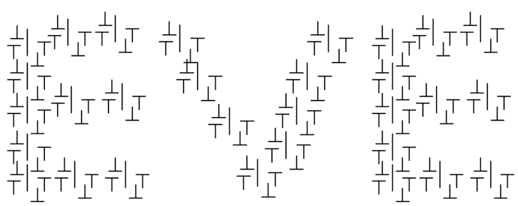
# Some XL Proposals to Help You Converge to a Better Statistical (Life) Philosophy

**A Conversation Between Xiao-Li Meng and Two Undergraduates**

*Jessica Hwang and Keli Liu* [1]

*Introductions are supposed to offer overviews, a big picture of sorts. But we thought it would be a bit tacky to offer up an 8.5×11in photo of Xiao-Li. So instead, we asked him to provide us with his Big Picture of statistics.*



*Sorry, Xiao-Li, we're not all as well versed as you in abstract art. You're going to have to provide a bit more explanation for what's going on.*

There are a few inference principles rooted in me. Whenever I see a new problem, I know it will reduce down to these principles somehow.

1. What's the optimal bias variance tradeoff?
2. What information is being lost?
3. What's the right conditioning?
4. What are the appropriate replications for evaluating inferential uncertainty?

*And somehow this all made it into the picture?*

Of course! But we're statisticians here. You'll need to do some inference, but you can use the interview for data.

*Seeing as how these four questions are rooted in him, it's no coincidence ($p < 0.05$) that our interview with Xiao-Li, conducted over Indian cuisine and (non-Indian) wine, would grow into this structure of its own accord. An interview is no different from any inference problem. Though a likelihood function for the true Xiao-Li Meng is beyond our computational limit (the question of how to implement approximate Bayesian computation for this problem is left open to further research/interviews), the non-random sample of advice, philosophizing, and jokes below allow for a design-based approach to inference on Xiao-Li and his Big Picture.*

## A Bias Variance Tradeoff for Statistics Education

*You have taught many undergraduates in courses like* Real Life Statistics: Your Chance for Happiness (or Misery). *Compare the undergraduate statistics curriculum here in the US to that in China.*

I can only speak of my experience more than 30 years ago. Back then, undergraduate education in China was much more specialized. You declared your major on the first day. I was in pure mathematics, and the only non-pure math course I took was *Mathematical Equations for Physics*, which was pretty much still pure math. It was all very strict, rigorous, and in-depth, but it was a narrow training. When I came to this country, I couldn't comprehend that undergrads were not admitted to a particular field. The liberal-arts concept took me a long time to appreciate.

Nowadays, China has also started thinking about general education. Fudan, where I did my undergrad, did not have the concept of a college. We had departments but no college. Now we have a college, and I heard that students receive general training for the first two years. In terms of pros and cons, when I came to Harvard, I was way more prepared than many other students in math, but I was seriously under-prepared for applied work.

*Don't you think it's easier to catch up in applied areas?*

Not necessarily. When I chose my first courses at Harvard, I asked one of my classmates from Fudan, who was already in US, for advice. He said, "Choose something easy and something hard." I followed that advice, except I got it completely

[1] We were undergraduate statistics students at Harvard University where we got to enjoy Xiao-Li's teachings, jokes, and wines (after turning 21 of course). In fact, we enjoyed the wine so much that we decided to use the excuse of studying statistics at Stanford University to get closer to Napa Valley.

Interviewers and the interviewee (and others) at another dining occasion (not the interview dinner). Xiao-Li and Professor Joe Blitzstein shares food, wine (but not bill) with undergraduate statistics students. Keli (front left) and Jessica (back left).

mixed up! I looked in the catalog, saw Probability Theory, and thought, “This has to be hard.” Instead the professor ended up replacing the TF's [teaching fellow's] solutions for the homework with mine. See, I wasn't just trying to get the right answer. Once you have serious mathematical training, you always try to give the most elegant, most compact answer. So he loved it. But the other course I took was Linear Regression, and that really gave me trouble. I got into it, thinking “How hard can it be to draw a line?” In China, we learned the least squares formula and how to derive it. But we never ran it, never learned how to draw residual plots, and that when the plots have a funny shape you need to transform the variables somehow. So I go “It's a funnel shape, so I need to take a log transformation”, but then of course, something else shows up, and it never stops! My homework was essentially a deck of computer output, with all the things that I tried. Augustine Kong, who taught the course, called me into his office and said, “Xiao-Li, explain to me what's going on here.” I had no idea what I was doing! Later, I learned that model evaluation and expansion is a really, really hard problem.

