Seminar, so hello everyone.

My name is Qingyuan Zhao, I'm currently a University Lecturer in Statistics in University of Cambridge.

I visited Yale Biostats briefly last year in February. And so it’s nice to see every guest very shortly this time.

And today I’ll talk about sensitivity analysis for observational studies, looking back and moving forward. So this is based on ongoing work with several people Bo Zhang, Ting Ye and Dylan Small at University of Pennsylvania, and also Joe Hogan at Brown University.

So sensitivity analysis is really a very broad term. For example, it can be any factor that can be changed in a model prior to its execution. So one example is structural.
or epistemic sources of uncertainty.
And this is sort of the things we’ll talk about.
So basically, what our talk about today is those things that we don’t really know.
I mean, we made a lot of assumptions about when proposing such a model.
So in the context of observational studies, a very common and typical question that requires sensitivity analysis is the following.
How do the qualitative and or the quantitative conclusions of the observational study change if the no unmeasured confounding assumption is violated?
So this is really common because essentially, in the vast majority of observational studies, it’s essential to assume this no unmeasured confounding assumption, and this is an assumption that we cannot test with empirical data, at least with just observational data.
So any, if you do any observational studies, you’re almost bound to be asked this question, that, what if this assumption doesn’t hold?
And I’d like to point out that this question fundamentally connected to missing not at random in the missing data literature.
So what I will do today is I’ll focus on sensitivity analysis for observational studies, but a lot of the ideas are drawn from the missing data literature.
And most of the ideas that I’ll talk about
Today can be also applied there and to related problems as well. So, currently, a state of the art of sensitivity analysis for observational studies is the following. There are many, many masters gazillions of methods of exaggeration, but certainly many methods that are specifically designed for different kinds of sensitivity analysis. It often also depends on how you analyze your data under unmeasured confounding assumption. There are various forms of statistical guarantees that have been proposed. And oftentimes, these methods are not always straightforward to interpret, at least for inexperienced researchers, it can be quite complicated and confusing. The goal of this talk is to give you a high level overview. So this is not a talk where I’m gonna unveil a lot of new methods. This is more of an overview kind of talk that just to try to go through some of the main ideas in this area. So in particular, what I wanted to address is the following two questions. What is the common structure behind all these sensitivity analysis methods? And what are some good principles and ideas we should follow? and perhaps extend when we have similar problems? The perspective of this talk will be global and frequentist. By that, I mean,
there’s an area in sensitivity analysis
called local sensitivity analysis,
where you’re only allowed to move your parameter
near its maximum likelihood estimate, usually.
But global sensitivity analysis refer to the method
that you can model your sensitivity parameter
freely in a space.
So that’s what we’ll focus on today.
I’ll take a frequentist perspective. So I won’t talk about Bayesian sensitivity analysis,
which is also a big area.
I’ll use this portal typical setup
in observational studies,
where you have iid copies of these observed data O,
which has three parts, x is the covariance,
A the binary treatment, Y is the outcome
and these observed data
that come from underlying full data, F,
which includes X and A
and the potential outcomes, Y(0) and Y(1).
Okay, so this is,
if you haven’t, if most of you probably have seen this
many, many times already,
but if you haven’t seen that this
is the most typical setup in observational studies.
And it kind of gets a little bit boring
when you see it so many times.
But what we’re trying to do
is to use this as the simplest example,
to demonstrate the structure and ideas.
And hopefully, if you understand these good ideas,
you can apply them to your problems that are maybe slightly more complicated than this. So here's the outline and I'll give a motivating example then I'll talk about three components in the sensitivity analysis. There the sensitivity model, the statistical inference and the interpretation. So the motivating example will sort of demonstrate where these three components come from. So this example is in the social sciences actually it's about child soldiering, a paper by Blattman and Annan, 2010. On the review of economics and statistics, so what they studied is this period of time in Uganda, from 1995 to 2004, where there was a civil war and about 60,000 to 80,000 youth were abducted by a rebel force. So the question is, what is the impact of child soldiering sort of this abduction by the rebel force, as on various outcomes, such as years of education, and in this paper to actually study the number of outcomes. The authors controlled for a variety of baseline covariates, like the children’s age, their household size, their parental education, et cetera. They were quite concerned about this possible unmeasured confounder.
That is the child’s ability to hide from the rebel. So it’s possible that maybe if this child is smart, and if he knows that he or she knows how to hide from the rebel, then he’s less likely to be abducted. And he’ll probably also be more likely to receive longer education just because maybe the skin is a bit more small, let’s say.

