Từ trước lúc bầu cử mấy tháng, có sự cạnh tranh ác liệt giữa ông ta và đối thủ chính của ông ta là ông McCain, và một người quan sát bên ngoài có thể nhận định là hai ông có khả năng được bầu cử ngang nhau (tức là xác suất được bầu của mỗi ông quãng 50%). Nhưng khi kết quả bầu cử được công bố trọn vẹn, thì xác suất được bầu của Obama chuyển thành 100% (tức là ông ta đã chắc chắn được bầu). Trước đó 1 năm, ông Obama là một người chưa được nhiều người biết đến và còn phải tranh cử với bà Clinton và các ứng cử viên khác trong Đảng của mình, và khi đó, đối với quan sát viên bên ngoài, xác suất được bầu làm tổng thống của Obama không phải 100%, cũng không phải 50%, mà nhỏ hơn thế nhiều.

Xác suất phụ thuộc vào thông tin. Lấy bài toán đố về trò chơi trên TV viết phía trên làm ví dụ. Gọi tên cửa mà người chơi chọn lúc đầu là A, cửa không có quà mà người hướng dẫn chương trình mở ra là B, và cửa còn lại là C. Vào thời điểm ban đầu, không có thông tin gì về cửa nào phía sau có quà, thông tin duy nhất là 1 trong 3 cửa có quà. Không có cơ sở gì để cho rằng cửa nào có nhiều khả năng có quà hơn cửa nào, bởi vậy vào thời điểm ban đầu ta coi $P(A) = P(B) = P(C) = 1/3$. Nhưng sau khi cửa B được mở ra, thì ta có thêm một thông tin mới, là cửa B không có quà. Như vậy thông tin mới này làm thay đổi xác suất của B: bây giờ ta có $P(B) = 0$. Không chỉ xác suất của B thay đổi, mà tổng xác suất của A và C bây giờ cũng thay đổi: $P(A) + P(C) = 1$ thay vì bằng $2/3$ như trước. Như vậy ít ra một trong hai số $P(A)$ hoặc $P(C)$ thay đổi, hoặc là cả hai. Xác suất $P(A)$ có thay đổi vì thông tin mới này không ? Câu trả lời là không (Giải thích vì sao không ?). Chỉ có $P(C)$ là thay đổi: sau khi người hướng dẫn chương trình mở cửa B, thì ta có $P(A) = 1/3$ và $P(C) = 2/3$. Như vậy người chơi nên đổi cửa A lấy cửa C thì dễ thắng hơn. Để thấy rõ hơn việc cánh cửa còn lại có nhiều khả năng có quà hơn là cánh cửa mà người chơi chọn ban đầu, thay vì chỉ có 3 cửa, ta hãy hình dung có 100 cửa. Sau khi bạn chọn 1 cửa, người dẫn chương trình mở 98 cửa không có quà trong số 99 cửa còn lại, chỉ để lại 1 cửa thôi. Khi đó, nếu được đổi, bạn sẽ giữ nguyên cửa của mình, hay là đổi lấy cái cửa còn lại kia ?

Xác suất phụ thuộc vào điều kiện. Chúng ta sẽ bàn về xác suất có điều kiện và công thức tính xác suất có điều kiện ở một phần sau. Điều đáng chú ý ở đây là, mọi xác suất đều có thể coi là xác suất có điều kiện, và đều phụ thuộc vào những điều kiện nào đó, có thể được nói ra hoặc không nói ra (điều kiện hiểu ngầm). Ví dụ, khi chúng ta nói "khi tung cái xúc sắc S, xác suất để hiện lên mặt có 3 chấm là 1/6", chúng ta hiểu ngầm S là một cái xúc sắc đều đặn, các mặt đều có khả năng xuất hiện như nhau. Nhưng nếu S là một cái xúc sắc méo mó, nhẹ bên này nặng bên nọ (điều kiện khác đi), thì hoàn toàn có thể là xác suất để khi tung hiện lên mặt có 3 chấm sẽ khác 1/6. Một ví dụ khác là xác suất xảy ra tai nạn khi lái ô tô: khi người lái xe khoe mạnh tỉnh táo, thì xác suất xảy ra tai nạn thấp, còn khi vẫn người lái đó bị say rượu hoặc buồn ngủ gật, thì xác suất xảy ra tai nạn cao hơn, v.v. Khi chúng ta biết thêm một điều kiện mới, tức là có thêm một thông tin mới, bởi vậy sự phụ thuộc vào điều kiện của xác suất cũng có thể coi là sự phụ thuộc vào thông tin.

Xác suất phụ thuộc vào người quan sát, hay là tính chủ quan của xác suất. Cùng là một sự kiện, nhưng hai người quan sát khác nhau có thể tính ra hai kết quả xác suất khác nhau, và cả hai đều "có lý", bởi vì họ dựa trên những thông tin và phân tích khác nhau. Ví dụ như, có chuyên gia tài chính đánh giá rằng cổ phiếu của hãng Vinamilk có nhiều khả năng đi lên trong thời gian tới, trong khi lại có chuyên gia tài chính khác đánh giá rằng cổ phiếu của hãng đó có nhiều khả năng đi xuống ít khả năng đi lên trong thời gian tới. Quay lại trò chơi truyền hình: với người chơi thì $P(A) = 1/3$, nhưng đối với người dẫn chương trình thì $P(A)$ không phải là 1/3, mà là 0 hoặc 1, vì người đó biết ở đằng sau cửa A có quà hay không.

1.1.4 Tính xác suất bằng thống kê

Đối với những hiện tượng xảy ra nhiều lần, thì người ta có thể dùng thống kê để tính xác suất của sự kiện xảy ra hiện tượng đó. Công thức sẽ là

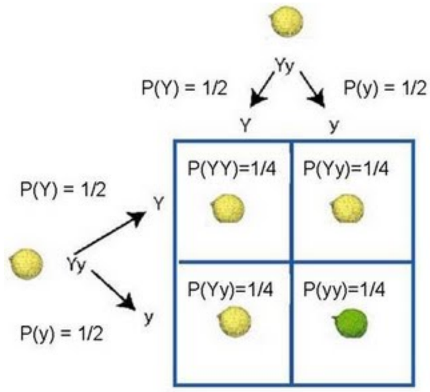
$$P(A) = \frac{N(A)}{N(total)} \tag{1.5}$$

Ở đây $N(total)$ là tổng số các trường hợp được khảo sát, và $N(A)$ là số các trường hợp được khảo sát thỏa mãn điều kiện *xảy ra A*.

Cơ sở toán học cho việc dùng thống kê để tính xác suất, là luật số lớn và các định lý giới hạn, mà chúng ta sẽ tìm hiểu ở phía sau trong sách này.

Ví dụ 1.4. Có một số số liệu sau đây về tai tạn ô tô và máy bay. Trong những năm 1989-1999, trên toàn thế giới, trung bình mỗi năm có khoảng 18 triệu chuyến bay, 24 tai nạn máy bay chết người, và 750 người chết trong tai nạn máy bay. Cũng trong khoảng thời gian đó, ở nước Pháp, trung bình mỗi năm có khoảng 8000 người chết vì tai nạn ô tô, trên tổng số 60 triệu dân. Từ các số liệu này, chúng ta có thể tính: Xác suất để một người ở Pháp bị chết vì tai nạn ô tô trong một năm là 8000/60000000 = 0,0133%. Xác suất để đi một chuyến bay gặp tai nạn chết người là 24/18000000 = 0,000133%, chỉ bằng 1/100 xác suất bị chết vì tai nạn ô tô trong 1 năm. Nếu một người một năm bay 20 chuyến, thì xác suất bị chết vì tai nạn máy bay trong năm bằng quãng $20 \times 0,000133\% = 0,00266\%$, tức là chỉ bằng 1/5 xác suất bị chết vì tai nạn ô tô trong năm.

Ví dụ 1.5. Ông Gregor Mendel (1822-1884) là một tu sĩ người Áo (Austria) thích nghiên cứu sinh vật. Ông ta trồng nhiều giống đậu khác nhau trong vườn của tu viện, và ghi chép tỉ mẩn về các tính chất di truyền và lai giống của chúng. Năm 1866 Mendel công bố một bài báo về các hiện tượng mà ông ta qua sát được, và lý thuyết của ông ta để giải thích các hiện tượng. Một trong những quan sát trong đó là về màu sắc: Khi lai đậu hạt vàng với đậu hạt xanh (thế hệ thứ nhất) thì các cây lai (thế hệ thứ hai) đều ra đậu hạt vàng, nhưng tiếp tục lai các cây đậu hạt vàng thế hệ thứ hai này với nhau, thì đến thế hệ thứ ba xác suất ra đậu hạt xanh là 1/4. Con số 1/4 là do Mendel thống kê thấy tỷ lệ



Hình 1.1: Lý thuyết di truyền của Mendel và xác suất trong lai giống đậu

đậu hạt xanh ở thế hệ thứ ba gần bằng 1/4. Từ đó Mendel xây dựng lý thuyết di truyền để giải thích hiện tượng này: màu của đậu được xác định bởi 1 gen, và gen gồm có hai phần. Thế hệ đầu tiên, cây đậu hạt vàng có gen thuần chủng "YY" còn hạt xanh có gen "yy" (tên gọi "Y" và "y" ở đây là tùy tiện). Khi lai nhau, thì một nửa gen của cây này ghép với một nửa gen của cây kia để tạo thành gen của cây con. Các cây thế hệ thứ hai đều có gen "Yy", và màu hạt của gen "Yy" cũng là vàng. Đến thế hệ thứ ba, khi lai "Yy" với "Yy" thì có 4 khả năng xảy ra : "YY", "Yy", "yY" và "yy". ("Yy" và "yY" là giống nhau về gen, nhưng viết như vậy là để phân biệt là phần "Y" đến từ cây thứ nhất hay cây thứ hai trong 2 cây lai với nhau). Về lý thuyết, có thể coi 4 khả năng trên là có xác suất xảy ra bằng nhau. Bởi vậy xác suất để cây thế hệ thứ ba có gen "yy" (hạt màu xanh) là 1/4. Trong rất nhiều năm sau khi công bố, công trình của Mendel không được các nhà khoa học khác quan tâm đến, nhưng ngày nay Mendel được coi là cha tổ của di truyền học.

1.2 Mô hình toán học của xác suất

1.2.1 Không gian xác suất

Không gian xác suất là một khái niệm toán học nhằm trừu tượng hóa 3 tiên đề phía trên về sự nhất quán của xác suất.

Định nghĩa 1.1. *Một* **không gian xác suất** *là một tập hợp* Ω*, cùng với:*

1) Một họ $\mathcal{S}$ các tập con của Ω, thỏa mãn các tính chất sau: $\Omega \in \mathcal{S}$, và nếu $A, B \in \mathcal{S}$ thì $A \cup B \in \mathcal{S}$, $A \cap B \in \mathcal{S}$ và $\overline{A} := \Omega \setminus A \in \mathcal{S}$. Một họ như vậy được gọi là một **đại số** *các tập con của Ω. Trong trường hợp Ω là một tập có vô hạn các phần tử, thì chúng ta sẽ đòi hỏi thêm điều kiện sau: Nếu $A_i, i = 1, 2, 3, \ldots$ là một dãy vô hạn các phần tử của $\mathcal{S}$, thì hợp $\bigcup_{i=1}^{\infty} A_i$ cũng thuộc họ $\mathcal{S}$. Với thêm điều kiện này, $\mathcal{S}$ được gọi là một* **sigma-đại số**. *Các phần tử của $\mathcal{S}$ được gọi là là tập hợp con* **đo được** *của không gian xác suất.*

2) Một hàm số thực $P : \mathcal{S} \to \mathbb{R}$ trên $\mathcal{S}$, được gọi là **phân bố xác suất** *hay* **độ đo xác suất** *trên Ω, thỏa mãn các tính chất sau:*
i) Với mọi $A \in \mathcal{S}$, ta có

$$0 \leq P(A) \leq 1. \tag{1.6}$$

ii)

$$P(\emptyset) = 0, \ P(\Omega) = 1. \tag{1.7}$$

iii) Nếu $A \cap B = \emptyset$ thì

$$P(A \cup B) = P(A) + P(B). \tag{1.8}$$

Tổng quát hơn, nếu $A_i, i = 1, 2, 3, \ldots$ là một dãy các tập hợp con đo được không giao nhau thì

$$P(\bigcup_i A_i) = \sum_i P(A_i). \tag{1.9}$$

Ghi chú 1.1. 1) Không gian xác suất Ω còn được gọi là **không gian mẫu** (sample space), và nó là mô hình toán học trừu tượng cho vấn đề tính toán xác suất đang được quan tâm. Mỗi phần tử của Ω có thể được gọi là một **sự kiện thành phần** (elementary event). Nếu A là một phần tử của Ω thì ta cũng có thể viết $P(A)$ và hiểu là $P(\{A\})$, trong đó $\{A\}$ là tập con của Ω chứa duy nhất một phần tử A. Mỗi sự kiện là một tập con của Ω, và có thể gồm nhiều (thậm chí vô hạn) sự kiện thành phần. Không nhất thiết tập con nào của Ω cũng đo được (tức là nằm trong họ $\mathcal{S}$), và chúng ta sẽ chỉ quan tâm đến những tập con đo được.

2) Trong toán học, một đại số là một tập hợp với các phép tính cộng, trừ, và phép nhân (không nhất thiết phải có phép chia). Các tính chất của họ $\mathcal{S}$ trong định nghĩa không gian xác suất khiến nó là một đại số theo nghĩa như vậy: Phần tử 0 trong $\mathcal{S}$ là tập rỗng, phần tử đơn vị trong $\mathcal{S}$ là tập Ω, phép nhân trong $\mathcal{S}$ là phép giao: $A \times B := A \cap B$, và phép cộng trong $\mathcal{S}$ là phép $A + B := (A \cup B) \setminus (A \cap B) = (A \setminus B) \cup (B \setminus A)$. Đại số này có số đặc trưng bằng 2, tức là $2A = A + A = 0$ với mọi A (và bởi vậy phép cộng và phép trừ chẳng qua là một). Chúng ta muốn $\mathcal{S}$ là một đại số chính là để cho việc làm các phép tính số học với xác suất được thuận tiện.

Hình 1.2: A. N. Kolmogorov

3) Đẳng thức (1.9) được gọi là **tính chất sigma** của xác suất. Trong toán, chữ cái hy lạp sigma thường dùng để ký hiệu tổng, với hữu hạn hay vô hạn các thành phần. Tính chất sigma là *tính chất cộng tính vô hạn*: khi có một dãy vô hạn các tập con không giao nhau, xác suất của hợp của chúng cũng bằng tổng vô hạn của các xác suất của các tập con. Tính chất sigma chính là tính chất cho phép chúng ta *lấy giới hạn* trong việc tính toán xác suất. Chẳng hạn như, nếu $A_1 \subset A_2 \subset \ldots$ là một dãy tăng các tập con của Ω, và $A = \lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$, thì ta có thể viết $P(A) = \lim_{n\to\infty} P(A_n)$, bởi vì

$$\begin{aligned} P(A) = P(A_1 \cup \bigcup_{n=1}^{\infty}(A_{n+1} \setminus A_n)) &= P(A_1) + \sum_{n=1}^{\infty} P(A_{n+1} \setminus A_n) \\ = P(A_1) + \lim_{n\to\infty} \sum_{k=1}^{n} P(A_{k+1} \setminus A_k) &= A_1 + \lim_{n\to\infty}(P(A_{n+1}) - P(A_1)) \end{aligned} \quad (1.10)$$

Phép toán *lấy giới hạn* là phép toán cơ bản nhất của giải tích toán học, và mọi phép toán giải tích khác như đạo hàm, tích phân, v.v. đều có thể được định nghĩa qua phép lấy giới hạn. Bởi vậy, tính chất *sigma* chính là tính chất cho phép chúng ta sử dụng giải tích toán học trong việc nghiên cứu xác suất. Các nhà toán học cổ điển trong thế kỷ 18 và 19 đã dùng các phép tính vi tích phân trong xác suất, tức là đã dùng tính chất sigma. Về mặt trực giác, tính chất sigma là mở rộng hiển nhiên của tính chất cộng tính (1.8). Tuy nhiên, nói một cách chặt chẽ toán học, đẳng thức (1.9) không suy ra được từ đẳng thức (1.8), và phải được coi là một tiên đề trong xác suất. Tiên đề này được đư ra bởi nhà toán học người Nga Andrei Nikolaievitch Kolmogorov (1903-1987), người xây dựng nền tảng cho lý thuyết xác suất hiện đại.

Bài tập 1.2. Chứng minh rằng, với 3 tập con A, B, C (đo được) bất kỳ trong một không gian xác suất, ta có:

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

1.2.2 Phân bố xác suất Bernoulli

Hình 1.3: Bia mộ của "mathematicus incomparabilis" J. Bernoulli ở Basel

Không gian xác suất đơn giản nhất mà không tầm thường là không gian sinh bởi đúng 1 sự kiện A và phủ định $\overline{A}$ của nó: $\Omega = \{A, \overline{A}\}$. Phân bố xác suất trên Ω trong trường hợp này được xác định bởi đúng một số $p = P(A)$. Phân bố này được gọi là **phân bố Bernoulli**, theo tên của Jacob Bernoulli (1654-1705), một nhà toán học người Thụ Sĩ.

Ví dụ 1.6. Một vận động viên bắn súng, nhằm vào đích bắn 1 phát súng. Có hai sự kiện đối lập nhau có thể xảy ra là A = "bắn trúng" và $\overline{A}$ = "bắn trượt". Giả sử xác suất bắn trúng là 95%. Khi đó ta có không gian xác suất $\Omega = \{A, \overline{A}\}$ với phân bố xác suất Bernoulli với $p = P(A) = 95\%$. Xác suất của $\overline{A}$ (sự kiện "bắn trượt") bằng $1 - p = 1 - 95\% = 5\%$.

Ví dụ 1.7. (Cái kim của Buffon). Bá tước George-Louis Leclerc de Buffon (1707-1788) là một nhà khoa học tự nhiên lớn, nghiên cứu về thực vật, động vật, trái đất, lịch sử tự nhiên, v.v. Thời trẻ, ông ta đặc biệt thích toán học, và vào năm 1733 có trình lên Viện Hàm lâm Pháp một công trình nhan đề "Sur le jeu du franc-carreau" (về chò trời franc-careau, là một trò chơi cá cược thịnh hành thời đó: người ta tung 1 đồng tiền vào 1 ô vuông và cá cược nhau xem vị trí nó sẽ nằm chỗ nào). Trong công trình này, các phép toán vi tích phân được Buffon đưa vào lý thuyết xác suất. Buffon còn là người nghĩ ra phương pháp sau đây để tính số π: Lấy 1 tờ giấy to và 1 cái kim. Kẻ các đường thẳng song song trên tờ giấy, cách đều nhau một khoảng cách đúng bằng chiều dài của cái kim. Tung cái kim một cách ngẫu nhiên lên trên tờ giấy. Có hai khả năng xảy ra: 1) kim nằm đè lên 1 đường thẳng trong các đường được kẻ; 2) kim nằm lọt vào giữa hai đường thẳng. Buffon tính ra rằng, sự kiện "kim nằm đè lên 1 đường thẳng" có xác suất bằng $1/\pi$. Như vậy hai sự kiện "nằm đè lên 1 đường thẳng" và "nằm lọt vào giữa hai đường thẳng" hợp thành một không gian xác suất Bernoulli với $p = 1/\pi$. Tung kim n lần, và gọi số lần kim nằm đè lên 1 đường thẳng trong số n lần tung là b_n. Khi đó, theo luật số lớn, b_n/n tiến tới $p = 1/\pi$ khi n tiến tới vô cùng. Bởi vậy để xấp xỉ tính số π, có thể làm như sau: tung kim thật nhiều lần, đếm số lần kim đè lên trên 1 đường thẳng, rồi lấy số lần tung chia cho số đó. Phương pháp tung kim của Buffon chính là tiền thân của phương pháp Monte-Carlo trong toán học.

Hình 1.4: Tượng của Buffon ở Jardin des Plantes, Paris

1.2.3 Phân bố xác suất đều

Định nghĩa 1.2. *Phân bố xác suất P trên không gian xác suất hữu hạn với N phần tử $\Omega = \{A_1, \ldots, A_N\}$ được gọi là* **phân bố xác suất đều** *nếu như $P(A_1) = \ldots = P(A_N) = 1/N$.*

Tất nhiên, mỗi không gian xác suất với một số hữu hạn các phần tử chỉ có duy nhất một phân bố xác suất đều trên đó.

Ghi chú 1.2. Khái niệm phân bố đều không mở rộng được lên các không gian xác suất có số phần tử là vô hạn và đếm được, bởi vì 1 chia cho vô cùng bằng 0, nhưng mà tổng của một chuỗi vô hạn số 0 vẫn bằng 0 chứ không bằng 1.

Các phân bố xác suất đều là các phân bố quan trọng hay gặp trong thực tế. Lý do chính dẫn đến phân bố xác suất đều là *tính đối xứng*, *cân bằng*, hay *hoán vị được* của các sự kiện thành phần.

Ví dụ 1.8. Lấy một bộ bài tú lơ khơ mới có 52 quân, đặt nằm sấp. Khi đó xác suất để rút một con bài trong đó ra một cách tùy ý được con "2 Cơ" (hay bất kỳ "số" nào khác) bằng 1/52. Vì sao vậy ? Vì các con bài khi đặt nằm sấp thì giống hệt nhau, không thể phân biệt được con nào với con nào, số nào cũng có thể được viết dưới bất kỳ con bài nào, và nếu chuyển chỗ 2 con bài trong bộ bài với nhau thì trông bộ bài vẫn hệt như cũ (đấy chính là tính "đối xứng", "hoán vị được"). Người quan sát không có thông tin gì để có thể nhận biết được số nào dễ nằm ở phía dưới con bài nào hơn trong các con bài đăng nằm sấp, và khi đó thì phải coi rằng xác suất của các số là như nhau. Nếu như có những con bài "được đánh dấu" (chơi ăn gian), thì tất nhiên đối với người biết chuyện đánh dấu, không còn phân bố xác suất đều nữa.

Công thức để tính xác suất của một sự kiện trong một phân bố xác suất đều rất đơn giản: Nếu như không gian xác suất Ω với phấn bố xác suất đều có N phần tử, và sự kiện được biểu diễn bằng một tập con A của Ω với k phần tử, thì xác suất của A bằng k/N:

$$P(A) = \frac{\#A}{\#\Omega} = \frac{k}{N} \tag{1.11}$$

Ví dụ 1.9. Giả sử một gia đình có 3 con. Khi đó xác suất để gia đình đó có 2 con trai 1 con gái là bao nhiêu. Chúng ta có thể lập mô hình xác suất với 4 sự kiện thành phần: 3 trai, 2 trai 1 gái, 1 trai 2 gái, 3 gái. Thế nhưng 4 sự kiện thành phần đó không "cân bằng" với nhau, và bởi vậy không kết luận được rằng xác suất của "2 trai 1 gái" là 1/4. Để có không gian xác suất với phân bố đều, ta có thể lập mô hình xác suất với 8 sự kiện thành phần như sau:

$$\Omega = \{TTT, TTG, TGT, TGG, GTT, GTG, GGT, GGG\}.$$

(Chẳng hạn, GGT có nghĩa là con thứ nhất là con gái, con thứ hai là con gái, con thứ ba là con trai). Sự kiện "2 trai một gái" là hợp của 3 sự kiện thành phần trong mô hình xác suất này: TTG, TGT, GTT. Như vậy xác suất của nó bằng 3/8.

Bài tập 1.3. Có một nhóm n bạn, trong đó có hai bạn Vôva và Lily. Xếp các bạn trong nhóm thành một hàng dọc một cách ngẫu nhiên. Hỏi xác suất để Vôva ở vị trí ngay sau Lily trong hàng là bao nhiêu ?

Bài tập 1.4. Một nhóm có 5 người, với 5 tên khác nhau. Mỗi người viết tên của một người khác trong nhóm một cách ngẫu nhiên vào giấy. Tính xác suất để có 2 người trong nhóm viết tên của nhau.

Bài tập 1.5. Giả sử trong một giải bóng đá đấu loại trực tiếp có 8 đội A,B,C,D,E,F,G,H tham gia: vòng 1 có 4 trận, vòng 2 có 2 trận, vòng 3 (vòng cuối cùng) có 1 trận. Giá sử xác suất để mỗi đội thắng mỗi trận đều là 1/2, và các đội bắt thăm để xem đội nào đấu với đội nào ở vòng đầu, các vòng sau thì được xếp theo kết quả vòng trước. Tính xác suất để đội A có đấu với đội B trong giải.

1.2.4 Mô hình xác suất với vô hạn các sự kiện

Mọi vấn đề xuất phát từ thực tế đều chỉ có một số hữu hạn các sự kiện thành phần. Nhưng khi mà số sự kiện thành phần đó lớn, thì người ta có thể dùng các mô hình toán học với vô hạn phần tử để biểu diễn, cho dễ hình dung và tiện tính toán.

Ví dụ 1.10. Nếu ta quan tâm đến lượng khách hàng trong một ngày của một siêu thị, thì có thể dùng tập hợp các số nguyên không âm $\mathbb{Z}_+$ làm không gian xác suất: mỗi số $n \in \mathbb{Z}_+$ ứng với một sự kiện "số khách trong ngày là n". Vấn đề tiếp theo là chọn phân bố xác suất nào trên $\mathbb{Z}_+$ cho hợp lý (phản ánh khá chính xác thực tế xảy ra, đồng thời lại tiện cho việc tính toán) ? Ví dụ người ta có thể dùng phân bố xác suất sau trên $\mathbb{Z}_+$, gọi là phân bố Poisson (đọc là Poa-Sông): $P(n) = e^{-\lambda}\dfrac{\lambda^n}{n!}$ với mọi $n \in \mathbb{Z}_+$. (Chú ý rằng $\displaystyle\sum_n P(n) = \sum_n e^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda}\sum_n \frac{\lambda^n}{n!} = e^{-\lambda}e^{\lambda} = 1$, như vậy các tiên đề về xác suất được thỏa mãn). Phân bố Poisson ứng với hai giả thuyết: lượng khách hàng trung bình trong một ngày là λ, và các khách hàng đi đến siêu thị một cách ngẫu nhiên và độc lập với nhau. Chúng ta sẽ tìm hiểu kỹ hơn về phân bố Poisson trong những phần sau.

