

# LỢI ÍCH CỦA PHÂN TÍCH KHOA HỌC LUẬN TRONG DẠY HỌC XÁC SUẤT THỐNG KÊ MỘT PHÂN TÍCH KHOA HỌC LUẬN VỀ PHÂN PHỐI CHUẨN

ĐÀO HỒNG NAM\*

## TÓM TẮT

*Trong bài báo này, chúng tôi thực hiện một phân tích khoa học luận nhằm điểm lại những nét chính trong quá trình hình thành và phát triển của luật phân phối chuẩn, giả thiết được áp dụng phổ biến khi thực hiện các kiểm định thống kê và làm nên nền tảng của phân tích thống kê, làm rõ những đặc trưng khoa học luận cơ bản và chỉ ra những lợi ích của phân tích khoa học luận trong dạy học khái niệm này.*

**Từ khóa:** khoa học luận, tri thức, phân phối chuẩn.

## ABSTRACT

### *Benefits of epistemological analysis of Probability Statistics in teaching the concept of normal distribution*

*In this paper, we conduct an epistemological analysis to review the main aspects in the process of formation and development of the Normal Distribution Law- the assumption commonly applied when performing the statistical tests, serving as the basis of statistical analysis, to clarify epistemological characteristics of this law and point out the benefits of the epistemological analysis in teaching this concept.*

**Keywords:** epistemology, knowledge, normal distribution.

## 1. Cơ sở lý thuyết

### 1.1. Về thuật ngữ “khoa học luận”: nguồn gốc và sự tiến triển

Thuật ngữ *khoa học luận* xuất hiện vào thế kỉ XIX, nó được hình thành từ hai từ gốc Hi Lạp *épistèmè* (khoa học) và *logos* (nghiên cứu về). Khoa học luận là một bộ phận quan trọng của triết học khoa học (philosophy of sciences)

Phân tích khoa học luận một tri thức (TT) là một nghiên cứu lịch sử hình thành TT đó, phân tích nhằm làm rõ:

- Những điều kiện, những trở ngại cho sự nảy sinh TT khoa học và sự “tiến triển” của TT hay kiến thức;

- Nghĩa của TT, những vấn đề mà TT đó cho phép giải quyết;

- Những quan niệm có thể gắn liền với TT.

Phân tích khoa học luận giúp ta hiểu rõ mối liên hệ giữa quá trình xây dựng TT trong cộng đồng khoa học với việc dạy và học TT này.

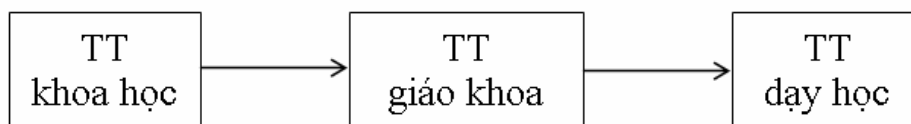
Thuật ngữ “tiến triển” được hiểu theo nghĩa rộng: nó có thể liên quan đến sự biến đổi tình trạng kiến thức của một hệ thống, một thể chế hay một cá thể. Nó không chỉ chú ý đến những tư tưởng tiến bộ mà còn đến cả những trì trệ, những bước lùi.

### 1.2. Khoa học luận và chuyển đổi *didactic*

\* NCS, Trường Đại học Sư phạm TP HCM

TT khoa học toán học là đối tượng nghiên cứu của các nhà toán học, được lưu hành trong cộng đồng các nhà toán học. Từ TT khoa học toán học đến TT toán học mà học sinh (HS) học được có sự **chuyển đổi didactic**. Đó là sự chuyển

đổi từ TT khoa học thành TT giáo khoa được thể hiện trong chương trình hay trong các sách giáo khoa (SGK). Từ TT giáo khoa, thầy giáo chuyên đổi thành TT dạy học. (hình 1)



**Hình 1.** Sự chuyển đổi didactic giữa các cấp độ TT

*1.2.1. Tri thức khoa học*

Sự ra đời của một TT bác học là kết quả của một hoạt động khoa học gắn liền với lịch sử cá nhân của nhà nghiên cứu. Để giải quyết một vấn đề toán học nào đó, các nhà toán học phải khám phá ra những phương pháp, những kiến thức và chọn một trong số các kiến thức đủ mới, đủ hay, có thể thông báo cho cộng đồng khoa học. Để trình bày một TT toán học, các nhà toán học tạo cho những kiến thức này một dạng khái quát nhất có thể được, theo những quy tắc diễn đạt thông dụng đang lưu hành trong cộng đồng khoa học mà không trình bày lịch sử của TT đó, không nêu lại quá trình tìm tòi, phát minh của mình, bỏ qua những sai lầm gặp phải. Khi đó, TT khoa học đã được phi hoàn cảnh hóa, phi cá nhân hóa và phi thời gian hóa.

*1.2.2. Tri thức giáo khoa*

Trong những TT toán học được tích lũy qua lịch sử, các nhà thiết kế chương trình chọn ra một số vấn đề làm đối tượng dạy học. Để trở thành TT có thể dạy được cho một bộ phận công chúng, TT lại tiếp tục bị biến đổi sao cho đảm bảo được sự tương hợp giữa hệ thống dạy

học với môi trường của nó. TT này được mô tả chính thức trong chương trình học hay thể hiện trong SGK. Các nhà biên soạn tạo nên SGK nhằm tìm cách trình bày lại những TT được chọn để có thể dạy được cho một bộ phận công chúng xác định, phù hợp với thể chế dạy - học, đối tượng dạy - học.

Để TT được sắp xếp theo một thứ tự hợp logic và người học có thể lĩnh hội được, nhiều khi tác giả phải viết lại các định nghĩa, các tính chất, biến đổi các phép chứng minh, thậm chí nhiều tính chất toán học chỉ được công nhận mà không chứng minh nếu việc chứng minh quá phức tạp và khó hiểu đối với HS. Quá trình tái sáng tạo này có thể dẫn tác giả đến chỗ sáng tạo ra một số đối tượng mới. Hệ quả là sự xuất hiện một sự chênh lệch khá lớn giữa TT khoa học với TT xuất hiện trong chương trình và SGK.

