Statistics within business in the era of big data

Gareth M. James

Department of Data Sciences and Operations, University of Southern California, United States

ARTICLE INFO

ABSTRACT

The last decade has seen a dramatic increase in the availability of business data. Here I discuss some of the corresponding opportunities and challenges for business related statistical applications and the role that statisticians can play within a business school.

1. Introduction

Over the last century the field of statistics has made significant contributions to efforts aimed at developing rational and objective interpretations of data, and information in general. Furthermore, there can be little doubt that the prominence and stature of the field has grown considerably over the last 20 years. This growth has been primarily driven by developments outside statistics, in particular technological advances that have both created vast quantities of new and complicated types of data, and correspondingly provided us with the mechanism for its analysis.

The late nineties, when I graduated, was a good time to be a statistician. The first wave of new data, the sequenced human genome, had just arrived. Suddenly biologists were inundated with enormous quantities of data, of a type that had never been studied before. This opened up the possibility to address significant new research questions and statisticians made major contributions towards addressing these problems. This continued a strong and productive tradition of statisticians working closely with the biological and medical sciences.

However, when I graduated I was not headed to a statistics or biostatistics department, but instead to a statistics group within the Marshall School of Business here at USC. For various reasons, the relationship between statistics and the business disciplines has historically been more strained. At that time, while elsewhere the field was flourishing, statistics departments within business schools were closing or merging with other departments, and the number of statistics faculty was on the decline.

Ten years ago represented the lowest point for statistics within Marshall and likely business schools in general. Several statistics faculty retired and there was a serious discussion about simply closing the group. Fortunately, our dean made a farsighted decision and elected to rebuild statistics within Marshall. Although at the time not an obvious call, in retrospect this turned out to be fantastic strategic decision. We have recruited extremely strong faculty almost every year since, started a successful Ph.D. program, and increased the number of students we teach in our electives by several hundred percent. Our department, which used to comfortably fit on one floor, is now overflowing on two floors, and the statistics group has become one of the largest, and strongest, business groups in the country. Marshall is now a very friendly environment to be a statistician.

This expansion was of course fueled by the second wave of new data, this time from industry and government. Whether it was Netflix offering one million dollars to build a better movie recommendation system, Amazon or Facebook performing A/B testing on millions of customers, Uber trying to set prices for both drivers and riders in a complicated two-sided dynamic market, Waze optimizing their traffic recommendations, Target predicting which customers were pregnant based

---

E-mail address: gareth.james@marshall.usc.edu.

on purchase history or countless other examples, many companies suddenly had an urgent need for statistical and related expertise. Businesses soon discovered that their survival depended on their ability to efficiently process these vast quantities of data. This in turn drove employers to demand stronger quantitative skills in the students they hired. A decade ago I launched a new MBA elective class, from which my "An Introduction to Statistical Learning" book eventually evolved. My colleagues roundly predicted that the class enrollment would provide a real world operationalization of the "measure zero" concept and worked to make the prophecy self-fulfilling by scheduling it at 8 am, but it ended up being well attended. At the time several students reported obtaining internships, and even job offers, because they represented a tiny minority of MBA students who could confidently discuss concepts such as Boosting or Support Vector Machines during an interview.

Fast forward a decade and the pendulum has swung in the opposite direction. Now, in many industries, it has become a necessary condition for our students to be able to converse fluently on such topics. As a result, despite a vast increase in our elective options, students are struggling to get into the statistics and related classes that they need. At the same time there is strong demand for our Ph.D. students, partly driven by the need for statisticians to teach all these new classes. In addition, more business school faculty are recognizing that new statistical methodologies are advantageous, and in many cases essential, to the analysis of these new types of data.

Working as a statistician within a business school is a perfect match to my interests. I am proud to represent both areas but there are also some obvious tensions that I encounter on a regular basis. I believe that statistical applications in the biological and medical sciences are extremely important. However, that topic has, and will continue to be, well covered. Hence, my goal here is to highlight some of the challenges and opportunities that statistics, and statisticians, face within business and business schools. In the following sections I discuss implications for research as well as for Ph.D., masters and undergraduate programs.