Ví dụ 1.11. Ta biết rằng có một xe ô tô X đang đậu ở trên một khúc phố Z, và ta quan tâm đến vị trí của X trên phố đó. Ta có thể mô hình X bằng 1 điểm, Z bằng một đoạn thẳng và lấy đoạn thẳng đó làm không gian xác suất: $\Omega = [a, b]$, $a, b \in \mathbb{R}$, $a < b$. (Mô hình xác suất liên tục này có số phần tử là continuum, không đếm được). Sự kiện "ô tô đỗ ở chỗ nào đó trên khúc phố" chuyển thành sự kiện "điểm x nằm trong một đoạn thẳng con nào đó trên đoạn thẳng $\Omega = [a, b]$". Ta có thể chọn phân bố xác suất đều trên $\Omega = [a, b]$ theo nghĩa sau: xác suất của mỗi đoạn thẳng con trên Ω tỷ lệ thuận với độ dài của đoạn thẳng con đó, và bằng chiều dài của đoạn thẳng con đó chia cho chiều dài của Ω: $P([c, d]) = (d - c)/(b - a)$.

1.2.5 Ánh xạ giữa các không gian xác suất

Cùng một vấn đề tính toán xác suất, ta có thể lập nhiều mô hình không gian xác suất khác nhau. Ví dụ, mô hình xác suất đơn giản nhất cho sự kiện "bị ốm" sẽ là mô hình Bernoulli $\Omega_1 = \{S, H\}$ với 2 sự kiện S = "bị ốm" (sick) và H = "không bị ốm" (healthy). Như ta cũng có thể chia nhỏ sự kiện bị ốm ra thành rất nhiều sự kiện con, ví dụ như "ốm bệnh A", "ốm bệnh B", "ốm cả bệnh A lẫn bệnh B", v.v. và sự kiện "không bị ốm" cũng có thể chia thành nhiều sự kiện con, ví dụ như "rất khỏe", "không ốm nhưng mà yếu", v.v. Khi chia nhỏ như vậy, ta được mô hình xác suất với một không gian xác suất $\Omega_2 = \{S_1, S_2, \ldots, H_1, H_2, \ldots\}$ với nhiều phần tử hơn. Hai không gian đó liên quan với nhau bởi một ánh xạ $\phi : \Omega_1 \to \Omega_2$, $\phi(S_i) = S, \phi(H_i) = H$. Tất nhiên, khi ta chia nhỏ sự kiện S ra thành nhiều sự kiện (không giao nhau) $S_1, S_2, \ldots$, thì không phải vì thế mà xác suất của nó thay đổi. Nói cách khác, ta phải có

$$P(S) = P(\phi^{-1}(S)) = P(\cup_i S_i) = \sum_i P(S_i) \tag{1.12}$$

Tính chất trên là tính chất bảo toàn xác suất của ánh xạ ϕ. Nói một cách tổng quát, ta có định nghĩa sau:

Định nghĩa 1.3. *Một ánh xạ $\phi : (\Omega_1, P_1) \to (\Omega_2, P_2)$ từ một không gian xác suất (Ω_1, P_1) vào một không gian xác suất (Ω_2, P_2) được gọi là một* **ánh xạ bảo toàn xác suất** *nếu nó bảo toàn độ đo xác suất, có nghĩa là với mọi tập con $B \subset \Omega_2$ đo được, ta có*

$$P_1(\phi^{-1}(B)) = P_2(B) \tag{1.13}$$

Nếu hơn nữa, ϕ là một song ánh modulo những tập có xác suất bằng 0, có nghĩa là tồn tại các tập con $A \in \Omega_1$, $B \in \Omega_2$ sao cho $P_1(A) = P_2(B) = 0$ và $\phi : \Omega_1 \setminus A \to \Omega_2 \setminus B$ là *song ánh bảo toàn xác suất), thì ϕ được gọi là một* **đẳng cấu xác suất** *, và ta nói rằng (Ω_1, P_1) đẳng cấu xác suất với (Ω_2, P_2).*

Ví dụ 1.12. Đặt 4 bạn Al, Ben, Cam, Don ngồi vào 4 ghế A, B, C, D một cách hoàn toàn ngẫu nhiên. Tính xác suất để Al được đặt ngồi vào ghế A. Có 4 ghế, và xác suất để Al ngồi vào mỗi nghế trong 4 ghế đó coi là bằng nhau (vì không có cớ gì để coi là khác nhau), bởi vậy xác suất để Al ngồi vào ghế A là 1/4. Nhưng cũng có thể lý luận tỷ mẩn hơn như sau: có tổng cộng $4! = 24$ cách đặt 4 bạn ngồi vào 4 ghế, trong đó có $3! = 6$ cách có Al ngồi vào ghế A. Bởi vậy xác suất để Al ngồi vào ghế A là $6/24 = 4$. Hai cách giải cho cùng một đáp số, nhưng sử dụng hai không gian xác suất khác nhau: không gian thứ nhất có 4 phần tử, còn không gian thứ hai có 24 phần tử. Có một phép chiếu tự nhiên bảo toàn xác suất từ không gian thứ hai lên không gian thứ nhất.

Định lý 1.1. *Nếu (Ω_1, P_1) là một không gian xác suất, và $\phi : \Omega_1 \to \Omega_2$ là một ánh xạ tùy ý, thì tồn tại một độ đo xác suất P_2 trên Ω_2, sao cho ánh xạ $\phi : (\Omega_1, P_1) \to (\Omega_2, P_2)$ là ánh xạ bảo toàn xác suất.*

Chứng minh. Có thể xây dựng P_2 theo công thức sau: với mỗi tập con $B \subset \Omega_2$, nếu tồn tại $P_1(\phi^{-1}(B))$ thì ta đặt

$$P_2(B) := P_1(\phi^{-1}(B)) \tag{1.14}$$

Độ đo xác suất P_2 định nghĩa theo công thức trên được gọi là **push-forward** của P_1 qua ánh xạ ϕ, hay còn gọi là **phân bố xác suất cảm sinh** từ P_1 qua ánh xạ ϕ. □

Bài tập 1.6. Chứng minh rằng quan hệ đẳng cấu xác suất giữa các không gian xác suất là một quan hệ tương đương.

1.2.6 Tích của các không gian xác suất

Nếu M và N là hai tập hợp, thì tích của chúng (hay còn gọi là tích trực tiếp, hay tích Descartes), ký hiệu là $M \times N$, là tập hợp các cặp phần tử (x, y), $x \in M, y \in N$. Trong trường hợp $M = (\Omega_1, P_1)$ và $N = (\Omega_2, P_2)$ là hai không gian xác suất, thì tích $\Omega_1 \times \Omega_2$, cũng có một độ đo xác suất P, được xác định một cách tự nhiên bởi P_1 và P_2 bằng công thức sau: Nếu $A_1 \subset \Omega_1$ và $A_2 \subset \Omega_2$ nằm trong các sigma-đại số tương ứng của P_1 và P_2 thì:

$$P(A_1 \times A_2) = P_1(A_1) \times P_2(A_2). \tag{1.15}$$

Sigma-đại số của P chính là sigma đại số sinh bởi các tập con của $\Omega_1 \times \Omega_2$ có dạng $A_1 \times A_2$ như trên. Khi ta nói đến tích trực tiếp của hai không gian xác suất, ta sẽ hiểu là nó đi kèm độ đo xác suất được xác định như trên.

Tương tự như vậy, ta có thể định nghĩa tích trực tiếp của n không gian xác suất, hay thậm chí tích trực tiếp của một dãy vô hạn các không gian xác suất.

Định lý 1.2. *Hai phép chiếu tự nhiên từ tích $(\Omega_1, P_1) \times (\Omega_2, P_2)$ của hai không gian xác suất xuống (Ω_1, P_1) và (Ω_2, P_2) là hai ánh xạ bảo toàn xác suất.*

Ví dụ 1.13. Lấy 1 đồng xu tung 3 lần, mỗi lần hiện lên S (sấp) hoặc N (ngửa). Không gian xác suất các sự kiện ở đây là không gian các dãy 3 chữ cái mà mỗi chữ cái là S hay N: $\Omega = \{SSS, SSN, SNS, SNN, NSS, NSN, NNS, NNN\}$. Ký hiệu $(\Omega_k = \{S_k, N_k\}, P_k)$ là không gian xác suất của mặt hiện lên trong lần tung thứ k. Ta giả sử các kết quả của các lần tung là độc lập với nhau (tức là kết quả lần trước không ảnh hưởng đến kết quả của các lần sau), khi đó Ω có thể coi là tích trực tiếp của các không gian xác suất $(\Omega_k = \{S_k, N_k\}, P_k)$. Giả sử đồng xu là "cân bằng", hai mặt sấp ngửa có xác suất hiện lên giống nhau trong mỗi lần tung. Khi đó các không gian $(\Omega_k = \{S_k, N_k\}, P_k)$ là đẳng cấu với nhau và với một không gian xác suất Bernoulli với tham số $p = 1/2$. Ta có thể viết: $\Omega = \{S, N\}^3$

Ví dụ 1.14. Trong ví dụ trên, nếu thay vì chỉ tung đồng xúc sắc có 3 lần, ta hình dùng la ta tung vô hạn lần (trong thực tế không làm được như vậy, nhưng cứ giả sử ta có vô hạn thời gian và làm được như vậy). Khi đó mỗi sự kiện được có thể được đánh dấu bằng một dãy vô hạn các chữ cái mà mỗi chữ là S hoặc N, và không gian xác suất là

$$\Omega = \{S, N\}^{\mathbb{N}}$$

Ta có thể xây dựng một ánh xạ bảo toàn xác suất sau từ $\{S, N\}^{\mathbb{N}}$ vào đoạn thẳng $[0, 1]$ với phân bố xác suất đều trên đó:

$$\phi((M_i)_{i \in \mathbb{N}}) := \sum_{i=1}^{\infty} \chi(M_i)/2^i$$

Ở đây mỗi M_i là S hoặc N, và $\chi(N) = 0, \chi(S) = 1$. Ánh xạ

$$\phi : \{S, N\}^{\mathbb{N}} \to [0, 1]$$

xây dựng như trên không phải là một song ánh, nhưng nó là một đẳng cấu xác suất !

Hình 1.5: Blaise Pascal (1623-1662)

Ví dụ 1.15. **Bài toán Méré**. Hiệp sĩ de Méré (tên khai sinh là Antoine Gombaud (1607-1684), là nhà văn và nhà triết học người Pháp) là một nhân vật lịch sử nghiện đánh bạc. Ông ta hay chơi xúc sắc, và nhận thấy rằng trong hai sự kiện sau:
A = "Tung một con xúc sắc 4 lần, có ít nhất 1 lần hiện lên 6", và
B = "Tung một đôi xúc sắc 24 lần, có ít nhất 1 lần hiện lên một đôi 6",
thì B ít xảy ra hơn A. Tuy nhiên ông ta không giải thích được tại sao. Theo ông ta thì đáng nhẽ hai sự kiện đó phải có khả năng xảy ra bằng nhau, vì $24 = 6 \times 4$. Ông ta bèn hỏi bạn mình là nhà toán học và triết học Blaise Pascal (1623-1662), vào năm 1654. Pascal lúc đó đã "từ bỏ toán", nhưng có nhận lời suy nghĩ về câu hỏi của de Méré. Sau đó Pascal viết thư trao đổi với Pierre de Fermat (159?-1665), một luật sư đồng thời là nhà toán học ở vùng Toulouse (Pháp). Hai người cùng nhau phát minh ra *lý thuyết xác suất cổ điển*, và giải được bài toán của de Méré. Kết quả là: $P(A) = 1 - P(\overline{A}) = 1 - (1-1/6)^4 \approx 0,5177$, và $P(B) = 1 - P(\overline{B}) = 1 - (1-(1/6)^2)^{24} \approx 0,4914$.

Bài tập 1.7. Chứng minh định lý 1.2.

1.2.7 Phân bố nhị thức

Phân bố nhị thức là một trong những phân bố hay gặp nhất, và nó là một ví dụ về sự xuất hiện các phép toán tổ hợp trong xác suất thống kê.

Định nghĩa 1.4. Phân bố nhị thức *với các tham số n, p ($n \in \mathbb{N}, 0 \leq p \leq 1$) là phân bố xác suất*

$$P(k) = C_n^k p^k (1-p)^{n-k} \tag{1.16}$$

Hình 1.6: Fermat và "nàng toán". Tượng ở Toulouse

trên tập hợp $\Omega = \{0, 1, 2, \ldots, n\}$.

Ở đây, $C_n^k = \dfrac{n!}{k!(n-k)!}$ là nhị thức Newton. Ý nghĩa tổ hợp của C_n^k là: nó là số các tập con có đúng k phần tử trong một tập hợp có n phần tử, hay nói cách khác, nó là số cách chọn ra một nhóm con với k phần tử, từ một nhóm có n phần tử.

Nhắc lại rằng ta có công thức đại số quen thuộc sau:

$$(x+y)^n = \sum_{k=0}^{n} C_n^k x^k y^{n-k}. \tag{1.17}$$

Nếu thay x bằng p và y bằng $1-p$ trong công thức trên, thì ta có $\sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} = 1$, chứng tỏ định nghĩa phân bố xác suất nhị thức trên phù hợp với các tiên đề về xác suất.

Ý nghĩa của phân bố nhị thức như sau: Khi ta làm n lần một phép thử nào đó, và mỗi lần thì xác suất xảy ra kết quả A nào đó là p (ví dụ: một người bắn súng n lần, xác suất trúng đích mỗi lần là p), và giả sử là kết quả của các lần thử khác nhau độc lập với nhau (lần thử này không ảnh hưởng đến lần thử khác), thì tổng số lần xảy ra kết quả A trong số n lần đó là một số nguyên nằm giữa 0 và n, và với mỗi $k = 0, 1, 2, \ldots, n$, xác sất của sự kiện "số lần ra kết quả A là k" bằng $C_n^k p^k (1-p)^{n-k}$.

Thật vậy, nếu ta lấy không gian xác suất cho mỗi phép thử là không gian $\{A, \overline{A}\}$, thì không gian xác suất các trường hợp của n lần thử là $\{A, \overline{A}\}^n$ (các phần tử của không gian này là các dãy n kết quả, mà mỗi kết quả là A hoặc $\overline{A}$. Có C_n^k phần tử của không gian $\{A, \overline{A}\}^n$ có chứa đúng k kết quả A và $(n-k)$ kết quả $\overline{A}$. Xác suất của mỗi phần tử đó là $p^k(1-p)^k$ theo công thức tích của xác suất. Bởi vậy xác suất của sự kiện "kết quả A xảy ra k lần" số phần tử của sự kiện này (hiểu như là một tập con của không gian xác suất) nhân với xác suất của một phần tử (vì các phần tử này có cùng xác suất), và bằng $C_n^k p^k (1-p)^{n-k}$.

Bài tập 1.8. Hai vận động viên Nam và Tiến chơi một trận tennis. Ai thắng được 3 set trước thì thắng cả trận. Giả sử xác suất để Nam thắng mỗi set là 40% (để Tiến thắng mỗi set là 60%, và kết quả của set này không ảnh hưởng đến set khác). Hỏi xác suất để Nam thắng trận tennis là bao nhiêu ?

1.3 Xác suất có điều kiện

1.3.1 Định nghĩa xác suất có điều kiện

Như chúng ta đã biết, xác suất của một sự kiện có thể phụ thuộc vào nhiều yếu tố, điều kiện khác nhau. Để chỉ ra một cách cụ thể hơn về việc xác suất của một sự kiện A nào đó phụ thuộc vào một điều kiện B nào đó ra sao, người ta đưa ra khái niệm xác suất có điều kiện. Điều kiện B cũng có thể hiểu là một sự kiện, tức là sự kiện “có B”.

Định nghĩa 1.5. *Giả sử (trong một không gian xác suất nào đó) điều kiện B có xác suất khác không, $P(B) > 0$, thì* **xác suất của sự kiện A dưới điều kiện** *B, ký hiệu là $P(A|B)$, được định nghĩa như sau:*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{1.18}$$

Một hệ quả trực tiếp của định nghĩa xác suất có điều kiện là công thức tích sau đây:

$$P(A \cap B) = P(A|B).P(B). \tag{1.19}$$

Tất nhiên, ta cũng có thể coi B là sự kiện, A là điều kiện, và khi đó ta có $P(A \cap B) = P(B|A).P(A)$

Ví dụ 1.16. Một lớp học có 30 bạn, trong đó có 17 bạn nữ và 13 bạn nam. Có 3 bạn tên là Thanh, trong đó có 1 bạn nữ và 2 bạn nam. Thầy giáo gọi ngẫu nhiên 1 bạn lên bảng. Xác suất để bạn đó có tên là Thanh sẽ là 1/10. Nhưng với điều kiện “đó là bạn nữ” thì xác suất để bạn đó tên là Thanh là 1/17. Sự kiện ở đây là $A =$ “tên là Thanh”, và điều kiện là B = "nữ". Không gian xác suất Ω có 30 phần tử, với phân bố xác suất đều. A có 3 phần tử, B có 17 phần tử, và $A \cap B$ có 1 phần tử. Bởi vậy: $P(A) = \frac{\#A}{\#\Omega} = 3/30 = 1/10$; $P(A|B) = P(A \cap B)/P(B) = (1/30)/(17/30) = 1/17$. Chú ý rằng, trong ví dụ này ta có $P(A|B) \neq P(A)$. Vẫn ví dụ này, nếu thầy giáo gọi 1 bạn có tên là Thanh lên bảng, thì xác suất để bạn đó là bạn nữ là bao nhiêu ? Lời giải: trong 3 bạn Thanh có 1 bạn là nữ, bởi vậy xác suất là 1/3. Sử dụng công thức $P(A \cap B) = P(B|A).P(A)$ với xác suất có điều kiện, ta cũng có $P(B|A) = P(A \cap B)/P(A) = (1/30)/(1/10) = 1/3$. (Câu hỏi: Vì sao hai cách giải khác nhau lại ra kết quả giống nhau ?)

Ghi chú 1.3. Có thể giải thích ý nghĩa triết lý và toán học của định nghĩa xác suất có điều kiện như sau: Sự kiện A cùng với điều kiện B chính là sự kiện $A \cap B$, tức là "cả A và B cùng xảy ra". Ta có thể coi A và B là hai tập con của một không gian xác suất Ω ban đầu. Các tập con của B chính là các sự kiện với điều kiện B được thỏa mãn. Khi chúng ta đặt điều kiện B, thì tức là chúng ta đã hạn chế không gian xác suất từ Ω xuống còn B, và hạn chế các sự kiện A xuống còn $A \cap B$. Xác suất của A với điệu kiện B chính là xác suất của $A \cap B$ trong không gian xác suất mới B với một độ đo xác suất P_1: $P(A|B) = P_1(A \cap B)$. Độ đo xác suất P_1 không tùy ý, mà nó được sinh ra bởi độ đo xác suất P ban đầu, theo nguyên tắc "bình quân": nếu C và D là hai tập con của B (tức là 2 sự kiện thỏa mãn điều kiện B) với cùng xác suất, $P(C) = P(D)$, thì ta cũng phải coi rằng chúng có cùng xác suất có điều kiện: $P_1(C) = P_1(D)$. Một cách tổng quát hơn, ta có công thức tỷ lệ thuận: $P(C)/P(D) = P_1(C)/P_1(D)$ nếu C và D là hai tập con của B. Từ đó suy ra: $P(A \cap B)/P(B) = P_1(A \cap B)/P_1(B) = P_1(A \cap B) = P(A|B)$ (bởi vì $P_1(B) = 1$).

Ví dụ 1.17. Theo một con số thống kê ở Mỹ năm 2007, có khoảng 40% các vụ tai nạn xe cộ gây chết người là có người lái say rượu. Giá sử tỷ lệ số người say rượu khi lái xe là 4%. Hỏi việc xay rượu khi lái xe làm tăng khả năng gây tai nạn chết người lên bao nhiêu lần ? Nói cách khác, chúng ta muốn tính tỷ lệ $P(A|S)/P(A)$, ở đây A là sự kiện "lái xe xảy ra tai nạn chết người", S là điều kiện "người lái say rượu". Từ công thức $P(A \cap S) = P(A|S).P(S) = P(S|A).P(A)$ ta có $P(A|S)/P(A) = P(S|A)/P(S) = 40\%/4\% = 10$, tức là việc say rượu khi lái xe có thể làm tăng khả năng gây tai nạn xe cộ chết người lên 10 lần.

Bài tập 1.9. Có hai sự kiện A và B với xác suất lớn hơn 0. Khi nào thì ta có $P(A|B) = P(B|A)$?

Bài tập 1.10. Ta biết rằng một nhà nọ có 3 con mèo, trong đó có ít nhất 1 con là mèo cái. Hỏi rằng xác suất để cả 3 con mèo đều là mèo cái là bao nhiêu ?

1.3.2 Sự độc lập và phụ thuộc của các sự kiện

Thế nào là hai sự kiện độc lập với nhau ? Về mặt triết lý, hai sự kiện độc lập là hai sự kiện không liên quan gì đến nhau. Ví dụ, tôi không liên quan gì đến đội bóng đá Barcelona. Đội đó đá thắng hay thua tôi cũng không quan tâm, không ảnh hưởng gì đến việc tôi có phải đi chợ hay không. Hai sự kiện "tôi đi chợ" và "đội Barcelona thắng" có thể coi là độc lập với nhau. Nếu hai sự kiện A và B độc lập với nhau, thì việc có xảy ra hay không sự kiện B không ảnh hưởng gì đến việc có xảy ra hay không sự kiện A. Nói cách khác, xác suất của A với điều kiện B không khác gì xác suất của A khi không tính đến điều kiện B. Đấy chính là định nghĩa trong lý thuyết xác suất về sự độc lập của hai sự kiện:

Định nghĩa 1.6. *Sự kiện A được gọi là* **độc lập** *với sự kiện B nếu như*

$$P(A) = P(A|B) = P(A \cap B)/P(B), \tag{1.20}$$

hay viết cách khác:

$$P(A \cap B) = P(A).P(B) \tag{1.21}$$

Ghi chú 1.4. Công thức $P(A|B) = P(A)$ tương đương với công thức $P(A \cap B) = P(A).P(B)$ và tương đương với $P(B|A) = P(B)$. Điều đó có nghĩa là quan hệ *độc lập* là một quan hệ đối xứng: nếu A độc lập với B thì B độc lập với A, và chúng ta có thể nói là A và B độc lập với nhau. Trong công thức $P(A|B) = P(A)$ ta phải giả sử là $P(B) \neq 0$. Kể cả khi $P(B) = 0$ thì công thức $P(A \cap B) = P(A).P(B)$ vẫn có thể dùng làm định nghĩa được, và khi đó nó hiển nhiên đúng: một sự kiện có xác suất bằng 0 thì độc lập với mọi sự kiện khác.

Tổng quát hơn, giả sử ta có một họ $\mathcal{M}$ (hữu hạn hoặc vô hạn) các sự kiện.

Định nghĩa 1.7. *Họ $\mathcal{M}$ được gọi là một* **họ các sự kiện độc lập**, *nếu như với bất kỳ số tự nhiên k nào và bất kỳ k sự kiện $A_1, \ldots, A_k$ khác nhau nào trong họ $\mathcal{M}$ ta cũng có:*

$$P\left(\bigcap_{i=1}^{k} A_i\right) = \prod_{i=1}^{k} P(A_i). \tag{1.22}$$

Nếu như $P(A \cap B) = P(A).P(B)$ với bất kỳ hai sự kiện khác nhau nào trong họ $\mathcal{M}$ (tức là đẳng ta chỉ yêu cầu đẳng thức trên đúng trong trường hợp $k = 2$, thì họ $\mathcal{M}$ được gọi là họ các sự kiện độc lập từng đôi một.

Ghi chú 1.5. Tất nhiên nếu ta có một họ các sự kiện độc lập, thì các sự kiện trong họ độc lập từng đôi một với nhau. Nhưng điều ngược lại không đúng: Có những họ không độc lập, mà trong đó các sự kiện độc lập từng đôi một với nhau !

Ví dụ 1.18. Tung 1 xúc sắc 2 lần, được 2 số ký hiệu là a, b. Xét 3 sự kiện sau: X là sự kiện "$a + b$ là số chẵn", Y là sự kiện "$a = 1$" và Z là sự kiện "$b = 4$". Ở đây không gian xác suất là không gian có $6^2 = 36$ phần tử, mỗi phần tử là một cặp số (a, b), mỗi số có thể nhận 1 trong 6 giá trị 1,2,3,4,5,6. Ta có thể giả sử không gian xác suất này có phân bố xác suất đều (2 lần tung độc lập với nhau). Khi đó dễ dàng kiểm tra rằng các sự kiện X, Y, Z độc lập từng đôi một với nhau, thế nhưng họ 3 sự kiên $\{X, Y, Z\}$ không phải là một họ độc lập: $P(X \cap Y \cap Z) = 0$ trong khi $P(X).P(Y).P(Z) = (1/2).(1/6).(1/6) \neq 0$

Nếu như hai sự kiện không độc lập với nhau, thì người ta nói là chúng phụ thuộc vào nhau. Do tính chất đối xứng, nếu sự kiện A phụ thuộc vào sự kiện B thì B cũng phụ thuộc vào A. Nếu như $P(A|B) > P(B)$ thì ta có thể nói là điều kiện B *thuận lợi* cho sự kiện A, và ngược lại nếu $P(A) < P(A|B)$ thì điều kiện B *không thuận lợi* cho sự kiện A.