*1.2.3. Tri thức dạy học*

Đây là TT được giảng dạy trên lớp học. Dựa vào trình độ từng đối tượng HS, cơ sở vật chất, phương tiện giảng dạy và phương pháp sư phạm của mình, thầy giáo sẽ chuyển tải những hiểu biết của họ về TT đó đến HS.

Về phương diện didactic, cái được hay không được của thầy giáo và HS là TT nhưng các phương án về TT đó thì khác nhau do vị trí của thầy giáo và HS trong môi quan hệ didactic.

### **1.3. Sự cần thiết của phân tích khoa học luận đối với việc nghiên cứu đối tượng cần dạy**

Muốn phân tích độ chênh lệch giữa TT bác học và TT được dạy phải căn cứ vào nội dung TT bác học trên quan điểm khoa học luận, nghĩa là trên những yếu tố do phân tích khoa học luận mang lại: nghĩa của TT, những vấn đề mà TT đó cho phép giải quyết, những trở ngại cho sự hình thành TT, những điều kiện cho phép TT nảy sinh,... Đây là những hiểu biết cần thiết cho việc thiết kế một môi trường để trong đó hoạt động học xảy ra.

Phân tích khoa học luận giúp ta vạch rõ các tham chiếu hợp thức của TT cần dạy, trả lại cho TT những nghĩa rộng hơn, điều mà việc nghiên cứu đơn thuần chương trình và SGK không thể mang lại. Những hiểu biết khoa học luận về TT cần dạy giúp nhà nghiên cứu và giáo viên nhìn nó ở một khoảng cách cần thiết, không hoàn toàn bị bó hẹp trong nội tại hệ thống dạy học, không chỉ xem xét nó dưới lăng kính của chương trình và SGK.

## **2. Một phân tích khoa học luận về phân phối chuẩn**

Luật phân phối chuẩn (PPC) là một phần quan trọng của lý thuyết xác suất (XS) và ứng dụng thống kê (TK). Có thể nói, nếu không có PPC thì không có phép kiểm định TK, cũng không có cả phân tích TK. Vai trò của luật PPC đã được chứng minh và khẳng định giá trị qua

nhiều thế hệ nghiên cứu học thuật, lịch sử phát triển của khái niệm PPC gắn liền với rất nhiều thành tựu quan trọng trong nhiều lĩnh vực toán học, đặc biệt là giai đoạn toán học hiện đại.

PPC không chỉ là một công cụ toán học hay chỉ có giá trị biểu tượng, mà là một quy luật vận hành của giới tự nhiên và xã hội đã được nhận thức và kiểm chứng. Theo đó, nguyên nhân tính phổ biến rộng rãi của PPC đã được giải thích. Theo Borel, PPC là quy luật của hiện tượng ngẫu nhiên, trong đó nhiều nguyên nhân tác động mà không nguyên nhân nào là quyết định. Theo Liapunov, đại lượng ngẫu nhiên là tổng một số lớn các đại lượng ngẫu nhiên độc lập và phương sai rất bé so với phương sai của tổng. Trong thực tiễn, mỗi đại lượng ngẫu nhiên được sinh ra bởi vô số nguyên nhân, trong đó không nguyên nhân nào chiếm ưu thế so với các nguyên nhân còn lại, nên phương sai của mỗi nguyên nhân là rất nhỏ không so sánh được với phương sai của đại lượng được xét. Do đó phần lớn các đại lượng ngẫu nhiên trong thực hành có PPC.

Vì những lý do trên, chúng tôi thực hiện một nghiên cứu khoa học luận nhằm điểu lại những nét chính trong quá trình hình thành và phát triển của luật PPC, trên cơ sở đó làm rõ những đặc trưng khoa học luận cơ bản và chỉ ra những điều kiện thiết yếu trong việc dạy học khái niệm này.

Về mục tiêu cụ thể, chúng tôi tiến hành khảo sát và tìm câu trả lời cho những câu hỏi về PPC như sau:

- Điều kiện nảy sinh và những đặc trưng cơ bản của khái niệm PPC là gì?

- Khái niệm PPC được hình thành và phát triển qua các giai đoạn nào?

- Bài toán và các đối tượng liên quan gắn với khái niệm PPC là gì?

- Có những cách tiếp cận nào liên quan đến khái niệm PPC? Sự tương đồng và khác biệt giữa những cách tiếp cận này là gì?

Khi trả lời những câu hỏi trên, chúng tôi hướng đến nhiệm vụ chuyên môn sư phạm là xây dựng nội dung và hệ thống khái niệm nền tảng cho XS-TK, làm cơ sở cho kỹ năng vận dụng hiệu quả các kiến thức, nâng cao chất lượng giảng dạy môn Xác suất và Thống kê Y học trong Đại học Y Dược (ĐHYD) TPHCM.

Chúng tôi cũng hướng đến nhiệm vụ giáo dục đào tạo là nâng cao trình độ, đào tạo và bồi dưỡng nhân lực ngành y tế trên cả ba mặt quy mô, chất lượng và hiệu quả, đáp ứng nhu cầu xã hội theo chiến lược phát triển của ngành y tế trong giai đoạn hiện nay.

Về giới hạn nghiên cứu, chúng tôi tập trung phân tích các khái niệm dùng trong phương pháp thống kê xác suất (probability-based statistical method) và thống kê có tham số (parametric statistics) trong khoảng thời gian lịch sử từ năm 1730 đến năm 1920. Đây cũng là những kiến thức cơ bản, mở đầu của học phần XS-TK được giảng dạy tại ĐHYD TPHCM.