So in their analysis, they follow the model proposed by Imbens, which is the following.

So basically, they assume this no unmeasured confounding after you conditional on this unmeasured confounder $U$. Okay, so $X$ are all covariates that $U$ controlled for, and $U$ is they assumed is a binary, unmeasured confounder. That’s just a coin flip. And then they assume the logistic model for the probability of being abducted and the normal linear model for the potential outcomes. So notice that here the linear these terms depends on not only the observed covariance, but also the unmeasured covariates $U$. And of course, we don’t measure this $U$. So we cannot directly fit these models. But what they did is they because they made some distribution assumptions on $U$. 
you can treat U as unmeasured variable. And then, for example, fit maximum likelihood estimate. So they’re treated these two parameters lambda and delta, as sensitivity parameters. So these are the parameters that you vary in a sensitivity analysis. So when they’re both equal to zero, that means that there is no unmeasured confounding. So you can actually just ignore this confounder U. So it corresponds to your primary analysis, but in a sensitivity analysis, you change the values of lambda and U and you see how that changes your result above this parameter beta, which is interpreted as a causal effect. Okay, so the results can be summarized in this one slide. I mean they’ve done a lot more definitely. But for the purpose of this talk, basically, what they found is that the primary analysis found that the average treatment effect is -0.76. So remember the outcome was years of education. So being abducted, has a significant negative effect on education. And then it did a sensitivity analysis, which can be summarized in this calibration plot. What is shown here is that these two axis are basically the two sensitivity parameters, lambda and delta. So what the paper did is they transform it
to the increase in R-squared. But that’s that can be mapped to lambda and delta, and then they compared this curve, so this dashed curve is where the values of lambda and delta such that the treatment in fact is reduced by half. And then they compare this curve with all the measured confounders, like year and a location, location of birth, year of birth, et cetera. And then you compare it with the corresponding coefficients of those variables in the model and they just plot these in the same figure. What is supposed to show is that look, this is the point where the treatment effect is reduced by half, and this is about the same strength as location or birth alone. So, if you think your unmeasured confounder is in some sense as strong as the location or the year of birth, then it is possible that the treatment infact is half of what it is estimated to be. Okay, so it’s a pretty neat way to present a sensitivity analysis. So in this example, you see, there’s three components of sensitivity analysis. First is model augmentation. And you need to expand the model used by primary analysis to allow for unmeasured confounding.
Second, you need to do statistical inference. So you vary the sensitivity parameter, estimate the effect, and then control some statistical errors. So what they did is, they essentially varied lambda and delta, and they estimated the average treatment effect under that lambda and delta. And the third component is to interpret the results. So this paper relied on that calibration plot for that purpose. But this is often quite a tricky because the sensitivity analysis is complicated as we need to probe different directions of unmeasured confounding. So the interpretation is actually not always straightforward and sometimes can be quite complicated. There did you have there do exist two issues with this analysis. So this is just the model and rewriting it. The first issue is that actually the sensitivity parameters lambda and delta, where we vary in a sensitivity analysis are identifiable from the observed data. This is because this is a perfect parametric model. And then it’s not constructed in any way so that these lambda and delta are not identifiable. In fact, in the next slide, I’m going to show you some empirical evidence that you can actually estimate these two parameters.
So, logically it is inconsistent for us to vary the sensitivity parameter. Because if we truly believe in this model and the data actually tell us what the values of lambda and delta is. So this is the similar criticism that for Hattman selection model, for example. The second issue is a bit subtle is that in a calibration plot, what they did is they use the partial R squared as a way to measure lambda and delta in a more interpretable way. But actually the partial R squared for the observed and unobserved confounders are not directly comparable. This is because they’re they use different reference model to start with. So, actually you need to be quite careful about these interpretation this calibration quotes. So, here is what I promised that suggests you can actually identify these two sensitivity parameters lambda and delta. So here the red dots are the maximum likelihood estimators. And then these solid curves this regions, or the rejection, or I should say acceptance region for the likelihood ratio test. So this is at level 0.50, this is 0.10, this is 0.05. There is a symmetry around the origin that’s
0:16:34 –> 0:16:37.27 because the U number is symmetric.
0:16:37.27 –> 0:16:40.68 So, lambda like delta is the same
0:16:40.68 –> 0:16:43.024 as minus lambda minus delta.
0:16:43.024 –> 0:16:44.47 But what you see
0:16:44.47 –> 0:16:47.13 is that you can actually estimate lambda and delta
0:16:47.13 –> 0:16:49.78 and you can sort of estimate it
0:16:49.78 –> 0:16:53.05 to be in a certain region.
0:16:53.05 –> 0:16:55.62 So, something a bit interesting here
0:16:55.62 –> 0:17:00.62 is that there’s more you can say about Delta,
0:17:01.05 –> 0:17:03 which is the parameter for the outcome,
0:17:04.059 –> 0:17:06.827 than the parameter for the treatment lambda.
0:17:07.13 –> 0:17:09.12 But in any case,
0:17:09.12 –> 0:17:10.64 it didn’t look like we can just vary
0:17:10.64 –> 0:17:12.79 this parameter lambda delta freely in this space
0:17:12.79 –> 0:17:16.03 and then expect to get different results
0:17:16.03 –> 0:17:18.719 for each each point.
0:17:18.719 –> 0:17:22.51 What we actually can get is some estimate
0:17:22.51 –> 0:17:24.91 of this sensitivity parameters.
0:17:27.92 –> 0:17:30.1 So the lesson here is that
0:17:30.1 –> 0:17:32.48 if you use a parametric sensitivity models,
0:17:32.48 –> 0:17:34.9 then they need to be carefully constructed
0:17:34.9 –> 0:17:37.143 to avoid these kind of issues.
0:17:40.32 –> 0:17:42.76 So next I’ll talk about the first component
0:17:42.76 –> 0:17:44.43 of the sensitivity analysis,
0:17:44.43 –> 0:17:46.693 which is your sensitivity model.
0:17:47.75 –> 0:17:50.68 So very generally,
0:17:50.68 –> 0:17:53.56 if you think about what is the sensitivity model,
0:17:53.56 –> 0:17:58.56 is essentially it’s a model for the full data F,
0:18:00.27 –> 0:18:03.14 that include some things that are not observed.
0:18:03.14 –> 0:18:04.78 So, what we are trying to do here
0:18:04.78 –> 0:18:07.65 is to infer the full data distribution
from some observed data, \( O \).