2. Research

Likely, the biggest challenge in developing a harmonious relationship between statistics and business schools revolves around the impact of statistical research on business applications. There appears to be clear agreement that business students need an exposure to statistical concepts, a sentiment which has only grown with the recent data analytics revolution. On the research front a number of my Marshall colleagues, especially in the Marketing area, have been very open to exploring new statistical ideas. However, for a variety of reasons, many business faculty, while implementing classical statistical methods in their own research, see little benefit in adopting more modern and innovative statistical approaches. The issue is prevalent and a key cause for many of the research tensions. Statisticians often also contribute to the problem by displaying little interest in educating their colleagues about the potential impact of these new methodologies.

In an attempt to address this issue, a colleague of mine and I are in the process of organizing a workshop titled "Business Research Applications Needing Data Science" (BRANDS). The emphasis of the workshop will be educating participants on the connections among analytics, data science and business research problems, with the goal of stimulating future collaborations. Hence, BRANDS will bring together researchers with expertise in data science, faculty working on important business research problems, and external industry leaders with the goal of generating cross-fertilization of knowledge across the areas. Providing faculty with exposure to real world problems and needs from industry has the potential to greatly inspire innovative research ideas and developments that will further advance the frontiers in both academia and industry.

I believe that business schools could benefit from exploring the successful "dry labs" model favored by biostatisticians, consisting of a senior faculty member working with junior faculty, postdocs, and Ph.D. students. A key goal of the BRANDS workshop is that it will ultimately lead to a BRANDS lab, where statistical faculty will interact with general business faculty and industry leaders in a cooperative and interactive environment. We believe that such a lab has the potential to help bridge the gap between statistics and general business research.

Another issue faced by business schools is that many of them are relatively isolated within their universities. Statisticians provide an opportunity to increase interactions across schools. In particular, the fields of statistics and computer science have grown closer over the last few decades as computational methods have become the norm within statistics. Several of my colleagues have joint research projects, and grants, with computer science faculty, which opens a conduit for interactions between the business and engineering schools.

Finally, perhaps somewhat surprisingly, business schools themselves often struggle to form productive connections with industry and, as a result, are concerned about the relevance of their research. Here too statisticians can help. In the last couple of years several faculty from my department have taken semester long sabbaticals within tech companies and returned to Marshall, not only with strong connections to these organizations, but also a much clearer understanding of the main issues they face, which in turn positively influences their research and teaching. Companies appear to understand the potential value of these interactions and are actively seeking them out. For instance, Uber runs symposiums bringing together academics from economics, statistics, computer science and related fields to meet with their data scientists to discuss potential joint research topics. Such events provide a perfect opportunity to bring relevance, as well as rigor, into our research. Despite some cultural differences, statistics faculty provide significant opportunities for new and innovative business research, provided successful mechanisms can be created to ensure productive interactions.

3. Ph.D. programs

After over a decade of actively recruiting faculty here at Marshall it has become clear that only a small percentage of students graduating from statistics departments have any experience with applications in business. Furthermore, many have never considered a position within a business school, and some are actively wary about such an appointment. Why is this? It is certainly not based on compensation, which with 9-month starting salaries of at least $150,000 USD are over 50% higher than within statistics departments. Neither is it research resources which generally include two months of guaranteed summer funding, research budgets of approximately $15,000 each year, and fully funded Ph.D. students, without a need to apply for grants. Teaching loads are also low, at around 2.5 semester or 3 quarter classes per year.