### 2.1. Phân tích khoa học luận lịch sử hình thành khái niệm PPC

Ban đầu PPC xuất hiện chỉ như một công cụ giải tích trợ giúp cho tính toán

XS. Sau một quá trình tích lũy phát triển lâu dài qua nhiều thế hệ các nhà nghiên cứu PPC mới được công nhận là khái niệm trọng tâm của XS-TK. Luật PPC được ứng dụng cho nhiều hiện tượng tự nhiên với các đơn vị đo khác nhau và tham số khác nhau, điều này gây khó khăn khi so sánh biến số. Vì vậy, cần xây dựng luật phân phối chuẩn tắc (PPCT) độc lập với đơn vị đo. Do PPC được xác định bởi 2 tham số trung bình và độ lệch chuẩn, nên PPCT có trung bình là 0 và độ lệch chuẩn là 1. Phép biến đổi chuẩn hóa biến ngẫu nhiên chính là hoán chuyển  $z$ , cho kết quả là chỉ số  $z$ . Một biến ngẫu nhiên  $Z$  được gọi là có PPCT khi nó có hàm mật độ xác suất  $f(z) = \frac{1}{\sqrt{2p}} e^{(-0.5z^2)}$ .

Hàm này hiển nhiên dương, nhưng không dễ dàng thấy được đó là hàm mật độ xác suất, vì muốn vậy cần tính được tích phân

$$\int_0^{\infty} e^{(-0.5z^2)} dz.$$

Vào năm 1774, Laplace đưa ra tính toán chặt chẽ đầu tiên đánh giá tích phân trên trong bài viết *Mesmoire sur la probabilités des causes par les esvenements*

$$\int_0^{\infty} e^{(-0.5z^2)} dz = \frac{\sqrt{2p}}{2}$$

Cuốn sách đầu tiên về lý thuyết xác suất, “*The Doctrine of Chances: or a method of calculating the probability of events in play*” được viết bởi Abraham de Moivre và được xuất bản 3 lần vào những năm 1718, 1738 và 1756. Trong đó, khái niệm mật độ XS chưa được đề cập mà chỉ xoay quanh vấn đề luật của

khai triển nhị thức  $(a+b)^n$ , nghiên cứu sâu các hệ số của hạng tử và chỉ ra khi  $n$  lớn, hệ số của hạng tử trung tâm xấp xỉ  $\frac{2}{\sqrt{2\pi n}}$ . Như vậy đây là dẫn nhập đầu tiên về PPC như là một giới hạn của dãy phân phối nhị thức.

Trong cuốn sách, vai trò của định lý giới hạn trung tâm được quan tâm, với định hướng ứng dụng trong khoa học bảo hiểm. Các định nghĩa và kết quả được trình bày với nhiều tính trực giác và thực nghiệm: “Phân phối XS của một số lần đặt mặt ngửa khi tung một đồng xu 1800 lần”. Nghiên cứu của Moivre chỉ giới hạn trong các yếu tố cơ bản nhất của lý thuyết XS, chứ chưa đề cập đến các vấn đề phương pháp của TK, ngay cả các ứng dụng cũng chỉ trong vấn đề về tính may rủi và thăng giáng của các hiện tượng. Đáng chú ý là đóng góp của ông nhằm xây dựng công cụ tính toán gần đúng khi tham số lớn, ông chỉ nhằm xây dựng phép xấp xỉ chuẩn tắc cho nhị thức Newton, đường cong chuẩn tắc chỉ đóng vai trò công cụ tính toán nhằm trình bày một xấp xỉ liên tục cho một đối tượng toán học rời rạc, chứ không phải là đường cong liên tục mật độ XS. Đây là phương pháp hàm sinh (generating function), được sử dụng xuyên suốt tác phẩm. Ông cũng chỉ ra sự quan trọng của cỡ mẫu  $n$ , chỉ ra độ lệch từ trung tâm phụ thuộc vào  $n$ . Ông cũng chỉ ra bản chất của luật PPC, khi XS của biến cố là 0.5, PPC là luật phân phối giới hạn của số lần xảy ra biến cố khi số các phép thử tăng lên vô hạn. Tuy nhiên, ông không đi xa hơn, để tiến đến luật số lớn và định lý giới hạn trung

tâm. Những thành quả này cần nhiều năm sau do một nhà cơ học triết học và toán học người Pháp khác khám phá. Thời kỳ của Moivre gắn liền với lý thuyết XS cổ điển, TK học chưa có sự phát triển về phương pháp riêng và nền tảng toán học mà chủ yếu tập trung vào mô tả. TK suy luận phải đợi một thời gian sau, với các công trình của Laplace và Gauss.

Các nghiên cứu liên quan đến PPC được định hình từ hai nguồn, nguồn thứ nhất tiếp nối các công trình của Moivre đưa ra các tính toán chuẩn hóa tham số và suy luận về tham số của nhị thức Newton, nguồn thứ hai xuất phát từ bài toán ngược trong cơ học nhằm suy đoán ước lượng hệ số của mô hình tuyến tính, gọi là phương pháp bình phương tối thiểu.

Năm 1782, Laplace với những đóng góp to lớn về lý luận và tính toán đã đưa ra khái niệm về hàm mật độ XS và chuẩn hóa các tham số của PPC. Công lao của Laplace là nêu lên vai trò trung tâm của luật PPC, đóng góp nhiều xây dựng về lý thuyết cho khái niệm PPC. Ông đã chỉ ra cách lập luận dựa vào XS để suy luận về tham số  $p$ , không những thế, ông đã mở rộng phương pháp suy luận tổng quát hóa kết quả cho rất nhiều các phân phối XS.

Năm 1812, ông hoàn tất công trình *Analytical theory of probabilities*, trong đó trình bày các kết quả căn bản với hình thức toán học chặt chẽ và toàn bộ lý thuyết sai số. Đó là một nền tảng vững để khẳng định vai trò của PPC. Nền tảng lý thuyết của PPC được khẳng định qua định lý giới hạn trung tâm, do Laplace phát biểu và chứng minh. Qua đó, PPC

được dùng để xấp xỉ và ước lượng các phân phối của tổng và trung bình của bất kì biến ngẫu nhiên của bất kì phân phối nào.