You see, whether theoretical or applied statistics is easier to pick up depends on one's previous training. I was trained as a pure mathematician, so anything with rigorous logic is easier for me to pick up. To this day, if you want me to learn something, if the thing has logic deep down, I may get stuck, but I won't feel like it's impossible for me. Since the applied part has more of an engineering or artistic aspect, I have no innate way of organizing my thoughts, so these are the hardest things for me. From this perspective, I appreciate my colleague Art Dempster's emphasis on distilling the logic behind common statistical practices.

*So what's the best system?*

I have no definitive answer. This is really a causal inference question. In the model I experienced, there are lots of things I should have learned but didn't. My hunch is that for most people, the better model is to start with more breadth. But you can find many examples where the top person in, say, mathematics focused solely on math, or the top person in physics was always a hard-line physicist. There's also the confounding factor that the best model may be non-stationary; the keys to success in the past might no longer work today. If we view education as having a T-shape, the question is, which T do we want?

*A T-shape?*

Yes, you need to build a base for breadth and a tower for depth. It would seem from an architectural standpoint that you should build the base first, to make a ⊥. But if you see the tower as a root instead, it can also be very sound structurally to first anchor into the ground, to make a T.

I'm actually fascinated by this question because it boils down to a bias variance tradeoff. Well, all problems do! With a broad base, you can view a real-life problem from all angles of the intellectual spectrum. To solve any hard problem that society faces takes more than one discipline's approach, so you incur severe bias from trying to pigeonhole it to your favorite toolkit. But without depth, you can't implement a meaningful solution. There's no precision in your efforts: it's too much hit or miss, reliant on chance, devoid of understanding. You have to be global in skill and thought, and yet be as familiar as a local would. So yes, the question really is, which T? How to strike a balance? You can even imagine that what we want isn't a T but a 工; this is the Chinese character for work. Actually, no matter which T, in the end, you still need to put in a lot of 工.

*Balance is the unifying principle behind T shaped education, but as students, we often get the suspicious feeling that balance is simply a buzzword that our elders like to throw around. What are we students supposed to do with it exactly?*

These days, career opportunities for the younger generation are much more diverse. You have a less clear idea of what you want to do, so you want to be more adaptive. When I went to college, I wanted to do mathematics, and that was the only thing I did. Only later in life did I realize that the fact I was missing everything else would hurt me. When you've learned something, even if you only do it once and completely forget about it, later it's much easier when you want to learn it again. That's just how the human brain works. Also, especially when you get older, there's a tendency to develop a "this is not my thing" mentality, so getting exposure to other topics at a young age serves as psychological preparation for learning later on.

*For statistics in particular, what do you think a modern curriculum should look like?*

As we get pressured to teach a lot of new stuff, I actually want to make sure that we still teach—maybe even more so—things like the likelihood principle and sufficiency reduction, the foundations and convey how these concepts continue to be relevant today. Especially since it is so easy nowadays to get *some* answer, we need to remember to ask whether the answer is any good. Do we know the limitations of the methodology we are using? What distinguishes the amateur from the expert is that the expert knows where the boundaries are: what is the best one can do, and what is the worst?

Another concept I would teach is principled corner-cutting. I'd like to teach a course that says, "Here is a real-life problem. If you have one week's worth of time, this is what you do. If you have one day, this is what you do." The less time you have, the more you cut, but you know exactly what you cut. Therefore, you can communicate to others why you did what you did, and how one could get more refined answers with more time or resources. It's like doing a Taylor expansion for real problems: get the linear term first, and if you have more time, get the quadratic term. Taylor expansion is about sorting out your priorities in life.

*Principled thinking has been a big emphasis of your research and teaching. We've both seen you extract some pretty deep principles from seemingly trivial mathematical expressions before. In fact, seeing you do it has helped to build our own understanding of what ``statistical thinking" is. For you, it was your advisor, Don Rubin, who embodied this statistical intuition. You've told stories where he knew you were wrong without even looking at your math. How does one go about developing this sort of intuition?*

It's a really gradual process, and maybe Don will be a bit surprised by how long it took! But the takeaway is that it eventually did take root. I'm a good example to show that intuition can be developed. I had a good analytical mind, but that doesn't mean I had a good intuitive mind. Now, like Don, I often can look at something and say, "This doesn't make sense," without carrying out the mathematics.

When I was learning those skills, I don't think I actively realized I was learning. It was only retrospectively that I realized, "Hmmm, now I can do what Don does." To me what this says is that an effective education is by doing. Don would ask "This doesn't make sense, right?" and he would guide me until I could see it too. I started thinking ahead and asking myself what he would be thinking. So if there is any teaching philosophy that can be abstracted here, it is teaching by doing—go through the thinking process with students and then let them practice a lot on their own. Nothing earth-shattering, but nevertheless effective, at least for someone like me.