So a sensitivity model is basically a family of distributions of the full data, is parameterized by two parameters \( \theta \) and \( \eta \).

So, I’m using \( \eta \) to stand for the sensitivity parameters and \( \theta \) is some other parameters that parameterize the distribution.

So the sensitivity model needs to satisfy two properties. So first of all, if we set the sensitivity parameter \( \eta \) to be equal to zero, then that should correspond to our primary analysis assuming no unmeasured confounders.

So I call this augmentation. A second property is that given the value of the sensitivity prior to \( \eta \), then we can actually identify this parameters data from the observed data.

So this is sort of a minimal assumption. Otherwise, this model is simply too rich, and so I call model identifiability. So the statistical problem in sensitivity analysis is that if I give you the value of \( \eta \) or the range of \( \eta \), can you use observed data to make inference about some causal parameter that is a function of the \( \theta \) and \( \eta \).

Okay, so this is a very general abstraction. But it’s a bit too general.
by understanding these observational equivalence causes.

So essentially, what we’re trying to do is we observe some data, but then we know there’s an underlying full data and instead of just modeling the observed data, we’re modeling the full data set. So that makes our model quite rich, because we’re modeling something that are all observed. For that purpose is useful to define this observationally equivalence relation between two full data distribution, which just means that their implied observed data distributions are exactly the same. So we write this as this approximate equal to this equivalence symbol. So then we can define the equivalence class of a distribution of a full data distribution, which are all the other full data distributions in this family that are observationally equivalent to that distribution. Then we can sort of classify these sensitivity models based on the behavior of these equivalence classes. So, what happened in the last example is that the full data distribution full data model is not rich enough. So these equivalence classes are just singleton’s and can actually identify the sensitivity parameter eta from the observed data. So, this makes this model testable in some sense with the choice of sensitivity parameter testable,
and this should generally be avoided in practice. Then there are the global sensitivity models where you can basically freely vary the sensitivity parameter \( \eta \). And for any \( \eta \) you can always find the \( \theta \) such that it is observational equivalent to where you started from.