Vào năm 1890, ông đã phát triển định luật Laplace, sau này gọi là định lí giới hạn trung tâm, nhấn mạnh vai trò quan trọng về mặt lí luận của PPC. Theo định lí này, phân phối của tổng các biến ngẫu nhiên xấp xỉ về PPC khi số biến ngẫu nhiên lớn. Laplace đã mở rộng kết quả của Moivre cho trường hợp phi đối xứng, khi XS của biến cố không là 0.5. Hàm Gauss  $f(x)=e^{-x^2}$  đã được soi rọi vai trò, ý nghĩa trong bước tiến gắn với tên tuổi của Laplace. Thời kì Laplace nở rộ các phương pháp tính toán giải tích và khai sinh các hàm quan trọng trong XS-TK, một ví dụ là hàm sai số:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^x e^{-t^2} dt = 1 - \frac{\Gamma\left(\frac{1}{2}, x^2\right)}{\sqrt{\pi}}.$$

Các định lí giới hạn cho phép tính gần đúng XS độ lệch của trung bình các đại lượng ngẫu nhiên với giá trị kì vọng, đây chính là cơ sở quan trọng của phép kiểm định TK và ước lượng. Kết quả này phản ánh định hướng nhận thức của thời kì khám phá các định luật của giới tự nhiên, tìm tòi phát hiện tính quy luật của các hiện tượng ngẫu nhiên.

Năm 1809, Gauss công bố độc lập các kết quả tính toán trong tác phẩm lí thuyết về chuyển động của các thiên thể theo quỹ đạo conic. Trong đó, nhiều kết quả quan trọng như: phương pháp bình phương tối thiểu, phương pháp hợp lí cực đại và PPC. Gauss không phát minh ra phương pháp bình phương tối thiểu,

nhưng đã chỉ ra phương pháp tương đương với phương pháp bình phương tối thiểu. Ông chỉ ra rằng, sử dụng phân phối các sai số của hệ số tuyến tính theo đường cong hình chuông úp và cực đại hóa tính hợp lí thông qua cực đại hóa phân phối hậu nghiệm của sai số là cách làm tương đương với sử dụng phương pháp bình phương tối thiểu. Kết quả này phù hợp với nhận định của Laplace vào năm 1810 cho rằng hiệu chỉnh các sai số đột biến có thể dựa vào định lí giới hạn trung tâm. Hai ông đã tiến đến rất gần phương pháp hồi quy tuyến tính, và đã xây dựng nền tảng lí luận toán học chặt chẽ cho PPC. Theo kí hiệu của Gauss

$$\varphi\Delta = \frac{h}{\sqrt{\pi}} e^{-hh\Delta\Delta},$$

trong đó  $\Delta$  là độ lớn của

sai số,  $h$  là độ chính xác của quan sát,  $\varphi\Delta$  là luật XS của sai số phép đo với độ lớn  $\Delta$ . Ông đặt giả thuyết rằng giá trị kì vọng là trung bình số học của các giá trị đo được, rồi chứng minh luật PPC của sai số là luật phân phối duy nhất hợp lí cho sự chọn lựa giá trị trung bình như là một đánh giá xấp xỉ cho tham số vị trí. Sử dụng luật phân phối này như một hình mẫu phổ biến cho sai số thực nghiệm, ông đã xây dựng phương pháp bình phương tối thiểu phi tuyến gia trọng.

Gauss cũng đưa ra khái niệm PPC

$$\text{với } \mu=0, \sigma^2 = \frac{1}{2}, f(x) = \frac{e^{-x^2}}{\sqrt{\pi}}.$$

Tuy nhiên các định lí giới hạn chỉ nêu lên mặt định lượng của luật số lớn, theo đó trung bình kết quả quan sát một số lớn các thử nghiệm độc lập trên cùng một đại lượng ngẫu nhiên có xu hướng

hội tụ về giá trị kì vọng và mất dần tính ngẫu nhiên. Các định lí này quá đặc biệt nên không giải thích được nguyên nhân tính phổ biến của PPC, cũng không chỉ ra được điều kiện nào là cơ sở để xây dựng PPC. Những hạn chế này phải đợi hơn một thế kỉ sau mới được Liapunov giải quyết và Berstein mở rộng cho vector ngẫu nhiên. Vì vậy, định lí giới hạn trung tâm gắn với tên tuổi của Liapunov, theo đó PPC là phân phối của tổng các đại lượng ngẫu nhiên với điều kiện Lindeberg.

Năm 1808, Adrain dùng phương pháp bình phương tối thiểu để chỉnh lí số liệu đo lường, nhằm giải bài toán ứng dụng được nêu lên bởi Robert Patterson và được bình luận, với sự gợi ý của Nathaniel Bouditah (1773 - 1838). Thông qua việc khảo sát đó, luật PPC của sai số được thiết lập, tạo ra nền tảng cho phép chứng minh chặt chẽ phương pháp bình phương tối thiểu, qua đó PPC giúp khẳng định tính giá trị và tính tin cậy của phương pháp bình phương tối thiểu. Các công trình của Adrain về PPC rất được quan tâm và được Cleveland Abbe, một nhà khí tượng học Mỹ phát minh lại vào năm 1871. Dựa vào sử dụng PPC, Abbe rất thành công trong công tác dự báo khí tượng thông qua xử lí số liệu quan trắc. PPC vẫn chưa được nhìn nhận thống nhất như một luật hay một mô hình có giá trị quy luật của thế giới khách quan, mà vẫn đóng vai trò một công cụ. Sự hình thành khái niệm khó khăn và lâu dài cũng phản ánh tính đa dạng và phức hợp và tính cơ bản của luật PPC. Vấn đề cơ bản lúc này không còn là khám phá tính quy luật của

các đại lượng ngẫu nhiên nữa, mà là làm sao giải thích tính quy luật của các đại lượng ngẫu nhiên. Điều cơ bản là chỉ ra điều kiện cơ sở để xây dựng PPC về mặt lí thuyết và chỉ ra giới hạn áp dụng giả thiết PPC trong ứng dụng.