*Is there anything you would change about the status quo? For example, everyone criticizes the p-value, yet it retains a hallowed status in the curriculum. What's up with that?*

The *p*-value is a perfectly fine statistical measure on its own. The problem is that it has been way overvalued. One the other hand, when something becomes as popular as *p*-values, it's inevitable that people will start to say there's something wrong with it. It's like when you declare that some actor or actress is the most beautiful in the world, someone will start thinking their ears are too big. This has also happened with bootstrap, multiple imputation, and the EM algorithm, so you have to put these things in perspective.

When teaching *p*-values, you want to peer at every level and see it inside-out. The beauty of statistics is that you can teach an entire course on one thing like *p*-values, and by the time someone learns this one thing, he or she will have learned a lot of other things, because everything is connected. The course could be titled "The value and the overvalue of the *p*-value".

*Any predictions for the future shape of statistics education?*

In any system reaching equilibrium, there's always compromise. This is the lesson of regression towards the mean. In the end, the right balance will likely be figured out not by any individual's wisdom, but by market forces balancing themselves out. We try, and err, and if we're too extreme, we'll regress towards the mean. Society is always evolving. If you think about MCMC, we're still in the warm-up phase.

It might also help to think about your question in miniature, or on the "nano"-scale. When designing the optimal exam question, one faces a nano-version of the breadth-depth tradeoff: the question should capture the big picture but also test for clarity of understanding regarding specific pixels. And of course it must meet the time constraints of the exam. My colleague Joe Blitzstein and I have discussed precisely this design issue (Blitzstein and Meng, 2010). Ironically, one of the sample exam problems on which we base our discussion —whether one can achieve an automated bias-variance tradeoff (one cannot)—illustrates my point exactly: there is no automatic way of finding the right T!

## Can Statisticians Transform without Information Loss?

*Do you think statistics is in need of a rebranding, in response to data science and machine learning?*

I haven't convinced myself either way. Recently I was asked to join a team of people who were writing an article arguing that we need to hire more people in data science. Ultimately I felt uncomfortable saying that we need data science without ever mentioning statistics. There were computer scientists and information scientists on this team, but none of them were arguing for their fields; they were all arguing for data science. So that made me pause, because I was more worried about statistics. Maybe there's a bigger picture for which we all need to unite, but if we do that, what's going to happen to statistics?

I have two thoughts, a selfish worry and a noble worry. My selfish worry is preserving my identity as a statistician. My more noble worry is, are we losing statistical thinking? Are concepts like conditioning, which we consider so fundamental, going to be marginalized? No pun intended. I understand why some philosophers get agitated, because philosophy used to be everything, but now you hear some people say, "Oh, here's a philosopher, they don't work on anything real."

There's a historical parallel too: one way to think about machine learning and data science is that they are spin-offs of statistics. But we ourselves are a spin-off of mathematics. Everyone talked about mathematical science, but a small group called themselves statisticians, and they grew and grew until they become us. Mathematicians must think, "These people never do serious and rigorous mathematics." But you also hear statisticians say that mathematicians don't deal with real-life problems, and that we have the answers because we have all these statistical principles. What I worry is, do the data scientists say, "These statisticians, hanging on these rigorous models, they don't understand real-life computations and complications"?

This history shows that we need to adapt to the new data environment in a proactive manner, but can we make this into an information-preserving transformation? My mathematical training makes me realize that the part of mathematics that doesn't work for us is such a small part. A statistician should never say, "Mathematics doesn't deal with real problems, so let's shut down the math department." That would be completely wrong. What we've done is to reorient mathematics towards practically relevant constructions and assumptions while keeping the logical thinking at its heart. How do we continue thinking about the Big Picture of statistics—such as bias variance tradeoff, minimizing information loss, conditioning, choosing the right replications—while revamping the parts of statistics inappropriate for modern applications?

*We're going to let you get away with answering our question with a question only because you've agreed to pay for dinner. Assuming that we've discovered this information-preserving transformation to apply to statisticians, how would you characterize the relationship between statistics and data science under this new regime?*

A very practical question is whether I should call myself a data scientist. There are two possible reactions to this: (1) sure, no brainer; (2) no you're not, because you don't have the computing skills that a data scientist needs. I'd agree on both counts. It depends on whether you view data science as an overarching theme or a professional label.

As for statisticians, we can take two rather different approaches. One is to protect our brand name rigorously, and not let people claim they are statisticians unless they go through regimented training and acquire a clearly defined set of skills. This is the route that lawyers have gone, with the bar exam. The other is to be generous and call any person who can do some statistics a statistician. This is another T-shaped argument: do you give your field a deep or broad definition?

To make an analogy to the medical profession, there are generalists and specialists. For a lot of problems, you don't need specialists because you have general practitioners. But for a difficult disease, you call the specialist. So maybe statisticians can be the specialists of quantitative science. Graduate training creates specialists; undergrad education creates generalists. No matter how many "data miners" there are, when they really can't solve a problem, they call the statistician. We want to be in that position.

