A new era of statistics and data science education in Japanese universities

Akimichi Takemura¹

Received: 31 December 2017 / Accepted: 9 April 2018 / Published online: 14 May 2018
© The Author(s) 2018, corrected publication October 2018

Abstract In April 2017, Shiga University launched an undergraduate program in the new faculty of data science. This faculty emphasizes the study and application of statistics and is the first of its kind in Japan. Shiga University also plans to launch a master's program in data science starting in April 2019. The inauguration of the faculty marks a new era of statistics and data science education in Japanese universities, in view of the fact that there were virtually no statistics faculties in Japanese universities before that of Shiga University. In April 2018, Yokohama City University will follow Shiga University with the opening of a new school of data science. We discuss the background of these developments and the prospects of statistics and data science in Japan.

Keywords Data as economic resource · Data scientist · Internet of Things · Project-based learning · Statistical science

1 The big data era and data science

The popularity and use of smartphones has increased dramatically over the last 10 years, since the release of the first iPhone by Apple in 2007, and is symbolic of the arrival of the big data era. These smartphones can receive good wireless reception almost anywhere and people are constantly using them. Smartphone users exchange emails and post on social networking services (SNS), query search engines, and order merchandise from online shops, with the transactions recorded in

JSPS Grant-in-Aid for Scientific Research No. 25220001.

✉ Akimichi Takemura
a-takemura@biwako.shiga-u.ad.jp

¹ Faculty of Data Science, Shiga University, 1-1-1 Banba, Hikone, Shiga 522-8522, Japan

databases. Hence, big data is now a reality and can be considered to be an economic resource. In its May 6, 2017 issue, the Economist magazine (2017) declared that “the world’s most valuable resource is no longer oil, but data”. It also discussed the need for regulating Internet giants, such as Google and Amazon, and its cover depicted Internet giants as oil rigs.

Wireless communication technology is progressing quickly and the fifth generation of mobile networks (5G) is expected to arrive in around 5 years. This 5G network will be 100 times faster than the current 4G technology. In addition, low power wide area (LPWA) technology will arrive within the next few years. As the name suggests, LPWA devices are highly efficient and can potentially be run for up to 1 year with a single battery and can communicate over long distances, around 50 km, although the communication speed is slow. LPWA networks will be useful for Internet of Things (IoT) applications. It is clear from the rapid progress of communication technology that more data will be generated and communicated by people and machines.

With the big data era at hand, there is a strong need to be able to analyze this data. Just as crude oil requires processing and refining before it becomes a valuable product, big data also needs processing and analysis before value can be extracted from it. As we discuss later, data science is the methodology for processing, analyzing, and extracting value from big data and its practitioners are called data scientists. Companies that have both data and data science skills have a competitive edge. This is the reason for the fast growth of Internet giants, and these companies are hiring many data scientists and investing heavily in algorithms for handling and analyzing big data. An article by Pierson in the October 2017 issue of AMSTAT news (Pierson 2017) shows a very large increase in the number of degrees conferred in statistics and biostatistics in the USA in recent years. In 2016, the number of master’s and bachelor’s degrees conferred in statistics was about 4000 and 3000, respectively, which is about five times as many as in 2008. Around 2008, Varian (2008, 2009) told the press: “I keep saying the sexy job in the next 10 years will be statisticians.” In 2012, Davenport and Patil (2012) talked about data scientist being the sexiest profession of the twenty-first century. The article by Pierson in AMSTAT news seems to confirm these predictions. In the USA, there are about 100 statistics departments and the histories of many of these departments are described in Agresti and Meng (2012).

The same phenomenon is being seen in China; there are now more than 300 statistics departments in China and the number is still growing (Wei 2017). China now has its own Internet giants, such as Tencent and the Alibaba Group, and these companies are also hiring many data scientists.