There appear to be several reasons for the relative lack of interest. First, many undergraduate students graduating with both strong quantitative skills and business interests choose to take up an industry position. Any such students who do complete a Ph.D., often opt to take these skills to firms such as Google, Amazon, Adobe, Microsoft Research, eBay, or a host of other companies who can provide similar highly stimulating opportunities to that of an academic position. Second, while many faculty within statistics departments collaborate on biological applications, a much smaller fraction work on business related problems. As a result, students who may seek to work in this area struggle to find a relevant advisor. This was certainly an issue when I completed my doctoral training, where I started with the goal of working on business problems but found that this presented me with few options. Instead I switched into a pure statistical research direction and only ended up at a business school partly by happenstance. Of course this issue is somewhat circular in that one reason that few statistics faculty work in this area is that they never had the opportunity during their doctoral training.

Third, many graduating students are concerned that there will be significant restrictions placed on the type of research that they can perform within a business school. In my experience this concern is misplaced. While I would never recommend such a position for someone with no interest in business applications, I have always felt complete freedom to work on any research problem that interests me, both pure statistical, and business application related. In fact, based on my interactions with faculty in biostatistics departments it appears that, for funding reasons, they often feel more constrained in the problems they can work on than do any of my Marshall colleagues. Finally, for some students there appears to be a certain prestige associated with an appointment in a statistics department and a fear of the unknown moving to a business school. Of course in the long run such considerations have little to do with one's everyday work environment and more than one faculty member who opted for a statistics department position over that of a business school has had a change of heart after a few years of additional experience.

One could point to statistics students graduating from business school programs as a possible source for faculty positions. However, business schools graduate a relatively small number of Ph.D. students compared to the sciences, so there are certainly not enough to satisfy the growing demand within these schools. Perhaps in the past there was less need for such students because of a relatively small number of positions. However, as discussed previously, there is now a growing demand within business schools for highly skilled statisticians. One negative consequence of this demand imbalance is that these schools sometimes end up recruiting statisticians who are less open to business applications. Unfortunately, this can cause tensions for the faculty member and a sense that they do not wish to contribute to the school's research mission, in a similar fashion to a statistician with no biology interests joining a biostatistics department.

What measures can be taken to address this issue? Since it seems likely that business schools will be unable to generate enough graduates it is important that demand is spurred among students graduating from statistics departments. These departments are generally seeing a growing interest in their Ph.D. programs and many of these students would be open to interesting business related research problems, provided they had faculty with whom to work. In this regard the core issue may well be the lack of such faculty within statistics departments. One potential solution is to encourage more joint appointments between statistics departments and business schools. Or alternatively, even without any formal appointment, by encouraging more collaborations with business school faculty, for instance, by establishing university wide data science institutes.

Government funding agencies such as the NSF can also play a role here. If these agencies were to encourage more proposals targeted at statistical applications in business it would likely have a significant impact. Finally, business schools themselves need to take a more active role in training future academics by partially funding students within statistics departments. This would likely have a similar effect to an increase in government funding. Working across schools is difficult but collaborations are often achieved between statistics departments and schools of medicine or public health, so a similar arrangement with business schools should be possible.

4. Masters programs

The dearth of statistics Ph.D. students with business interests is certainly not an issue at the masters level. In fact, the opposite is clearly the case, with enormous increases in demand for statistics related courses among both MBA students and more specialized one year MS programs in "Business Analytics" (MSBA). Marshall launched an MSBA program in Fall 2014 which received 1200 applications this year. We have seen similar dramatic increases in demand from the MBA students. The challenge here then is not to find qualified students but rather to train them appropriately.

In response to the changing needs of our students, Marshall has split our first year MBA statistics class into two parts. The first course covers more traditional statistical concepts such as probability, confidence intervals, hypothesis testing and linear

regression. However, the second course highlights analytics concepts, which include both modern statistical techniques such as non-linear regression, as well as more business related topics like spreadsheet and simulation modeling. Many of these students then take more advanced analytics elective classes in their second year.

In our MSBA program we teach a mixture of technical courses in statistics and other quantitative fields, pure business courses, and softer skills such as communication. This balance is in response to a generally expressed concern from the business community that students with a pure statistics training often have difficulties working in teams, and in communicating their conclusions in real world settings. Our approach appears to have received high marks from both students and employers, with a number one ranking for the program, and essentially 100% placement rate.