*How do we get to be in that position?*

We can learn from the example of mathematicians. When the financial crisis came, in many places there was talk about shutting down departments. Statistics is one of those departments that could have been shut down. Indeed, I recall receiving petition letters from statistical colleagues in several state universities asking for moral support because their departments were in danger of being merged or shut down. But no major university would consider shutting down the math department. Why is that? What have mathematicians done to make themselves so secure?

One thing math has done, which we can mimic, is their introductory calculus course, which is considered so fundamental that it's required for students virtually in any scientific field. So that's the first step that we can do, should do, and are doing: promote proper statistical education at the undergrad level and even at the K-12 levels whenever appropriate and feasible, so that we get to the point where people say, "How could you not have taken any statistics?"

But what's the second step that makes the math department not just a teaching unit in the eyes of the public? If you think about famous math problems like the Goldbach conjecture and the Riemann hypothesis, most people have little clue regarding what they actually entail but are fascinated by them nonetheless because they've become an index of human IQ. They've defeated human intelligence for hundreds of years. So the math department is viewed as a collection of geniuses who tackle this frontier. We as a field need to communicate to the general population that statistics is also an incredibly intellectually challenging field. With all these things we do to clarify teaching and make statistics accessible, have we accidentally trivialized statistics? We need to do a better job to create a brand name that can attract the utmost intellectual talents to solve the hardest problems in statistics. If statistics is "easy to learn", it's not because it lacks depth, but because it's rooted in everyday practical problems.

*How does big data figure into our identity? How should we be thinking about big data?*

Big data is a very catchy name, but not a good scientific term. For people studying asymptotics, you'd think they must be really happy because with big data, now all of their theory can be applied, right? But of course we know that's not how things work. The issue isn't whether a dataset is big; "big" is a proxy for "complex".

A key step of big data is preprocessing, and this is what I've been focused on in my work with Alex Blocker. Fisher said that statistics is just data reduction, and preprocessing is really a problem of data reduction. So from that perspective, Fisher already envisioned everything. Of course, when Fisher thought about data reduction, it was with respect to a single model. The problem of preprocessing in the big data era is that preprocessors and data users are often separate parties with different resources and models. What does a sufficient reduction mean here? One can easily imagine "The Potential and Perils of Preprocessing" in this multi-phase context (Blocker and Meng, 2013).

When we work on these kinds of research problems, we have all our principles to guide us, but when you give me a gigabyte of data, I don't know what to do. I think we need to acknowledge that there needs to be someone else there. But we statisticians have thought for a century about how to reduce data. If someone can think of a good way to do data reduction, I'm happy to call them a statistician, whether they have systematic training or not —to me that is statistical thinking. We should be a part of the team that thinks about how to do data reduction in *practical and principled* ways.

*Going back to what you said about not knowing what to do with a gigabyte of data, CS people might just say, "We know what to do, we don't need you." Isn't it our job to learn these new skills?*

Ideally, yes, to protect our field, we should all grow computational skills. But of course this also applies to CS: they already know how to do the computation, so they can also learn more statistical principles. In fact, people in data mining and machine learning want students to do more statistics. You can "try" all your algorithms, but you need theory, including statistical theory, to understand the "why". Understanding "why" is especially critical when dealing with big and complex data, where one typically cannot afford to waste time trying things that either cannot work or are easily dominated by other methods. In fact, in the absence of principled inferential methods, more data can even worsen an estimator's performance, a phenomenon that I have investigated with Xianchao Xie. Big data should allow us to explore more complex questions, but if we're not principled, we could easily find ourselves in the situation of "I got more data, my model is more refined, but my estimator is getting worse!" (Meng and Xie, 2013).

Some scientists I work with love using the Cox proportional hazards model. Someone else could have discovered this model just by trying things out without deep understanding. But the beauty of Cox developing it is that he distilled the idea of partial likelihood behind it, which has led to an entire industry on Cox regression in terms of both theory and applications. The question is whether there is enough follow-up to know what things we try are just a waste of time and which are sustainable. I was on a hiring committee looking at two statistical geneticists: one person had developed lots of software, and the other focused on thinking about the deeper issues. You need both types of researchers. The first type conducts the frontal assault, then leaves to fight another battle. But the battle has not been won; we need the finishers to solidify the theory, which provides us with both scientific assurance and a roadmap for further methodological developments.

## Conditioning on Ancillarities: If It Seems Useless, Just REsearch It!

*What was the most useless thing you learned in grad school that turned out to be useful or remained useless?*

When my professors thought something was useless, they told it to me straight. Things like mean imputation and complete case analysis. But fiducial inference also got lumped in there! I was told that fiducial is completely wrong, but now I believe there is something deep about it. That will probably be the most controversial statement in this interview.