Các nhà khoa học tin tưởng vào tính quy luật của đại lượng ngẫu nhiên đã được khám phá, tìm cách mở rộng quy mô ứng dụng và suy rộng các tính chất của hệ cơ học sang hệ thống xã hội của con người. Do thiếu sót cơ sở xây dựng PPC dẫn đến sai lầm trong ứng dụng, các nghiên cứu này ban đầu cũng không tránh khỏi sơ xuất.

Năm 1835, nhà TK xã hội học đầu tiên là Adolphe Quetelet (1796-1874), được coi là cha đẻ của ngành khoa học xã hội định lượng, nêu lên khái niệm con người trung bình, nhằm nghiên cứu con người ở tầm vĩ mô, trong đó khái niệm PPC được sử dụng cho rất nhiều phân phối từ các số liệu xã hội học. Quetelet thành công ở tầm tiếp cận vĩ mô nhưng thất bại ở tầm vi mô khi không thể rút ra suy luận cho con người cá nhân cụ thể. Các chỉ số TK không thể giải thích đầy đủ tính chất được xét, cũng không nhất thiết tương ứng với thực thể nào đó tồn tại trong tự nhiên và xã hội, con người trung bình là một khái quát hóa sai lầm. Đặc biệt là thất bại của ông đưa đến thành công mới cho PPC, khi các nhà vật lí lí thuyết đã sửa chữa sai lầm bằng cách sử dụng điểm mạnh của các định luật Quetelet và giảm thiểu điểm yếu khi áp dụng mô hình ứng xử của xã hội tự do vào vật lí lí thuyết các khí lí tưởng và phát triển vật lí TK.

Trong nhân chủng học, nhà di truyền học và khí tượng học Francis Galton (1822-1911) đã phát minh phương pháp hồi quy tuyến tính và phương pháp hệ số tương quan, ông này cũng sai lầm trong phương pháp luận khi suy rộng kết quả vật lý học đem áp dụng vào nghiên cứu não tướng học. Tuy rằng suy luận TK dựa vào dữ liệu, XS dựa vào giả thiết, nhưng kết quả XS được chứng minh là đúng, còn suy luận TK có thể sai nên không có sự thật tuyệt đối mà chỉ đề cập đến mức ý nghĩa. Sự sai lầm do pha trộn lẫn giữa yếu tố thiếu sót về giả thiết XS và mức có ý nghĩa TK bắt nguồn từ sự thiếu sót cơ sở của PPC. Thiếu sót này là về khoa học luận chứ không về toán học, các kết quả toán học được sử dụng đều phải được chứng minh. Galton để lại nhiều đóng góp có ý nghĩa, mô hình quincunx vẫn được dùng trong dạy học XS-TK, đặc biệt là các thuật ngữ chuyên môn TK y sinh học mang đậm dấu ấn của ông.

Các tư tưởng cách mạng trong vật lý TK được đẩy lên qua đột phá của Boltzman và Maxwell. Năm 1860, Maxwell nêu lên luật phân phối Maxwell: “Khi tổng số hạt là  $N$  thì số các hạt chuyển động phân bố theo một hướng, nằm giữa  $x$  và  $x + dx$ , là  $N \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{\alpha^2}} dx$ ”

Qua đó khẳng định rằng PPC không chỉ là một công cụ toán học phổ biến mà còn là một luật chi phối các hiện tượng tự nhiên, đồng thời là mô hình cho nhiều lớp hiện tượng có thể được mô tả chính xác dựa vào phân phối ấy. Thông qua định luật số lớn, PPC là một mô hình xấp

xi đơn giản cho nhiều hiện tượng phức tạp, và là một giả thiết áp dụng cho những lớp mô hình ứng dụng khác có xu hướng tập trung quanh một giá trị trung bình và giảm nhanh đột ngột khi ra xa giá trị trung bình.

Vào những năm cuối thế kỉ XIX, khái niệm PPC đã hoàn chỉnh và tìm được ứng dụng rộng lớn. Phương pháp toán học có giá trị sâu sắc lại được thừa kế và phát triển mạnh, qua đó PPC tìm được vị trí mới và được khẳng định đúng đắn hơn, mỗi lần lại vững chắc và xuất sắc hơn trước. Sang thế kỉ XX, sau khi khẳng định vai trò quan trọng và nội dung phong phú, PPC được thống nhất tên gọi. Karl Pearson, người phát minh ra kiểm định chi bình phương cho phân phối theo bảng contingency, là người đóng lại một thời kì đầy phức tạp và mở ra một thời kì mới với trật tự. Trước Pearson, các nhà nghiên cứu ở nhiều lĩnh vực khác nhau cùng phát triển phương pháp và kết quả nghiên cứu khoa học trên nhiều hướng đề cập khác nhau tiến đến một cơ sở cho TK học, trong đó PPC trải qua rất nhiều đấu tranh và nhiều thử thách trong nhiều lĩnh vực.

Năm 1894, ông đề xuất khái niệm độ lệch chuẩn. Năm 1900, ông phát minh phương pháp chi bình phương để kiểm định tính phù hợp giữa các phân phối. Năm 1920 ông đề ra tên gọi PPC như ngày nay. Ông cũng là tác giả của các phân tích độ nhọn của phân phối và họ đường cong Pearson. Cùng với trường phái Pearson, Yule đã đem lại sự thống nhất cho TK phân tích khi tổng hợp được phương pháp tương quan và hồi quy với



phương pháp bình phương tối thiểu và lý thuyết sai số vào năm 1897. Như vậy, không có gì lạ khi vai trò trung tâm quan trọng hàng đầu của PPC được chứng minh và sử dụng. Yule cũng là người phát minh ra khái niệm tương quan bội và tương quan từng phần, trong đó PPC nhiều chiều là công cụ không thể thiếu. Tên gọi “phân phối chuẩn tắc” cho PPC có  $\mu = 0$  và  $\sigma^2 = 1$ , được Hoel chỉ định vào năm 1947.