However, achieving the optimal allocation among these three areas is challenging. We have been adding units to the program as it has become clear, based on student and employer feedback, that there are still holes in their knowledge. Even among the more technical subjects there is some disagreement about how much material should be related to traditional statistical content versus other topics like operations management, or information systems. Unfortunately, many statistics faculty appear to be relatively passive in advocating for their subject. As a result, statistics often does not receive the credit that it may deserve. Marketing is important even within academia!

5. Undergraduate education

Arguably the single largest impact that statistics faculty have is through the required introductory statistics class that the vast majority of undergraduate students will take at some stage during their studies. Teaching this course to business majors possess a unique set of challenges. At USC these are strong students, but typically not as quantitatively orientated as STEM majors. Our students usually take the class in their first year, before they have been exposed to most of their other business courses, or any real world statistics applications. As a result, unlike our MBA students, who have generally had enough business experience to draw connections to the statistical tools, undergraduates struggle to understand the applicability of the concepts.

Given this disconnect between the timing of the statistics class and their other general business courses, Marshall elected to introduce a half semester class that students take as a capstone in their senior year. This relatively applied course, taught in a computer lab, integrates topics from statistics in a case based fashion with applications in finance, marketing, accounting, operations management, and other business disciplines. The goal is to help students understand how all the topics interrelate and how statistics can be used to address these problems. It was a difficult course to create and required a couple of iterations to perfect, but it now appears to be quite successful, and of great value to our graduates.

However, as I suspect is the case in many departments, our first year core course still relies heavily on classical statistical techniques. Many of these topics are relevant to today's data environment, but some methods, and certainly many of the more technical details, may need to be dropped to free up time for more modern approaches. Unfortunately, this is a difficult balancing act, and removing content is always painful, but our statistics classes have not evolved at the same rate as has the field as a whole, and that is a loss for our students. A well constructed class should be easier to teach now than it was 20 years ago, because there are now so many interesting real world business applications of statistical ideas.

Perhaps partly as a result of all of these issues, while we have seen an increase in elective enrollments in our undergraduate courses, there has not been nearly as dramatic a change in the landscape as has occurred for the graduate classes. I do not see any intrinsic reason why the demand should differ between the two groups since all students will be taking up positions in a heavily datacentric world. The main difference seems to be a lack of knowledge from the undergraduates about the applications of statistical techniques.

6. Conclusion

Statisticians have integrated well within the biological and medical sciences, but have had a more limited impact on business schools. However, with the sudden influx of large quantities of business related data, and the corresponding interest in statistical methods from students and researchers, there now appears to be a clear opportunity for a closer working relationship. This process could be accelerated if we are willing to adopt some of the previously discussed initiatives.

A key concern is the development of more high quality Ph.D. graduates with interests in business applications. This could be achieved by promoting more joint faculty appointments, and the provision of more funding options, either from business schools or government funding agencies, for statistics Ph.D. students with business interests. Currently, there is a great deal of demand for statistics, or analytics, masters programs, which presents a significant opportunity for statistics groups within business schools. However, it is important that statisticians are willing to work as full partners on the creation of such degrees, neither insisting that the program be overly technical, nor turning the process over entirely to other business faculty.

In many cases the introductory statistics class in both business, and non-business, undergraduate programs has not been adequately updated to reflect the many new innovations within the field. There is a tendency for these courses to overly emphasize pencil and paper calculations, at the expense of allocating time to more modern, and often computational, approaches. Finally, within the research domain both statisticians, and general business faculty, could benefit from the BRANDS lab concept. Exposing statistics faculty to research problems with which business faculty are struggling provides them with the background to develop useful methodological contributions. Similarly, exposing business faculty to many

of the new cutting edge statistical approaches provides them with the insights needed to use these methods in their own research.

It is an exciting time to be a statistician, both within a business school, and in the wider academic community. It is important that we use this opportunity to further cement the field as a fundamental partner in all data related applications, including those within business.