For most things that I initially thought were useless, I tried to convince myself that there must be some deep reason for their existence. It's an obvious lesson to draw from the story of ancillary statistics. How can a statistic whose distribution is free of the parameter possibly inform our inference? It can't! Not directly at least. But indirectly, by conditioning on it, we can recover useful information. It's the same with research. You think something's useless, but then later on you come back to it when you have a deeper understanding of it, because everything's connected, and it gives you a back door to what you want. You can cite this as my backdoor criterion.

I was once convinced that sharp null hypotheses are useless. But working with Augie Kong in statistical genetics, I saw that you do have point nulls which correspond to certain biological laws, such as Mendel's laws. At one point, I also bought the argument that minimaxity was completely stupid, but now I've seen situations where people do all sorts of crazy things with no regard for principle. You've got to give people some minimal constraint. Minimaxity does this by saying that you at least have to minimize the worst thing, and it is one of the very few principles that can be mathematically formulated in almost any situation.

There is actually a very important lesson here, statistically, and for your statistical education. Suppose that your "parameter of interest" is a full understanding of statistics. What you already know about statistics is your "observed data", while everything you don't know about statistics or don't fully grasp constitutes "missing data". What missing data should you augment your observed data with? On the one hand, there are topics you don't know and which you believe are important to know; you consider these to be a sufficient statistic for the parameter of interest. On the other hand, there are topics you don't know and don't really want to know; you consider these to be ancillary. Which data augmentation scheme is better, the sufficient or ancillary augmentation? As you can guess, unless you're omniscient, there will always be times when what you believe to be useless turns out to be a gem. We need to protect ourselves from our own hubris. Does this mean that for every three topics we pick from our "useful" list, we should work on something we believe to be useless? Incidentally and interestingly enough, my work with Yaming Yu and Xiaojin Xu in the context of MCMC (Yu and Meng, 2011; Xu et al., 2013) has shown that the best strategy is not to alternate but to interweave. That is to say, leverage what you believe to be useful to discover the value in what you currently see as useless. Build connections.

*You started your career with computation, but now you're working more on foundational issues. Was this a random walk, or was there a plan?*

I would say it was a random walk with a strategically placed starting point. Not by me, but by Don Rubin. I think I would be a very different researcher today if I had been at a different university and was given something to prove about UMVUEs. In statistics, the types of problems that students can work on are: theory, methodology, computation, and application. Can you guess, if a student comes to me and says "I'm ready to work but have no idea what I want to work on, give me something," which one I would want them to work on?

*Nope.*

The type I would start with is computation. The reason is very simple. When a student has never done any research, it's like a first-time fishing trip. The thing you want to do is to get them somewhere where they can immediately get some fish. You don't want them to stay all day and catch no fish; then they're gone and won't be hooked. Starting with theory is typically risky unless the student is really good at theory, because the student tends to get very frustrated and discouraged when the student can't prove or develop anything. As for methodology, for someone who's starting out, coming up with a good methodology that no one has thought of is very difficult. To really make a contribution in applied statistics, the student has to learn the language of the field of application. But for computation, one can always run a simulation: change the model, make the assumptions not fit, and see what happens. The student will immediately have some output in their hand. You get some fish right away. It may be ugly and occasionally it may even be a crab (yes, I caught a crab once on my fishing rod), but at least you got something moving. Don knew I had a strong mathematical foundation but was completely clueless about statistics. I had never even seen a histogram! So he gave me the EM algorithm, because I could run something, do something.

But from there, it was like a random walk. Once I started working on EM, I naturally got into missing data. One day Don was talking about multiple imputation and hypothesis testing, and by that time I'd become a little bit braver. I looked at one hypothesis testing procedure and said "Professor Rubin, I can do better than that." And he said, "Try." I went back, and he said "Oh yeah, you can get something!" That became my qualifying paper. I also worked with Professor Shaw-Hwa Lo on Kaplan-Meier (K-M) estimators and K-M processes. These all seem random, but when I was trying to put together a thesis, I needed a common theme and I wanted a fun title, so I called it "Toward complete results for incomplete-data problems".

Since EM is an iterative algorithm, it was very natural for me to work on the Gibbs sampler, data augmentation, and MCMC, and then bridge sampling and path sampling. From the MI side, a controversy arose about the Rubin variance estimator not being consistent, which got me thinking about modeling issues and uncongeniality, which then led me to Bayesian modeling, and now multi-phase inference. There's that saying, "From nothing, nothing comes." But for me, it's really been "From randomness and missingness, a lot of stuff comes." In fact, there's a great lesson here for multi-resolution inference: if it looks random or useless, then you have to look deeper. Noise at one resolution level becomes signal at another.

