

Quá khứ, hiện tại và tương lai của Thống kê toán học

C. R. Rao*

Đôi nét về giáo sư Calyampudi Radhakrishna Rao

C.R.Rao là nhà thống kê nổi tiếng trên thế giới, những công trình nghiên cứu của ông đã đặt nền tảng cho lý thuyết thống kê hiện đại. Rao đã được trao Huân chương khoa học của Liên bang Mỹ vào tháng 6/2002, đó là giải thưởng khoa học cao quý nhất của Mỹ trao cho những người có đóng góp lớn, lâu dài trong nghiên cứu khoa học. Ông cũng được tặng Huân chương Mahalanobis và Huân chương Wilks, là cựu chủ tịch của IMS, ISI và Hội Toán Kinh tế, đồng thời là viện sĩ Viện Hàn lâm quốc gia Mỹ, Anh, Ấn Độ và Lithuania. C.R.Rao còn là viện sĩ danh dự của ISI, Hội thống kê hoàng gia Anh, Hội Toán Kinh tế, Viện lý thuyết tổ hợp và ứng dụng thuộc đại học Cambridge, và là viện sĩ bầu chọn của IMS, ASA, AAAS và Viện Hàn lâm Khoa học và Nghệ thuật Mỹ.

Sau đây là tóm tắt bài giảng của C.R. Rao.

Thống kê học là gì?

Thống kê học có thể được định nghĩa một cách khái quát như là **khoa học, kỹ thuật hay nghệ thuật của việc rút ra thông tin từ dữ liệu quan sát**, nhằm giải quyết các bài toán từ thực tế cuộc sống. Việc rút ra thông tin đó có thể là kiểm định một giả thiết khoa học, ước lượng một đại lượng chưa biết hay dự đoán một sự kiện trong tương lai. Bởi vì một tập dữ liệu cụ thể có thể chứa những lượng bất định (uncertainty) nào đó, nên kết luận rút ra dựa trên tập dữ liệu đó có thể là sai. Nhưng nếu cần phải đưa ra một quyết định dựa trên dữ liệu có độ bất định

thì đâu là chiến lược tốt nhất? Chỉ đến đầu thế kỷ 20 người ta mới nhận ra được rằng vấn đề then chốt trong các bài toán này nằm ở chỗ định lượng lượng bất định đó. Một chuyên ngành mới của thống kê được mở ra trong những năm đầu thế kỷ 20 là nghiên cứu việc đưa ra quyết định tối ưu dựa trên độ bất định.

Phương pháp Khi - bình phương của Karl Pearson

Chúng ta định lượng lượng bất định đó như thế nào? Có nhiều tranh luận với những trường phái khác nhau về việc đưa ra cách biểu thị lượng bất định đó. Năm 1900, nỗ lực đầu tiên nhằm giải quyết vấn đề này thuộc về Karl Pearson, ông đã đưa ra khái niệm sai số tiêu chuẩn của ước lượng và **thống kê hợp lý Khi-bình phương cho bài toán kiểm định giả thiết. Đóng góp này của ông được đánh giá là một trong hai mươi khám phá quan trọng nhất của thế kỷ 20.**

Thời đại Fisher

Khoảng 20 năm sau, trong những năm 20 của thế kỷ 20, nền tảng toán học cho thống kê đã được **R.A. Fisher** xây dựng bằng cách đồng nhất các bài toán đặc tả (mô hình ngẫu nhiên cho dữ liệu được quan sát), ước lượng (xác định các tham số chưa biết) và phân phối (của những thống kê được tính toán từ dữ liệu) với 3 bài toán mang tính phương pháp luận của thống kê học. Các bài toán này là cơ sở cho những nghiên cứu thống kê lý thuyết trong suốt hầu hết thế kỷ 20 (sự bàn luận

* Giáo sư Calyampudi Radhakrishna Rao

hiện nay vẫn được tiếp tục trên các tạp chí chuyên ngành thống kê).

Fisher đã có nhiều đóng góp cho ngành thống kê học, những đóng góp đáng kể của ông là đưa ra khái niệm hợp lý, ước lượng hợp lý cực đại, thống kê đủ và đo lường thông tin trong mẫu. Nhưng ông chỉ làm việc với những mô hình đơn giản như là phân phối chuẩn, điều này làm hạn chế khả năng ứng dụng trong các bài toán của cuộc sống thực, mà theo như John Tukey thì đó là một “lời nguyền của thống kê”. Trong kiểm định giả thiết, Fisher nhấn mạnh quá nhiều vào **giả thiết không**, mà tiên nghiệm đó là không đúng trong hầu hết các tình huống, khi sử dụng những mức ý nghĩa thông thường như là 5% và 1% trong việc bác bỏ giả thiết. Bình luận về những công trình của Fisher và những lý thuyết mà sau này được Neyman và Pearson xây dựng, Tukey, Yates và Wolfowitz nhận xét rằng, trong thực hành bài toán **kiểm định giả thiết không**, chúng ta đang hỏi một câu hỏi sai và thu được câu trả lời lẩn thẩn.