Japan is lagging far behind the USA and China in data science, and so the Japanese government has recently started to emphasize the importance of data science. The Japan Revitalization Strategy 2016 (Prime Minister 2016) states “in the big data era, technologies for new business and services are based on the utilization of data. They include artificial intelligence, big data, IoT, etc.” The Strategy for Scientific and Technological Innovation 2015 (Cabinet Office 2015) states “Japan is in a very risky position compared to other countries, because of the severe lack of people knowledgeable in data analysis and statistical science.” One

of the main reasons for this lack in Japan is the absence of statistics departments in Japanese universities. We discuss this point in the next section.

Historically, Japanese industries were very successful after the Second World War until the end of the 1980s. The growth of the Japanese economy after the Second World War was remarkable. The reasons for this success in the 1980s have been widely discussed, e.g., in Ezra Vogel's book (1979). One of the reasons is the wide utilization of statistical quality control (SQC) techniques in manufacturing. Before 1950, Japanese products were cheap and of poor quality, then in July 1950, Edwards Deming came to Japan and gave a series of lectures on SQC. These lectures were very influential and Japanese manufacturing companies then began implementing SQC techniques to continually improve and stabilize product quality. It should also be mentioned that these techniques were particularly useful for the mass production of products such as cars or home electric appliances. It is somewhat ironic that on June 24, 1980, NBC broadcast a special program "If Japan Can, Why Can't We?", reintroducing Deming to the American public (NBC News Program 1980).

After the burst of Japan's economic bubble in 1991, the Japanese economy stagnated and the period until 2010 is often called "the lost two decades" of the Japanese economy. Mass production of commodities moved to China and other Southeast Asian countries. During these two decades, there were large innovations in the Information and Communication Technology (ICT) sector, for example the advent of the World Wide Web in 1993 and smartphones in 2007. However, Japanese manufacturers were not successful in keeping pace with these innovations.

2 Statistics and data science in Japan and other countries

Before Shiga University, there were almost no statistics departments in Japanese universities. It is natural to ask why this happened. One exception is the Department of Statistical Science, School of Multidisciplinary Sciences, The Graduate University of Advanced Studies, which was established in October 1988 and is closely connected to the Institute of Statistical Mathematics in Tokyo. This department is mainly concerned with its Ph.D. program and confers about 5 Ph.D. degrees in statistical science each year. It has been the only department in Japan awarding Ph.D. degrees in statistical science. It should also be mentioned that for a brief period before 1970, there was a department of statistics in the engineering school of Nihon University, led by Junjiro Ogawa. Details of this department are not well documented. Takeuchi (2018) of Osaka University collected some relevant material on this department. It seems that the department was active about 3 years from 1966 to 1969. Junjiro Ogawa and other professors of the department had some confrontation with the executive office of Nihon University and they resigned from Nihon University in 1969.

In Japanese universities, statisticians are scattered across various faculties, such as economics, mathematics, engineering, or education. This is in sharp contrast to the USA, Korea and China, where there are independent statistics departments. There have been some efforts to form independent statistics departments in Japanese universities, such as the one in Nihon University mentioned above. However, these efforts were not successful before the formation of the data science faculty of Shiga

University. There are many possible reasons for this. Japanese statisticians tended to emphasize the application of statistics to other fields, rather than pursuing theoretical statistics itself. There were some original contributions from Japanese statisticians, such as statistical quality control for the manufacturing industries and the Akaike information criterion for model selection. These innovations were motivated by the application of statistics to practical problems. Another reason may be that the voices of Japanese statisticians were not united. In fact, the Japanese statistics community was divided into many academic societies. This reflects the fact that Japanese academic statisticians are scattered across various faculties. As an umbrella organization for six academic societies of statistical science (Japanese Society of Applied Statistics, Japanese Society of Computational Statistics, The Biometric Society of Japan, The Behaviormetric Society, Japan Statistical Society, Japanese Classification Society), the Japanese Federation of Statistical Science Associations (JFSSA) was formed in 2005 to promote common interests of these societies, such as statistical education.