Công thức $P(A|B)P(B) = P(B|A)P(A)$ tương đương với công thức

$$P(A|B)/P(A) = P(B|A)/P(B), \tag{1.23}$$

có thể được suy diễn như sau: B thuận lợi cho A (tức là $P(A|B)/P(B) > 1$) thì A cũng thuận lợi cho B và ngược lại.

Ví dụ 1.19. Giả sử cứ 5 học sinh thì có 1 học sinh giỏi toán, cứ 3 học sinh thì có 1 học sinh giỏi ngoại ngữ, và trong số các học sinh giỏi toán thì cứ 2 học sinh có 1 học sinh giỏi ngoại ngữ (lớn hơn tỷ lệ trung bình). Khi đó trong số các học sinh giỏi ngoại ngữ, tỷ lệ học sinh giỏi toán là 30% (cũng lớn hơn tỷ lệ trung bình): $(1/2)/(1/3) = 30\%/(1/5)$.

Bài tập 1.11. Chứng minh rằng nếu một sự kiện A độc lập với sự kiện B, thì nó cũng độc lập với sự kiện $\overline{B}$.

Bài tập 1.12. Tìm một ví dụ với 3 sự kiện A, B, C sao cho A độc lập với hai sự kiện B và C, nhưng không độc lập với $B \cap C$.

Bài tập 1.13. Lấy một bộ bài tú lơ khơ 52 quân, và rút ra từ đó 2 lần mỗi lần 1 quân, để được 2 quân. Gọi A là sự kiện "quân rút ra đầu tiên là quân nhép" và B là sự kiện "quân rút ra thứ hai là quân cơ". Hỏi hai sự kiện A và B có độc lập với nhau không ?

1.3.3 Công thức xác suất toàn phần

Định nghĩa 1.8. *Một họ các tập con $B_1, \ldots, B_n$ của không gian xác suất Ω là một* **phân hoạch** *(partition) của Ω nếu như các tập B_i đôi một không giao nhau, và hợp của chúng bằng Ω:*

$$B_i \cap B_j = \emptyset \ \forall \ i \neq j \ , \ \cup_{i=1}^{n} B_i = \Omega. \tag{1.24}$$

Nếu như ta chưa biết xác suất $P(A)$ của một sự kiện A nào đó, nhưng biết các xát suất $P(B_i)$ của một phân hoạch $(B_1, \ldots, B_n)$ của không gian xác suất, và biết các xác suất có điều kiện $P(A|B_i)$, thì ta có thể dùng công thức sau, gọi là **công thức xác suất toàn phần** (total probability formula), để tính xác suất của A:

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i).P(B_i). \tag{1.25}$$

Trường hợp riêng của công thức trên là khi ta có hai sự kiện A, B, có thể sử dụng phân hoạch $(B, \overline{B} = \Omega \setminus B)$ hai thành phần của Ω để tính xác suất của A:

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

Bài tập 1.14. Theo một số liệu thống kê, năm 2004 ở Canada có 65,0% đàn ông là thừa cân[(1)], và 53,4% đàn bà thừa cân. Số đàn ông và đàn bà ở Canada coi như bằng nhau. Hỏi rằng, trong năm 2004, xác suất để một người Canada được chọn ngẫu nhiên là người thừa cân bằng bao nhiêu ?

1.3.4 Công thức Bayes

Công thức Bayes, mang tên của linh mục và nhà toán học người Anh Thomas Bayes (1702-1761), là công thức ngược, cho phép tính xác suất có điều kiện $P(B|A)$ khi biết xác suất có điều kiện $P(A|B)$ và một số thông tin khác. Dạng đơn giản nhất của công thức này là: Nếu A, B là hai sự kiện bất kỳ với xác suất khác 0 thì ta có:

$$P(B|A) = \frac{P(A|B).P(B)}{P(A)}. \tag{1.27}$$

[(1)] Theo định nghĩa của các tổ chức y tế, những người có chỉ số trọng lượng cơ thể (body mass index) ≥ 25 được gọi là thừa cân (overweight or obese), trên 30 được gọi là béo phì (obese), trên 40 là béo bệnh hoạn (morbidly obese). Chỉ số trọng lượng cơ thể được tính ra từ chiều cao và cân nặng theo công thức: BMI = trọng lượng (tính theo kg) chia cho chiều cao (tính theo mét) bình phương

Công thức trên là hệ quả trực tiếp của công thức $P(B|A).P(A) = P(A|B).P(B) = P(A \cap B)$ đã được bàn đến ở những phần trước. Kết hợp công thức trên với công thức xác suất toàn phần cho $P(A)$, ta được:

Định lý 1.3. *Giả sử* $(B_1, \ldots, B_n)$ *là một phân hoạch của không gian xác suất. Khi đó ta có* **công thức Bayes** *sau:*

$$P(B_k|A) = \frac{P(A|B_k).P(B_k)}{P(A)} = \frac{P(A|B_k).P(B_k)}{\sum_{i=1}^{n} P(A|B_i).P(B_i)}. \tag{1.28}$$

với mọi $k = 1, 2, \ldots, n$.

Hình 1.7: Thomas Bayes (1702-1761)

Công thức Bayes rất đơn giản nhưng nó có ý nghĩa rất sâu xa. Một trong những lỗi mà rất nhiều người mắc phải, là lẫn lộn giữa $P(A|B)$ và $P(B|A)$, coi hai con số đó như là bằng nhau. Nhưng công thức Bayes cho thấy hai con số đó có thể chênh lệch nhau rất nhiều, nếu như $P(A)$ và $P(B)$ chênh lệch nhau rất nhiều ! Dưới đây là một ví dụ minh họa điều đó.

Ví dụ 1.20. Đây là một bài toán được 3 nhà toán học Cassels, Shoenberger và Grayboys đem đố 60 sinh viên và cán bộ y khoa tại Harvard Medical School năm 1978[(2)]. Giả sử có một loại bệnh mà tỷ lệ người mắc bệnh là 1/1000. Giả sử có một loại xét nghiệm, mà ai mắc bệnh khi xét cũng ra phản ứng dương tính, nhưng tỷ lệ phản ứng dương tính nhầm (false positive) là 5% (tức là trong số những người không bị bệnh có 5% số người thử ra phản ứng dương tính). Hỏi khi một người xét nghiệm bị phản ứng dương tính, thì khả

[(2)]Nguồn: Cassels, Schoenberger and Grayboys, Interpretation by physicians of clinical laboratory results. New England Journal of Medicine, 299 (1978), 999-1000 năng mắc bệnh của người đó là bao nhiêu ? Theo bạn là bao nhiêu ? Hãy thử tự tìm câu trả lời trước khi đọc tiếp.

Nếu bạn trả lời 95% (= 100% - 5%), thì câu trả lời của bạn cũng giống câu trả lời của phần lớn những người khác được hỏi. Ta hãy thử phân tích kỹ thêm về câu hỏi này. Nếu ký hiệu K là sự kiên "không bị bệnh" và D là sự kiện phản ứng dương tính, thì con số 5% là con số $P(D|K)$ (xác suất có phản ứng dương tính khi mà không bị bệnh) chứ không phải $P(K|D)$ (xác suất không bị bệnh khi mà có phản ứng dương tính). Để tính $P(K|D)$, ta dùng công thức Bayes $P(K|D) = \dfrac{P(D|K).P(K)}{P(D|\bar{K}).P(\bar{K}) + P(D|K).P(K)}$. Ta có $P(D|K) = 5/100$, $P(K) = 1 - 1/1000 = 999/1000$, và $P(D|\bar{K}).P(\bar{K}) + P(D|K).P(K) = (1).(1/1000) + (5/100).(999/1000) = 51/1000$ (tính xấp xỉ), và bởi vậy: $P(K|D) = (5/100).(999/1000)/(51/1000) \approx 98\%$. Như vậy trong số những người xét nghiệm ra dương tính, có khoảng 98% số người là không bị bệnh. Nói cách khác, khi xét nghiệm ra dương tính, xác suất để thực sự mắc bệnh chỉ có 2% !

Bài tập 1.15. Được biết có 5% đàn ông bị mù màu, và 0,25% đàn bà bị mù màu. Giả sử số đàn ông bằng số đàn bà. Chọn 1 người bị mù màu một cách ngẫu nhiên. Hỏi rằng xác suất để người đó là đàn ông là bao nhiêu ?

1.4 Một số nghịch lý trong xác suất

Tính toán xác suất là một vấn đề nhiều khi hết sức tế nhị. Kể cả trong những bài toán tưởng chừng như rất đơn giản, cũng có thể tính ra kết quả sai mà khó phát hiện sai ở đâu. Phần này sẽ gồm một số "nghịch lý" trong xác suất để minh họa điều đó . Những nghịch lý này cho thấy chúng ta cần hết sức cẩn thận trong lúc lập mô hình tính toán xác suất, đặc biệt là xác suất có điều kiện, kiểm tra lại những điều tưởng chừng như hiển nhiên, để tránh sai lầm.

1.4.1 Nghịch lý 1 (Nghịch lý Simpson). Thuốc nào tốt hơn ?

Một người nghiên cứu muốn xác định xem giữa 2 loại thuốc cùng để chữa 1 bệnh, loại nào tốt hơn. Kết quả thống kê về lượng người chữa được khỏi bệnh, phân biệt theo giới tính, được viết dưới đây

Giới tính: Nữ	Thuốc I	Thuốc II
Chữa được	150	15
Không chữa được	850	285

Giới tính: Nam	Thuốc I	Thuốc II
Chữa được	190	720
Không chữa được	10	180

Dựa vào bảng thống kê trên, có 2 câu trả lời trái ngược nhau như sau cho câu hỏi thuốc nào tốt hơn:
1) Thuốc I đem cho 1200 người dùng, chữ được bệnh cho 340 người. Thuốc II đem cho 1200 người dùng, chữa được 735 người, như vậy thuôc II tốt hơn.
2) Đối với nữ, tỷ lệ chữa được bệnh của Thuốc I là 15%, của Thuốc II là 5%. Đối với nam, tỷ lệ chữa được bệnh của thuốc I là 95%, của thuốc II là 80%. Trong cả hai trường hợp thì tỷ lệ chữa được bệnh của thuốc I cao hơn, vậy nên thuốc I tốt hơn.

Trong hai câu trả lời trên câu trả lời nào đáng tin? Vì sao ? Nghịch lý nằm ở đâu ?

1.4.2 Nghịch lý 2. Hoàng tử có chị em gái không ?

Biết rằng cha mẹ của 1 hoàng tử có 2 con. Hỏi xác suất để hoàng tử đó có sister (chị gái hoặc em gái) là bao nhiêu ? Có 2 đáp án sau:
1) Hoàng tử có 1 người anh chị em ruột. Có hai khả năng: hoặc người đó là con trai, hoặc là con gái. Như vậy xác suất để người đó là con gái (tức là hoàng tử có sister) là 1/2.
2) Có 4 khả năng cho 1 gia đình có 2 con: {B,B}, {B,G}, {G,B}, {G,G}. (B = boy = con trai, G = girl = con gái, xếp theo thứ tự con thứ nhất - con thứ hai). Vì ta biết hoàng tử là con trai (đây là điều kiện) nên loại đi khả năng {G,G}, còn 3 khả năng {B,B}, {B,G}, {G,B}. Trong số 3 khả năng đó thì có 2 khả năng có con gái. Như vậy xác suất để hoàng tử có sister là 2/3.

Trong hai đáp án trên, ắt hẳn phải có (ít nhất) 1 đáp án sai. Thế nhưng cái nào sai, sai ở chỗ nào ?

1.4.3 Nghịch lý 3. Văn Phạm có phải là thủ phạm ?

Một người đàn ông tên là Văn Phạm bị tình nghi là thủ phạm trong một vụ án. Cảnh sát điều tra được những tin sau đây: 1) ngoài nạn nhân chỉ có 2 người có mặt lúc xảy ra vụ án, một trong hai người đó là Văn Phạm, người kia cảnh sát không hề biết là ai, và một trong hai người đó là thủ phạm; 2) thủ phạm phải là đàn ông. Hỏi xác suất để "Văn Phạm là thủ phạm" là bao nhiêu ?

Gọi người thứ hai mà cảnh sát không biết là ai là "X". X có thể là đàn ông hoặc đàn bà. Ta gọi sự kiện "Văn Phạm là thủ phạm" là A, sự kiện "X là đàn ông" là B, "thủ phạm là đàn ông" là C. Có hai cách giải khác nhau như sau:

1) Theo công thức xác suất toàn phần ta có $P(A) = P(A|B).P(B) + P(A|\overline{B}).P(\overline{B})$ Nếu X là đàn bà thì X không thể là thủ phạm và Văn Phạm phải là thủ phạm, bởi vậy $P(A|\overline{B}) = 1$. Nếu X là đàn ông thì một trong hai người, X hoặc Văn Phạm, là thủ phạm, bởi vậy $P(A|B) = 1/2$. X có thể là đàn ông hoặc đàn bà, và ta coi số đàn ông bằng số đàn bà, bởi vậy $P(B) = P(\overline{B}) = 1/2$. Từ đó ta có $P(A) = (1/2).(1/2) + 1.(1/2) = 3/4$, có nghĩa là xác suất để "Văn Phạm là thủ phạm" bằng 3/4.

2) Ta coi C là điều kiện, và muốn tính xác suất có điều kiện $P(A|C)$ (xác suất để Văn Phạm là thủ phạm, khi biết rằng thủ phạm là đàn ông). Theo công thức Bayes ta có $P(A|C) = \dfrac{P(C|A).P(A)}{P(C|A).P(A) + P(C|\overline{A}).P(\overline{A})}$. Ở trong công thức trên, $P(A)$ là xác suất của sự kiện "Văn Phạm là thủ phạm" nếu như chưa có điều kiện "thủ phạm là đàn ông". Vì một trong hai người Văn Phạm và X là thủ phạm, nên xác suất $P(A)$ không có điều kiện ở đây là $P(A) = 1/2$. Ta có $P(C|A) = 1$ vì tất nhiên nếu Văn Phạm là thủ phạm thì thủ phạm là đàn ông. Ngược lại, $P(C|\overline{A}) = 1/2$ (nếu X là thủ phạm, thì thủ phạm có thể là đàn ông hoặc đàn bà, khi mà chưa đặt điều kiện "thủ phạm là đàn ông"). Bởi vậy ta có: $P(A|C) = \frac{1.(1/2)}{1.(1/2)+(1/2).(1/2)} = \frac{1/2}{3/4} = 2/3$, tức là xác suất để Văn Phạm là thủ phạm bằng 2/3.

Hai cách giải trên cho 2 đáp số khác nhau, như vậy (ít nhất) một trong hai cách giải trên là sai. Cách giải nào sai và sai ở chỗ nào ?

1.4.4 Lời giải cho các nghịch lý

Nghịch lý 1. Vấn đề nằm ở chỗ Thuốc I được đem thử cho quá it nam, quá nhiều nữ so với thuốc II, nên khi lấy tổng số các kết quả của các phép thử thì nó thiên vị thuốc II và không phản ánh đúng tỷ lệ chữa được bệnh. Kết luận 1) là sai và kết luận 2) đáng tin hơn.

Nghịch lý 2. Nghịch lý này có trong 1 quyển giáo trình tiếng Anh về xác suất. Điều đáng ngạc nhiên là tác giả của giáo trình đó nói rằng đáp án thứ hai đúng (tức là xác suất = 2/3) và đáp án thứ nhất sai. Đọc kỹ đáp án thứ 2, ta thấy khả năng B,B thực ra không phải là một khả năng đơn, mà là một khả năng kép gồm có 2 khả năng trong đó: hoảng tử được nói đến hoặc là người con trai thứ nhất, hoặc là người con trai thứ hai. Như vậy phải tính B,B là 2 khả năng B=H,B và B, B=H (H là hoàng tử). Như thế tổng cộng vẫn có 4 khả năng, và xác suất vẫn là 2/4 = 1/2. Sai ở đây là sai trong cách đếm số khả năng. (Có câu hỏi khác: tại sao 4 khả năng này lại phải có xác suất bằng nhau ? Tại sao lại phải có phân bố xác suất đều ? Câu trả lời dành cho bạn đọc). Nếu ta đổi bài toán đi một chút thành: Một gia đình có 2 con, biết rằng ít nhất một trong hai con là con trai, thử hỏi xác suất để có con gái là bao nhiêu ? Trong bài toán này thì xác suất là 2/3 thật. Bạn đọc thử nghĩ xem sự khác nhau giữa hai bài toán nằm ở chỗ nào ?

Nghịch lý 3. Vấn đề ở đây nằm ở sự lẫn lỗn giữa các không gian xác suất trong lúc lập mô hình để tính xác suất. Trong cách giải thứ nhất, khi ta viết $P(A)$ để tính xác suất của sự kiện "Văn Phạm là thủ phạm", không gian xác suất của ta phải là không gian Ω_C tất cả các khả năng (với một trong 2 người Văn Phạm và X là thủ phạm) thỏa mãn điều kiện "thủ phạm là đàn ông", chứ không phải là không gian Ω của tất cả các khả năng có thể xảy ra (với một trong 2 người Văn Phạm và X là thủ phạm), bất kể thủ phạm là đàn ông hay đàn bà. Để cho khỏi lẫn lộn, thì trong cách giải thứ nhất ta phải viết $P_C(A) = P_C(A|B).P_C(B) + P_C(A|\overline{B}).P_C(\overline{B})$ Trong không gian Ω thì ta có $P(B) = 1/2$, tức là xác suất để X là đàn ông là 1/2. Nhưng trong không gian Ω_C dùng trong cách giải thứ nhất, thì ta phải dùng xác suất P_C của không gian đó, và $P_C(B)$ không phải là 1/2, mà thực ra là 2/3, và $P_C(\overline{B}) = 1/3$. Nói cách khác, khi biết rằng một trong hai người X và Văn Phạm là thủ phạm, và biết rằng thủ phạm là đàn ông, thì xác suất để X là đàn ông là 2/3 chứ không còn là 1/2 nữa ! (Vì sao vậy ?). Nếu ta sử dụng các con số xác suất này trong công thức tính xác suất toàn phần của A trong không gian Ω_C thì ta được: $p_C(A) = (1/2).(2/3) + 1.(1/3) = 2/3$ Tức là nếu ta sửa lỗi về xác suất của B đi, thì cách giải thứ nhất sẽ cho cùng đáp số 2/3 như cách giải thứ hai.

1.5 Luật số lớn

Luật số lớn là một trong những định luật cơ bản nhất của lý thuyết xác suất và thống kê. Ở dạng đơn giản nhất, nó có thể được phát biểu một cách nôm na như sau: khi một phép thử được lặp đi lặp lại rất nhiều lần, thì số lần cho ra một kết quả nào đó trong tổng số các lần thử sẽ phản ánh khá chính xác xác suất để xảy ra kết quả đó trong 1 lần thử. Ví dụ, giả sử ta có một đồng tiền với hai mặt sấp (S) và ngửa (N) với xác suất hiện lên bằng nhau và bằng 1/2 khi tung đồng tiền. Giả sử ta tung đi tung lại đồng tiền nhiều lần, và được một dãy các kết quả sấp ngửa, ví dụ như: S N S S N S N S N N S S ... Ta gọi $S(n)$ là tần số xuất hiện lên mặt sấp sau khi tung đồng tiền n lần, tức là số lần hiện lên mặt sấp sau khi tung đồng tiền n lần chia cho n, ví dụ như theo dãy trên: $S(1) = 1, S(2) = 1/2, S(3) = 2/3, S(4) = 3/4, S(5) = 3/5, S(6) = 2/3, S(7) = 4/7, S(8) =$

$5/8, S(9) = 5/9, S(10) = 1/2, S(11) = 6/11, S(12) = 7/12, \ldots$. Các con số $S(n)$ mà chúng ta thu được nói chung khác 1/2, nhưng luật số lớn nói rằng chúng ta có thể yên tâm rằng khi n tiến tới vô cùng thì $S(n)$ sẽ tiến tới 1/2: $\lim_{n\to\infty} S(n) = 1/2$.

Dưới đây chúng ta sẽ phát biểu luật số lớn một cách chặt chẽ thành định lý toán học và chứng minh nó, cho phân bố Bernoulli.

Giả sử có một phép thử nào đó có thể thực hiện được nhiều lần, và xác suất để xảy ra kết quả X trong một lần thử là một hằng số p, $0 < p < 1$. (Ví dụ: phép thử là "tung xúc sắc", kết quả là "hiện lên 1 chấm", xác suất là p=1/6). Ta gọi $X_{k,n}$ là sự kiện sau: khi thực hiện n lần phép thử thì X xuất hiện k lần trong số n lần thử. Chúng ta biết rằng xác suất của $X_{k,n}$ tuân theo phân bố nhị thức:

$$P(X_{k,n}) = C_n^k p^k (1-p)^{n-k}. \tag{1.29}$$

Lấy một số dương $\epsilon > 0$ tùy ý sao cho $0 < p - \epsilon < p + \epsilon < 1$. Gọi X_n^ϵ là sự kiện sau: khi làm phép thử n lần thì tần suất xuất hiện kết quả X chênh lệch so với xác suất p không quá ϵ, tức là $p - \epsilon \leq k/n \leq p + \epsilon$, trong đó k là số lần hiện lên kết quả X. Sự kiện X_n^ϵ là hợp của các sự kiện $X_{k,n}$ thỏa mãn bất đẳng thức $p - \epsilon \leq k/n \leq p + \epsilon$, do vậy:

$$P(X_n^\epsilon) = \sum_{n(p-\epsilon)\leq k\leq n(p+\epsilon)} C_n^k p^k (1-p)^{n-k}. \tag{1.30}$$

Định lý 1.4. *Với hai số dương p, ϵ bất kỳ thỏa mãn $0 < p - \epsilon < p + \epsilon < 1$, ta có*

$$\lim_{n\to\infty} \sum_{n(p-\epsilon)\leq k\leq n(p+\epsilon)} C_n^k p^k (1-p)^{n-k} = 1. \tag{1.31}$$

Có nghĩa là, xác suất $P(X_n^\epsilon)$ của sự kiện "sau n phép thử thì tần suất hiện kết quả X sai lệch so với xác suất p của X không quá ϵ" tiến tới 1 khi số phép thử n tiến tới vô cùng.

Định lý trên gọi là **dạng yếu của luật số lớn** cho phân bố Bernoulli. Dạng mạnh của luật số lớn, sẽ được xét tới trong chương sau, phát biểu là tần suất k/n tiến tới p khi n tiến tới vô cùng *hầu như chắc chắn* (tức là với xác suất bằng 1: tập những dãy vô hạn lần thử mà điều đó sai có xác suất bằng 0 trong không gian tất cả các dãy vô hạn lần thử).

Chứng minh. Chúng ta muốn chứng minh rằng hiệu

$$1 - P(X_n^\epsilon) = U_n + V_n, \tag{1.32}$$

trong đó

$$U_n = \sum_{0\leq k<n(p-\epsilon)} C_n^k p^k (1-p)^{n-k} \quad \text{và} \quad V_n = \sum_{n(p+\epsilon)<k\leq n} C_n^k p^k (1-p)^{n-k}, \tag{1.33}$$

tiến tới 0 khi n tiến tới vô cùng. Để đánh giá V_n, chúng ta có thể dùng thủ thuật sau đây: Gọi λ là một số dương bất kỳ, khi đó ta có

$$\begin{aligned} V_n &\leq \sum_{n(p+\epsilon)<k\leq n} e^{\lambda(k-n(p+\epsilon))} C_n^k p^k (1-p)^{n-k} \leq \sum_{0\leq k\leq n} e^{\lambda(k-n(p+\epsilon))} C_n^k p^k (1-p)^{n-k} \\ &= e^{-\lambda n\epsilon} \sum_{0\leq k\leq n} C_n^k (e^{\lambda(1-p)}p)^k (e^{-\lambda p}(1-p))^{n-k} = e^{-\lambda n\epsilon}(e^{\lambda(1-p)}p + e^{-\lambda p}(1-p))^n \\ &= [e^{-\lambda\epsilon}(e^{\lambda(1-p)}p + e^{-\lambda p}(1-p))]^n = f(\lambda)^n, \end{aligned} \tag{1.34}$$

với $f(\lambda) = e^{-\lambda\epsilon}(e^{\lambda(1-p)}p + e^{-\lambda p}(1-p))$. Chú ý rằng hàm số $f(\lambda)$ có $f(0) = 1$ và đạo hàm $f'(0) = -\epsilon < 0$. Như vậy nếu ta chọn $\lambda > 0$ đủ nhỏ thì ta có $0 < f(\lambda) < 1$, dẫn tới $\lim_{n\to\infty} f(\lambda)^n = 0$. Vì $V_n \leq f(\lambda)^n$ nên ta cũng có $\lim_{n\to\infty} V_n = 0$. Một cách hoàn toàn tương tự, ta có thể chứng minh $\lim_{n\to\infty} U_n = 0$, suy ra $\lim_{n\to\infty} U_n + V_n = 0$. Phần chứng minh này là bài tập dành cho bạn đọc. □

Chúng ta có thể mở rộng luật số lớn cho phân bố Bernoulli thành luật số lớn cho một không gian xác suất bất kỳ vời hữu hạn các phần tử như sau. Giả sử có một phép thử, mà cứ một lần thử thì hiện lên một trong các kết quả $A_1, \ldots, A_s$, với các xác suất $P(A_i) = p_i$ tương ứng. ($\sum_{i=1}^s p_i = 1$, và $\Omega = \{A_1, \ldots, A_s\}$ lập thành một không gian xác suất hữu hạn với các xác suất $P(A_i) = p_i$). Làm phép thử đó n lần (các lần thử độc lập với nhau), và gọi k_i là số lần hiện lên kết quả A_i trong số n lần thử đó. Gọi $B_{n,i}^\epsilon$ là sự kiện $|\frac{k_i}{n} - p_i| < \epsilon$.