Những năm 1920 là thời kỳ then chốt cho TK học dựa trên lý thuyết XS. Đây cũng là thời kỳ kinh tế tư bản chủ nghĩa phát triển, cùng với nó là sự thay đổi mạnh mẽ về tư tưởng. Trong giai đoạn này, tư tưởng kinh tế học mới do John Maynard Keynes đề xuất dựa trên xây dựng khái niệm XS dựa vào niềm tin và liên kết với hành vi của cá nhân thực thể kinh tế đã đem lại hình ảnh mới cho PPC, với vai trò là tiếng ồn hay tác động tổng hợp của các nhân tố trong hệ thống phức hợp tạo ra sự giao thoa tác động và tính bất định. Nhà TK vĩ đại nhất thế kỷ Ronald Fisher (1890-1962), đồng thời là nhà di truyền học, đã thay đổi lịch sử, đem lại bước ngoặt phát triển cho ngành phân tích TK. Phương pháp kiểm định ý nghĩa TK của Fisher kết hợp với kiểm định giả thuyết TK của Pearson là mô hình cơ sở của nghiên cứu lâm sàng.

Câu chuyện về PPC đã và đang tiếp diễn sẽ còn tiếp tục, khái niệm PPC ngày càng tìm thêm ứng dụng sâu sắc hơn. Trong những năm 50, lý thuyết của quá trình ngẫu nhiên phát triển mạnh, PPC được sử dụng lại và phát triển mở rộng dưới nhiều tên gọi khác nhau mô tả

những quá trình ngẫu nhiên: PPC nhiều chiều, quá trình Gauss, chuyển động Brown, v.v. Nhiều ứng dụng sâu vào các chuyên ngành hẹp của toán học như lý thuyết số được triển khai cũng dựa trên PPC. Sự thống nhất này là tất yếu của một quá trình mở rộng, phát triển mạnh mẽ khái niệm PPC.

## 2.2. Các giai đoạn nảy sinh và phát triển

Theo Thomas Kuhn, trong mỗi giai đoạn phát triển của khoa học, kiến thức được tổ chức theo khuôn mẫu khoa học, đó là bộ khung của những quan điểm được công nhận về đối tượng được nghiên cứu. Tiến trình phát triển của khoa học có những giai đoạn đan xen của sự phát triển tích lũy dần dần và nhảy vọt cách mạng, tại những điểm nhảy vọt là những công trình khám phá có tính cách mạng thay đổi khuôn mẫu cũ tạo nên bước ngoặt phát triển, tại đó bắt đầu những bước đột phá tư tưởng.

Trong trường hợp riêng của lịch sử hình thành phát triển khái niệm PPC, chúng tôi nhận thấy có những giai đoạn nhảy vọt mang tính cách mạng. Sự nhảy vọt thứ nhất diễn ra với Gauss và Laplace, tạo ra tiền đề cơ sở toán học cho khái niệm, đây chính là sự khai sinh của khái niệm. Sự khai sinh khái niệm về mặt toán học này gắn liền với bước nhảy vọt về đối tượng phương pháp từ lý thuyết XS sang lý thuyết TK. Sự nhảy vọt thứ hai diễn ra với Pearson, khái niệm được chuẩn hóa với tên gọi hiện đại, kiến thức được tổ chức theo chuẩn mực của một khoa học TK độc lập với các ngành khác, đây là sự khai sinh lần thứ hai của khái

niệm, song hành với sự khai sinh ngành khoa học nó được vinh dự phục vụ, phân tích TK có cơ sở vững vàng, tách khỏi TK mô tả. Vì vậy ở đây sự phân chia các giai đoạn hình thành và phát triển khái niệm phụ thuộc vào quan điểm của nhà nghiên cứu, khi nhìn PPC như khái niệm thuần túy toán học được phát triển trong tiến hóa các ngành toán học, hay là một khái niệm TK học được gieo mầm và lớn lên trong liên ngành phức hợp toán học, cơ học, xã hội học, v.v để khai sinh như khái niệm trung tâm của XS-TK học.

### 2.2.1. Giai đoạn những năm 1730 đến 1770

Trong giai đoạn này, PPC mới được phát hiện và mang tính công cụ. TK ứng dụng giai đoạn này cũng thô sơ, chủ yếu là các vấn đề dân số và nhân khẩu. Vì vậy PPC được đề cập đến trong các kỹ thuật tính toán, chưa có dáng vẻ một vấn đề toán học, không được coi là đối tượng để phát triển nghiên cứu, cũng không được đặt tên. PPC xuất hiện dưới hình thức công cụ tính toán đại số, chưa thật sự là một đối tượng toán học được hình thành bằng phương pháp giải tích toán học. Về mặt khái niệm, PPC chỉ được xem như một luật hàm sinh xấp xỉ liên tục các hệ số nhị thức Newton rời rạc, do đó phụ thuộc rất nhiều vào ý thích của các nhà nghiên cứu, chưa được xem như một công cụ toán học trợ giúp cho các nghiên cứu ứng dụng và các tính toán khoa học. Vì chưa phải là một đối tượng toán học, PPC chưa được nghiên cứu sâu về toán học, thêm nữa mối liên quan của PPC với định lý giới hạn trung tâm và phương pháp bình phương tối thiểu chưa

được xác định rõ ràng. Phương pháp bình phương tối thiểu cũng mới ra đời, chưa được khẳng định là phương pháp chủ chốt của tính toán khoa học và các ngành khoa học ứng dụng. Định lý giới hạn trung tâm cũng được chứng minh dựa vào ý tưởng khai thác phương pháp này và mở rộng cho các bài toán cơ học.