However, probably the biggest reason for the lack of a dedicated faculty was that there was no “statistics industry” and people were not sure whether graduates from a statistics faculty would have good employment opportunities in Japan. The faculties and departments of Japanese universities are organized to reflect the organization of industry. For example, graduates of economics faculties usually enter the financial sector, e.g., banks and insurance companies, and graduates of law faculties typically become public officials or lawyers. Similarly, electrical engineering departments have close ties with electric companies and mechanical engineering departments have close ties with car and other manufacturing companies. In contrast to this “vertical” segmentation of faculties and departments according to the segmentation of industries, statistics is a methodology that is useful for many fields. We may thus call statistics a horizontal field, where horizontal means transdisciplinary or transversal. Computer science also has this horizontal characteristic, because information technology is useful in many fields. However, there is also a manufacturing industry for computers and related technology. Hence, there are some computer science departments in Japanese universities, although these departments are not as ubiquitous as e.g., electrical engineering.

In Japan, horizontal fields and techniques were not considered of primary importance. Students were supposed to first learn specific fields, such as economics and mechanical engineering, and then learn statistics only if it became necessary for research or product development. Some good Japanese applied statisticians taught themselves statistics, because a formal and systematic education in statistics was not available, mainly due to the absence of statistics departments in Japanese universities.

Vertical segmentation can also be seen in the Japanese high school education system. High school students intending to go to university are divided early on into a humanities-oriented course and a science-oriented course. Students in the humanities-oriented course take entrance examinations for the faculties of law, economics or literature and students in the science-oriented course take entrance examinations for engineering and science faculties.

As a result of this vertical segmentation of university education, in Japan there are many people with deep vertical skills, who are experts in their own field. However,

there are few people with horizontal skills. For example, Japanese managers are humanity oriented and do not necessarily have a good understanding of the technical side of their companies. Similarly, engineers typically do not understand the business side of their own companies. In the big data era, where technologies for new businesses and services are based on the utilization of data, it seems that horizontal skills are more effective than vertical skills. This is one reason Japan is lagging behind other countries in the big data era. However, Japanese people are finally becoming aware that we need more people with horizontal skills, such as data scientists, as shown in government reports (Prime Minister 2016; Cabinet Office 2015) and this has opened opportunities to establish data science faculties in Japanese universities.

At this point, I touch on the difference between statistics and data science. As I discuss in the next section, data science is an interdisciplinary field combining statistics, computer science, and domain knowledge to extract value from data (e.g., The data science 2015). In this sense, data science is a broader field than traditional statistics. However, big data and data science are buzzwords and there are some doubts about the longevity of these fields. In the USA, where there are established departments of statistics, many statisticians argue that statisticians have already been doing data science for a long time. Donoho (2017) provides some deep insights into the development of statistics and exploratory data analysis in the last 50 years since Tukey (1962). Cleveland (2001) and Wu (1997) proposed the use of the term data science about 20 years ago, as Hayashi (1998) and Shibata (2001) did in Japan. Baumer (2015) discusses opportunities and challenges for statisticians in teaching data science to undergraduate students. In Japan, where there were virtually no statistics departments until recently, the situation is somewhat different and data science is more of an opportunity than a challenge for statistics.

Another important difference between traditional statistics and data science is the increasing importance of unstructured data in data science, such as text data, image data and sound data. Text messages on SNS are important sources of information regarding how people think and act. Imaging devices are becoming ubiquitous and there is an increasing need to analyze image data in real time. For example, imaging devices can be used to monitor manufacturing processes and then abnormalities can be detected in the processes from the images in real time. In the case of unstructured data, numerical features with appropriate dimensions have to be computed before the application of traditional statistical methods. The construction of appropriate features may be more important than the statistical methods used.

As big data becomes more widely available, some people claim that traditional statistical sampling is irrelevant. However, big data often contains biases, because the underlying population may not be correctly reflected in big data. For example, when text messages from a particular SNS are analyzed to determine people's opinion on a given political issue, it should be kept in mind that the messages only come from people using the service. A more important point is that big data is observational data, unlike data from randomized controlled trials, and it is difficult to derive a causal interpretation of the data. Basic statistical notions such as population, bias, sampling and randomization remain important for the analysis of big data.

