

## BÁO CÁO VỀ TƯƠNG LAI CỦA THỐNG KÊ HỌC

*Bruce G. Lindsay, Jon Kettenring và David O. Siegmund*

**T**rong thời gian qua, Thống kê học đã không ngừng phát triển và đạt được những thành tựu nhất định, đặc biệt là công cụ hỗ trợ đắc lực cho sự phát triển của các ngành Khoa học khác. Tuy nhiên, bên cạnh đó, bản thân Khoa học Thống kê vẫn còn gặp phải những vấn đề và thách thức cần giải quyết. Nhận thấy được tầm quan trọng của điều này, tháng 5-2002, Quỹ Khoa học Quốc Gia (NSF) đã tổ chức một cuộc hội thảo, qua đó đưa ra một báo cáo toàn diện về những cơ hội, thách thức cũng như tương lai phát triển của Khoa học Thống kê. Đây là một tài liệu hết sức hữu ích và thiết thực cho những người làm công tác Thống kê nhằm trang bị thêm những hiểu biết về hoạt động của Thống kê học trên thế giới, đồng thời xác định mục tiêu và nhiệm vụ để nắm bắt được cơ hội và vượt qua những khó khăn, thách thức hiện tại của Thống kê học nói chung cũng như Thống kê Việt Nam nói riêng. Dưới đây là bài viết được các tác giả Bruce G. Lindsay, Jon Kettenring và David O. Siegmund đúc rút từ báo cáo chính thức của Hội thảo. Thông tin Khoa học thống kê trân trọng giới thiệu cùng độc giả bài dịch từ bài viết nói trên.

**Tóm lược:** Tháng 5-2002, một hội thảo đã được tổ chức tại Quỹ Khoa học Quốc Gia nhằm thảo luận về những cơ hội và thách thức trong tương lai đối với cộng đồng thống kê. Sau hội thảo, Ủy ban khoa học đã đưa ra một báo cáo mở rộng, trong đó thể hiện sự đồng thuận của cộng đồng Thống kê. Bài viết này được trích từ báo cáo đầy đủ của hội thảo.

**Từ khóa:** Research funding, National Science Foundation, challenges, opportunities, statistical education.

### LỜI MỞ ĐẦU

Rõ ràng là hiện nay Thống kê học đang có rất nhiều cơ hội. Lấy ví dụ, ba lĩnh vực hoạt động chính của Khu vực Ưu tiên các ngành Khoa học Toán học (MSPA) thuộc NSF gồm có: xử lý số liệu thô, mô hình hóa các hệ thống phức tạp và xử lý lượng bất định. Đó cũng chính là ba mối quan tâm hàng đầu của Thống kê học. Từ trước tới nay, chưa bao giờ các kiến thức thống kê lại trở nên quan trọng và thực sự hữu ích cho những đột phá về khoa học đến vậy.

Ngày nay, rất nhiều số liệu được thu thập trong các lĩnh vực khoa học khác nhau. Tuy nhiên, nếu chúng ta không có các kế hoạch thu thập số liệu phù hợp thì rất khó tránh khỏi nhiều số liệu sẽ trở nên vô ích. Nếu không có các phương pháp khoa

học và các công cụ hữu hiệu nhằm thu thập, khai thác và phân tích các bộ số liệu, bất kể là quy mô của các bộ số liệu thế nào chăng nữa thì chúng ta đều không thể nắm bắt được các quá trình sản xuất số liệu hết sức phức tạp và khó hiểu.

Và để nắm được cơ hội lớn đó, cộng đồng Thống kê sẽ phải giải quyết rất nhiều thách thức phát sinh. Trong đó, ngoài một số những thách thức về mặt trí tuệ thì còn có những thách thức về cơ sở hạ tầng phát sinh từ những xu thế biến đổi của các tác nhân bên ngoài. Tài liệu này ghi nhận nỗ lực của cộng đồng Thống kê trong việc xác định và giải quyết những thách thức và tác động đó dựa trên một hội thảo bàn về tương lai của Thống kê học được tổ chức tại NSF vào tháng 5-2002.

### 1.1 Hội thảo về tương lai của Thống kê học

Hội thảo được tổ chức theo đề nghị của NSF và được tổ chức bởi một ủy ban khoa học gồm có 9 thành viên. Ủy ban này cũng đã đưa ra một Báo cáo Tổng kết với định hướng và trợ giúp của các đại biểu cũng như nhiều chuyên gia khác. Có thể xem báo cáo đầy đủ trực tuyến tại địa chỉ website: [http://www.stat.psu.edu/~bgl/nsf\\_report.pdf](http://www.stat.psu.edu/~bgl/nsf_report.pdf).

Bài viết này chỉ là bản tóm lược đã được cập nhật của Báo cáo.