### 2.2.2. Giai đoạn những năm 1780 đến 1860

Khoa học thực nghiệm phát triển mạnh, các nhu cầu đo lường và xử lý số liệu cũng tăng lên. Giải tích toán học là môi trường tồn tại và phương pháp khai thác sức mạnh ứng dụng của PPC. Bên cạnh việc phục vụ tính toán xấp xỉ, PPC có vai trò kép, vừa là một phân phối của kết quả quan trắc vừa là một phân phối dùng làm chuẩn để so sánh kết quả. Trong giai đoạn này, PPC đã qua tích lũy mà được sử dụng rộng rãi, có nhiều đóng góp quan trọng trong nhiều lĩnh vực tính toán khoa học như lý thuyết đo lường, vật lý thiên thể, khí động học. Về mặt lý thuyết, PPC được sử dụng trong phép chứng minh chặt chẽ và là một công cụ tính toán trợ giúp cho phương pháp bình phương tối thiểu và lý thuyết đo lường hiệu chỉnh sai số. Vị trí trong toán học như một phân phối XS liên tục cơ bản nhất được chứng minh nhờ định lý giới hạn trung tâm, theo đó trong những điều kiện nhất định, tổng các biến ngẫu nhiên có trung bình và độ lệch chuẩn hữu hạn tiến đến tiệm cận PPC khi số biến tăng.

Phương pháp bình phương tối thiểu và lý thuyết sai số đạt được vị trí quan trọng, được sử dụng rộng rãi trong ứng dụng, được chứng minh chặt chẽ về mặt

toán học. Đây chính là thời cơ để PPC được xem xét như một đối tượng toán học. Đây cũng là thời kì nảy sinh nhiều vấn đề xung quanh PPC. Lí thuyết tương quan và hồi quy là một mảng phát triển rất mạnh và giàu tính ứng dụng thực tiễn cũng như ý nghĩa phương pháp luận khoa học, được phát triển đã khẳng định vai trò trung tâm của PPC. Trong giai đoạn này PPC được biết đến và khai thác dưới nhiều khía cạnh, đặt tên theo khía cạnh được khai thác. Các nhà nghiên cứu không thống nhất ý kiến về khái niệm và biểu diễn công thức của PPC. PPC xuất hiện dưới nhiều cái tên khác nhau như luật sai số, luật Gauss, luật Laplace, tùy theo mối quan tâm và lĩnh vực của các chuyên gia.

### 2.2.3. Giai đoạn cuối thế kỉ 18 đến 1920

Khái niệm PPC đã được phát triển trong khung lí thuyết toán học nhất quán và chặt chẽ, vai trò của PPC được chứng minh qua thực tế ứng dụng. PPC trở thành công cụ quan trọng được ứng dụng hết sức đa dạng trong nhiều lĩnh vực khoa học, không những trong nội bộ toán học mà cả các ngành khoa học tự nhiên, xã hội và con người. Bước phát triển mới liên kết, thống nhất phương pháp hồi quy tương quan với phương pháp bình phương tối thiểu và lí thuyết sai số đem lại ý nghĩa và phương pháp phân tích TK hiện đại, qua đó khẳng định vai trò của PPC. TK phân tích và TK tham số đã khẳng định được vị trí, tạo phân ngành đa dạng trong nội bộ TK học, tạo tiền đề cho TK phi tham số. Việc không sử dụng giả thiết PPC không phải là sự phủ định vai trò của PPC, mà chỉ nêu lên giới hạn của

phổ khái niệm. Điều này hợp quy luật phát triển.

Ngành khoa học TK mới ra đời, ngành toán học cũng được hưởng lợi ích và đang tích lũy để nhảy vọt với cuộc cách mạng do Kolmogorov vào những năm 1930.

Tên gọi PPC cũng được quy định thống nhất. Từ đây những dạng mở rộng của PPC được nghiên cứu và ứng dụng rộng rãi, PPC nhiều chiều liên quan đến tương quan đa bội và hồi quy đa bội.

## 2.3. Phạm vi tác động, bài toán và đối tượng liên quan

### 2.3.1. Phạm vi tác động

PPC có phạm vi tác động rất lớn và trở nên phổ biến trong khoa học tự nhiên, kĩ thuật và khoa học xã hội nhân văn từ lí thuyết đến ứng dụng.

Luật PPC tỏ ra phù hợp chính xác với nhiều định luật của tự nhiên. Hầu hết các hiện tượng tự nhiên được mô tả bằng luật PPC hay được chuyển biến mô tả để tuân theo luật PPC. Luật PPC được xấp xỉ cho nhiều lớp hiện tượng, nhiều áp dụng định lí giới hạn trung tâm và đánh giá sai số theo giải tích toán trong y sinh học, kĩ thuật. Luật PPC là nền tảng của các phân tích TK, là mô hình cho nhiều phân phối XS khác. Các bài toán kiểm định TK tham số không thể giải quyết được nếu thiếu giả thiết về PPC của quần thể khảo sát.

### 2.3.2. Các bài toán liên quan

Các bài toán liên quan đến PPC được phân chia theo phạm vi tác động.

- Sự hình thành PPC: phân phối nhị thức, luật số lớn, luật giới hạn phân phối,

xử lý sai số đo đạc và bình phương tối thiểu.

- Sự mở rộng PPC: PPC đa chiều với các vector ngẫu nhiên

- Ứng dụng của PPC:

Trong XS-TK, các bài toán liên quan bao gồm: kiểm định giả thuyết TK, ước lượng tham số và thiết kế nghiên cứu.

Trong mô hình toán học: Tiếng ồn trắng, mô hình dân số dưới tác động ngẫu nhiên của môi trường, mô hình dòng điện dưới tác động ngẫu nhiên trong mạng điện, mô hình Black - Sholes

Trong toán học thuần túy: Lí thuyết số và tổ hợp, luật phân phối số nguyên tố.