There is some concern as to whether data science is really a science or not. We can argue that data science is a science whose objective is the understanding of big

data. But, the phrase big data itself is not clearly defined. People will be more comfortable with the expression “data-driven science”, in view of the fact that now almost all scientific research is based on the analysis of large amounts of data. This tendency for data-driven developments in scientific research is well discussed in Hey et al. (2009) and called the fourth paradigm of scientific research. We should also note that statisticians are comfortable with the expression “statistical science”, which refers to the set of scientific fields that use statistics and related methodologies heavily. Statistical science and data-driven science are almost synonyms, although the former emphasizes the methodology and the latter emphasizes the data. As Donoho (2017) discusses, data science is currently more motivated by commercial rather than intellectual developments. Although I feel that it is a good thing to have commercial and business motivations for data science, “data-driven business” would be a more appropriate phrase than data science for business-focused applications.

3 Establishment of a data science department in Shiga University

With the social background described in the previous sections, Shiga University proposed the formation of a new faculty of data science in 2014. Before the establishment of the data science faculty, Shiga University consisted of only two faculties: economics and education. For a long time, Shiga University has been trying to add a new faculty, which is more science oriented than economics and education. When the need for data science became clearer in 2014, the president of Shiga University at that time, Prof. Takamitsu Sawa, convinced the university that a data science faculty was the way to go for Shiga University. As a national university, the university had to negotiate with the Ministry of Education, Culture, Sports, Science and Technology. Since the government already acknowledged the importance of data science, the negotiations went rather smoothly and the opening of the new data science faculty was officially approved in August 2016. The faculty accepts 100 students each year.

The basic idea of the faculty is that the field of data science consists of data engineering (computer science), data analysis (statistics) and the extraction of value from data by utilizing domain knowledge. This combination is often depicted in the form of a Venn diagram (The data science 2015). The curriculum of the data science faculty is also based on this idea. Students first learn basic programming skills and statistics. Then, they learn how to apply these skills to real data.

The data science faculty of Shiga University offers a full range of courses on statistics and computer science. In statistics, in addition to descriptive and inferential statistics, courses are offered on multivariate analysis, time series analysis, Bayesian methods, survival analysis, model selection, simulation, etc. In the computer science courses, knowledge of the Python and R programming languages is required. Furthermore, courses on data structure and algorithms, information theory, visual programming, artificial intelligence, etc., are also provided. More details on these courses can be found on the curriculum map in

<https://www.ds.shiga-u.ac.jp/en/>.

Unlike computer science and statistics, skills for extracting value from data cannot be taught only with lectures. They are gained by students through project-based learning in practical sessions. In the data science faculty of Shiga University, we obtain data sets, such as Point of Sales (POS) data, from companies for this project-based learning. For this purpose, we have collaborative agreements with more than 30 companies and other institutions.

For educational purposes, real data often has to be anonymized or partially aggregated. Some IT companies provide a data analytics platform rather than the data itself. The platform is typically a web interface which allows users to summarize and visualize data easily. This turns out to be a convenient scheme for project-based learning, if the platform is flexible enough to allow students to explore a data set from various viewpoints.

As another possibility, we encourage students to participate in data science competitions, such as those on the Kaggle platform. There are similar competitions in Japan, including the sports data analysis competition organized by the Japan Statistical Society.

In preparation for the opening of the data science faculty of Shiga University, I contacted more than 100 companies for possible cooperation. The real value in contacting and interviewing these companies was that we could gain insights into trends in data science in Japanese companies. Many companies now have lots of data, but do not have people with the appropriate skills for analyzing this data. Many large companies have recently set up data science departments, but have difficulty finding suitable personnel for the department. These companies are interested in hiring graduates from our faculty.

4 Prospects for statistics and data science in Japan

The arrival of the big data era has opened a window of opportunity for statistics in Japan, where there were previously almost no statistics department. Statistics education in Japan must adjust to these new challenges. The Japanese government and Japanese industry have become very aware of the need for statisticians and data scientists. The job prospects for our graduates will be good for some time to come. Also, there is a strong need for updating the data science skills of current company employees. This is the reason for launching a master's program in data science at Shiga University. Yokohama City University has plans to open a similar course.