Định lý 1.5. *Với mọi $\epsilon > 0$ ta có*

$$\lim_{n\to\infty} P(B_{n,1}^\epsilon \cap B_{n,2}^\epsilon \cap \ldots \cap B_{n,s}^\epsilon) = 1. \tag{1.35}$$

Ghi chú 1.6. (Một chút lịch sử(3)). Luật số lớn được biết đến ở dạng trực giác, "càng thí nghiệm nhiều lần thì kết quả thống kê càng chính xác", từ hàng nghìn năm trước đây. Nhà toán học và thiên văn học người Ấn Độ Brahmagupta (598-668), và sau đó nhà toán học người Italia Gerolamo Cardano (1501-1576), có phát biểu nó mà không chứng minh. Người đầu tiên đưa ra chứng minh toán học cho luật số lớn có lẽ là Jacob Bernoulli năm 1713, và luật số lớn còn được gọi là Định lý Bernoulli. Cái tên *luật số lớn* (la loi des grands nombres) được Siméon Denis Poisson viết ra năm 1835, và ngày nay người ta hay gọi theo tên đó.

Bài tập 1.16. Suy ra định lý 1.5 từ định lý 1.4.

(3)Xem: http://en.wikipedia.org/wiki/Law_of_large_numbers.

1.6 Bài tập bổ sung cho Chương 1

Bài tập 1.17. Tung một đồng tiền cân bằng cho đến khi mặt ngửa hiện lên 3 lần. Gọi A là sự kiện "cần tung 6 lần". Hãy lập một không gian xác suất cho vấn đề xác suất này, và tính xác suất của sự kiện A.

Bài tập 1.18. (Bài tập của ngành bảo hiểm). Một công ty bảo hiểm ô tô có 20000 người đăng ký bảo hiểm. Những người đăng ký bảo hiểm được công ty phân loại theo 3 tiêu chuẩn:
i) Trẻ hay già,
ii) Đàn ông hay đàn bà.
iii) Có vợ/chồng hay độc thân.
Được biết, trong số những người đăng ký bảo hiểm, có 6300 người trẻ, 9600 người là đàn ông, 13800 người có vợ/chồng, 2700 đàn ông trẻ, 6400 đàn ông có vợ, 2900 người trẻ có vợ/chồng, 1100 người là đàn ông trẻ có vợ. Hỏi xác suất để một người đăng ký bảo hiểm ô tô của hãng được chọn một cách ngẫu nhiên là một phụ nữ trẻ độc thân bằng bao nhiêu?

Bài tập 1.19. Một anh chàng có 2 cô bạn gái A và B, và không biết là thích cô nào hơn. Anh ta hay đi thăm các cô bạn một cách ngẫu nhiên: ra bến xe buýt, nếu gặp xe buýt đi tuyến đường đến nhà cô A trước thì đi lên xe đó thăm cô A, còn nếu gặp xe đi tuyến đường đến nhà cô B trước thì đi thăm cô B. Cả hai tuyến đường đều có xe đều đặn 10 phút một xe. Sau một thời gian dài, anh ta nhận ra rằng mình đi thăm cô bạn A nhiều gấp 3 lần cô bạn B. Có thể giải thích bằng xác suất tại sao ?

Bài tập 1.20. (Số may rủi). Giả sử có một loại xổ số chỉ có 100 số, từ 00 đến 99, mỗi lần quay có 1 số trúng giải.
i) Tính xác suất để sao cho trong 100 lần quay có một số trúng giải ít nhất 3 lần.
ii) Tính xác suất sao cho trong 100 lần quay, không có lần nào số 68 trúng giải.

Bài tập 1.21. Một lớp học có 36 học sinh. Hỏi rằng xác suất để có hai học sinh của lớp có cùng ngày sinh nhật là bao nhiêu ? (Viết công thức để tính số đó, và thử ước lượng xem số đó gần số nào hơn trong 3 số này: 0, 50%, 1?)

Ví dụ 1.21. Có n người chơi trò tung mũng trong một dạ hội: mỗi người cầm 1 cái mũ của mình, tung vào giữa phòng. Sau đó mỗi người nhặt lấy một cái mũ trong số các mũ được tung một cách ngẫu nhiên. Chứng minh rằng xác suất để không có người nào nhặt được đúng mũ của chính mình là

$$\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \ldots + \frac{(-1)^n}{n!}.$$

Khi n tiến tới vô cùng thì số này tiến tới e^{-1}.

Bài tập 1.22. (Bổ đề Borel–Cantelli). Giả sử $(A_n)_{n\in\mathbb{N}}$ là một dãy các tập con đo được trong một không gian xác suất (Ω, P). Gọi B_∞ là tập hợp các phần tử của Ω mà nằm trong một số vô hạn các tập con A_n của dãy. Chứng minh rằng:
i) Nếu $\sum_{n=1}^\infty P(A_n) < \infty$ thì $P(B_\infty) = 0$.
ii) Nếu tồn tại một số ϵ và vô hạn các tập con A_n của dãy thỏa mãn điều kiện $P(A_n) \geq \epsilon$, thì $P(B_\infty) \geq \epsilon$.
(Gợi ý: Đặt $B_k =$ tập các phần tử của Ω nằm trong ít nhất k tập con A_n của dãy. Khi đó $P(B_\infty) = \lim_{k\to\infty} B_k$. Trong trường hợp thứ nhất, chứng minh rằng $k.B_k \leq \sum_{n=1}^\infty P(A_n)$ với mọi k. Trong trường hợp thứ hai, chứng minh rằng $P(B_k) \geq \epsilon$ với mọi k).

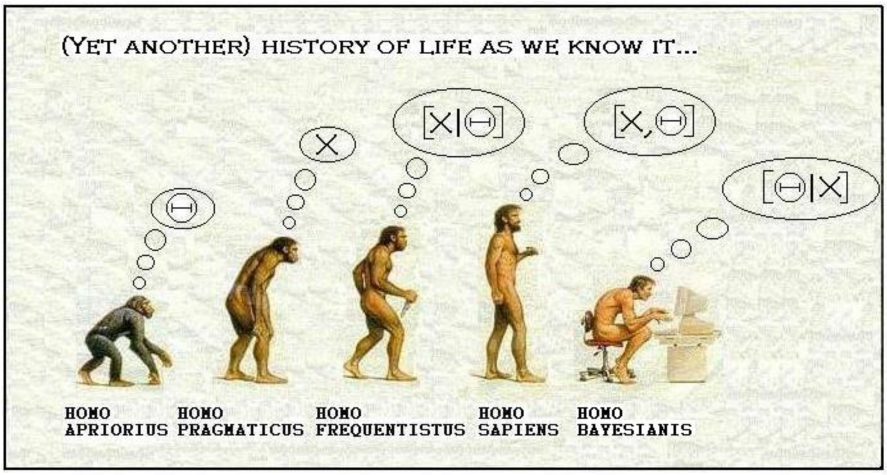
Bài tập 1.23. (Tủ của Bertrand). Có 3 ngăn kéo, 1 ngăn có 2 đồng tiền vàng, 1 ngăn có 2 đồng tiền bạc, và 1 ngăn có 1 đồng tiền vàng và 1 đồng tiền bạc. Rút ra một ngăn kéo một cách ngẫu nhiên, và lôi ra từ ngăn kéo đó một đồng tiền một cách ngẫu nhiên. Giả sử được 1 đồng tiền vàng. Hỏi xác suất để ngăn kéo được rút ra là ngăn kéo chứa hai đồng tiền vàng bằng bao nhiêu?

Bài tập 1.24. Có ba người A, B, C bị bắt vào tù. Có lệnh thả hai trong số ba người này ra. Cai tù nhận được lệnh, nhưng đến hôm sau mới được công bố và thi hành lệnh. Người tù A bảo cai tù: hãy nói cho tôi biết tên 1 người được thả trong hai người B và C đi. Cai ngục trả lời: anh đang có xác suất được thả là 2/3. Nếu tôi nói tên một người được thả trong số hai người B và C, thì giữa anh và người còn lại chỉ còn một người được thả nữa thôi, bởi vậy xác suất để anh được thả sẽ giảm xuống còn $1/2$. Tôi không muốn xác suất để anh được thả bị giảm đi, bởi vậy tôi sẽ không nói tên. Hỏi rằng người cai ngục lý luận như vậy có đúng không ?

Bài tập 1.25. Hai kẻ trộm đeo mặt nạ, bị cảnh sát đuổi bắt, bèn vứt mặt nạ đi và trà trộn vào một đám đông. Cảnh sát bắt giữ toàn bộ đám đông, tổng cộng 60 người, và dùng máy phát hiện nói dối (lie detector) để điều tra xem ai trong đám đông là kẻ trộm. Biết rằng đối với kẻ trộm, xác suất bị máy nghi là có tội là 85%, nhưng đối với người vô tội, thì xác suất để bị máy nghi nhầm thành có tội là 7%. Giả sử X là một nhân vật trong đám đông bị máy nghi là có tội. Tính xác suất để X là kẻ trộm.

Bài tập 1.26. (Bò điên). Năm 2001 Cộng Đồng Châu Âu có làm một đợt kiểm tra rất rộng rãi các con bò để phát hiện những con bị *bệnh bò điên* (bovine spongiform encephalopathy). Không có xét nghiệm nào cho kết quả chính xác 100%. Một loại xét nghiệm, mà ở đây ta gọi là xét nghiệm A, cho kết quả như sau: khi con bò bị bệnh bò điên, thì xác suất để ra phản ứng dương tính trong xét nghiệm A là 70%, còn khi con bò không bị bệnh, thì xác suất để xảy ra phản ứng dương tính trong xét nghiệm A là 10%. Biết rằng tỷ lệ bò bị mắc bệnh bò điên ở Hà Lan là 1,3 con trên 100000 con. Hỏi rằng khi một con bò ở Hà Lan phản ứng dương tính với xét nghiệm A, thì xác suất để nó bị mắc bệnh bò điên là bao nhiêu ?

Bài tập 1.27. (Giá dầu hỏa). Giá dầu hỏa có những lúc dao động rất mạnh, có khi đi lên hơn 100% trong vòng 1 năm. Giả sử rằng, nếu tính giá theo USD của năm 2009 (sau khi đã điều chỉnh theo tỷ lệ lạm phát), thì giá dầu hỏa không bao giờ xuống dưới 10 USD một thùng (dưới mức đó người ta ngừng sản xuất dầu hỏa vì không còn lãi gì nữa) và không bao giờ lên quá 300 USD một thùng (trên mức đó người ta dùng các loại năng lượng khác rẻ hơn). Hỏi họ các sự kiện Gx sau đây (x=0,1,...9) có thể là một họ độc lập các sự kiện được không : Gx = "năm 201x giá dầu hỏa tăng lên ít nhất 50% tính từ đầu năm đến cuối năm, tính theo USD của năm 2009". Giải thích tại sao ?



Hình 1.8: Tranh vui về sự tiến hóa của loài người

Chương 2

Biến Ngẫu Nhiên

2.1 Biến ngẫu nhiên và phân bố xác suất của nó

2.1.1 Biến ngẫu nhiên là gì ?

"Biến" là cái có thể thay đổi. "Ngẫu nhiên" là khi người ta chưa xác định được cái gì đó, thì người ta gọi nó là ngẫu nhiên. Cái gì khi đã xác định được, thì thành "định tính", hết ngẫu nhiên. Một biến có thể là ngẫu nhiên với người này, nhưng không ngẫu nhiên với người khác, tùy theo lượng thông tin nhận được. Ví dụ, số thứ tiếng ngoại ngữ mà ông A nói được là một số xác định, không ngẫu nhiên đối với ông A, nhưng nó là một số không xác định, ngẫu nhiên với một ông B nào đó.

Biến ngẫu nhiên có thể nhận giá trị trong mọi phạm trù (hiểu từ phạm trù ở đây theo nghĩa thông thường chứ không phải theo nghĩa phạm trù toán học), ví dụ như màu sắc, hình dạng, phương hướng, v.v. Tuy nhiên, bằng các ánh xạ (không ngẫu nhiên), chúng ta có thể chuyển việc nghiên cứu mọi biến ngẫu nhiên về việc nghiên cứu các biến ngẫu nhiên nhận giá trị là các số. Bởi vậy ở đây, khi nói đến một biến ngẫu nhiên mà không nói cụ thể nó nhận giá trị ở đâu, chúng ta sẽ hiểu là các giá trị của nó là các con số.

Ví dụ 2.1. Tại thời điểm đóng cửa thị trường chứng khoán Mỹ hôm 04/09/2009, giá cổ phiếu của hãng phần mềm máy tính Oracle (mã chứng khoán: ORCL) là 21,97 USD. Nó đã được xác định và không còn ngẫu nhiên. Thế nhưng tại thời điểm đó, thì giá cố phiếu của Oracle cho lúc cuối ngày 18/09/2009 chưa được biết, và nó là một biết ngẫu nhiên đối với thị trường chứng khoán. Người ta cho rằng giá của nó vào ngày 18/09/2009 có thể lên trên 23 USD, mà cũng có thể xuống dưới 21 USD. Điều này thể hiện qua việc, tại thời điểm cuối ngày 04/09/2009 , quyền mua ORCL trước ngày 19/09/2009 với giá 23 USD

(September 2009 call option at strike price 23) có giá 0,25 USD (nếu như ai cũng biết chắc rằng giá của ORCL vào thời điểm 18/09/2009 sẽ không vượt quá 23 thì cái quyền mua đó sẽ phải có giá bằng 0 vì không có giá trị gì), đồng thời quyền bán (put option) ORCL với giá 21 có giá là 0,30 USD. (Các thông tin về giá cả cổ phiếu và option có thể xem trên rất nhiều các trang web về chứng khoán).

Tương tự như với các số và các hàm số, ta có thể làm nhiều phép toán khác nhau với các biến ngẫu nhiên: cộng, trừ, nhân, chia, lấy giới hạn, tích phân, hàm hợp, v.v. Qua các phép toán như vậy, chúng ta có thể sinh ra các biến ngẫu nhiên mới từ các biến ngẫu nhiên cho trước.

Ví dụ 2.2. Một học sinh thi vào đại học phải thi 3 môn. Điểm của mỗi môn có thể coi là 1 biến ngẫu nhiên. Tổng số điểm cũng là một biến ngẫu nhiên, và nó là tổng của 3 biến ngẫu nhiên phía trước.

Ví dụ 2.3. Tốc độ V của một xe ô tô đang chạy trên đường có thể coi là một biến ngẫu nhiên. Nếu xe đang chạy mà phải phanh gấp lại vì phía trước có nguy hiểm, thì từ thời điểm người lái xe bóp phanh cho đến thời điểm xe dừng lại, xe phải chạy thêm mất một quãng đường có độ dài D nữa. D cũng có thể coi là một biến ngẫu nhiên. Nó không phải là tỷ lệ thuận với V, mà là tỷ lệ thuận với bình phương của V. Tức là biến ngẫu nhiên D có thể được sinh ra từ biến ngẫu nhiên V theo công thức: $D = k.V^2$. Hệ số k ở đây phụ thuộc vào điều kiện của đường và điều kiện của xe; nó có thể coi là xác định nếu ta biết các điều kiện này, còn nếu không thì có thể coi là một biến ngẫu nhiên khác. Ví dụ, trong điều kiện bình thường, thì $k = 0,08m^{-1}.s^2$: một xe đang chạy với tốc độ 36km/h = 10m/s thì từ lúc bóp phanh đến lúc dừng lại chạy thêm mất $0,08 \times 10^2 = 8$ mét, nhưng nếu xe đang chạy với tốc độ $108km/h = 3 \times 36km/h$, thì từ lúc bóp phanh đến lúc dừng lại sẽ chạy thêm mất những $8 \times 3^2 = 72$ mét.

2.1.2 Mô hình toán học của biến ngẫu nhiên

Giả sử có một biến ngẫu nhiên X. Chúng ta giả sử là có nhiều tình huống khác nhau có thể xảy ra, và trong mỗi tình huống thì X sẽ nhận được một giá trị nào đó. Như vậy một biến ngẫu nhiên có thể được mô hình hóa bằng một hàm số $X : \Omega \to \mathbb{R}$. Ở đây Ω là không gian đại diện cho các tình huống có thể xảy ra. Các tình huống, hay các nhóm các tình huống (các tập hợp con của Ω) là các sự kiện, và chúng ta có thể gắn cho mỗi sự kiện một xác suất về khả năng xảy ra. Điều đó có nghĩa là Ω có thể coi là một không gian xác suất, ký hiệu là (Ω, P) với một độ đo xác suất P. Chúng ta luôn giả sử rằng, với mọi cặp số $a, b \in \mathbb{R}, a < b$, tồn tại xác suất $P(a < X \leq b)$ của sự kiện $(a < X \leq b)$, hay nói cách khác, tập hợp $\{\omega \in \Omega | a < X(\omega) \leq b\}$ là tập đo được. Các hàm $X : \Omega \to \mathbb{R}$ thỏa mãn điều kiện này được gọi là **hàm đo được** trên (Ω, P). Từ đó chúng ta có định nghĩa toán học sau:

Định nghĩa 2.1. *Một* **biến ngẫu nhiên** *(random variable) với giá trị thực là một hàm số đo được trên một không gian xác suất:*

$$X : (\Omega, P) \to \mathbb{R}. \tag{2.1}$$

Định nghĩa 2.2. *Nếu ta có hai biến ngẫu nhiên X, Y (với cùng một mô hình không gian xác suất), thì ta sẽ nói rằng $X = Y$ theo nghĩa xác suất, hay $X = Y$* **hầu khắp mọi nơi** *, nếu như sự kiện "$X = Y$" có xác suất bằng 1 (tức là tập hợp các trường hợp mà ở đó $X \neq Y$ có xác suất bằng 0, có thể bỏ qua).*

Ví dụ 2.4. Một thí sinh đi kiểm tra trắc nghiệm, được giao 5 câu hỏi một cách ngẫu nhiên. Được biết 3 câu đầu thuộc loại vừa, và xác suất để thí sinh làm đúng cho mỗi câu là 80%, 2 câu sau thuộc loại khó, và xác suất làm đúng mỗi câu là 50%. Mỗi câu làm đúng thì được tính 1 điểm. Không gian Ω các tình huống ở đây gồm $2^5 = 32$ phần tử, mỗi phần tử có thể được ký hiệu bằng 1 dãy 5 chữ cái mà mỗi chứ cái là D (đúng) hoặc S (sai). Từ thông tin phía trên có thể suy ra xác suất của mỗi phần tử của Ω, ví dụ như $P(DDSSD) = 80\%.80\%.20\%.50\%.50\% = 4/125 = 3,2\%$. Biến ngẫu nhiên là tổng số điểm, tức là hàm $X : \Omega \to \{0, 1, 2, 3, 4, 5\}$, X của một dãy chữ cái bằng số lần chữ cái D xuất hiện trong dãy.

Ví dụ 2.5. Nếu A là một sự kiện, thì ta có thể định nghĩa **hàm chỉ báo** χ_A của A như sau: $\chi_A = 1$ khi A xảy ra và $\chi_A = 0$ khi A không xảy ra. Nếu ta có một sự kiện, thì hàm chỉ báo của nó là một biến ngẫu nhiên chỉ nhận hai giá trị 0 và 1, và ngược lại, nếu ta có một biến ngẫu nhiên F chỉ nhận 2 giá trị 0 và 1, thì nó là hàm chỉ báo của sự kiện $\{F = 1\}$. Nếu ta biểu diễn A như là một tập con của một không gian xác suất Ω, thì hàm chỉ báo của A được biểu diễn như là hàm chỉ báo của tập A trong Ω:

$$\chi_A(\omega) = \begin{cases} 1 & \text{khi } \omega \in A \\ 0 & \text{khi } \omega \in \overline{A} = \Omega \setminus A \end{cases}. \tag{2.2}$$

2.1.3 Phân bố xác suất của biến ngẫu nhiên

Nhắc lại rằng, nếu ta có một không gian xác suất (Ω, P) và một ánh xạ $X : (\Omega, P) \to \Lambda$ từ Ω lên một không gian Λ nào đó, thì phép push-forward theo X sẽ biến Λ thành một không gian xác suất, với độ đo xác suất cảm sinh $P_X = X^*P$: theo định nghĩa, nếu B là một tập con của Λ sao cho tồn tại $P(X^{-1}(B))$ thì

$$P_X(B) = P(X^{-1}(B))$$

Trong trường hợp $X : (\Omega, P) \to \mathbb{R}$ là một biến ngẫu nhiên, tính chất đo được của X (trong định nghĩa của biến ngẫu nhiên) nói rằng tồn tại $P(X^{-1}(]a,b])) = P(a < F \leq b)$ với mọi đoạn thẳng nửa mở $]a,b]$ trên $\mathbb{R}$. Sigma-đại số $\mathcal{B}$ sinh bởi các đoạn thẳng nửa mở trên $\mathbb{R}$ được gọi là **signma-đại số Borel** của $\mathbb{R}$. Khi nói đến một phân bố xác suất trên $\mathbb{R}$, chúng ta sẽ coi rằng sigma-đại số tương ứng chính là sigma-đại số Borel, bởi vì nói chung chúng ta sẽ chỉ quan tâm đến xác suất của các đoạn thẳng, và các tập con của $\mathbb{R}$ xây dựng được từ các đoạn thẳng bằng các phép giao, hợp, lấy phần bù. Do đó ta có định nghĩa sau:

Định nghĩa 2.3. Phân bố xác suất *(hay còn gọi là* **phân phối xác suất***) của một biến ngẫu nhiên X (trên $\mathbb{R}$) là phân bố xác suất P_X trên $\mathbb{R}$, với sigma-đại số là sigma-đại số Borel $\mathcal{B}$ của $\mathbb{R}$, cho bởi công thức sau:*

$$P_F(B) = P(X^{-1}(B)) \tag{2.3}$$

với mọi tập con B của $\mathbb{R}$ nằm trong sigma-đại số $\mathcal{B}$.

Định lý sau cho phép hiểu rõ hơn về sigma-đại số Borel:

Định lý 2.1. *i) Mọi đoạn thẳng mở (bị chặn hay không bị chặn) đều là phần tử của sigma-đại số Borel. Ngược lại, sigma-đại số sinh bởi các đoạn thẳng mở cũng chính bằng sigma-đại số Borel.*
ii) Mọi đoạn thẳng đóng đều là phần tử của sigma-đại số Borel. Ngược lại, sigma-đại số sinh bởi các đoạn thẳng đóng cũng chính bằng sigma-đại số Borel.

Chứng minh. Giả sử $]a,b[$ là một đoạn thẳng mở bị chặn của $\mathbb{R}$, với $a < b$. Khi đó tồn tại một dãy số đơn điệu tăng $a = b_0 < b_1 < b_2 < \ldots$ với $\lim_{n\to\infty} b_n = b$, và ta có thể viết $]a,b[= \bigcup_{n=1}^{\infty}]b_{n-1}, b_n]$, từ đó suy ra $]a,b[\in \mathcal{B}$, bởi vì $]b_{n-1}, b_n] \in \mathcal{B}$ với mọi n. Trong trường hợp $b = +\infty$ ta vẫn có thể làm hệt như trên để chứng minh rằng $]a, +\infty[\in \mathcal{B}$. Khi $a = -\infty$, thì tồn tại một dãy số đơn điệu giảm $b = b_0 > b_1 > b_2 > \ldots$ với $\lim_{n\to\infty} b_n = -\infty$, và ta có thể viết $]-\infty, b[=]b_1, b_0[\cup \bigcup_{n=1}^{\infty}]b_{n+1}, b_n]$, từ đó suy ra $]-\infty, b[\in \mathcal{B}$. Đối với một đoạn thẳng đóng $[a,b]$, ta có $]-\infty, a[\in \mathcal{B}$, $]b, +\infty[\in \mathcal{B}$, và $[a,b] = \mathbb{R} \setminus (]-\infty, a[\cup]b, +\infty[)$, từ đó suy ra $[a,b] \in \mathcal{B}$. Các khẳng định ngược lại (các tập đóng sinh ra sigma-đại số $\mathcal{B}$, và các tập mở cũng sinh ra sigma-đại số $\mathcal{B}$) nhường cho bạn đọc làm bài tập. □

Định nghĩa 2.4. Hàm phân phối *xác suất của phân bố xác suất P_X trên $\mathbb{R}$ của một biến ngẫu nhiên X là hàm $\mathcal{F}_X : \mathbb{R} \to [0,1]$ cho bởi công thức*

$$\mathcal{F}_X(x) := P(X \le x) = P_X(]-\infty, x]). \tag{2.4}$$

Tất nhiên, hàm phân phối được xác định duy nhất bởi phân bố xác suất. Điều ngược lại cũng đúng: Nếu ta biết hàm phân phối $\mathcal{F}_X$, thì ta có thể tính được xác suất P_X của các đoạn thẳng đóng và nửa mở của $\mathbb{R}$ qua các công thức sau

$$P_X(]a,b]) = \mathcal{F}_X(b) - \mathcal{F}_X(a), \tag{2.5}$$

$$P_X([a,b]) = \mathcal{F}_X(b) - \lim_{x \to a-} \mathcal{F}_X(x), \tag{2.6}$$

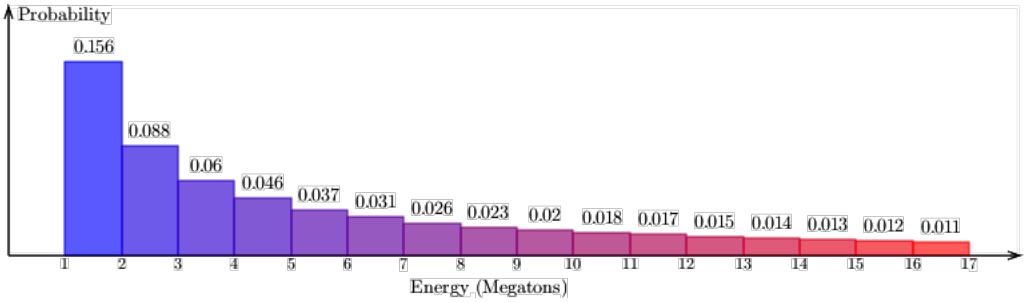
và từ đó tính được xác suất của các tập con khác của $\mathbb{R}$.

