Trends in Applied Mathematics and its Adoption in the Finance Industry

or why you should give blockchains and big data a pass

John Dodson

October 30, 2020

Introduction

Over the course of the twentieth century, applied mathematics has gradually assimilated the subjects of

- probability,
- statistics, and
- information.

An outside observer of trends in STEM in finance might focus on the industry's embrace of computing technology during the Moore's Law era.

- I claim these quieter innovations in math are ultimately more impactful, because
- they help firms to organize information technologies and financial innovation to create lasting value for clients.

Probability (MSC 60)

Probability became recognized as a field in mathematics in the mid-1800s based on results such as the law of large numbers, the central limit theorem, the Glivenko-Cantelli theorem about empirical measures, and, probably most significantly, the main results in extreme value theory in the middle of the 20th century.

Prior to the work of Andrey Kolmogorov in the 1930s to make an explicit connection to measure theory, probability was considered by mathematicians as just a class of counting exercises. Laplace's illumination of the central limit theorem in the 1810s was the first hint that probability might ultimately offer deep insights into the nature of mathematics. The "law of small numbers" about the universality of the Pareto distribution was probably the result that formally secured probability's place in mathematics.

Statistics (MSC 62)

Statistics, like probability, was a subject long before it was a *mathematical* subject. It began to attract the attention of mathematicians with the notions of the "null hypothesis" and experiment design, which were a few of the many contributions of Ronald Fisher.

Fisher also showed that it was valuable to analyze estimators as functionals, leading to the maximum likelihood estimator and ultimately the Cramér-Rao bound and an explicit notion of robustness in terms of the boundedness of the Gateaux derivative.

The deep nature of statistical dependence was explored, notably on the empirical side by Maurice Kendall and on the theoretical side by Abe Sklar, in the mid 20th century.

Information (MSC 94)

The formal study of information includes the theory of encoding and recovering digital messages in secure, noisy, lossy, or lossless formats, which is of course of great interest to electrical engineers. But it also includes topics related to Claude Shannon's fundamental concept of entropy, including the contributions of Solomon Kullback, that expose another deep connection between information and probability. There is also a connection to filtrations, which can be thought of as noisy signals.

Information and Category Theory

There have been recent innovations in adjacent concepts within category theory (MSC 18) about ontology, the nature of information itself. See for example the works of Robin Milner, David Spivak, and John Baez, which has influenced the design of programming languages and database systems.

Bitcoin

Blockchain

- Distributed blockchain ledger is a database architecture.
- Relative to other database architectures, it offers the possibility of a "trust-free" implementation for creating a virtual scarce resource: a crypto asset.
- in practice, all commercial implementations to date are private and only allow trusted entities to participate in validation.

Bitcoin

Bitcoin, in contrast, is an example of a public blockchain.

- Other then through source-code control, there is no authority to regulate the activity of participants or validators.

Bitcoin

it was expected that Bitcoin would exhibit the properties of a currency (i.e., a medium of exchange, store of value, and unit of account)

It has not.

It seems Bitcoin is more similar to a precious commodity such as gold. And that is how it is currently regulated in the US, as a commodity subject to CFTC provisions.

According to Katharina Pistor at Columbia, the issue seems to originate in the fundamental incompleteness of contacts and the lack of an authority to resolve this incompleteness in a manner that is perceived as just. In mathematical terms, there is a gap in the Bitcoin ontology.

Bitcoin

Let me give three specific examples of failure related to the incompleteness of contracts:

- Merchant chargebacks

With no authority able impose an edit on the blockchain, even cancel/correct transactions have to go through an expensive validation. Who should pay for this, the merchant or the customer? Does it need to be negotiated each time?

- Tether

Tether is a public blockchain that was meant to be like a digital trust certificate for USD. But the originators refused to co-operate with any trusted authorities or auditors, and users soon came to doubt that the underlying trust was fully-funded, or even whether it existed. Maybe it does; maybe it doesn't. In any case, the doubt has forced users to price this uncertainty; and the product has ceased to function as a USD proxy.