Khoảng 50 đại biểu tham dự Hội thảo được chọn làm đại diện cho toàn ngành Thống kê. Trong số đó, các đại biểu không thuộc Hoa Kỳ chiếm số lượng đáng kể, tuy nhiên Hội thảo cũng như Báo cáo đầy đủ lại chủ yếu tập trung vào Thống kê trong lãnh thổ Hoa Kỳ.

Ủy ban khoa học cũng đã quyết định: để có tác động tối đa, bản Báo cáo nên hướng vào tất cả các đại biểu. Chính vì thế, mục tiêu của chúng ta không chỉ là nhắm tới các nhà nghiên cứu Thống kê mà còn là những người đóng vai trò hỗ trợ hết sức quan trọng như các cộng tác viên, giám đốc bộ phận, trưởng khoa các trường Đại học hay các cơ quan cấp vốn.

Hội thảo được xây dựng nhằm tập trung vào các khía cạnh của lĩnh vực Thống kê và đặc biệt liên quan tới NSF, theo đó nội dung Hội thảo không bao gồm Thống kê Sinh học. Thống kê Sinh học là một phân ngành lớn và rất phát triển của ngành Thống kê nói chung, trong đó có nhiều bộ phận liên kết với các trường Y trên khắp Hoa Kỳ. Do đó, dù không được tập trung xem xét trong Hội thảo, chúng ta vẫn nên chỉ rõ: nghiên cứu về Thống kê Sinh học là một phần chủ đạo để tìm hiểu về ngành Thống kê.

Cũng nên lưu ý rằng Báo cáo 98-95 của NSF (1998) được nhiều người biết đến với tên gọi Báo cáo Odom (và được trích dẫn khá nhiều trong bài

viết này) là một tài liệu quan trọng cho Báo cáo của chúng tôi vì Báo cáo này đóng vai trò không nhỏ trong hoạch định chính sách của NSF. Báo cáo Odom được viết nhằm đưa ra một đánh giá tổng quan về nhu cầu của các ngành Khoa học Toán học, trong đó có ngành Thống kê. Là một tài liệu hay và quan trọng, đồng thời hỗ trợ cho nhiều điểm chính trong Báo cáo của chúng tôi nên Báo cáo Odom được trích dẫn khá nhiều trong bản báo cáo đầy đủ. Ví dụ như, bản Báo cáo xác định ba hoạt động chính của các nhà Toán học như sau:

- (1) Xây dựng các khái niệm trong toán học cơ bản;
- (2) Tương tác với các lĩnh vực sử dụng toán học, chẳng hạn như khoa học, kỹ thuật, công nghệ, tài chính và an ninh quốc gia;
- (3) Thu hút và phát triển thế hệ các nhà toán học tiếp theo.

Sau khi thay Toán học bằng Thống kê, việc chia ra ba hoạt động chính như trên cũng có thể dùng để mô tả các hoạt động chủ yếu của các nhà Thống kê. Tuy nhiên, có một điểm khác biệt cơ bản giữa Toán học và Thống kê chính là ở sự cân bằng giữa mục (1) và (2), điều này sẽ được bàn tới ở phần sau.

### 1.2. Thống kê học là gì

Một động lực quan trọng để tổ chức Hội thảo và đưa ra báo cáo kết quả chính là do nhận thức được rằng: vai trò của ngành Thống kê thường chỉ được các ngành khoa học khác biết đến một cách hết sức khiêm tốn. Phần lớn người ta bắt đầu quan tâm về ngành Thống kê từ khi có việc sử dụng và phát triển các công cụ toán học và máy tính phức tạp vượt ra khỏi tầm hiểu biết của mọi người, trừ một số nhà khoa học. Vì vậy, cải thiện tình hình này chính là một trong những mục tiêu mà Hội thảo hướng đến.

Đáp ứng yêu cầu đó, nhà Thông kê học nổi tiếng đang làm việc tại trường Đại học Oxford, Tiến sĩ D. R. Cox đã bắt đầu Hội thảo với những điều cơ bản và trả lời cho câu hỏi “Thông kê học là gì?”. Câu hỏi này đã được trả lời nhiều lần trong suốt quá trình Hội thảo để phục vụ cho các đại biểu. Ta có thể tóm lược một số các điểm chính sau đây.

Thông kê học là một ngành liên quan đến việc nghiên cứu tính biến thiên, lượng bất định cũng như đưa quyết định khi có lượng bất định. Đây là những vấn đề rất quan trọng trong các ngành khoa học và kỹ thuật nên Thông kê học vốn là một khoa học liên ngành. Mặc dù ngành Thông kê không có những chủ thể khoa học tồn tại dưới dạng vật chất cụ thể (như đá, mây, sao hay DNA) nhưng nó vẫn thống nhất với hệ thống kiến thức chung và có sự kế thừa tri thức.