Nhu cầu và vấn đề đào tạo các nhà thống kê

Sau chiến tranh thế giới thứ 2, có rất nhiều yêu cầu được đặt ra cho các nhà thống kê trong các lĩnh vực như quản lý, công nghiệp, thực nghiệm nông nghiệp, các công ty dược phẩm, tham vấn pháp luật Thống kê được xem như là một cách thức chính để thu nhận thông tin. Các trường đại học đã bắt đầu mở các khoa thống kê để đào tạo các nhà thống kê và cổ vũ, khuyến khích nghiên cứu thống kê. Nhưng cũng có nhiều những lời phê bình cho rằng nhiều nghiên cứu thống kê không bắt nguồn từ các bài toán thực tế, và rằng các sinh viên thống kê được đào tạo ra không quen với việc phân tích dữ liệu thực tế để đưa ra những câu trả lời làm thỏa mãn khách hàng. Chúng ta nghe thấy những lời phát biểu kiểu như: “Thống kê đã không còn gắn với các lĩnh vực khoa học mà nó đã từng được nuôi dưỡng (như toán sinh, vận trù học), điều này làm

cho thống kê trở nên bó hẹp hơn”; “Đây là kỷ nguyên vàng cho thống kê, mặc dù có thể không cho các nhà thống kê” (Mosteller); “Không có khám phá quan trọng nào bị bỏ qua do thiếu kiến thức thống kê” (F N David). Những lĩnh vực ứng dụng quan trọng của thống kê như nhận dạng mô hình, xử lý tín hiệu, khai thác dữ liệu do các nhà thống kê đầu tiên đưa ra đã được các nhà khoa học máy tính và các kỹ sư đón nhận và nghiên cứu phát triển sâu hơn.

Tương lai của thống kê

Thống kê thường bị phê phán rằng các phương pháp của nó phải được đặt trên các mô hình có sẵn. Đã có sự thay đổi trong khoảng 30 năm gần đây: những phương pháp mới mang tính thuật toán đã được sử dụng như là bootstrap, phân lớp, cây hồi quy và mạng thần kinh không dùng các mô hình hiển, nhưng tính hiệu quả của các phương pháp này gần tốt như các phương pháp dựa trên những mô hình quen thuộc.

Thống kê phát triển nhanh chóng khi nó được đánh giá như là một công cụ hữu ích trong tất cả các nghiên cứu đòi hỏi làm thực nghiệm, tạo dữ liệu, thu thập thông tin và rút ra kết luận. Không giống như các ngành khoa học khác, thống kê không phát triển từ thống kê. Nó cần sự thúc đẩy từ những bài toán mới phát sinh trong tất cả các lĩnh vực hoạt động tích cực của loài người. Tương lai của thống kê nằm ở sự giao tiếp trao đổi hợp tác giữa các nhà thống kê với các cán bộ nghiên cứu trong các nhánh học hỏi khác. Vai trò của một nhà thống kê không nhất thiết phải là vai trò của một nhà kỹ thuật biết áp dụng các kỹ thuật thông thường để trả lời nhưng câu hỏi đặc biệt. Được đào tạo chuyên sâu, vai trò này cần phải năng động khi trợ giúp nhà chuyên môn phát biểu lý thuyết hoặc giả thuyết nhằm đưa đến những tiến bộ trong nhận thức.

(Xem tiếp trang 32)

Quá khứ, hiện tại và tương lai...

(Tiếp theo trang 31)

Trước đây, các phương pháp thống kê được xây dựng trong phạm vi mẫu nhỏ. Ngày nay, cùng với sự phát triển của công nghệ ghi chép tự động, và nguồn thực nghiệm ngày càng tăng, nên có nhiều các tập dữ liệu cỡ lớn. Điều này tạo ra các bài toán mới về quản lý, lưu trữ và xử lý dữ liệu. Với sự thay đổi nhanh chóng của công nghệ, những bài toán mới xuất hiện đòi hỏi các cách thức mới trong công tác thực nghiệm và đo lường, như là dữ liệu vi mảng trong nghiên cứu gen, xử lý ảnh kỹ thuật số để nhận diện, nhận dạng mã nén và hệ thống cảnh báo sớm của các hoạt động khủng bố. Thêm vào đó, chúng ta cũng có một lượng lớn dữ liệu kiểu giao dịch như

là các hồ sơ trong cửa hàng tạp phẩm, giao dịch ngân hàng, v.v... mà từ đó những thông tin hữu ích có thể được rút ra, nhờ những phương pháp có tên là khai thác dữ liệu (data mining). Các nhà thống kê đã và đang có những khả năng tạo ra các ý tưởng phương pháp mới và mở rộng phạm vi của thống kê khi giải quyết những bài toán mới./.

ST: PĐQ

Nguồn:

*Bài giảng của Viện Thống kê Thế giới IMS tại Malaysia (theo IMS Bulletin, Vol. 35, 2, 2006, pp.4-5)
<http://diendantoanhoc.net>*