Định lý 2.2. *Hàm phân phối $\mathcal{F}_X$ của một phân bố xác suất tùy ý trên $\mathbb{R}$ thỏa mãn 4 tính chất sau:*
1) Đơn điệu không giảm: $\mathcal{F}_X(x) \ge \mathcal{F}_X(y)$ với mọi $x \ge y$,
2) Liên tục bên phải: $\lim_{\epsilon \to 0+} \mathcal{F}_X(x+\epsilon) = \mathcal{F}_X(x)$ với mọi x,
3) $\lim_{x\to-\infty} \mathcal{F}_X(x) = 0$,
4) $\lim_{x\to+\infty} \mathcal{F}_X(y) = 1$.
Ngược lại, mọi hàm số thực trên $\mathbb{R}$ thỏa mãn 4 tính chất trên là hàm phân phối của một phân bố xác suất trên $\mathbb{R}$

Chứng minh. Tính chất thứ nhất là hiển nhiên: nếu $x < y$ thì $\mathcal{F}_X(y) - \mathcal{F}_X(x) = P(x < X \le y) \ge 0$. Tính chất thứ hai có thể phát biểu cách khác như sau: nếu $x_1 > x_2 > \ldots$ là một dãy số đơn điệu giảm với $x_n \to x$ khi n tiến tới vô cùng thì ta có $\lim_{n\to\infty} \mathcal{F}_X(x_n) = \mathcal{F}_X(x)$. Để thấy điều đó, ta có thể viết $\mathcal{F}_X(x_n) - \mathcal{F}_X(x) = P_X(]x, x_n]) = P_X(\bigcup_{k=n}^{\infty}]x_{k+1}, x_k]) = \sum_{k=n}^{\infty} P_X(]x_{k+1}, x_k])$. Chuỗi số dương $\sum_{k=1}^{\infty} P_X(]x_{k+1}, x_k])$ là một chuỗi hội tụ, và bởi vậy *phần đuôi* $\sum_{k=n}^{\infty} P_X(]x_{k+1}, x_k])$ của nó tiến tới 0 khi n tiến tới vô cùng. Tính chất thứ 3 và tính chất thứ 4 có thể chứng minh một cách hoàn toàn tương tự. Khẳng định ngược lại là bài tập dành cho bạn đọc. □

Bài tập 2.1. Đồ thị 2.1 là biểu đồ phân bố xác suất (partial histogram, thiếu phần "đuôi") của mức năng lượng tỏa ra, tính theo đơn vị năng lượng megaton, của các thiên thạch lớn đâm vào bầu khí quyển của trái đất[(1)]. Hãy tính xác suất để một thiên thạch lớn đâm vào bầu khí quyển của trái đất có mức năng lượng tỏa ra không vượt quá 7 megaton.

[(1)]Số liệu của NASA năm 1994. Một thiên thạch lớn là một thiên thạch tỏa ra năng lượng ít nhất 1 megaton, bằng 1 quả bom hạt nhân nhỏ.



Hình 2.1: Năng lượng của các thiên thạch đâm vào bầu khí quyển trái đất

2.1.4 Các loại phân bố xác suất trên $\mathbb{R}$

Trong nhiều công việc tính toán với biến ngẫu nhiên, ta có thể quên đi không gian xác suất ban đầu của biến ngẫu nhiên đó, mà chỉ cần biết đến phân bố xác suất trên $\mathbb{R}$ của nó. Các phân bố xác suất trên $\mathbb{R}$ có thể được chia làm 3 loại sau: rời rạc, liên tục, và hỗn hợp (nửa rời rạc nửa liên tục).

Định nghĩa 2.5. *Một phân bố xác suất P_X trên $\mathbb{R}$ được gọi là* **liên tục** *nếu như hàm phân phối xác suất $\mathcal{F}_X$ là hàm liên tục trên $\mathbb{R}$. Nó được gọi là* **liên tục tuyệt đối** *nếu như như tồn tại một hàm số $\rho_X : \mathbb{R} \to \mathbb{R}_+$ khả tích và không âm, sao cho với mọi $a \in \mathbb{R}$ ta có*

$$\mathcal{F}_X(a) = P_X(]-\infty, a]) = \int_{-\infty}^{a} \rho_X(x)dx$$

Hàm $\rho_X : \mathbb{R} \to \mathbb{R}_+$ thoả mãn điều kiện như trên gọi là **hàm mật độ** *của P_X.*

Ghi chú 2.1. Hàm mật độ của một phân bố xác suất liên tục tuyệt đối P_X trên $\mathbb{R}$ là duy nhất theo nghĩa xác suất: nếu P_X có hai hàm mật độ ρ_1 và ρ_2, thì $\rho_1 = \rho_2$ hầu khắp mọi nơi trên $\mathbb{R}$, tức là tập $\{x \in \mathbb{R}, \rho_1(x) \neq \rho_2(x)\}$ có độ đo Lebesgue bằng 0. Một phân bố xác suất có thể là liên tục mà không liên tục tuyệt đối. (Bài tập: xây dựng ví dụ). Tuy nhiên, trong thực tế, khi người ta nói đến một phân bố xác suất liên tục trên $\mathbb{R}$, thường được hiểu là nó liên tục tuyệt đối, tức là được cho bởi một hàm mật độ. Chú ý rằng hàm mật độ chính bằng đạo hàm của hàm phân phối xác suất (hầu khắp mọi nơi). Rất nhiều vấn đề trong thực tế có thể được mô hình hóa bằng các biến ngẫu nhiên với phân bố xác suất liên tục, ví dụ như nhiệt độ của nước biển, giá dầu hỏa, sản lượng điện, trọng lượng của trứng gà, v.v.

Định lý 2.3. *Giả sử X có phân bố xác suất liên tục với hàm mật độ ρ_X, và $f : \mathbb{R} \to \mathbb{R}$ là một đơn ánh khả vi liên tục trên $\mathbb{R}$ trừ một số hữu hạn các điểm. Khi đó $Y = f(X)$*

cũng có phân bố xác suất liên tục, với hàm mật độ cho bởi công thức sau:

$$\rho_Y(y) = \frac{\rho_X(x)}{|f'(x)|} \text{ tại điểm } y = f(x) \tag{2.7}$$

Công thức trên chẳng qua là công thức đổi biến trong tích phân, và sinh ra từ công thức $df(x) = f'(x)dx$.

Một điểm $x \in \mathbb{R}$ được gọi là một điểm **hạt** của một phân bố xác suất P_X nếu như $P_X(x) > 0$. Bổ đề sau cho thấy một phân bố xác suất là liên tục khi và chỉ khi nó không có điểm hạt:

Định lý 2.4. *Giả sử $\mathcal{F}_X$ là hàm phân phối xác suất của một phân bố xác suất P_X trên $\mathbb{R}$.*
i) Với mọi $x \in \mathbb{R}$ ta có

$$P_X(x) = \mathcal{F}_X(x) - \lim_{y \to x-} \mathcal{F}_X(y). \tag{2.8}$$

i) Hàm $\mathcal{F}_X$ là hàm liên tục trên $\mathbb{R}$ khi và chỉ khi $P_X(x) = 0$ với mọi $x \in \mathbb{R}$

Chứng minh. i) Nếu $x_0 < x_1 < x_2 < \ldots$ là một dãy số đơn điệu tăng có giới hạn là x, thì ta có

$$\lim_{n\to\infty} \mathcal{F}_X(x_n) = \mathcal{F}_X(x_0) + \lim_{n\to\infty} P_X(]x_0, x_n]) = \mathcal{F}_X(x_0) + \lim_{n\to\infty} \sum_{k=1}^{n} P_X(]x_{k-1}, x_k])$$
$$= \mathcal{F}_X(x_0) + \sum_{k=1}^{\infty} P_X(]x_{k-1}, x_k]) = \mathcal{F}_X(x_0) + P_X(\bigcup_{k=1}^{\infty}]x_{k-1}, x_k]) = \mathcal{F}_X(x_0) + P_X(]x_0, x[)$$
$$= P_X(]-\infty, x[) = P_X(]-\infty, x]) - P_X(x) = \mathcal{F}_X(x) - P_X(x),$$

từ đó suy ra công thức trong bổ đề. Để chứng minh phần thứ hai của bồ đề trên, nhắc lại rằng hàm phân phối xác suất luôn luôn liên tục bên phải. Bởi vậy nó liên tục khi và chỉ khi nó liên tục bên trái, tức là khi và chỉ khi $P_X(x) = \mathcal{F}_X(x) - \lim_{y\to x-} \mathcal{F}_X(y) = 0$ với mọi $x \in \mathbb{R}$.

Trong trường hợp phân bố xác suất P_X không liên tục, gọi

$$K_X = \{x \in \mathbb{R} | P_X(x) > 0\} \tag{2.9}$$

là tập hợp các điểm hạt của nó (tức là tập hợp các điểm gián đoạn của hàm phân phối xác suất). Khi đó K_X là tập hữu hạn hoặc cùng lắm là đếm được, vì $P_X(A) = \sum_{x\in K_X} P_X(x) \leq 1$.

Định nghĩa 2.6. *Một phân bố xác suất P_X được gọi là* **rời rạc** *nếu như nó tập trung trên tập hợp các điểm hạt của nó: $P_X(A_X) = 1$, $P_X(\mathbb{R} \setminus A_X) = 0$.*

Ví dụ 2.6. Phân bố xác suất trên $\mathbb{R}$ của biến ngẫu nhiên "điểm kiểm tra" trong ví dụ "bài kiểm tra trắc nghiệm" ở mục trước là một phân bố rời rạc tập trung ở 6 điểm: 0,1,2,3,4,5. (Bài tập: tính các xác suất của 6 điểm đó).

Giả sử P_X là một phân bố xác suất bất kỳ trên $\mathbb{R}$, với hàm phân phối $\mathcal{F}_X$. Khi đó ta có thể viết:

$$\mathcal{F}_X(x) = \mathcal{D}_X(x) + \mathcal{C}_X(x) \tag{2.10}$$

với $\mathcal{D}_X(x) = P_X(]-\infty, x] \cap K_X)$ gọi là **phần rời rạc** của $\mathcal{F}_X$, và $\mathcal{C}_X(x) = \mathcal{F}_X(x) - \mathcal{D}_X(x)$ gọi là **phần liên tục** của $\mathcal{F}_X$. Phân bố P_X được gọi là **hỗn hợp** nếu như cả hai phần rời rạc và liên tục đều khác 0. Nếu phần liên tục không phải là liên tục tuyệt đối (không viết được dưới dạng tích phân của một hàm không âm), thì ta có thể tách nó tiếp thành tổng của *phần liên tục tuyệt đối* và *phần liên tục kỳ dị*, nhưng chúng ta sẽ không đi vào chi tiết ở đây.

Ví dụ 2.7. Trong xe ô tô thường có kim chỉ mức xăng, dao động trong khoảng từ 0 (0%, tức là hết xăng) đến 1 (100%, bình xăng đầy). Mức xăng được kim chỉ vào có thể coi là một biến ngẫu nhiên nhận giá trị trong đoạn thẳng $[0, 1]$ với phân bố xác suất liên tục. Tuy nhiên, ở một số xe ô tô cũ, kim bị hỏng, có lúc nó chỉ đúng mức xăng nhưng có lúc nó bị tắc ở chỗ số 0 tuy rằng xe còn xăng. Khi đó, phân bố xác suất không còn là liên tục nữa mà là hỗn hợp, với "hạt" tại điểm 0.

Bài tập 2.2. Giả sử biến ngẫu nhiên X có phân bố xác suất liên tục với hàm mật độ ρ_X sau : $\rho_X(x) = 0$ khi $|x| > 1$ và $\rho_X(x) = 1 - |x|$ khi $|x| \leq 1$. Tìm hàm mật độ của phân bố xác suất của biến ngẫu nhiên $Y = \arcsin(x)$.

Bài tập 2.3. Giả sử biến ngẫu nhiên X có phân bố xác suất liên tục và đối xứng, theo nghĩa X và $-X$ có cùng phân bố xác suất. Chứng minh rằng hàm phân phối xác suất của X thỏa mãn tính chất $\mathcal{F}_X(-x) + \mathcal{F}_X(x) = 1$ với mọi $x \in \mathbb{R}$. Điều này còn đúng không nếu phân bố xác suất của X không liên tục ?

2.2 Một số phân bố xác suất thường gặp

Nhắc lại rằng, phân bố nhị thức với các tham số n, p là phân bố xác suất $P(k) = C_n^k p^k (1-p)^n$. trên không gian $\Omega = \{0, 1, \ldots, n\}$. Nó cũng có thể được coi như một phân bố rời rạc trên $\mathbb{R}$ tập trung tại các điểm $0, 1, \ldots, n$ với các xác suất như trên. Tương tự như vậy, phân bố Bernoulli với tham số p có thể được coi như một phân bố xác suất trên $\mathbb{R}$ tập trung tại hai điểm $0, 1$ (hoặc hai điểm nào đó khác), với các xác suất $P(1) = p$ và $P(0) = 1 - p$. Phân bố Bernoulli và phân bố nhị thức là những phân bố rất hay gặp trong thực tế. Ở đây, chúng ta sẽ thảo luận thêm một số phân bố rời rạc và liên tục phổ biến khác trên $\mathbb{R}$.

2.2.1 Phân bố hình học và phân bố nhị thức âm

Định nghĩa 2.7. Phân bố hình học *với tham số p $(0 \leq p \leq 1)$ là phân bố xác suất rời rạc tập trung tại tập hợp các số tự nhiên, cho bởi công thức sau:*

$$P(k) = p(1-p)^{k-1} \quad \forall\, k \in \mathbb{N}. \tag{2.11}$$

Ý nghĩa của phân bố hình học là: nó là phân bố xác suất của "số lần thử cho đến khi thành công", nếu như xác suất thành công của mỗi lần thử là p.

Ví dụ 2.8. Một người chơi trò tung vòng vào cổ chai, tung đến bao giờ trúng thì thôi. Xác suất để tung trúng mỗi lần là p. Gọi T là số lần phải tung cho đến khi tung trúng. Khi đó T là một biến ngẫu nhiên nhận giá trị trong $\mathbb{N}$. Xác suất để sao cho tung $k-1$ lần đầu trượt, nhưng lần thứ k trúng, là $(1-p)^{k-1}p$. Như vậy phân bố xác suất của T chính là phân bố hình học với tham số p.

Nếu thay vì tính số lần thử cho đến khi có 1 lần thành công, ta tính tổng số lần thử *thất bại* k cho đến khi có tổng cộng r lần thành công ($r \in \mathbb{N}$) thì ta có một biến ngẫu nhiên mới, nhận giá trị trong $\mathbb{Z}_+$, với phân bố xác suất sau:

$$P(k) = C^k_{k+r-1} p^r (1-p)^k$$

Nhị thức Newton C^k_{k+r-1} trong công thức trên là số cách chọn ra $r-1$ phần tử từ tập hợp $\{1, 2, \ldots, k+r-1\}$. (Mỗi cách chọn như vậy ứng với một tình huống, với k lần thất bại và $r-1$ lần thành công trong số $k+r-1$ lần thử đầu tiên, và lần thử thứ $k+r$ thành công). Các nhị thức Newton C^k_{k+r-1} còn có thể viết dưới dạng $C^k_{k+r-1} = \frac{(k+r-1).....(r+1).r}{k!} = (-1)^k \frac{(-r).(-r-1).....(-r-k+1)}{k!} = (-1)^k C^k_{-r}$, và chúng xuất hiện trong khai triển Taylor sau:

$$(1-q)^{-r} = \sum_{k=0}^{\infty} (-1)^k C^k_{-r} q^k$$

Trong khai triển Taylor trên, nếu đặt $q = 1 - p$ và nhân cả hai vế với p^r, thì ta được

$$1 = \sum_{k=0}^{\infty} (-1)^k C^k_{-r} p^r (1-p)^k = \sum_{k=0}^{\infty} P(k)$$

Chú ý rằng khai triển Taylor trên có giá trị (và hội tụ khi $|q| < 1$) cả khi mà $r > 0$ không phải là số nguyên. Các công thức trên dẫn đến định nghĩa sau:

Định nghĩa 2.8. *Giả sử $0 < p < 1$ và $r > 0$. Khi đó phân bố xác suất rời rạc cho bởi công thức*

$$P(k) = C^k_{k+r-1}p^r(1-p)^k = (-1)^k C^k_{-r}p^r(1-p)^k \tag{2.12}$$

với mọi $k \in \mathbb{Z}_+$ được gọi là **phân bố nhị thức âm** *với các tham số r và p.*

Tất nhiên, phân bố hình học có thể coi là trường hợp đặc biệt của phân bố nhị thức âm, với $r = 1$ (và trên $\mathbb{N}$ thay vì trên $\mathbb{Z}_+$, tức là có cộng thêm 1 vào biến ngẫu nhiên).

Bài tập 2.4. Kiểm tra công thức sau: hàm phân phối xác suất của phân bố hình học với tham số p cho bởi công thức $\mathcal{F}(x) = 0$ nếu $x < 0$ và $\mathcal{F}(x) = 1 - (1-p)^{[x]}$ nếu $x \geq 0$. Ở đây $[x]$ là phần nguyên của số x.

2.2.2 Phân bố Poisson

Định nghĩa 2.9. *Một biến ngẫu nhiên X được gọi là có* **phân bố Poisson** *(đọc là Poa-Sông) với tham số λ, nếu như các giá trị của nó là các số nguyên không âm, và với mọi $k \in \mathbb{Z}_+$ ta có:*

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}. \tag{2.13}$$

Ghi chú 2.2. Phân bố Poisson mang tên của nhà toán học và vật lý người Pháp Siméon Denis Poisson (1781–1840). Trong lý thuyết xác suất, Poisson được biết đến nhiều nhất bởi phân bố Poisson, và *quá trình Poisson* (một quá trình ngẫu nhiên ứng với phân bố này). Tên gọi *luật số lớn* (của các luật số lớn, mà chúng ta sẽ tìm hiểu trong Chương 4) cũng là do Poisson đặt ra.

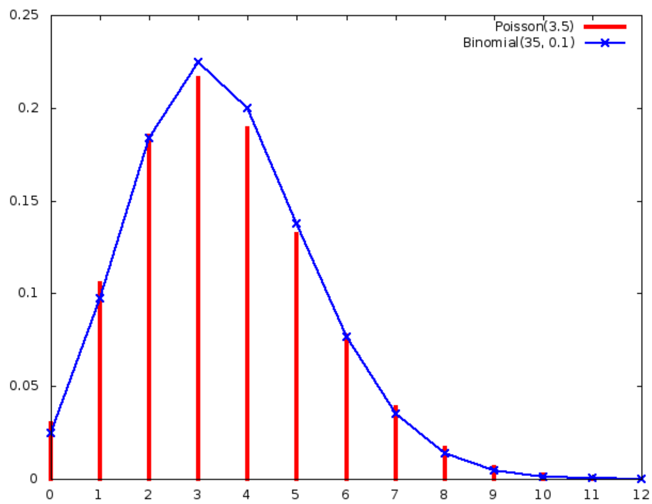
Phân bố Poisson là giới hạn của phân bố nhị thức với các tham số $p = \lambda/n$ và n, khi n tiến tới vô cùng. Thật vậy, ta có

$$\begin{aligned} C^k_n(\lambda/n)^k(1-\lambda/n)^{n-k} &= \frac{n!}{k!(n-k)!}(\lambda/n)^k(1-\lambda/n)^{n-k} \\ &= \frac{\lambda^k}{k!}\frac{n(n-1)\dots(n-k+1)}{n^k}(1-\lambda/n)^{-k}(1-\lambda/n)^n. \end{aligned}$$

Khi n tiến tới vô cùng thì $(n(n-1)\dots(n-k+1)/n^k)(1-\lambda/n)^{-k}$ tiến tới 1 (k ở đây là cố định) và $(1-\lambda/n)^n$ tiến tới $e^{-\lambda}$, bởi vậy ta có

$$\lim_{n\to\infty} C^k_n.(\lambda/n)^k(1-\lambda/n)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}. \tag{2.14}$$

Hình 2.2: Siméon Denis Poisson



Hình 2.3: Các phân bố Poisson(3.5) và Binomial(35,0.1)

Xem đồ thị minh họa trên hình 2.3 cho trường hợp $\lambda = 3,5$, $n = 35$, $p = 0,1$.

Mô hình phân bố Poisson là mô hình thường được dùng cho các biến ngẫu nhiên dạng "số sự kiện xảy ra trong một khoảng thời gian nào đó".

Ví dụ 2.9. Biến ngẫu nhiên "số vụ tai nạn giao thông xảy ra trong một ngày" ở một vùng nào đó có thể được mô hình hóa bằng phân bố Poisson. Ta sẽ giả sử các tai nạn giao thông xảy ra một cách ngẫu nhiên, độc lập với nhau, và trung bình mỗi ngày có λ vụ tai nạn. Ta sẽ chia 24 tiếng đồng hồ trong ngày thành n khoảng thời gian (n là một số rất lớn), để sao cho có thể coi rằng trong mỗi khoảng thời gian có nhiều nhất 1 vụ giao thông xảy ra, và khả năng xảy ra tai nạn giao thông trong mỗi khoảng thời gian bằng λ/n. Khi đó tổng số tai nạn xảy ra trong ngày tuân theo phân bố nhị thức với các tham số $n, p = \lambda/n$, và khi cho n tiến tới vô cùng ta được phân bố Poisson. Tất nhiên phân bố Poisson không thể là phân bố xác suất chính xác của vấn đề (vì số người là hữu hạn, và số tai nạn bị chặn trên bởi số người chứ không lớn tuỳ ý được), nhưng nó là phân bố gần đúng thuận tiện cho việc tính toán.

2.2.3 Phân bố đều (trường hợp liên tục)

Định nghĩa 2.10. *Giả sử a và b là hai số thực, với $b > a$. Khi đó* **phân bố đều** *(uniform distribution) trên đoạn thẳng $]a, b[$ là phân bố xác suất liên tục với hàm mật độ $\rho(x)$ sau:*

$$\rho(x) = \begin{cases} \frac{1}{b-a} & \text{khi } a \leq x \leq b \\ 0 & \text{khi } x < a \text{ hoặc } x > b \end{cases} . \quad (2.15)$$

Phân bố xác suất đều trên đoạn thẳng $]a, b[$ hay được ký hiệu là $U(a, b)$.

Ghi chú 2.3. Trong định nghĩa trên, thay vì lấy đoạn thẳng mở $]a, b[$, có thể lấy đoạn thẳng đóng $[a, b]$ hoặc đoạn thẳng nửa mở $]a, b]$ hoặc $[a, b[$ cũng được. Về mặt xác suất không có gì thay đổi.

Ví dụ 2.10. Vị trí của một người đi bộ trên một đoạn đường có thể được mô hình hóa bằng một biến ngẫu nhiên với phân bố đều, nếu như ta không có thông tin gì ngoài thông tin người đi bộ đang ở trên đoạn đường đó.

Khái niệm phân bố đều có thể mở rộng lên trường hợp nhiều chiều: không gian xác suất là một miền trong $\mathbb{R}^n$ ($n \geq 2$), và xác suất của một miền con tỷ lệ thuận với thể tích (n chiều) của miền con đó.

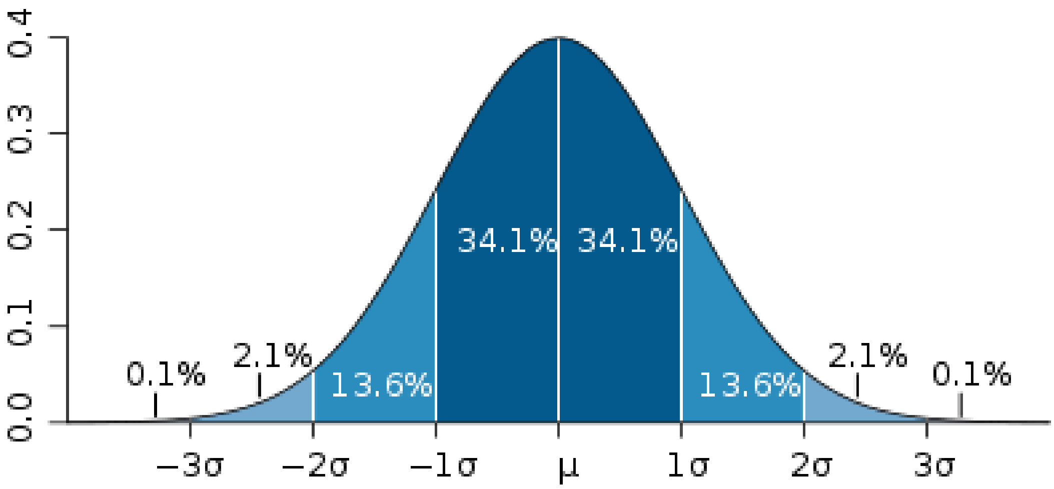
Bài tập 2.5. Giả sử X có phân bố đều $U(0, 1)$, và Y là một biến ngẫu nhiên bất kỳ. Chứng minh rằng tồn tại một hàm số g sao cho $g(X)$ và Y có cùng phân bố xác suất. (Bài tập này có ý nghĩa thực tế trong việc làm giả lập (simulation) các biến ngẫu nhiên: dùng random number generator (chương trình tạo số ngẫu nhiên) trên máy tính để giả lập một biến ngẫu nhiên với phân bố đều $U(0, 1)$, rồi qua đó giả lập được mọi phân bố xác suất, qua các hàm số thích ứng).

2.2.4 Phân bố normal

Định nghĩa 2.11. *Phân bố xác suất* **normal** *(còn gọi là phân bố* **chuẩn** *, hay phân bố* **Gauss***) trên $\mathbb{R}$ với trung điểm μ và độ lệch chuẩn σ là phân bố liên tục với hàm mật độ sau:*

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}). \tag{2.16}$$

Ký hiệu thường dùng để chỉ phân phối xác suất normal là: $N(\mu, \sigma^2)$. Phân bố normal $N(0,1)$ (với $\mu = 0$, $\sigma^2 = 1$) được gọi là **phân bố normal chuẩn tắc** .