Một đặc điểm khác biệt của ngành Thông kê và hệ phương pháp mà ngành Thông kê phát triển chính là tập trung vào các nguyên tắc thận trọng để rút ra những kết luận khoa học từ số liệu. Cách tiếp cận theo nguyên tắc như vậy đã tách biệt số liệu thống kê ra khỏi phạm vi thao tác, tổ chức và phân tích số liệu. Trong đó có một nguyên tắc chung đó là ta nên định lượng lượng bất định để rút ra các kết luận khoa học dựa trên số liệu. Các công cụ thống kê như: hệ số tin cậy, mức ý nghĩa hay khoảng tin cậy được tạo ra để có những định lượng giá trị dễ hiểu. Các công cụ này sẽ giúp hạn chế việc đưa ra các kết luận sai lầm dựa trên số liệu nếu chúng được sử dụng phù hợp.

Tất nhiên, các nhà Thông kê không có nhiều công cụ như trong ngành Toán học. Rõ ràng là phần lớn các ứng dụng thống kê cũng như nhiều nghiên cứu thống kê do các nhà khoa học trong các ngành khác thực hiện. Vai trò chủ yếu của các nghiên cứu thống kê là nhằm phát triển các công cụ mới trong các ngành khoa học. Trong những phần sau của bản

báo cáo chúng ta sẽ chứng minh những khả năng nghiên cứu Thông kê cực kỳ thú vị có được trong những năm gần đây. Đặc biệt, khả năng thu thập và lưu trữ số liệu đã mở ra nhu cầu về những phương pháp tiếp cận hoàn toàn mới đối với các vấn đề về phân tích số liệu.

### 1.3. Phạm vi nghiên cứu của ngành Thông kê

Phạm vi nghiên cứu của công tác Thông kê gần như trải rộng trên tất cả hoạt động khoa học. Trong Hội thảo chúng ta đã tập trung vào hai phạm vi chủ yếu. Trước hết, phần trọng tâm ta gọi là nòng cốt. Người đưa ra luận điểm này chính là Tiến sĩ Iain Johnstone của trường Đại học Stanford.

Tiếp theo đó là 5 phạm vi ứng dụng bao gồm: (1) Sinh học (Warren Ewens, trường Đại học Pennsylvania), (2) Thông kê Kỹ thuật và Công nghiệp (Vijay Nair, trường Đại học Michigan), (3) Địa chất và Môi trường (Richard Smith, trường Đại học North Carolina), (4) Công nghệ Thông tin (Werner Steutzle, Trường Đại học Washington), (5) Xã hội và Kinh tế (Joel Horowitz, trường Đại học Northwestern). Các hạng mục này được lựa chọn để phù hợp tương đối với các bộ phận khác nhau của NSF, tại đó có hỗ trợ các nghiên cứu. (Trong quá trình viết, Báo cáo có thêm vào một lĩnh vực nữa là Vật lý học do vai trò của Thông kê học trong ngành này đang ngày một tăng).

Ngoài ra còn có các bài phát biểu của Chris Heyde (trường Đại học Quốc gia Australian và trường Đại học Columbia) về “Thông kê học trong bối cảnh quốc tế” và của James Berger (trường Đại học Duke) về “Vai trò và đóng góp của các Viện nghiên cứu đối với Thông kê học”.

Cần lưu ý rằng các vấn đề trong Thông kê không được chia ra thành các bộ phận nghiên cứu riêng biệt, chẳng hạn như “đại số, giải tích, hình học Topo và hình học” trong Toán học hay “vô cơ, hữu cơ, vật lý và lý sinh” trong Hóa học. Thay vào đó, có

phần trọng tâm nghiên cứu ta gọi là nòng cốt và một loạt các nghiên cứu định hướng ứng dụng được chia theo các lĩnh vực khoa học khác nhau.

Phần sau của bài viết sẽ đưa ra một cái nhìn tổng quan về hoạt động nghiên cứu trên mỗi lĩnh vực đó, trừ Xã hội học và Kinh tế học không được đề cập trong báo cáo đầy đủ. Khó khăn phải đối mặt khi biên soạn về Xã hội học và Kinh tế học không chỉ vì chúng khá tách biệt với các lĩnh vực còn lại mà còn vì chúng quá phức tạp. Các cán bộ nghiên cứu trong các lĩnh vực này thường không thuộc các bộ phận thống kê mà thay vào đó là thuộc các phòng ban như kinh tế, tâm lý hay xã hội. Các lĩnh vực này cũng bao gồm một số các lĩnh vực thành phần, trong đó có các tài liệu thống kê chuyên ngành riêng biệt, chẳng hạn như Trắc nghiệm học tâm lý hay Kinh tế lượng. Như vậy, chúng ta cũng đã thấy được đây là lĩnh vực hết sức rộng lớn và phức tạp, vượt quá sự cho phép về thời gian và các nguồn lực của chúng ta.