*We heard that you have this secret archive of research ideas stashed away somewhere in Harvard's underground tunnels. Can you give us a sneak peek?*

I don't actually have a physical file. Andrew Gelman and I used to keep the "Monica File". Don't ask me or Andy why we called it the Monica File, because we promised each other we would never let the other reveal the secret. We'd get together and talk about the problems we wanted to work on and he'd add it to the file. But we haven't talked in a while so I don't know what happened to the Monica File. What I do have is a list of titles for all the papers that I want to write and the list is still growing.

*You haven't written the papers but you already know the titles? Isn't that sort of ancillary to the content?*

I always write the titles first, though I often refine them as I move along. Then conditional on that, I write the paper. When I wrote my paper commemorating the 50th anniversary of COPSS, my goal was to lure the brightest minds to statistics by showing them the intellectual richness of our hardest questions. I had three large classes of delicious questions in mind and we all know that a Nobel Prize is a good attention grabber, so a natural choice was to call it "A trio of inference problems that could win you a Nobel Prize in statistics (if you help fund it)" (Meng, 2014). It's amazing how well the real world fits the titles I make for it—I say this only half jokingly. To share with you just a sample of this sample platter of open research questions:

- Multi-Resolution Inference. This relates to the desire for personalized medicine. What happens when our data resolution is lower than the resolution of our estimand? How far above the data resolution should we try to estimate?
- Multi-Phase Inference. How can we clean/process raw data to preserve as much statistical information as possible while also ensuring computational efficiency? This is a huge problem because upstream processing can have severe impacts on downstream analysis.
- Multi-Source Inference. We now have large administrative datasets not collected for inference purposes at all. There was no random sampling. Which is better: a 85% non-random sample or 5% random sample? It took decades for the merit of random samples to take root. During the time of Laplace, people thought that was a crazy idea. We're in the opposite situation today. We have data that almost resembles a census, but is susceptible to all sorts of selection bias. How do we think about these problems? I think this is one of the major areas where the next revolution in statistics will come.

Of course, I also have some non-trio-themed antipasti to whet your intellectual appetites. One thing I've noticed is that since everything's connected, there's no such thing as a bad flavor combination, so feel free to mix and match to your heart's (or more appropriately mind's) desire.

- Partial Bayes. Let's say you have some information but not enough to put a prior on everything. What are principles for partial Bayes methods? What are the analogous partial risk calculations?
- Return of Robustness. A big concern now is confidentiality. How do we protect data yet preserve the information inside? We want to be able to purposely inject noise, yet still extract the relevant signal. This is precisely the purpose of robust procedures. So much work is needed to understand, say, when a sample median beats a sample mean for estimating the population mean using data that have been randomly "de-classified" .
- Index of Non-Parametricity. My former student Paul Baines and I have found an example where the MLE, $L_2$ regression, $L_1$ regression, Cox regression and quantile regression all estimate the same exact parameter. This allows us to measure how non-parametric each of these methods are. But we still need to figure out how to do this with some generality.
- Regression towards the mean is an $L_2$ phenomenon. I want to study the entire $L_p$ family. Is there a concept called regression towards the median? Why or why not? $L_2$ is the only Hilbert space in this class. What's so fundamental, statistically, about being a Hilbert space?
- Non-Markovian Monte Carlo. There are all these adaptive MCMC procedures. I have a problem with the term "adaptive MCMC". When we move from i.i.d. to MCMC sampling, we don't call it "adaptive rejection sampling", which if you think about it, perfectly

describes Metropolis-Hastings. We think it's a big conceptual advance, so we grant it a new name. By giving up the i.i.d structure and using one-step dependence, we can open up a door and do a lot more. Adaptive MCMC says to give up one-step dependence and use the entire history, that is, to give up the Markovian property altogether. So why don't we call it Non-Markovian Monte Carlo? Why stick with the narrow connotations of modified MCMC when we have something fundamentally different? Of course it is much harder, at least for developing theory, but that's called research and good progress has been already made.

- Bayesian Analysis for Finite Population Inference. Putting down a prior on individual parameters is okay, but if the entire population is unknown and we put priors on individual units, we will run into lots of paradoxes. It's sort of like the Neyman-Scott problem. When does the likelihood principle fall apart? I used to believe that a simple information argument would allow us to understand this and that Bayes with some proper prior would fix the problem. But now I'm starting to doubt that. Of course, this gets me very worried because I'm the one talking about principled corner-cutting, right? So if there's no existing principle for me to hold onto, then I have to work one out.