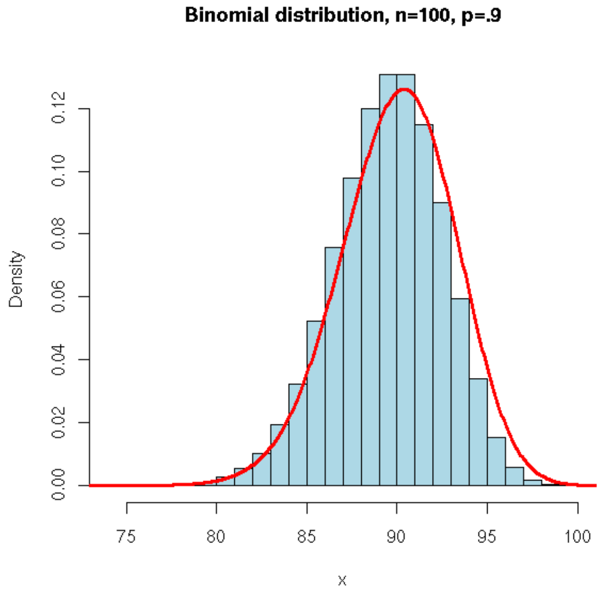


Hình 2.4: Hàm mật độ của phân bố normal

Đồ thị của hàm mật độ của phân bố normal có hình cái chuông, và bởi vậy phân bố normal còn được gọi một cách nôm na là **phân bố hình cái chuông**. Trung điểm của cái chuông này chính là điểm $x = \mu$, và độ cao của chuông chính bằng $\frac{1}{\sigma\sqrt{2\pi}}$. Nếu σ càng nhỏ thì chuông càng cao và càng "hẹp", và ngược lại σ càng lớn thì chuông càng thấp và càng bè ra.

Hình vẽ minh họa 2.4 cho thấy hầu hết xác suất của một phân bố normal nằm trong đoạn $[\mu - 3\sigma, \mu + 3\sigma]$. Chỉ có không đến 0,3% nằm ngoài đoạn đó. Nói cách khác, nếu X là một biến ngẫu nhiên có phân bố xác suất normal với các tham số μ, σ, thì với xác suất 99,7% ta có thể tin rằng giá trị của X nằm trong đoạn $[\mu - 3\sigma, \mu + 3\sigma]$: $P(\mu - 3\sigma < X < \mu + 3\sigma) = 99,7\%$.

Phân bố normal là một trong những phân bố xác suất quan trọng nhất, vì nhiều phân bố xác suất gặp trong thực tế có dáng điệu khá giống phân bố normal, ví dụ như phân bố của chiều cao của đàn ông, phân bố của chỉ số IQ (chỉ số trí tuệ), phân bố của giá chứng khoán trong tương lai, v.v. Khi n tiến tới vô cùng và p cố định, thì dáng điệu của phân bố nhị thức với các tham số n, p cũng ngày càng gần giống phân bố normal. Ví dụ, lấy $p = 0,9$. Khi n nhỏ thì phân bố nhị thức với các tham số n và $p = 0,9$ có dáng điệu khác xa phân bố normal, nhưng khi $n = 100$, thì dáng điệu của phân bố nhị thức trông đã rất gần giống phân bố normal, như thể hiện trên Hình 2.5.



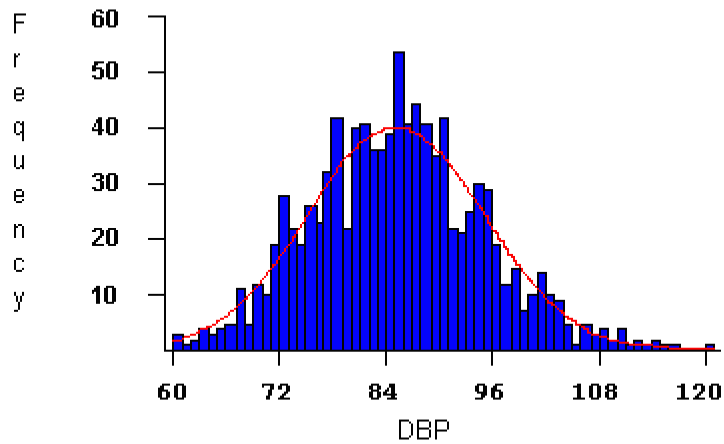
Hình 2.5: Phân bố nhị thức với $n = 100$, $p = 0,9$

Các định lý giới hạn trung tâm mà chúng ta sẽ đề cập đến trong Chương 4 sẽ cho chúng ta cơ sở lý thuyết để hiểu tại sao có nhiều phân bố xác xuất trong thực tế trông giống phân bố normal.

Ví dụ 2.11. Hình 2.6 là **biểu đồ tần số** (histogram) của huyết áp của người, trong một thí nghiệm đo huyết áp 1000 người. **Tần số** (frequency) của một giá trị tức là số lần xuất hiện giá trị đó trong dãy số các kết quả. Nếu chúng ta coi không gian xác suất ở đây là có 1000 phần tử, với xác suất của một phần tử là 1/1000, thì bảng tần số trên cho ta bảng phân bố xác suất rời rạc của biến ngẫu nhiên "huyết áp" H: xác suất của sự kiện $H = x$ bằng **tần suất** (relative frequency[2]) của x. Tần suất là tần số chia cho tổng số (tức là chia cho 1000 ở đây). Vì đồ thị có hình gần giống hình cái chuông, nên ta thấy phân bố

[2]Từ *frequency* tiếng Anh vừa có nghĩa là tần suất vừa có nghĩa tần số. Để phân biệt, tần suất có khi được gọi là relative frequency, hoặc là frequency rate.

xác suất của biến "huyết áp" trong thí nghiệm này có thể được xấp xỉ khá tốt bằng một phân bố normal.



Hình 2.6: Biểu đồ tần số huyết áp

Ghi chú 2.4. Để có một phân bố xác suất gần giống phân bố normal, cần phải có một sự “thuần nhất” nào đó trong biến ngẫu nhiên. Ví dụ, nếu ta có 1 thùng táo chín cùng một giống táo, thì khi xét biến ngẫu nhiên “đường kính của quả táo” trên thùng táo đó, ta có thể được một phân bố gần giống phân bố normal. Nhưng nếu ta trộn 2 thùng táo thuộc 2 giống táo khác nhau, một giống táo to một giống táo nhỏ, thì phân bố xác suất của biến “đường kính” trong đống táo trộn lẫn này không còn là normal được nữa, mà nó phải có 2 “đỉnh”, 1 đỉnh ứng với đường kính trung bình của giống táo to và 1 đỉnh ứng với đường kính trung bình của giống táo nhỏ.

Bài tập 2.6. Giả sử X là một biến ngẫu nhiên tuân theo luật phân bố normal $N(\mu, \sigma^2)$. Chứng minh rằng biến ngẫu nhiên $Z = (X - \mu)/\sigma$ tuân theo phân bố normal chuẩn tắc $N(0, 1)$.

2.2.5 Phân bố lũy thừa

Định nghĩa 2.12. Phân bố lũy thừa *(exponential distribution) với tham số λ là phân bố xác suất liên tục tuyệt đối trên $\mathbb{R}$ cho bởi hàm mật độ sau:*

$$\rho(x) = \begin{cases} \lambda e^{-\lambda x} & khi\ x > 0 \\ 0 & khi\ x \leq 0 \end{cases}. \tag{2.17}$$

Hàm phân bố xác suất $\mathcal{F}$ của phân bố này như sau: $\mathcal{F}(x) = 0$ khi $x \leq 0$, và khi $x > 0$ thì

$$\mathcal{F}(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}. \quad (2.18)$$

Phân bố lũy thừa có thể được xem như là *dạng liên tục* của phân bố hình học: phân bố hình học là rời rạc còn phân bố lũy thừa là liên tục, nhưng hàm phân phối xác suất của hai phân bố này có dáng điệu tương tự nhau.

Phân bố lũy thừa có thể được dùng để làm mô hình xác suất cho những biến ngẫu nhiên kiểu "khoảng cách giữa hai lần xuất hiện", ví dụ như: khoảng cách thời gian giữa hai cú điện thoại gọi đến, khoảng cách giữa hai gen đột biến kế tiếp trên một dải DNA, v.v.

Bài tập 2.7. Giả sử biến ngẫu nhiên X có phân bố lũy thừa với tham số λ, và $c > 0$. Chứng minh rằng cX cũng có phân bố lũy thừa với tham số λ/c.

Bài tập 2.8. Giả sử biến ngẫu nhiên X có phân bố lũy thừa với tham số λ, và s và t là hai số dương. Chứng minh rằng

$$P(X > s + t | X > s) = P(X > t)$$

Giải thích tại sao đẳng thức này gọi là tính chất *không có trí nhớ* (lack of memory property) của phân bố lũy thừa.

Bài tập 2.9. Giả sử X là một biến ngẫu nhiên liên tục với hàm phân phối xác suất liên tục $f = \mathcal{F}_X$. Chứng minh rằng:
i) $f(X)$ có phân bố xác suất đều trên đoạn thẳng $[0, 1]$.
ii) $-\ln f(X)$ có phân bố lũy thừa.

2.2.6 Phân bố Pareto

Vilfredo Pareto (1848–1923) là một nhà kinh tế người Italia. Ông ta quan sát thấy rằng, phân bố tài sản trên thế giới rất không đều, và "80% tài sản là do 20% người làm chủ" (80% nhân dân còn lại chỉ làm chủ 20% tài sản). Quan sát này mang tên *nguyên tắc Pareto* hay *nguyên tắc 80-20* (có khi nó còn trở thành *nguyên tắc 90-10*). Pareto đưa ra mô hình phân bố xác suất liên tục hóa sau cho biến ngẫu nhiên "giá trị tài sản của một người":

Hình 2.7: Vilfredo Pareto

Định nghĩa 2.13. Phân bố Pareto *với tham số* $\alpha > 0$ *là phân bố liên tục trên* $\mathbb{R}$ *với hàm mật độ sau:*

$$\rho(x) = \begin{cases} \dfrac{\alpha}{x^{\alpha+1}} & \textit{khi } x \geq 1 \\ 0 & \textit{khi } x < 1 \end{cases}. \tag{2.19}$$

Phân bố Pareto còn được dùng làm mô hình phân bố xác suất gần đúng cho rất nhiều biến ngẫu nhiên khác, vị dụ như: kích thước của các hạt cát, các thiên thạch, các khu dân cư, dự trữ dầu hỏa của các mỏ dầu, mức độ thiệt hại của các vụ tai nạn, v.v.

Bài tập 2.10. Chứng minh rằng nếu X có phân bố Pareto với tham số α, và $Y = X^s$ với $s > 0$, thì Y cũng có phân bố Pareto, và tìm tham số của phân bố này.

Bài tập 2.11. Giả sử X có phân bố xác suất đều $U(0,1)$. Chứng minh rằng $Y = 1/(1-X)$ có phân bố Pareto với tham số $\alpha = 1$.

2.3 Kỳ vọng của biến ngẫu nhiên

2.3.1 Trường hợp rời rạc

Khi ta có một biến ngẫu nhiên, ta có thể nghiên cứu các tính chất, đặc trưng của nó, để rút ra các thông tin, kết luận nào đó. Một trong những đặc trưng quan trọng nhất là giá trị kỳ vọng.

Định nghĩa 2.14. *Giá trị* **kỳ vọng** *của một biến ngẫu nhiên X, ký hiệu là $\mathbb{E}(X)$, chính là trung bình cộng của biến ngẫu nhiên đó trên không gian xác suất các tình huống.*

Từ định nghĩa có thể suy ra được rằng, hai biến ngẫu nhiên có cùng phân bố xác suất trên $\mathbb{R}$ thì có cùng kỳ vọng. Bởi vậy, thay vì nói về kỳ vọng của một biến ngẫu nhiên, ta cũng có thể nói về *kỳ vọng của một phân bố xác suất trên* $\mathbb{R}$.

Trong trường hợp không gian xác suất các tình huống là một tập hợp hữu hạn hoặc đếm được, $\Omega = \{\omega_1, \omega_2, \ldots\}$ với các xác suất $P(\omega_i)$ ($\sum_i P(\omega_i) = 1$), thì công thức tính giá trị kỳ vọng (trung bình cộng) của một biến ngẫu nhiên $X : \Omega \to \mathbb{R}$ là

$$\mathbb{E}(X) = \sum_i X(\omega_i)P(\omega_i). \tag{2.20}$$

Ví dụ 2.12. Trò chơi đề (một trò đánh bạc): trong 100 số đề sẽ chỉ có 1 số thắng, 99 số thua. Thắng thì được 70 lần tiền đặt cọc. Thua thì mất tiền đặt cọc. Nếu đặt cọc T tiền, thì kỳ vọng số tiền nhận lại được là $99\% \times 0 + 1\% \times 70.T = 0,7.T$. Kỳ vọng lãi (lỗ) là $0,7.T - T = -0,3.T$. Tức là đặt cọc T tiền chơi đề, thì kỳ vọng là bị thua $0,3.T$.

Ví dụ 2.13. Giá trị kỳ vọng của phân bố Poisson $P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$ là $\mathbb{E}(X) = \lambda$. Thật vậy,

$$\mathbb{E}(X) = \sum_k kP(X = k) = \sum_k ke^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k\geq 1}\frac{\lambda^k}{(k-1)!} = e^{-\lambda}\lambda\sum_{k\geq 1}\frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\lambda e^{\lambda} = \lambda.$$

Ví dụ 2.14. Giá trị kỳ vọng của phân bố hình học $P(T = k) = p(1-p)^{k-1}$ là

$$\mathbb{E}(T) = \sum_{k=1}^{\infty} k.p.(1-p)^{k-1} = 1/p.$$

Điều này phù hợp với suy luận trực giác rằng, nếu xác suất để ném vòng một lần trúng cổ chai là p, thì trung bình phải ném vòng $1/p$ lần mới trúng cổ chai.

Ghi chú 2.5. Trong trường hợp không gian xác suất rời rạc $\Omega = \{\omega_1, \omega_2, \ldots\}$ có vô hạn các sự kiện, khi định nghĩa kỳ vọng, chúng ta đòi hỏi chuỗi $\sum_{i=1}^{\infty} X(\omega_i).P(\omega_i)$ phải là chuỗi *hội tụ tuyệt đối*, có nghĩa là chuỗi $\sum_{i=1}^{\infty} |X(\omega_i)|.P(\omega_i)$ phải hội tụ. Trong trường hợp chuỗi $\sum_{i=1}^{\infty} X(\omega_i).P(\omega_i)$ không hội tụ tuyệt đối, thì kỳ vọng không được xác định hoặc là bằng vô cùng. Lý do để đòi hỏi điều kiện hội tụ tuyệt đối là, chúng ta muốn tổng của chuỗi $\sum_{i=1}^{\infty} X(\omega_i).P(\omega_i)$ phải hữu hạn và không phụ thuộc vào thứ tự của các số trong tổng, tức là nếu có thay đổi cách đánh số các sự kiện, thì vẫn phải ra cùng một tổng. Các chuỗi thỏa mãn điều kiện này chính là các chuỗi hội tụ tuyệt đối.

Định lý 2.5. *Một số tính chất cơ bản của giá trị kỳ vọng:*
i) Kỳ vọng của một hằnh số c (biến ngẫu nhiên chỉ nhận 1 giá trị) chính là hằng số đó:

$$\mathbb{E}(c) = c. \tag{2.21}$$

ii) Tuyến tính: Nếu X, Y là hai biến ngẫu nhiên và a, b là hai hằng số thì

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y). \tag{2.22}$$

iii) Đơn điệu: Nếu $X \geq 0$ thì $\mathbb{E}(X) \geq 0$. Tổng quát hơn,

$$X \geq Y \Rightarrow \mathbb{E}(X) \geq \mathbb{E}(Y). \tag{2.23}$$

Định lý trên đúng trong trường hợp tổng quát, khi mà các giá trị kỳ vọng được xác định. Chứng minh của nó trong trường hợp rời rạc tương đối hiển nhiên.

Khi chúng ta sử dụng hai mô hình không gian xác suất khác nhau để nghiên cứu cùng một biến ngẫu nhiên, thì không phải vì thế mà kỳ vọng của nó thay đổi. Nói một cách chính xác hơn, ta có:

Định lý 2.6. *Giả sử $X : (\Omega, P) \to \mathbb{R}$ là một biến ngẫu nhiên với không gian xác suất Ω, và $\phi : (\Omega_1, P_1) \to (\Omega, P)$ là một ánh xạ bảo toàn xác suất từ một không gian xác suất (Ω_1, P_1) lên (Ω_1, P_1). Đặt $X_1 = X \circ \phi : (\Omega_1, P_1) \to \mathbb{R}$ là biến ngẫu nhiên giống X nhưng với không gian xác suất (Ω_1, P_1). Khi đó*

$$\mathbb{E}(X_1) = \mathbb{E}(X). \tag{2.24}$$

Định lý trên cũng đúng trong trường hợp tổng quát. Chứng minh của nó tương đối hiển nhiên trong trường hợp Ω và Ω_1 là các không gian xác suất rời rạc, và là bài tập dành cho bạn đọc.

Bài tập 2.12. Một doanh nghiệp đầu tư phát triển một sản phẩm mới, xác suất thành công là 30%. Chi phí đầu tư bỏ ra là 100 nghìn USD. Nếu không thành công thì mất chi phí đầu tư mà không thu về được gì, nhưng nếu thành công thì thu về được 1 triệu (trước khi trừ đi chi phí đầu tư). Tính kỳ vọng lợi nhuận từ vụ đầu tư này.

Bài tập 2.13. Xây dựng một ví dụ đơn giản với hai biến ngẫu nhiên X, Y rời rạc sao cho $\mathbb{E}(XY) \neq \mathbb{E}(X)\mathbb{E}(Y)$.

Bài tập 2.14. Trong một rổ có 99 quả bóng đánh số từ 1 đến 99. Lôi ra từ trong rổ 5 quả bóng một cách ngẫu nhiên. Gọi X là số nhỏ nhất hiện lên trên 5 quả bóng được lôi ra,

và Y là số lớn nhất hiện lên.
i) Tính phân bố xác suất của các biến ngẫu nhiên X và Y.
ii) Chứng minh rằng, với mọi $m, n \in \mathbb{N}, m \leq n$, ta có $\sum_{k=m}^{n} C_k^m = C_{n+1}^{m+1}$.
iii) Dùng ii) để tính $\mathbb{E}(X)$.

Bài tập 2.15. Một người tập bóng rổ, đứng từ một chỗ ném bóng vào rổ 6 lần. Xác suất ném trúng mỗi lần là 2/3. Gọi X là số lần ném trúng, Y là số lần ném trượt, và $Z = X - Y$. Hãy tính kỳ vọng $\mathbb{E}(Z)$ của Z bằng hai cách khác nhau: một cách thông qua phân bố xác suất của Z, và một cách không dùng đến phân bố xác suất của Z.

Bài tập 2.16. (Entropy). Giả sử có 1 trò chơi giữa hai người A và B như sau: A chọn 1 số tự nhiên trong các số từ 1 đến 2^n (n là một số cố định nào đó), và B phải tìm xem là số nào. B có thể hỏi A bất cứ câu hỏi nào về số mà A chọn, và A sẽ trả lời "có" hoặc "không" cho các câu hỏi của B.
i) Chỉ ra một chiến thuật (mộ cách hỏi), để sau khi hỏi đúng n lần, B tìm được số mà A chọn. (Số n ở đây được gọi là *entropy*, hay là *lượng thông tin*).
ii) Chứng minh rằng, với bất kỳ chiến thuật nào của B, thì kỳ vọng về số lần phải hỏi cho đến khi tìm được số mà A chọn là một số lớn hơn hoặc bằng n.
(Đầu tiên hãy thử làm cho các trường hợp $n = 2$, $n = 3$, rồi làm cho trường hợp tổng quát).

2.3.2 Trường hợp tổng quát: tích phân trên không gian xác suất

Trong trường hợp tổng quát, công thức tính giá trị kỳ vọng được viết dưới dạng **tích phân Lebesgue** của X trên không gian xác suất (Ω, P):

$$\mathbb{E}(X) = \int_{\Omega} X dP. \tag{2.25}$$

Định nghĩa của tích phân Lebesgue như sau. Giả sử có một hàm số $F : (\Omega, P) \to \mathbb{R}$ đo được trên một không gian xác suất (Ω, P) với độ đo xác suất P. Nhắc lại rằng, tính chất đo được có nghĩa là tồn tại $P(F^{-1}(]a, b]))$ với mọi $a, b \in \mathbb{R}, a < b$.

Trước hết ta xét trường hợp F là một hàm bị chặn: tồn tại một số dương $M \in \mathbb{R}_+$ sao cho $|F(\omega)| < M$ với mọi $\omega \in \Omega$.

Một phân hoạch (sự chia nhỏ) của đoạn thẳng $]-M, M]$ là một dãy số $a_0 = -M < a_1 < a_2 < \ldots < a_n = M$ hữu hạn đơn điệu tăng nào đó, sao cho số đầu bằng $-M$ và số cuối bằng M. Nói cách khác, ta chia đoạn thẳng $]-M, M]$ thành một hợp không giao nhau của các đoạn thẳng nửa mở $]a_i, a_{i+1}]$. Khi có một phân hoạch như vậy, ký hiệu là σ, ta có thể lập hai số sau:

$$I_\sigma(F) = \sum_{i=0}^{n} a_i.P(F^{-1}(]a_i, a_{i+1}])), \tag{2.26}$$

và

$$J_\sigma(F) = \sum_{i=0}^{n} a_{i+1}.P(F^{-1}(]a_i, a_{i+1}])). \tag{2.27}$$

Ký hiệu Σ là tập hợp tất cả các phân hoạch của đoạn thẳng $]-M, M]$. Dễ thấy rằng

$$I_\sigma(g) \leq J_\delta(F) \ \ \forall\, \sigma, \delta \in \Sigma.$$

(Bài tập: Chứng minh bất đẳng thức trên). Hơn nữa, nếu phân hoạch σ thỏa mãn tính chất $a_{i+1} - a_i < \epsilon$ với mọi i, thì ta cũng có $J_\sigma(g) - I_\sigma(g) < \epsilon$. Từ đó suy ra $\sup_{\sigma\in\Sigma} I_\sigma(F) = \inf_{\delta\in\Sigma} J_\delta(F)$. Theo định nghĩa, tích phân Lebesgue của F trên (Ω, P) chính là giá trị chung đó:

$$\int_\Omega F dP = \sup_{\sigma\in\Sigma} I_\sigma(F) = \inf_{\delta\in\Sigma} J_\delta(F). \tag{2.28}$$

Trong trường hợp F không bị chặn, thì đầu tiên ta thay F bằng các hàm bị chặn

$$F_{M,N}(\omega) := \min(\max(-N, F(\omega)), M), \tag{2.29}$$

$(M, N > 0)$, rồi định nghĩa

$$\int_\Omega F dP = \lim_{M,N\to+\infty} \int_\Omega F_{M,N} dP \tag{2.30}$$

nếu như giới hạn đó tồn tại. Trong trường hợp giới hạn đó tồn tại và hữu hạn, thì ta nói F là hàm khả tích. **Khả tích** có nghĩa là định nghĩa được tích phân, và các cách định nghĩa khác nhau (qua các cách lấy giới hạn khác nhau) cho cùng một kết quả hữu hạn. Hàm F khả tích khi và chỉ khi giá trị tuyệt đối của nó có tích phân hữu hạn: $\int_\Omega |F| dP < \infty$. (Đây là một định lý trong giải tích, chứng minh không khó).

Trong trường hợp Ω là một miền trong $\mathbb{R}^n$ với thể tích bằng 1, phân bố xác suất P là phân bố đều trên đó (xác suất của một miền con của Ω là thể tích của miền con đó), và F là một hàm liên tục bị chặn, thì tích phân Lebesgue trùng với tích phân (Riemann) nhiều chiều thông thường. Trong trường hợp tổng quát, thì tích phân Lebesgue là mở rộng của khái niệm tích phân Riemann.

Tất nhiên, trong trường hợp $\Omega = \{\omega_1, \omega_2, \ldots\}$ là một không gian xác suất rời rạc, ta có

$$\int_\Omega F dP = \sum_i F(\omega_i).P(\omega_i), \tag{2.31}$$

và (nếu Ω có vô hạn phần tử) F khả tích khi và chỉ khi chuỗi $\sum_i F(\omega_i).P(\omega_i)$ hội tụ tuyệt đối.