Bổ sung cho báo cáo này cần phải đề cập đến là cuốn “Thống kê trong thế kỷ XXI” (của các tác giả Raftery, Tanner và Wells, xuất bản năm 2002) với dung lượng 70 trang do các học giả hàng đầu hiện nay viết. Các nhà Thống kê nên sử dụng cuốn sách này như là một bản tóm tắt có giá trị về thông tin, bao gồm cả tình hình hiện tại cũng như định hướng trong tương lai cho các nghiên cứu trong một loạt các chuyên đề Thống kê.

#### 1.4. Cộng đồng thống kê

Xét về bản chất công việc, các nhà Thống kê làm việc trong nhiều môi trường công việc khác nhau. Tại Hoa Kỳ, nhiều nhà Thống kê làm việc trong các bộ phận thống kê và các bộ phận này chủ yếu thuộc các trường đại học chuyên nghiên cứu. Hiện nay có 86 chương trình cấp tiến sĩ chuyên ngành Thống kê, Thống kê Sinh học và Sinh trắc học. Các chương trình này hướng vào nghiên cứu sau đại học, bao gồm cả liên kết với các ngành khác cũng như có

các khóa học bậc đại học.

Các bộ phận thống kê chủ yếu được hình thành sau khi tách ra khỏi các bộ phận toán học trong nửa sau thế kỷ 20. Do có mối liên hệ lâu dài đó mà Thống kê thường được xem như là một bộ phận của Toán học. Có thể thấy quan điểm này tồn tại ngay trong NSF. Tại đây Xác suất- thống kê là một chương trình thuộc Bộ phận các Khoa học Toán học, bên cạnh các ngành Toán học thuần túy như Hình học Topo hay Số học.

Tuy vậy, một trong những kết luận quan trọng được các đại biểu đưa ra chính là: ngành Thống kê đã ngày càng trở nên khác biệt hơn so với các ngành Toán học khác. Mục tiêu của các nhà Thống kê và định hướng của khoa học hiện đại chính là nhằm hướng tới một tương lai, trong đó các công cụ Khoa học máy tính và thông tin ít nhất là trở nên quan trọng với ngành Thống kê như các lý thuyết về xác suất vậy.

Ngoài ra còn rất nhiều các nhà Thống kê đang làm việc trong các bộ phận khác (mà không phải là các bộ phận thống kê). Thực tế này thậm chí còn tồn tại ngay trong các trường đại học có các bộ phận thống kê. Tại đây có thể tìm thấy các nhà Thống kê làm việc trong khoa kinh doanh hay các bộ phận nghiên cứu khoa học và xã hội. Tại các trường không có bộ phận thống kê riêng biệt, ví dụ như trong các trường đại học bốn năm thì thường có các nhà Thống kê làm việc trong bộ phận Toán học, nơi cần có các nhà Thống kê phục vụ cho công tác giảng dạy. Cuối cùng, cũng có nhiều nhà Thống kê làm việc trong các bộ phận thống kê Sinh học và các trường Y.

*(xem tiếp trang 6)*

***(tiếp theo trang 19)***

Nằm ngoài giới học thuật nhưng vẫn có mối liên kết chặt chẽ với cộng đồng này chính là một số lượng lớn hơn nhiều những nhà Thống kê làm việc cho chính phủ, các doanh nghiệp cũng như nhiều đối tượng sử dụng số liệu thông kê khác. Nói về lĩnh vực Thông kê, Báo cáo Odom có nêu:

*Sự tương tác giữa giới học thuật và người dùng số liệu trong các ngành công nghiệp và trong chính phủ được thúc đẩy mạnh mẽ, do đó các ý tưởng trên lý thuyết và những khó khăn trong ứng dụng thực tiễn được phổ biến nhanh chóng, đồng thời tạo ra truyền thống của công tác liên ngành.*

Các nhà Thống kê làm việc trong các cơ quan chính phủ từ Cục điều tra dân số, Viện Tiêu chuẩn và Công nghệ Quốc gia cho tới Viện Y tế Quốc gia. Họ làm việc trong nhiều ngành khác nhau, thường là về công tác quản lý chất lượng. Đặc biệt, ngành công nghiệp dược phẩm dẫn đầu về việc tuyển dụng các nhà Thống kê - những người chịu trách nhiệm tiên hành thiết kế và phân tích các thử nghiệm cần thiết để phát triển các loại thuốc.

**(Còn tiếp)**

Quỳnh Trang (dịch), Đoàn Dũng (hiệu đính)  
Nguồn: A Report on the Future of Statistics  
<http://www.biostat.jhsph.edu/~kbroche/Stat%20-%20PDF/lindsay%20et%20al%20future%20of%20stat.pdf>