Bitcoin

- The Bitcoin Cash fork

In this case, the fault was with the one authority that Bitcoin users do depend on: the managers of the source-code. Since there is no mechanism for determining and imposing a just settlement with respect to a design modification that was not backwards-compatible, certain users were able to refuse to participate in the update and the marketplace split. Imagine if Bitcoin were, in fact, a currency, and you had taken out two loans. One lender decided to stay with the legacy Bitcoin and another decided to migrate to the updated Bitcoin...

The lesson: computer scientists and engineers will continue to improve database technology and the functionality of virtual materialization for its own sake. But in finance this may just be a solution in search of a problem.

Big Data

The Era of Distributed Computing

Let's talk about Big Data.

In early computers, processors were limited to 4-byte words and 32-bit addresses. So

- the maximum useable memory was 16 GB.

Additional limitations of certain file systems led to an

- upper bound of 4 GB on any file

for many scientific computing users.

This forced scientists working in data-intensive fields or with simulations to invent ways of distributing simulation modeling or the statistical analysis of large datasets across networks of computers, which is the hallmark of "big data" in my opinion.

Big Data

The Era of Distributed Computing

In the era of 64-bit architectures, available commercially since the mid-90's, our only practical limitation in working with large datasets is CPU time.

- The era of big data in the sense of distributed computing for statistical analysis is over.

Distributed Computing Today

Distribution is still employed for speeding-up simulation models and for querying large datasets efficiently (such as web searches).

But one rarely encounters distributed computing anymore in a statistical or data analytics context.

Big Data

The nature of statistical problems

I believe the reasons for this is that models of interest in finance either involve

- exponential families,

in which case the Rao-Blackwell theorem allows for a significant simplification of the estimation procedure; or they involve

- hidden states

and are subject to the curse of dimensionality.

In the former case, distributed analysis (aside from trivial parallelization) is typically unnecessary.

In the latter, it is futile.

Big Data

Rao-Blackwell theorem

The Rao-Blackwell theorem says that any statistical estimator can be improved by taking a conditional expectation with respect to a sufficient statistic. For data modeled by a distribution in the exponential family, the count of sufficient statistics is guaranteed to be finite in the limit of sample size N.

In practice, this means that almost all parametric models can be estimated from data by first calculating a relatively small number of sufficient statistics and then modeling those. And the complexity of the latter task is independent of the sample size. For a very large and complicated dataset, one may want to consider map-reduce methods for distributing the calculation of the sufficient statistics, but once values have been obtained the remainder of the analysis can usually be easily completed on a single CPU.

Big Data

Hidden States

What about data that does not fit within a parametric model based on the exponential family?

non-Markovian Processes

An example we see often is finance is where the distribution of process increments is dependent on its own history. Such "non-Markovian" processes may arise from low-dimensional hidden-factor models, such as GARCH, but the un-conditional distribution cannot be distilled down to sufficient statistics in terms of the observations. So then we are left with analyzing conditional distributions, which means that rather than having a single sample with N points we have N samples with one point each.

So we have to start making assumptions either about the presence and nature of the hidden factors or about which features of the data we can ignore.

Roles for Quants

History

According to Harvey Stein at Bloomberg, the history of the quants begins with the Berlin Wall and Tiananmen Square events in 1989 closing off academic and research career opportunities for a generation of Western STEM graduate students.

1980's The pioneers: Equity options market-making.

1990's The wave: Interest rate derivatives.

2000's securitization (CDOs), derivatives counterparty credit risk (xVA), then liquidity risk and contagion networks as the mortgage crisis unfolded. By the end of decade, swaps and securitizations have largely disappeared.

2010's Everyone left in banking shifted into stress testing and model validation. By the end of the decade, banking regulators have backed away from stress testing; and regulated firms have largely outsourced model validation.

Roles for Quants

Present

Banking has significantly reduced its employment of quants. Non-banking (private equity, lending marketplaces, payment networks, central clearing, crypto assets) has picked up some of the slack

There are entry-level opportunities in

- performance and exposure measurement
- model library maintenance
- user support ("desk quants")
- documentation

As Greg Sobczak at Chicago Trading Corp. said at a student event a few years ago, "At the entry level, you either need to be able to write good papers, or you need to be able to write good code."