Tương tự như tích phân Riemann thông thường, tích phân Lebesgue trên không gian xác suất có tính chất đơn điệu, tuyến tính, và giao hoán với phép lấy giới hạn của một dãy hàm hội tụ đều:

Định lý 2.7. *Giả sử F, G và F_n là các hàm đo được trên một không gian xác suất (Ω, P).*
i) Nếu $F \geq 0$ (hầu khắp mọi nơi trên Ω) thì $\int_\Omega F dP \geq 0$. Tổng quát hơn, nếu $F \geq G$ thì $\int_\Omega F dP \geq \int_\Omega G dP$.
ii) Nếu F_n hội tụ đều đến F trên Ω, có nghĩa là $\lim_{n\to\infty} \sup_{\omega\in\Omega} |F_n(\omega) - F(\omega)| = 0$, thì $\lim_{n\to\infty} F_n dP = \int_\Omega F dP$.
iii) Với hai số thực a, b bất kỳ, ta có

$$\int_\Omega (aF + bG) dP = a \int_\Omega F dP + b \int_\Omega G dP. \tag{2.32}$$

Hai khẳng định đầu tiên của định lý trên suy ra trực tiếp từ định nghĩa của tích phân Lebesgue. Khẳng định thứ ba có thể kiểm tra trực tiếp dễ dàng trong trường hợp F và G chỉ nhận một số hữu hạn các giá trị. Trong trường hợp tổng quát, ta có thể xấp xỉ F và G bằng các hàm chỉ nhận một số hữu hạn các giá trị, sau đó lấy giới hạn. □

Định lý sau, gọi là *định lý hội tụ bị chặn Lebesgue* (Lebesgue dominated convergence theorem), là một định lý hay được sử dụng trong việc nghiên cứu các tích phân Lebesgue:

Định lý 2.8 (Lebesgue). *Giả sử $F_n : (\Omega, P) \to \mathbb{R}$ là một dãy hàm đo được trên không gian xác suất (Ω, P) thỏa mãn hai điều kiện sau:*
i) $|F_n| \leq G$ với mọi n, trong đó G là một hàm khả tích trên Ω.
ii) F_n hội tụ hầu khắp mọi nơi đến một hàm đo được F trên Ω, có nghĩa là tập các điểm $\omega \in \Omega$ sao cho $\lim_{n\to\infty} F_n(\omega) = F(\omega)$ có độ đo bằng 1.
Khi đó ta có

$$\int_\Omega F dP = \lim_{n\to\infty} \int_\Omega F_n dP. \tag{2.33}$$

Sơ lược chứng minh. Vì $|F_n| \leq G$ nên ta cũng có $|F| \leq G$ hầu khắp mọi nơi. Lấy một số $\delta > 0$ nhỏ tùy ý. Đặt $A_n = \{\omega \in \Omega | |F_n(\omega) - F(\omega)| > \delta\}$. Ta có

$$|\int_\Omega F dP - \int_\Omega F_n dP| \leq \int_\Omega |F - F_n| dP \leq \int_{\Omega\setminus A_n} \delta dP + \int_{A_n} 2G dP \leq \delta + 2\int_{A_n} G dP.$$

Để chứng minh $|\int_\Omega FdP - \int_\Omega F_n dP|$ tiến tới 0 khi n tiến tới vô cùng, ta chỉ cần chứng minh $\int_{A_n} GdP$ tiến tới 0 khi n tiến tới vô cùng với mọi δ. Vì F_n hội tụ hầu khắp mọi nơi đến F trên Ω, nên tập hợp các điểm mà nằm trong vô số các tập A_n có độ đo bằng 0. Do đó $P(A_n)$ tiến tới 0 khi n tiến tới vô cùng (xem khẳng định thứ hai của Bài tập 1.22), từ đó suy ra $\int_{A_n} GdP$ tiến tới 0 khi n tiến tới vô cùng. □

Định lý 2.6 về sự bảo toàn giá trị kỳ vọng dưới ánh xạ bảo toàn xác suất có thể được phát biểu lại dưới dạng định lý về sự bảo toàn tích phân Lebesgue dưới dánh xạ bảo toàn xác suất:

Định lý 2.9. *Giả sử $F : (\Omega, P) \to \mathbb{R}$ là một hàm khả tích trên không gian xác suất (Ω, P), và $\phi : (\Omega_1, P_1) \to (\Omega, P)$ là một ánh xạ bảo toàn xác suất. Khi đó $F \circ \phi$ là hàm khả tích trên (Ω_1, P_1) và*

$$\int_{\Omega_1} (F \circ \phi) dP_1 = \int_\Omega FdP. \tag{2.34}$$

Chứng minh của định lý trên suy ra trực tiếp từ định nghĩa tích phân Lebesgue. □

2.3.3 Kỳ vọng của phân bố xác suất trên $\mathbb{R}$

Đôi khi, ta sẽ ký hiệu tích phân $\int_\Omega FdP$ thành $\int_{\omega\in\Omega} F(\omega)dP$, hoặc là $\int_\Omega F(\omega)dP(\omega)$, để chỉ rõ hơn về việc lấy tích phân theo biến nào.

Theo định nghĩa, **kỳ vọng của một phân bố xác suất** P_X trên $\mathbb{R}$ là $\displaystyle\int_{x\in\mathbb{R}} xdP_X$.

Định lý 2.10. *i) Kỳ vọng của một biến ngẫu nhiên X bằng kỳ vọng của phân bố xác suất P_X của biến ngẫu nhiên đó:*

$$\mathbb{E}(X) = \int_{x\in\mathbb{R}} xdP_X. \tag{2.35}$$

ii) Nếu P_X là một phân bố liên tục với hàm mật độ ρ_X, thì ta có:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x\rho_X(x)dx. \tag{2.36}$$

iii) Nếu g là một hàm số thực thì

$$\mathbb{E}(g(X)) = \int_{x\in\mathbb{R}} g(x)dP_X = \int_{-\infty}^{\infty} g(x)\rho_X(x)dx. \tag{2.37}$$

Khẳng định đầu tiên của định lý trên chẳng qua là trường hợp đặc biệt của tính chất bảo toàn kỳ vọng qua ánh xạ bảo toàn xác suất. Thật vậy, ta có thể viết $X = Id \circ X$, trong đó Id là hàm đồng nhất trên $\mathbb{R} : Id(x) = x$. Do đó kỳ vọng của X bằng kỳ vọng của hàm Id trên $\mathbb{R}$ với phân bố xác suất P_X, và ta có công thức (2.35). Khẳng định thứ hai là hệ quả của khẳng định thứ nhất trong trường hợp liên tục tuyệt đối. Khẳng định thứ ba cũng suy ra từ tính chất bảo toàn kỳ vọng qua ánh xạ bảo toàn xác suất, tương tự như khẳng định thứ nhất. □

Ví dụ 2.15. Giá trị kỳ vọng của phân bố xác suất normal $N(\mu, \sigma^2)$ bằng μ.

Ví dụ 2.16. Giả sử giá 1kg vàng vào thời điểm T là 35000 (USD). Tại thời điểm T, thì giá 1kg vàng cho thời điểm $T+1$ chưa được biết, và có thể coi là một biến ngẫu nhiên X. Giả sử rằng X có phân bố (gần như) normal với kỳ vọng 35000 và độ lệch chuẩn 400. Hỏi rằng, vào thời điểm T, giá trị của quyền mua 1kg với giá 35000 tại thời điểm $T+1$ là bao nhiêu ? Quyền mua (call) vàng là một chứng khoán phái sinh, cho phép người sở hữu nó mua vàng với giá cố định trước, tại một thời điểm trong tương lai, nhưng không bắt buộc phải mua. Gọi giá trị của quyền mua này tại thời điểm $T+1$ là Y. Khi đó $Y = \max(0, X - 35000)$, tức là nếu giá vàng lúc đó dưới 35000 thì giá trị của quyền mua bằng 0, còn nếu giá vàng trên 35000 thì giá trị của quyền mua bằng sự chênh lệch giữa giá vàng và giá ghi trong quyền mua. Giá trị của quyền mua này tại thời điểm T được coi bằng kỳ vọng của Y. Như vậy giá trị này bằng

$$\mathbb{E}(Y) = \frac{1}{400\sqrt{2\pi}} \int_{35000}^{\infty} (x-35000). \exp\left(-\frac{(x-35000)^2}{2.400^2}\right) dx = \frac{400}{\sqrt{2\pi}} \int_0^{\infty} \exp(-z) dz \approx 160.$$

Bài tập 2.17. Giả sử Y là một biến ngẫu nhiên liên tục với hàm mật độ sau: $\rho_Y(x) = c \sin x$ khi $x \in]0, \pi[$, và $\rho_Y(x) = 0$ tại các điểm khác.

i) Hãy tính c.

ii) Hãy tính $\mathbb{E}(Y)$

iii) Thử nghĩ một vấn đề có thể xảy ra trong thực tế với phân bố xác suất này.

Bài tập 2.18. Tính kỳ vọng của phân bố Pareto (2.19) với tham số $\alpha > 1$. (Khi $\alpha \leq 1$ thì kỳ vọng bằng vô cùng).

2.3.4 Giá trị kỳ vọng hình học

Trong các tài liệu về xác suất ít khi nhắc tới kỳ vọng hình học. Nhưng khái niệm này cũng rất quan trọng, bởi vậy chúng ta sẽ đề cập nó ở đây. Giá trị kỳ vọng ứng với trung bình cộng, còn giá trì kỳ vọng hình học ứng với trung bình nhân. Một ví dụ đơn giản sau đây cho thấy sự quan trọng của trung bình nhân trong thực tế.

Ví dụ 2.17. Giả sử giá nhà dao động trong 4 năm như sau. Năm đầu tiên giảm 15%, năm thứ hai tăng 35%, năm thứ ba giảm 20%, năm thứ tư tăng 20%. Hỏi xem trong 4 năm đó giá nhà tăng lên (hay giảm đi) trùng bình mỗi năm bao nhiêu % ? Nếu ta lấy trung bình cộng thì được (-15% + 35% - 20% + 20%)/ 4 = 5% một năm. Nhưng con số đó có phản ánh chính xác sự đi lên của giá nhà trong 4 năm không ? Nếu gọi giá lúc đầu là X, thì sau năm đầu giá là (1-15%)X, sau năm thứ hai giá là (1+35%)(1-15%)X, sau năm thứ ba giá là (1-20%)(1+35%)(1-15%)X, sau 4 năm giá là (1+20%)(1-20%)(1+35%)(1-15%)X = 1,1016 X. Tức là sau 4 năm giá nhà chỉ tăng lên có 10,16%, chứ không phải 20% (= 4 lần 5%) như là người ta tưởng ! Để có cái nhìn chính xác về mức độ tăng trưởng trung bình hàng năm trong giai đoạn 4 năm, cần phải lấy trung bình nhân của các con số 1+20%, 1-20%, 1+35%, 1-15% rồi trừ đi 1. Kết quả là 2,449% một năm.

Như chúng ta biết, nếu có một dãy các số dương $a_1, \ldots, a_n$, $a_i > 0$ với mọi i, thì ngoài giá trị trung bình cộng $(\sum a_i)/n$, chúng ta còn có thể nói đến trung bình nhân: $(\prod_i a_i)^{1/n}$. Từ tiếng Anh cho trung bình nhân là geometric mean, nếu dịch từng chữ ra tiếng Việt thì là "trung bình hình học", còn trung bình cộng là "trung bình số học". Trung bình nhân có thể được định nghĩa qua trung bình cộng và qua hàm logarithm ln, và hàm ngược của hàm ln, tức là hàm exp:

$$(\prod_i a_i)^{1/n} = \exp(\sum_i (\ln a_i)/n). \tag{2.38}$$

Hàm ln là hàm lõm trên nửa đường thẳng dương (đạo hàm bậc hai của nó bằng $-1/x^2$ là một hàm âm), bởi vậy ta có:

$$\frac{\sum_i \ln a_i}{n} \leq \ln\left(\frac{\sum_i a_i}{n}\right)$$

Lấy exp của hai vế của bất đẳng trên, ta được bất đẳng thức quen thuộc sau: Trung bình nhân luôn luôn nhỏ hơn hoặc bằng trung bình cộng:

$$\left(\prod_i a_i\right)^{1/n} \leq \frac{\sum_i a_i}{n}. \tag{2.39}$$

Dấu bằng xảy ra khi và chỉ khi tất cả các số a_i bằng nhau.

Nếu thay vì một dãy các số dương, ta có một biến ngẫu nhiên X mà các giá trị đều dương, thì ta cũng có thể làm tương tự như trên, và kết quả gọi là giá trị kỳ vọng hình học của X:

Định nghĩa 2.15. *Nếu X là một biết ngẫu nhiên chỉ nhận các giá trị dương, thì* **giá trị kỳ vọng hình học** *của X, ký hiệu là $\mathbb{G}(X)$, được cho bở công thức sau:*

$$\mathbb{G}(X) = \exp(\mathbb{E}(\ln X)) = \exp(\int_\Omega \ln(X) dP). \tag{2.40}$$

Định lý 2.11. *Giá trị kỳ vọng hình học luôn nhỏ hơn hoặc bằng giá trị kỳ vọng:*

$$\mathbb{G}(X) \leq \mathbb{E}(X). \tag{2.41}$$

Dấu bằng xảy ra khi và chỉ khi F là hằng số hầu khắp mọi nơi trên không gian xác suất, tức là tồn tại một số thực dương c sao cho $P(X = c) = 1$.

Định lý trên là trường hợp riêng của bất đẳng thức Jensen phát biểu như sau:

Định lý 2.12 (Bất đẳng thức Jensen). *Nếu f là một hàm lồi, và X là một biến ngẫu nhiên bất kỳ, thì*

$$\mathbb{E}(f(X)) \geq f(\mathbb{E}(X)). \tag{2.42}$$

Ví dụ 2.18. Giả sử có một cơ hội đầu tư như sau. Khả năng thắng/thua là 50%/50%, sau 1 tháng biết kết quả. Nếu thắng thì lãi 100%, nếu thua thì lỗ 50% tiền bỏ ra. (Trên thị trường chứng khoán có những trường hợp tương tự như vậy, ví dụ như 1 hãng công nghệ sinh học khi đang đợi kết quả thí nghiệm lâm sàng của một loại thuốc chữa ung thư, nếu thành công thì giá trị cổ phiếu của hãng có thể tăng hơn gấp đôi, nếu thất bại thì giá trị cũng có thể mất trên 50%). Hỏi đối với người đầu tư thì có nên đầu tư vào những cơ hội như vậy không, và nếu nên thì nên đầu tư với nhiều nhất nhiêu % vốn đầu tư (để đạt kỳ vọng lợi nhuận cao nhất, giả sử là không có các cơ hội đầu tư khác)?

Trước hết, ta có thể tính giá trị kỳ vọng của lợi nhuận của đầu tư theo cơ hội trên, với 1 đơn vị vốn bỏ ra. Gọi L là biến "lợi nhuận", ta có 2 khả năng: hoặc $L = 1$ hoặc $L = -1/2$, mỗi khả năng có xác suất 50%. Như vậy kỳ vọng lợi nhuận trên 1 đơn vị vốn bỏ ra là: $\mathbb{E}(L) = 50\%.1 + 50\%.(-1/2) = 0,25$ Kỳ vọng lợi nhận ở đây là dương và khá lớn (bằng 25% vốn bỏ ra), nên đây là cơ hội nên đầu tư, trừ khi có những cơ hội khác tốt hơn. (Lãi 25% trong một tháng có thể gọi là siêu lợi nhuận).

Câu hỏi thứ hai là nhà đầu tư nên đầu tư vào đó nhiều nhất là bao nhiêu phần trăm vốn đầu tư ? Nếu giả sử đầu tư toàn bộ 100% vốn. Khi đó có 2 khả năng, hoặc là tổng số vốn tăng lên gấp đôi, hoặc là giảm đi còn 1 nửa, với xác suất của mỗi khả năng là 50%. Nhưng nếu một nhà đầu tư mà làm như vậy 2 lần liên tiếp, 1 lần thắng một lần thua, thì sau hai lần số vốn lại về như cũ không tăng trưởng được gì cả. Muốn đảm bảo cho vốn tăng trưởng "về lâu về dài", cái cần tính đến không phải là giá trị kỳ vọng của vốn sau mỗi lần đầu tư, mà là giá trị kỳ vọng hình học. Nếu giả sử chỉ có 1 cơ hội đầu tư duy nhất như trên, thì giá trị kỳ vọng hình học của vốn có được sau khi đầu tư Y tiền vào đó trên tổng số X tiền sẽ là: $\sqrt{(X - Y/2)(X + Y)}$ Để tối ưu hóa giá trị kỳ vọng hình học tức là tìm Y sao cho $\sqrt{(X - Y/2)(X + Y)}$ đặt cực đại, với X cho trước. Kết quả là $Y = X/2$, và khi đó giá trị kỳ vọng hình học của vốn sau khi đầu tư là $\sqrt{(X - X/4)(X + X/2)} = 1,061.X$ Như vậy, kỳ vọng lợi nhuận của một cơ hội đầu tư như trên, tính trên toàn bộ vốn của nhà đầu tư, chỉ có không quá 6,1% chứ không phải 25%.

Định lý 2.13. *Giá trị kỳ vọng hình học có những tính chất sau:*
Tính đơn điệu: nếu $F \geq G$ thì $\mathbb{G}(F) \geq \mathbb{G}(G)$
Tính thuần nhất: Nếu c là hằng số thì $\mathbb{G}(cF) = c\mathbb{G}(F)$
Tính lõm: $(\mathbb{G}(F) + \mathbb{G}(G))/2 \leq \mathbb{G}((F + G)/2)$. Dấu bằng xảy ra khi và chỉ khi F và G tỷ lệ thuận với nhau, tức là tồn tại một hằng số dương c sao cho $G = cF$ hầu khắp mọi nơi.

Ghi chú 2.6. Tính lõm của giá trị kỳ vọng hình học chính là cơ sở của nguyên tắc **đa dạng hóa tài sản** (diversification) trong đầu tư: Bằng cách đa dạng hóa tài sản (đầu tư một phần vào F và một phần vào G , thay vì chỉ đầu tư vào F hay chỉ đầu tư vào G) có thể làm tăng giá trị kỳ vọng hình học của danh mục đầu tư (ít ra là trong trường hợp F và G có cùng kỳ vọng hình học về tăng trưởng).

Bài tập 2.19. Chứng minh bất đẳng thức $(\mathbb{G}(F) + \mathbb{G}(G))/2 \leq \mathbb{G}((F + G)/2)$, cho trường hợp không gian xác suất là một không gian hữu hạn phần tử có phân bố xác suất đều.

2.4 Phương sai, độ lệch chuẩn, và các moment

2.4.1 Phương sai và độ lệch chuẩn

Định nghĩa 2.16. Độ lệch chuẩn *(standard deviation) của một biến ngẫu nhiên X là*

$$\sigma(X) = \sqrt{\mathbb{E}((X - \mathbb{E}(X))^2)}. \quad (2.43)$$

Phương sai *(variance) của X, ký hiệu là $var(X)$, chính là bình phương của độ lệch chuẩn của X, tức là bằng $\mathbb{E}((X - \mathbb{E}(X))^2)$.*

Sử dụng tính tuyến tính của giá trị kỳ vọng, ta có thể biến đổi công thức của phương sai như sau: $\mathbb{E}((X-\mathbb{E}(X))^2) = \mathbb{E}(X^2-2\mathbb{E}(X).X+\mathbb{E}(X)^2) = \mathbb{E}(X^2)-2\mathbb{E}(X).\mathbb{E}(X)+\mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. Như vậy, ta có công thức sau:

$$var(X) = \sigma(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \quad (2.44)$$

Độ lệch chuẩn có tính thuần nhất bậc một: $\sigma(cX) = c\sigma(X)$, còn phương sai thì thuần nhất bậc hai: $var(cX) = \sigma(cX)^2 = c^2 var(X)$. Ý nghĩa của độ lệch chuẩn là: nó là thước đo độ lệch của các giá trị của X so với giá trị trung bình của nó. Định nghĩa của phương sai cho thấy nó luôn luôn lớn hơn hoặc bằng 0, và bằng 0 khi và chỉ khi X là hằng số hầu khắp mọi nơi, tức là nó không bị lệch đi đâu cả so với giá trị trung bình của nó.

Câu hỏi cho những người tò mò: Tại sao người ta lại hay dùng phương sai và độ lệch chuẩn làm thước đo cho độ lệch giữa các giá trị của một biến ngẫu nhiên X với giá trị kỳ vọng của nó, chứ không dùng một đại lượng kiểu như $\mathbb{E}(|X - \mathbb{E}(X)|)$?

Ví dụ 2.19. Nếu F nhận hai giá trị a và $-a$ ($a > 0$), mỗi giá trị với xác suất 50%, thì giá trị kỳ vọng của F là 0, phương sai của F là $a^2.50\% + (-a)^2.50\% = a^2$, và độ lệch chuẩn chính là a.

Ví dụ 2.20. Nếu F có phân bố normal $N(\mu, \sigma^2)$, thì giá trị kỳ vọng của F chính là μ, còn độ lệch chuẩn của F chính là σ. (Bài tập: chứng minh điều đó bằng các biến đổi tích phân, xuất phát từ công thức $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})dx = 1$)

Ghi chú 2.7. Đối với các biến ngẫu nhiên với vô hạn các giá trị, thì các đại lượng đặc trưng của chúng như kỳ vọng, phương sai, và các đại lượng khác, không phải lúc nào cũng tồn tại hay hữu hạn. Ví dụ, phân bố xác suất rời rạc $P(k) = C/k^2$ với mọi $k \in \mathbb{N}$, với $C = 1/(\sum 1/n^2) = 6/\pi^2$, không có kỳ vọng và sai phương hữu hạn. Ta chỉ sử dụng các đại lượng đặc trưng khi chúng tồn tại và hữu hạn.

Bài tập 2.20. Chứng minh rằng:
i) Độ lệch chuẩn của phân bố hình học với tham số p ($P(k) = p(1-p)^{k-1}$ với mọi $k \in \mathbb{N}$) là $\sigma = \dfrac{\sqrt{1-q}}{q}$.
ii) Độ lệch chuẩn của phân bố Poisson với tham số λ ($P(k) = e^{-\lambda}.\lambda^k/k!$ với mọi $k \in \mathbb{Z}_+$) là $\sigma = \sqrt{\lambda}$.

Bài tập 2.21. Giả sử X là một biến ngẫu nhiên với $\mathbb{E}(X) = 2$, và có phân bố xác suất liên tục với hàm mật độ ρ_X có dạng sau: $\rho_X(x) = ax^2 + b$ nếu $0 < x < 1$, và $\rho_X(x) = 0$ ở những điểm còn lại. Hãy tính a, b, và $var(X)$.

Bài tập 2.22. Một phòng thí nghiệm phải kiểm tra một lượng N rất lớn các mẫu máu người (mỗi mẫu của 1 người) để tìm ra các mẫu có chứa một loại kháng thể X. Thay vì xét nghiệm từng mẫu một, người ta làm như sau: Chia các mẫu thành từng nhóm, mỗi nhóm có k mẫu. Trộn các mẫu máu trong cùng một nhóm với nhau (lấy một ít máu từ mỗi mẫu) để được 1 mẫu hỗn hợp, rồi xét nghiệp mẫu hỗn hợp đó. Nếu kết quả xét nghiệm là âm tính (mẫu hỗn hợp không có kháng thể X) thì coi như cả k mẫu trong nhóm đều không có kháng thể X, còn nếu mẫu hỗn hợp có kháng thể X, thì làm tiếp k xét nghiệm, mỗi xét nghiệm cho từng mẫu của nhóm. Giả sử xác suất để 1 mẫu máu có kháng thể X

là một số p, và các mẫu máu độc lập với nhau. Gọi S là tổng số lần phải xét nghiệm.
i) Xác suất để một mẫu máu hỗn hợp có chứa kháng thể X là bao nhiêu ?
ii) Tính kỳ vọng và phương sai của S, khi tổng số mẫu máu phải kiểm tra là $N = km$.
iii) Với những giá trị nào của p thì tồn tại một số k thích hợp nào đó sao cho phương pháp xét nghiệm trên tiết kiệm được số lần xét nghiệm (kỳ vọng của S nhỏ hơn N) ? Tìm giá trị của k tối ưu, như là hàm của p.

2.4.2 Các moment của một biến ngẫu nhiên

Định nghĩa 2.17. *Nếu X là một biến ngẫu nhiên, và k là một số tự nhiên, thì đại lượng $\mathbb{E}(X^k)$ được gọi là* **moment** *(hay* **mô men***) bậc k của X, và đại lượng $\mathbb{E}((X - \mathbb{E}(X))^k)$ được gọi là* **moment trung tâm** *bậc k của X.*

Ghi chú 2.8. Có nhiều từ thuật ngữ gốc nước ngoài, mà trong tiếng Việt không có từ "thuần Việt" tương ứng, chỉ dịch phiên âm, ví dụ như mô men (moment), véc tơ (vector), mô đun (module), v.v. Trong những trường hợp như vậy, ở đây chúng ta sẽ để nguyên từ theo tiếng Anh, thay vì dùng phiên âm tiếng Việt.

Như phía trên chúng ta đã thấy, moment bậc 1 của X chính là giá trị kỳ vọng của nó, moment trung tâm bậc 1 của X thì luôn bằng 0, moment trung tâm bậc 2 của X chính là phương sai của nó, và nó có thể được biểu diễn qua các moment của X theo công thức:

$$\mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \tag{2.45}$$

Tương tự như vậy, các moment trung tâm bậc cao hơn của X cũng có thể khai triển dưới dạng đa thức của các moment của X.

Nếu ký hiệu P_X là phân bố xác suất trên $\mathbb{R}$ của X, thì ta có thể viết moment bậc k của X theo công thức sau:

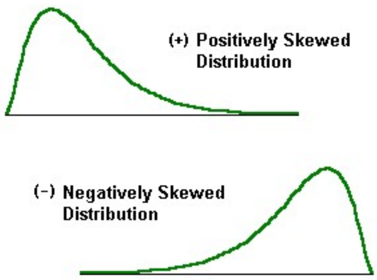
$$\mathbb{E}(X^k) = \int_{x\in\mathbb{R}} x^k dP_X. \tag{2.46}$$

Nếu như phân bố xác suất P_X la một phân bố xác suất liên tục với hàm mật độ ρ_X thì ta có thể viết:

$$\mathbb{E}(X^k) = \int_{-\infty}^{+\infty} x^k \rho_X(x) dx. \tag{2.47}$$

Các moment của một biến ngẫu nhiên cho ta các thông tin về dáng điệu của phân bố xác suất của biến ngẫu nhiên đó. Ví dụ, nếu moment trung tâm bậc 2 nhỏ, thì có nghĩa là các giá trị của X nói chung ít bị sai lệch so với giá trị kỳ vọng của nó, hay nói cách khác phần lớn xác suất của phân bổ xác suất của X tập trung trong một khoảng nhỏ xung quanh điểm giá trị kỳ vọng. Ngược lại, nếu moment trung tâm bậc 2 lớn, thì phân bố xác suất của X nói chung sẽ "dàn trải" hơn ra xa điểm giá trị kỳ vọng.

Moment trung tâm bậc 3 của X được gọi là **hệ số bất đối xứng** (skewness), hay còn có thể gọi là **độ xiên** của phân bố xác suất của X: Nếu X có phân bố xác suất đối xứng quanh điểm giá trị kỳ vọng (có nghĩa là X và $2\mathbb{E}(X) - X$ có cùng phân bố xác suất), thì moment trung tâm bậc 3 của nó bằng 0. Nếu như moment trung tâm bậc 3 lớn hơn 0 thì phân bố xác suất của X được gọi là *xiên về bên phải*, còn nếu moment trung tâm bậc 3 nhỏ hơn 0 thì phân bố xác suất của X được gọi là *xiên về bên trái*.