The success of data science at Shiga University and Yokohama City University is being closely watched by other universities in Japan. With the decline and aging of the Japanese population, many universities are facing financial difficulties. National universities have to operate under tight budgets and it is difficult to form a new faculty in national universities. However, if Shiga University and Yokoyama City University are successful, then other universities will follow.

Recently, artificial intelligence (AI) is much talked about and there is an inflated expectation that AI will make many professions obsolete, including data scientists. Since the progress in the field of AI is so fast, we cannot predict what will happen. However, current AI technologies are based on improvements in predictive

modeling based on big data. These models are complicated and tend to be black box models. As noted above, big data is observational data and predictive modeling cannot give insight into causal interpretations. We will need knowledgeable people to interpret big data for the foreseeable future.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

Acknowledgements The author is grateful for useful comments by a reviewer. This work was supported by JSPS KAKENHI Grant 25220001.

References

- Agresti, A., & Meng, X.-L. (2012). *Strength in numbers: The rising of academic statistics departments in the U.S.* New York: Springer.
- Baumer, B. (2015). A data science course for undergraduates: Thinking with data. *The American Statistician*, 69(4), 334–342.
- Cabinet Office. (2015). Government of Japan. *Strategy for Scientific and Technological Innovation 2015*, June 19, 2015.
- Cleveland, W. S. (2001). Data Science: An action plan for expanding the technical areas of the field of statistics. *International Statistical Review*, 69, 21–26.
- Davenport, T., & Patil, T. H. D. J. (2013). Data scientist: The sexiest job of the 21st century. October 2012 issue. *Harvard Business Review*. 2012.
- Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4), 745–766.
- Hayashi, C. (1998). What is data science? Fundamental concepts and a heuristic example. In C. Hayashi, K. Yajima, H. H. Bock, N. Ohsumi, Y. Tanaka, & Y. Baba (Eds.) *Proceedings of the fifth conference of the international federation of classification societies (IFCS-96). Data science, classification, and related methods*. Tokyo: Springer
- Hey, T., Tansley, S., & Tolle, K. (2009). The fourth paradigm: Data-intensive scientific discovery. Microsoft Research.
- NBC News Program. If Japan Can, Why Can't? June 24, 1980.
- Pierson, S. (2017). Bachelor's, master's statistics and biostatistics degree growth strong through 2016. *AMSTAT News*, 472, 14–16.
- Prime Minister of Japan and His Cabinet. (2016). *Japan Revitalization Strategy 2016*. June 2, 2016.
- Shibata, R. (2001). *Data Literacy.*, Data Science Series Tokyo: Kyoritsu Shuppan. (in Japanese).
- Takeuchi, Y. (2018). Curriculum of the department of statistics in the engineering school of Nihon University. Presentation at the spring meeting of Japan Statistical Society, March 2018 (In Japanese).
- The data science Venn diagram. Drew Conway Data Consulting, LLC. 2015. <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>
- The Economist. (2017). The world's most valuable resource is no longer oil, but data. *The Economist*, May 6th 2017.
- Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 33, 1–67.
- Varian, H. (2008). Statistics—dream job of the next decade. Keynote Presentation at the 2008 Almaden Institute: <http://www.youtube.com/watch?v=D4FQsYTbLoI>
- Varian, H. (2009). How the web challenges managers. *The McKinsey Quarterly*, January 2009.
- Vogel, E. F. (1979). *Japan as number one: Lessons for America*. Harvard: Harvard University Press.
- Wei, Y. (2017). Big data analytics education in China. A talk at Shiga University. January 19, 2017.
- Wu, C. F. J. (1997). Statistics = data science? Inaugural lecture for the H. C. Carver Chair in Statistics at the University of Michigan. 1997. <http://www2.isye.gatech.edu/jeffwpresentations/datascience.pdf>.