Trong vật lí: Mô hình và lí thuyết khí lí tưởng, mô hình và lí thuyết chuyển động Brown, lí thuyết đo lường và sai số, v.v.

Trong khoa học xã hội: Thiết kế nghiên cứu, xử lý và phân tích số liệu, ước lượng tham số, kiểm định TK, v.v.

### 2.3.3. Các đối tượng có liên quan

Sự hình thành và phát triển của khái niệm PPC có liên quan chặt chẽ đến các công cụ và phương pháp giải tích toán học, khung khái niệm và lí thuyết của XS và TK. Các đối tượng toán học được đề cập đến chủ yếu và được phân chia theo hai loại TT, nhằm tổ chức khung TT cho việc dạy và học.

- Hàm số: Hàm của tập (biến ngẫu nhiên), hàm của hàm (hàm của biến ngẫu nhiên), hàm đặc biệt (hàm mật độ XS, hàm tích lũy XS, hàm moment, hàm Gauss), hàm định nghĩa bằng tích phân (tích phân Gamma, PPC, phân phối

Student, phân phối Chi bình phương, phân phối Fisher).

- Không gian và topo: Không gian mẫu, không gian XS, không gian tham số  
- Phép tính vi tích phân: Độ đo XS, phép tính tích phân định hạn, phép tính tích phân mở rộng, tích phân Laplace, tích phân entropy, tin lượng Fisher.

## 2.4. Các cách tiếp cận khái niệm PPC

### 2.4.1. Theo giới hạn

- Chuyển từ rời rạc sang liên tục, PPC là giới hạn của phân phối nhị thức khi mẫu lớn:  $B(n, p) \sim N(np, np(1-p))$  với  $n$  lớn và  $p \approx 0,5$ .

- Chuyển từ tham số của phân phối XS liên tục, PPC xem như xấp xỉ của phân phối:

+ Phân phối Poisson:  $P(\lambda) \sim N(\lambda, \lambda)$  với  $\lambda = n \times p$ ;

+ Phân phối Chi bình phương:  $\chi^2(k) \sim N(k, 2k)$  với  $k$  lớn;

+ Phân phối Student:  $t(n) \sim N(0; 1)$  khi  $n$  lớn.

### 2.4.2. Dùng đồ thị, biểu đồ

- Chuyển từ công cụ tính toán, phần mềm tính toán: Dùng máy tính đếm số liệu và biểu diễn trực quan trên đồ thị, biểu đồ theo TK mô tả.

- Xấp xỉ hình dạng đồ thị của phân phối, đặt giả thuyết phân phối đã cho là xấp xỉ PPC.

### 2.4.3. Theo giải tích toán

- Hàm Gauss, Q, erf,...

- Tích phân Laplace, hệ số chuẩn hóa  $\sqrt{\pi}$

- Định nghĩa hàm số của PPC, PPCT.

- Định nghĩa XS như diện tích dưới đường cong phân phối XS.

### 3. Kết luận

Phân tích khoa học luận là một phần quan trọng trong các công trình nghiên cứu của didactic toán. Nó mang lại các yếu tố dường như không thể thiếu cho việc tìm hiểu quá trình hình thành và phát triển của TT trên nhiều phương diện khác nhau.

Thông qua phân tích các kết quả nghiên cứu lịch sử toán học, phân tích quá trình hình thành và phát triển của

khái niệm PPC và PPCT, chúng tôi rút ra những đặc trưng logic về quá trình phát triển cũng như chỉ ra một số đặc trưng khoa học luận cơ bản trong từng thời kì, từ đó xây dựng nên bức tranh đại cương thống nhất về lịch sử logic của khái niệm.

Phân tích cũng chỉ ra những cách tiếp cận khác nhau đến PPC và PPCT, những cách tiếp cận này có những đặc trưng có thể dùng trong thiết kế dạy học khái niệm PPC ở ĐHYD TP HCM.

### TÀI LIỆU THAM KHẢO

1. Annie Bessot, Claude Comiti, Lê Thị Hoài Châu, Lê Văn Tiến (2009), *Những yếu tố cơ bản của Didactic toán*, Nxb Đại học Quốc gia TP HCM.
2. Lê Thị Hoài Châu (2010), “Những chướng ngại, khó khăn trong dạy học khái niệm xác suất”, *Tạp chí Khoa học Đại học Sư phạm TP HCM*, (24).
3. Nguyễn Phú Lộc (2007), *Xu hướng dạy học không truyền thống*, Trường Đại học Cần Thơ.
4. Đào Hồng Nam (2010), “Mối quan hệ thể chế với PPC trong việc dạy và học Xác suất – Thống kê ở trường Đại học Y Dược TP HCM”, *Tạp chí Khoa học Đại học Sư phạm TP HCM*, (24).
5. Abraham de Moivre (1756), *The Doctrine of Chances: or a method of calculating the probability of events in play*, London.
6. Dorier J.-L. (1996), *Recherche en historique et en didactique des mathématiques sur l’algèbre linéaire*.
7. Dutka J. (1990), *Robert Adrain and the method of least squares*, *Archive or History of. Exact Sciences*, vol. 41, pp.171-184.
8. Gauss C. F. (1857), *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium*, Translation by Charles Henry Davis, Boston.
9. Maxwell J.C. (1860), *Illustrations of the Dynamical Theory of Gases*, *Philosophical Magazine*, pp.19-32.
10. Stigler S.M. (1977), *An attack on Gauss*, *Legendre*, *Historia Math.*4, pp.31-35.
11. Stigler S.M. (1978), *Mathematical statistics in the early States*, *Annals of Statistics*, pp.239-265.

(Ngày Tòa soạn nhận được bài: 01-11-2011; ngày chấp nhận đăng: 16-12-2011)