Hình 2.8: Phân bố bất đối xứng

Ví dụ 2.21. Moment trung tâm bậc 3 của một phân bố normal bằng 0.

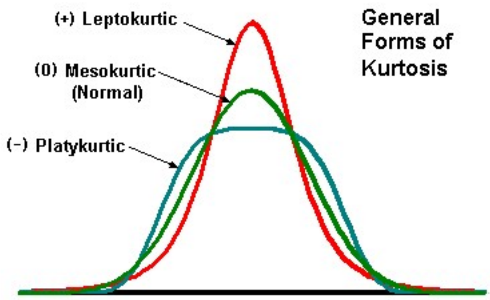
Ví dụ 2.22. Giả sử có một biến ngẫu nhiên X với phân bố xác suất rời rạc sau: $P(X = -2) = 1/2, P(X = 1) = 1/4, P(X = 3) = 1/4$. Khi đó giá trị kỳ vọng của X bằng 0, moment trung tâm bậc 3 của X bằng moment bậc 3 của X và bằng: $(1/2).(-2)^3 + (1/4).1^3 + (1/4).3^3 = 3 > 0$. Đồ thị phân bố xác suất của X (với 3 đoạn thẳng nhô lên ở 3 điểm -2,1,3 trên trục hoành) bị "lệch về bên phải" so nếu lấy điểm giá trị kỳ vọng $(= 0)$ làm trung điểm.

Moment trung tâm bậc 4 của X liên quan đến cái gọi là kurtosis[(3)] của X. Theo định nghĩa, **kurtosis** (hay còn gọi là **hệ số nhọn**) của một biến ngẫu nhiên là đại lượng

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3, \tag{2.48}$$

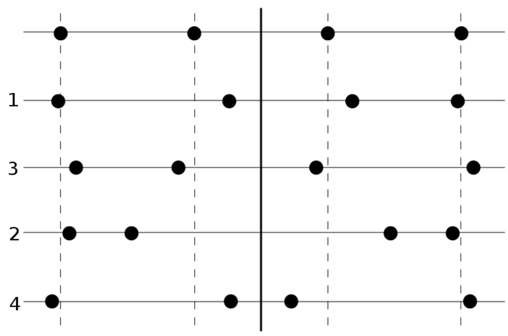
trong đó μ_4 là moment trung tâm bậc 4, còn σ là độ lệch chuẩn. Tỷ lệ μ_4/σ^4 được gọi là **moment chuẩn hóa** bậc 4. Lý do của việc chuẩn hóa này là: các moment chuẩn hóa của các phân bố normal đều là hằng số và không phụ thuộc vào độ lệch chuẩn. Moment

(3) kurtosis là một từ gốc tiếng Hy lạp, chỉ độ nhọn



Hình 2.9: Kurtosis

chuẩn hóa của bậc 4 của một phân bố normal chính bằng 3, bởi vậy kurtosis của một phân bố normal bằng 0. Khi một phân bố xác suất có kurtosis dương (phân bố như vậy gọi là phân bố **leptokurtic** hay **nhọn vượt chuẩn**) thì có nghĩa là nó "nhọn" hơn phân bố normal có cùng độ lệch chuẩn, còn khi kurtosis âm (phân bố như vậy gọi là phân bố **platykurtic**) thì có nghĩa là nó "bẹt" hơn phân bố normal có cùng độ lệch chuẩn. Nếu kurtosis bằng 0 thì phân bố được gọi là **mesokurtic**. (Xem hình 2.9).



Hình 2.10: Thay đổi các moment bậc 1 đến bậc 4

Ví dụ 2.23. Hình 2.10 là ví dụ minh họa về việc dịch chuyển 4 điểm a, b, c, d của một phân bố xác suất đều rời rạc $P(a) = P(b) = P(c) = P(d) = 1/4$, từ vị trí ban đầu $a = -3, b = -1, c = 1, d = 3$, sao cho làm tăng 1 trong 4 moment bậc 1, bậc 2, bậc 3, bậc 4 trong khi giữ nguyên 3 moment còn lại.

Tất nhiên, nếu hai biến ngẫu nhiên có cùng phân bố xác suất trên $\mathbb{R}$, thì tất cả các moment của chúng đều bằng nhau. Điều ngược có đúng không, hay nói cách khác, dãy các moment $\mathbb{E}(X^k), k = 1, 2, 3, \ldots$ của một biến ngẫu nhiên xác định hoàn toàn phân bố xác suất của biến ngẫu nhiên đó không ? Đây là một câu hỏi toán học thú vị. Có những ví dụ về các phân bố xác suất liên tục khác nhau nhưng có tất cả các moment như nhau.

Tuy nhiên, trong trường hợp các không gian xác suất chỉ có hữu hạn phần tử (mà thực ra tất cả các vấn đề trong thực tế đều chỉ có hữu hạn các khả năng xảy ra, và các mô hình liên tục với vô hạn khả năng chỉ là các mô hình mô phỏng gần đúng), thì ta có:

Mệnh đề 2.14. *Nếu X và Y là hai biến ngẫu nhiên chỉ nhận một số hữu hạn các giá trị, và có $\mathbb{E}(X^k) = \mathbb{E}(Y^k)$ với mọi $k \in \mathbb{N}$, thì phân bố xác suất của chúng trên $\mathbb{R}$ bằng nhau.*

Bài tập 2.23. Chứng minh mệnh đề trên.

Bài tập 2.24. Tính kỳ vọng và các moment trung tâm của phân bố lũy thừa với tham số λ.

2.4.3 Bất đẳng thức Chebyschev và bất đẳng thức Markov

Những bất đẳng thức tương đối đơn giản sau đây của Chebyschev và Markov liên quan đến các moment sẽ có ích trong việc đánh giá phân bố xác suất của các biến ngẫu nhiên.

Định lý 2.15. *(Bất đẳng thức Chebyschev cho kỳ vọng) Với mọi biến ngẫu nhiên X chỉ nhận các giá trị không âm, và mọi số dương $a > 0$ ta có*

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}. \tag{2.49}$$

Chứng minh. Gọi X_a là biến ngẫu nhiên sau: $X_a = a$ khi $X \geq a$ và $X_a = 0$ khi $X < a$. Khi đó $X \geq X_a$, và X_a chỉ nhận hai giá trị 0 và a. Bởi vậy

$$\mathbb{E}(X) \geq \mathbb{E}(X_a) = 0.P(X_a = 0) + a.P(X_a = a) = a.P(X \geq a),$$

từ đó suy ra điều phải chứng minh. □

Định lý 2.16. *(Bất đẳng thức Markov cho các moment tuyệt đối) Với mọi biến ngẫu nhiên X, số dương $a > 0$, và số tự nhiên k, ta có*

$$P(|X| \geq a) \leq \frac{\mathbb{E}(|X|^k)}{a^k}. \tag{2.50}$$

Chứng minh. Suy ra từ bất đẳng thức Chebyschev cho biến ngẫu nhiên $|X|^k$ và hằng số a^k. □

Định lý 2.17. *(Bất đẳng thức Chebyschev cho phương sai) Nếu X là một biến ngẫu nhiên có phương sai $var(X)$ hữu hạn và $a > 0$ bất kỳ, ta có*

$$P(|X - \mathbb{E}(X)| \geq a) \leq \frac{var(X)}{a^2}. \tag{2.51}$$

Hình 2.11: Pafnouti Lvovitch Chebyschev (1821-1894)

Hình 2.12: Andrei Andreevitch Markov (1856-1922)

Chứng minh. Suy ra từ bất đẳng thức Markov cho biến ngẫu nhiên $X - \mathbb{E}(X)$ và cho $k = 2$. $\square$

Ghi chú 2.9. Pafnouti Lvovitch Chebyschev (1821-1894) là một nhà toán học người Nga. Ngoài lý thuyết xác suất, ông ta còn nghiên cứu nhiều về số học và đại số. Các đa thức U_n bậc n thỏa mãn $U_n(\cos(x)) = \dfrac{\sin((n+1)x)}{\sin x}$ được gọi là *đa thức Chebyschev*, và chúng xuất hiện nhiều trong toán học và ứng dụng. Andrei Andreevitch Markov (1856-1922) cũng là một nhà toán học người Nga, và là học trò của Chebyschev. Các *xích Markov* (Markov chains) đặc biệt quan trọng trong lý thuyết xác suất về các quá trình ngẫu nhiên (stochastic processes). Các quá trình ngẫu nhiên nằm ngoài khuôn khổ của cuốn sách này, nhưng sẽ được bàn đến trong một cuốn sách tiếp theo.

2.5 Hàm đặc trưng, hàm sinh, và biến đổi Laplace

Thay vì xét các moment $\mathbb{E}(X^k)$ của một biến ngẫu nhiên X, ta có thể xét các giá trị đặc trưng dạng $\mathbb{E}(\exp(yX))$ trong đó y là một tham số nào đó. Khi ta biến đổi y trong một miền nào đó trên $\mathbb{R}$ hoặc $\mathbb{C}$, sẽ ta được một hàm các giá trị đặc trưng của X. Sự liên quan giữa hàm này và các moment được thể hiện qua đẳng thức sau (xảy ra nếu như ta có các điều kiện về hội tụ):

$$M_X(y) = \mathbb{E}(\exp(yX)) = \mathbb{E}(\sum_k (y^k/k!).X^k) = \sum_k \mathbb{E}(X^k).(y^k/k!) \tag{2.52}$$

Hàm $M_X(y) = \mathbb{E}(\exp(yX))$ được gọi là **hàm sinh moment** của X.

2.5.1 Hàm đặc trưng

Trong biểu thức $M_X(y) = \mathbb{E}(\exp(yX))$, nếu ta lấy $y = is$, (ở đây $i = \sqrt{-1}$) ,với $s \in \mathbb{R}$, thì ta có $\exp(yX) = \exp(isX) = \cos(sX) + i\sin(sX)$ là một biến ngẫu nhiên bị chặn (có giá trị tuyệt đối bằng 1), và ta có thể yên tâm về sự tồn tại của $\mathbb{E}(\exp(isX))$. Từ đó có định nghĩa sau:

Định nghĩa 2.18. Hàm đặc trưng *của một biến ngẫu nhiên thực X là hàm $\Phi_X : \mathbb{R} \to \mathbb{C}$ được cho bởi công thức*

$$\Phi_X(s) = \mathbb{E}(\exp(isX)) = \int_{x\in\mathbb{R}} e^{isx} dP_X. \tag{2.53}$$

Ví dụ 2.24. Hàm đặc trưng của một số phân bố xác suất quen thuộc:
i) Hàm đặc trưng của một hằng số c (tức là biến ngẫu nhiên chỉ nhận mỗi giá trị c) là $\Phi_c(s) = e^{ics}$.
ii) Hàm đặc trưng của phân bố nhị thức với các tham số n, p là hàm $(1 - p + pe^{is})^n$.
iii) Hàm đặc trưng của phân bố xác suất đều trên một đoạn thẳng $[a, b]$ là hàm $\dfrac{e^{ibs} - e^{-ias}}{i(b-a)s}$.
iv) Hàm đặc trưng của phân bố xác suất lũy thừa với tham số 1 (với mật độ $\rho(x) = e^{-x}$ khi $x > 0$) là hàm $\dfrac{1}{1 - is}$.
v) Hàm đặc trưng của phân bố xác suất normal chuẩn tắc $N(0,1)$ là hàm $\Phi(s) = \exp(-s^2/2)$.

(Bài tập: Hãy suy ra các công thức trên từ định nghĩa của hàm đặc trưng và của các phân bố xác suất).

Định lý 2.18. *Một số tính chất của hàm đặc trưng:*
i) $\Phi_X(0) = 1$
ii) $|\Phi_X(s)| \leq 1$ *với mọi* $s \in \mathbb{R}$
iii) Nếu $Y = aX + b$ *với* a, b *là các hằng số, thì* $\Phi_Y(s) = e^{\sqrt{-1}bs}\Phi_X(as)$.
iv) Φ_X *liên tục đều trên* $\mathbb{R}$.
v) Nếu $\mathbb{E}(|X|^k) < \infty$ *với một số tự nhiên* k *nào đó, thì hàm đặc trưng* Φ_X *khả vi liên tục* k *lần trên* $\mathbb{R}$, *và*

$$\mathbb{E}(X^k) = \frac{1}{(\sqrt{-1})^k}.\Phi_X^{(k)}(0), \tag{2.54}$$

trong đó $\Phi_X^{(k)}$ *là ký hiệu đạo hàm bậc* k *của* Φ_X.

Chứng minh. Ba tính chất đầu tiên tương đối hiển nhiên, suy ra ngay từ định nghĩa. Tính chất thứ tư là bài tập dành cho những bạn đọc quen với khái niệm liên tục đều. Để chứng minh tính chất cuối cùng, chúng ta nhớ rằng phép lấy giá trị kỳ vọng là một phép lấy giá trị trung bình, có thể hiểu như là một phép lấy tổng (của một chuỗi), và do đó nó giao hoán với phép lấy đạp hàm (khi một số điều kiện hội tụ nào đó được thỏa mãn). Áp dụng nguyên tắc giao hoán đó vào định nghĩa của hàm đặc trưng, ta có đạo hàm bậc k của hàm đặc trưng là:

$$\Phi_X^{(k)}(s) = \frac{d^k}{ds^k}\Phi_X(s) = \mathbb{E}(\frac{d^k}{ds^k}\exp(isX)) = \mathbb{E}((iX)^k\exp(isX)) = i^k\mathbb{E}(X^k\exp(isX)). \tag{2.55}$$

Đặt $s = 0$, ta được $\Phi_X^{(k)}(0) = i^k\mathbb{E}(X^k)$. □

2.5.2 Tìm lại phân bố xác suất từ hàm đặc trưng

Chúng ta có công thức giới hạn sau đây, cho phép tìm lại được phân bố xác suất từ hàm đặc trưng của nó:

Định lý 2.19. *Gọi* P_X *là phân bố xác suất của một biến ngẫu nhiên* X *tùy ý, và* Φ_X *là hàm đặc trưng của nó. Khi đó với mọi* $a, b \in \mathbb{R}$, $a < b$, *ta có:*

$$\lim_{R\to\infty}\frac{1}{2\pi}\int_{-R}^{R}\frac{e^{-ias} - e^{-ibs}}{is}\Phi_X(s)ds = P_X(]a,b[) + \frac{P_X(a) + P_X(b)}{2}. \tag{2.56}$$

Chúng ta sẽ chấp nhận định lý trên mà không chứng minh. Nếu bạn đọc đã biết qua về giải tích Fourier thì có thể tự chứng minh nó không quá khó khăn (nó tương tự như định lý Dirichlet cho chuỗi Fourier). Nếu không thì có thể xem chẳng hạn trong Chương 9 của quyển sách của Koralov và Sinai [5].

Trong trường hợp X có phân bố xác suất liên tục với hàm mật độ ρ_X, thì ta có thể viết

$$\Phi_X(s) = \int_{-\infty}^{+\infty} e^{isx}\rho_X(x)dx. \tag{2.57}$$

Trong giải tích, phép tính trên gọi là phép biến đổi Fourier. Có nghĩa là, hàm đặc trưng chính là **biến đổi Fourier** của hàm mật độ.

Chia cả hai vế của công thức (2.56) cho $b - a$, và cho b tiến tới a, ta được công thức sau, gọi là phép **biến đổi ngược Fourier**, để tính hàm mật độ từ hàm đặc trưng:

$$\rho_F(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-isx}\Phi_F(s)ds. \tag{2.58}$$

Trong trường hợp X là biến ngẫu nhiên nguyên (chỉ nhận giá trị trong $\mathbb{Z}$), thì hàm đặc trưng Φ_X của X chính là **chuỗi Fourier** với các hệ số là các xác suất $P_X(k) = P(X = k)$, $k \in \mathbb{Z}$:

$$\Phi_X(s) = \sum_{k\in\mathbb{Z}} P_X(k)\exp(iks), \tag{2.59}$$

và ta có thể tính ra $P_X(k)$ từ Φ_X theo công thức quen thuộc để tính các hệ số của một chuỗi Fourier:

$$P_X(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-iks}\Phi_X(s)ds \tag{2.60}$$

Ghi chú 2.10. Joseph Fourier (1768–1830) là nhà toán học và vật lý Pháp. Trong khoảng thời gian 1798–1801 Fourier đi theo Napoléon, cùng với 35000 lính Pháp và một đoàn các nhà khoa học, sang chinh chiến ở Ai Cập (Egypt) và tìm hiểu nền văn minh Ai Cập. Khi ở Ai Cập, Fourier trở thành người điều hành Viện Hàn lâm Ai Cập do Napoléon lập ra, và sau đó điều hành luôn cả các công việc hành chính và ngoại giao ở Ai Cập, gần như là quan toàn quyền. Fourier tỏ ra rất có tài về chính trị và ngoại giao, có thể đàm phán, hòa giải các bên đối lập. Sau khi Pháp đầu hàng Anh ở Ai Cập năm 1801 và Fourier trở về Pháp, được cử làm tỉnh trưởng (préfet) vùng Isère. Trong thời gian ở Ai Cập, Fourier phát minh ra chuỗi Fourier, khi nhìn thấy các lớp sóng cát (dunes) ở sa mạc. Chuỗi Fourier và biến đổi Fourier là một thứ *công cụ vạn năng*, không chỉ quan trọng trong xác suất, mà còn xuất hiện khắp nơi trong toán học và vật lý.

Trong trường hợp tổng quát, một phân bố xác suất cũng được xác định một cách duy nhất bởi hàm đặc trưng của nó:

Hình 2.13: Joseph Fourier (1768–1830)

Định lý 2.20. *Hai biến ngẫu nhiên có cùng phân bố xác suất khi và chỉ khi chúng có cùng hàm đặc trưng.*

Chứng minh. Giả sử hai phân bố xác suất P_X và P_Y có cùng hàm đặc trưng Φ. Công thức (2.56) dẫn đến:

$$P_X(]a,b[) + \frac{P_X(a)+P_X(b)}{2} = P_Y(]a,b[) + \frac{P_Y(a)+P_Y(b)}{2}$$

với mọi $a < b$. Ta có thể chọn a và b là những điểm liên tục của $\mathcal{F}_X$ và $\mathcal{F}_Y$, rồi cho a tiến tới $-\infty$, ta được: $\mathcal{F}_X(b) = \mathcal{F}_Y(b)$ tại mọi điểm b mà là điểm liên tục của cả $\mathcal{F}_X$ và $\mathcal{F}_Y$. Giả sử $x \in \mathbb{R}$ là một điểm tùy ý. Nhắc lại rằng số điểm gián đoạn của một hàm phân phối xác suất trên $\mathbb{R}$ là không quá đếm được. Vì thế tồn tại một dãy các điểm $x_n > x$ sao cho x_n tiến tới x khi n tiến tới vô cùng, và x_n là điểm liên tục của $\mathcal{F}_X$ và $\mathcal{F}_Y$ với mọi n. Nhắc lại rằng, các hàm phân phối xác suất có tính chất liên tục bên phải. Do đó ta có: $\mathcal{F}_X(x) = \lim_{n\to\infty} \mathcal{F}_X(x_n) = \lim_{n\to\infty} \mathcal{F}_Y(x_n) = \mathcal{F}_Y(x)$. Như vậy, hai hàm phân phối xác suất $\mathcal{F}_X$ và $\mathcal{F}_Y$ trùng nhau, do đó hai phân bố xác suất P_X và P_Y cũng trùng nhau. □

Bài tập 2.25. Ta sẽ gọi một biến ngẫu nhiên X là *đối xứng* nếu như X và $-X$ có cùng phân bố xác suất. Hãy xây dựng những ví dụ biến ngẫu nhiên đối xứng, và chứng minh rằng một biến ngẫu nhiên là đối xứng khi và chỉ khi hàm đặc trưng Φ_X của nó là một hàm thực (tức là $\Phi_X(s) \in \mathbb{R}$ với mọi $s \in \mathbb{R}$).

2.5.3 Hàm sinh xác suất và biến đổi Laplace

Trong biểu thức $\mathbb{E}(\exp(yX))$, nếu đặt $y = \ln z$, thì ta được hàm sau, gọi là **hàm sinh xác suất**:

$$G_X(z) = \mathbb{E}(z^X) \tag{2.61}$$

Hàm sinh xác suất hay được dùng khi mà các giá trị của biến ngẫu nhiên đều là số nguyên không âm. Khi đó hàm sinh xác suất có dạng đa thức hoặc chuỗi Taylor có bán kính hội tụ lớn hơn hoặc bằng 1:

$$G_X(z) = \sum_k P_X(k).z^k, \tag{2.62}$$

và ta có $P(X = k) = \frac{1}{k!}\frac{d^k G_X(z)}{dz^k}\Big|_{z=0}$ với mọi $k \in \mathbb{Z}_+$.

Từ quan điểm của giải tích phức, hàm đặc trưng $\Phi_X(s)$ và hàm sinh $G_X(z)$ gần như là một, có thể chuyển từ hàm này sang hàm kia bằng cách đổi biến. Bởi vậy, tất nhiên các moment của một biến ngẫu nhiên cũng có thể suy ra được từ hàm sinh xác suất của biến ngẫu nhiên đó. Ta có định lý sau:

Định lý 2.21. *Giả sử X là một biến ngẫu nhiên với hàm sinh xác suất G . Khi đó:*
1) $\mathbb{E}(X) = G'(1)$
2) $var(X) = \sigma^2(X) = G''(1) + G'(1) - (G'(1))^2$
3) $\mathbb{E}(X(X-1)\dots(X-k+1)) = G^{(k)}(1)$ *với mọi $k \in \mathbb{N}$. Ở đây $G^{(k)}$ là đạo hàm bậc k của G.*

Ví dụ 2.25. Hàm sinh xác suất của một biến ngẫu nhiên X với phân bố Poisson với tham số λ là hàm $G_X(z) = \exp((z-1)\lambda)$ Thật vậy, ta có: $G_X(z) = \mathbb{E}(z^X) = \sum_k z^k P(X = k) = \sum_k e^{-\lambda}\lambda^k z^k/k! = e^{-\lambda}e^{\lambda z} = e^{\lambda(z-1)}$. Từ đó suy ra $\mathbb{E}(X) = G'_X(1) = \lambda, G''_X(1) = \lambda^2$ và $var(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Trong trường hợp biến ngẫu nhiên X chỉ nhận các giá trị thực không âm, người ta hay dùng **hàm Laplace** $L_X(t) : \mathbb{R}_+ \to \mathbb{R}$, nhân được từ biểu thức $\mathbb{E}(\exp(yX))$ bằng cách đặt $t = -y$:

$$L_X(t) = \mathbb{E}(\exp(-tX)). \tag{2.63}$$

Ở đây ta coi biến t nằm trong tập các số thực không âm. Với giả sử rằng F chỉ nhận các giá trị không âm, ta luôn có $0 < (\exp(-tF)) \leq 1$, từ đó suy ra các giá trị của $L_F(t)$ là số dương và bị chặn trên bởi 1.

Trong trường hợp F có phân bố xác suất liên tục với hàm mật độ ρ_F thỏa mãn điều kiện $\rho_F(x) = 0$ với mọi $x < 0$ (có nghĩa là F không nhận các giá trị âm), thì ta có

$$L_F(t) = \int_0^\infty \exp^{-tx} \rho_F(x)dx, \tag{2.64}$$

và hàm $L_F(t)$ được gọi là **biến đổi Laplace** của hàm mật độ $\rho_F(x)$.

Tương tự như đối với hàm sinh và hàm đặc trưng, các đạo hàm của hàm $L_F(t)$ tại điểm $t = 0$ cũng cho ta các moment của F.

Hình 2.14: Pierre-Simon Laplace (1949-1827)

Ghi chú 2.11. [(4)] Pierre-Simon Laplace (1749-1827) là nhà toán học, thiên văn học và vật lý người Pháp, một trong những nhà khoa học có thế lực nhất ở châu Âu thời đại ông ta. Ông ta nghiên cứu rất nhiều thứ, từ xác suất (định lý giới hạn trung tâm, biến đổi Laplace) đến giải tích điều hòa, cơ học, âm thanh, truyền nhiệt, các thiên thể, v.v. Laplace chính là người đặt ra giả thuyết về lỗ đen (black hole) và về sự co lại do trọng lượng (gravitational collapse) trong vật lý thiên văn. Laplace còn có tham vọng về chính trị, là thành viên của thượng nghị viện. Có lúc làm Bộ trưởng Bộ nội vụ dưới thời Napoléon, nhưng sau 6 tuần thì bị cách chức vì không được việc. Laplace bị nhiều người cùng thời không ưa vì tính bạc bẽo, ích kỷ, có khi còn vơ cả công trình của người khác thành của mình, và thay đổi quan điểm chính trị như chong chóng "theo chiều gió". Nhưng về mặt khoa học, Laplace là một con người vĩ đại của thế kỷ 18-19. Biến đổi Laplace được gọi như vậy là do Laplace đưa vào để nghiên cứu xác suất, cùng với hàm sinh xác suất. Biến đổi Laplace còn xuất hiện ở nhiều nơi khác trong vật lý và toán học. Leonhard Euler (1707–1783) có lẽ là người đầu tiên nghĩ ra biến đổi này.

[(4)] Xem wikipedia: http://fr.wikipedia.org/wiki/Pierre-Simon_Laplace.

Bài tập 2.26. Chứng minh rằng hàm sinh xác suất của một biến ngẫu nhiên với phân bố hình học với tham số p là hàm $G(z) = \dfrac{pz}{1 - z + pz}$. Từ đó suy ra kỳ vọng và phương sai của phân bố hình học.

Bài tập 2.27. Tính hàm sinh xác suất và hàm Laplace của phân bố nhị thức với các tham số n, p.

Bài tập 2.28. Chứng minh định lý 2.21 cho trường hợp F nhỉ nhận một số hữu hạn các giá trị.