*For some of these problems, the mathematics seem quite intimidating. What's your advice to students who want to work on them?*

The math is often not the key thing. It's really to formulate the problem in such a way so that mathematics can actually be applied. That's the hardest part. I've been trying to formulate multi-resolution inference by borrowing strategies from the wavelets literature, such as the concept of primary resolution. I wanted to define the primary resolution for analysis but I couldn't do it. So how did I move on? I said, since I'm not ready to give a precise definition, I'll take a page out of Fisher's book. He'd always say something like "Without attaching specific meaning to this thing, here's what I'm going to do with it." And I can see why Fisher did that. He had a good feeling that something had to happen. He couldn't do what he wanted with mathematical exactness but he didn't let that stop him, because he knew that what he could do had statistical meaning and exactness. Of course, ultimately the right math will come (and needs to come), after one crystallizes the statistical idea one wants to develop. As you know, I now have a better mathematical description of multi-resolution than when I started, but of course there is still a long way to go.

*We've heard that you have a special attachment to if-and-only-if statements. Why is this? Would you require answers to the above questions in if-and-only-if format?*

If and only if you want true understanding. When you are trained as a mathematician, the best type of result is if-and-only-if. That tells you that you've considered every possibility. It's like when you go to court and they ask you to tell the truth, the whole truth, and nothing but the truth. If I can prove something if-and-only-if, I feel fundamentally satisfied, because I have nailed it down.

There are people who say that all these if-and-only-if results are tautological, by definition. If two statements are completely equivalent and you understand one, then you understand the other, right? But that's not true. If-and-only-if thinking is how humans evolve. To understand something that we cannot think about intuitively, we invent a tool, such as mathematics, that allows us to think *otherwise*. By believing the logic of mathematics, we can transfer the thing we couldn't grasp into something that allows us to build intuition. What's remarkable is that we've invented this tool to convince ourselves. If-and-only-if is just so fundamental to my way of thinking. And it's one of those ways of thinking that I don't think you can pick up by just taking a course or two in calculus.

## The Key to Work-Life Balance?: Finding the Right Replication

*At the 2013 JSM, you were on a panel hosted by COPSS that was aimed at junior researchers. One of the topics that came up was work-life balance. What advice would you give to young statisticians about work-life balance?*

At different stages of my life, I have had different balances. After we had kids, my wife and I were incredibly busy, and I recall asking my wife, "What were we doing back then when it was only the two of us?" We must have had tons of time! But we didn't see it that way.

How do you psychologically justify to yourself that you have work-life balance? Part of work-life balance is out of your control. Of course, in the grand scheme of things, everything is under your control, since you can just not do anything. But in real life we have to worry about things such as outside expectations, our perceived duties to others. You internalize these outside expectations and, in doing so, turn "work" into life. Then there's no way to go back.

For example, shortly after I was appointed as the department chair, I also agreed to be a co-editor of *Statistica Sinica*. Some people thought that I was insane or that I was driven by a desire for another title. But, actually, I initially said no to the invitation as well as to another editorship from a larger journal. I eventually said yes to *Sinica* because I believed it would be more manageable timewise and because I learned that the available pool of people who could and were willing to serve as *Sinica* editor at that time was much smaller than those who could and were willing to serve for the other journal. Hence, I felt that my contributions would be more valuable as a co-editor for *Sinica* than for the other journal. I saw the *Sinica* editorship not as a burden but as a duty. Duties are part of our lives. At least that's how I rationalized my decision and found a psychological balance in myself.

Time then passed quickly and I was ready to step down as *Sinica* co-editor. And I still remember that the idea of retirement excited me because I thought I was going to have so much more time, since I had been spending 3-4 hours a day on editorial work. But not even a week later, I found myself completely back to my full work schedule!

*What happened?*

That's exactly what I asked! But it's very simple. It's like all the water is waiting outside, and as soon as you have some vacuum, the water—pfooosh!—rushes back in. It took me several years to reach this point, but it's good psychological conditioning: you have to realize that you'll never be able to finish everything before you move onto the next thing. You're always going to be multitasking, you're always going to be behind. So you just change your mindset. You no longer think about projects as your replications; instead, you think about days as your replications: *I finished another day. I finished another day.*

Incidentally, we encounter this problem in statistics all the time, where a huge challenge is finding the right replication with which to evaluate our procedure. Bayesians complain that Frequentist replications are irrelevant to the data at hand; Frequentists counter that Bayesian replications are synthetic, dependent on prior choice. I'm hoping that finding the right replication for life should give us hope that we can one day reconcile Frequency and Bayes, since that's a simpler problem, right? If projects are your replication, you'll think, "Look at how many projects I didn't get done!" You feel like you never finish anything. But you'll always finish the day, day after day after day …

And thank god there's no negative time. No matter how busy you are, you only have 24 hours. You can sleep less, but it's still 24 hours. So the fact that all 24 hours are filled means that you have to set priorities.

*How should young statisticians set these priorities?*

It's not necessarily a balance of doing one thing versus another, but it's a psychological balance. This is what I tell young faculty: you'll struggle with balancing research, teaching, and service to your department, and you'll be thinking, I'm doing too much of one thing and too little of another. But what you have to do is to put them on an equal footing. If you are doing research on three projects, they obviously compete with each other for your time, but you complain less when you want to work on all of them because you think they're all important. The same thing goes for teaching: think of teaching as another research project. As you know, in our work together, it was a teaching challenge—can we explain Simpson's Paradox in terms of the simple concept of comparing apples and oranges?—which led us to a deeper understanding of multi-resolution inference and the appropriate resolution for conditioning. We can now claim to have a "Fruitful Resolution to Simpson's Paradox" (Liu and Meng, 2014), but this tree was not watered by research in the narrow sense.

Don't think that you shouldn't be teaching and should be doing research instead—every act you choose to dedicate yourself to or choose to neglect comes to define your life as a scholar and, more fundamentally, as a person. It's cliche, but if you think of your life as a book, then every day is a new page, and you don't want every page to be the same. A novel is a more intriguing read when the protagonist is multifaceted.

## Time to Call It an Eve

*Can we go back to the picture on the first page for a moment? We feel like your meaning is non-identified given just the data from the interview. How about some prior information?*

I wouldn't say that it's non-identified. Rather, I wanted to give a real life inference problem where the MLE is not unique. As for prior information, that's my favorite equation in statistics, EVE's law [law of total variance], which is precisely about the bias variance tradeoff!

$$\mathrm{V}(Y) = \mathbf{E}[\mathbf{V}(Y|X)] + \mathbf{V}[\mathbf{E}(Y|X)]$$

*Shouldn't it be EVVE then?*

Well, there was some information loss to make the acronym fit the name. Remember principled corner-cutting? Whoever came up with that acronym must understand it well! It was actually quite labor-intensive to create each letter, so removing that extra V generated a ton of computational savings. This is the type of statistical thinking we need for Big Data**!**

*Shall we bid each other a good eve then?*

But you haven't even asked me the secret to eternal youth yet!

*Oh, well, we assumed that your answer would just be wine. Were we wrong?*

Only partly. You definitely need wine! But if you drink by yourself then all you have is a drinking problem and not the elixir of life. What matters is who you share the wine with. I've been blessed to be able to dine and wine with hundreds of young talents around the globe, including virtually all my students and young colleagues. Their energy and excitement beat out the antioxidants in wine any day, and their inspirations and aspirations are simply intoxicating! When I was at Chicago, I was fortunate enough to get some advice from Professor George Tiao. He said, ``Xiao-Li, you keep yourself young by always surrounding yourself with young people.'' He told me that more than 20 years ago, and I cannot say I fully appreciated the wisdom then. But it's wisdom that's aged like a fine wine.

So cheers and thank you, my fountain of youth!

## Bibliography


Blitzstein, J. and X.-L. Meng (2010). Nano-project qualifying exam process: An intensified dialogue between students and faculty. *The American Statistician 64*(4), 282--290.

Blocker, A. W. and X.-L. Meng (2013). The potential and perils of preprocessing: Building new foundations. *Bernoulli 19*(4), 1176--1211.

Liu, K. and X.-L. Meng (2014). Comment: A fruitful resolution to Simpson's Paradox via multiresolution inference. *The American Statistician 68*(1), 17--29.

Meng, X.-L. (2014). A trio of inference problems that could win you a Nobel Prize in statistics (if you help fund it). In X. Lin, D. L. Banks, C. Genest, G. Molenberghs, D. W. Scott, and J.-L. Wang (Eds.), *Past, Present and Future of Statistical Science*, pp. 535--560. CRC Press.

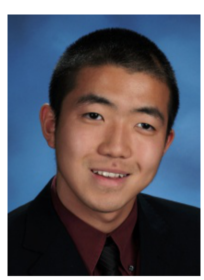
Meng, X.-L. and X. Xie (2013). I got more data, my model is more refined, but my estimator is getting worse! Am I just dumb? (with discussions). *Econometric Reviews 33*, 218--250.

Xu, X., X.-L. Meng, and Y. Yu (2013). Thank god that regressing Y on X is not the same as regressing X on Y: Direct and indirect residual augmentations. *Journal of Computational and Graphical Statistics 22*(3), 598--622.

Yu, Y. and X.-L. Meng (2011). To center or not to center: That is not the question---an Ancillarity--Sufficiency Interweaving Strategy (ASIS) for boosting MCMC efficiency (with discussions). *Journal of Computational and Graphical Statistics 20*(3), 531--570.



*Jessica Hwang*
*PhD Student*
*Department of Statistics*
*Stanford University*
jjhwang@stanford.edu


*Keli Liu*
*PhD Student*
*Department of Statistics*
*Stanford University*
keli.liu25@gmail.com