And then even nicer models the separable model where basically, this \( \eta \), the sensitivity parameter doesn’t change the observation of the observed data distribution. So for any \( \theta \) and \( \eta \), \( \theta \) and \( \eta \) is equivalent to \( \theta \) and zero.

These are really nice models to work with.

So understand the difference between global models and separable models. Basically, it’s just that they have different shapes of the equivalence classes. So for separable models, these equivalence classes, needs to be perpendicular to the \( \theta \) axis. But that’s not needed for global sensitivity models. So I’ve talked about what a sensitivity model means and some basic properties of it, but haven’t talked about how to build them. Generally, in this setup, there’s three ways to build a sensitivity model. And then they essentially correspond with different factorizations of the full data distribution. So there’s a simultaneous model.
that tries to factorize distribution this way.
So introduces unmeasured confounder, U,
and then you need to model
these three conditional probabilities.
There’s also the treatment model
that doesn’t rely on this unmeasured confounder U.
But whether you need to specify is the distribution
of the treatment given the unmeasured cofounders and x.
And once you’ve specified that you can use Bayes formula
to get this part.
And then there’s the outcome model that factorizes
this distribution in the other way.
So this is basically the propensity score
and the third turn is what we need to specify
it’s a sensitivity parameter.
So in the missing data literature,
second model kind of model
is usually called selection model.
And the third kind of models usually called
pattern mixture model,
and there are other names that have been given to it.
And basically different sensitivity models,
they amount to different ways of specifying these
either non identifiable distributions,
which are these ones that are underlined.
A good review is this report by a committee
organized by the National Research Council.
This ongoing review paper that we’re writing
also gives a comprehensive review of many models
that have been proposed using these factorizations.
Okay, so that’s about the sensitivity model. The next component is statistical inference. Things get a little bit tricky here, because there are two kinds of inference or two modes of inference we can talk about in this study.

So, the first mode of inference is point identify inference. So you only care about a fixed value of the sensitivity parameter $\eta$. And the second kind of inference is partial identified inference, where you perform the statistical inference simultaneously for a range of security parameters $\eta$. And that range $H$ is given to you.

And in these different modes of inferences, it comes differences to core guarantees. So for point identified inference usually let’s say for interval estimators, you want to construct confidence intervals. And these confidence intervals depend on the observed $\theta$ and the sensitivity parameter which your last to use in a point of identified inference and it must cover the true parameter with one minus alpha probability for all the distributions in your model. Okay that’s the infimum. But for partial identified inference, you’re only allowed to use an interval that depends on the range, $H$. 


So, it cannot depend on a specific values of the sensitivity parameter, because you only know eta is in this range H. It need to satisfy this very similar criteria. So I call this intervals that satisfy this criteria in the sensitivity interval. But in the literature people have also called this uncertainty interval and or just confidence interval. But to make it different from the first case, we’re calling a sensitivity interval here. So you can see that these two equations, two criterias look very similar, besides just that this interval needs to depend on the range instead of a particular value of the sensitivity parameter. But actually, they’re quite different. This is usually much wider. The reason is, you can actually write an equivalent form of this equation one, because this only depends on the observed data and the range H. Then for every theta in that, sorry for every eta in that range H, is missing here, eta in H and also that’s observationally equivalent to a two distribution. This interval also needs to cover the corresponding theta parameter. So in that sense, this is a much stronger guarantee that you have. So, in terms of the statistical methods,
Point identified inference is usually quite straightforward. It’s very similar to our primary analysis. So, primary analysis just assumes this eta equals to zero, but this sensitivity analysis assumes eta is known. So usually you just can just plug in this eta in some way as an offset to your model. And then everything works out in almost the same way as a primary analysis. But for partially identified analysis, things become quite more challenging. And there are several methods several approaches that you can take. So, essentially there are two big classes of methods, one is bound estimation, one is combining point identified inference. So, for bound estimation, it tries to directly make inference about the two ends of this partial identify region. So, this set this is the region of the parameter beta that are sort of indistinguishable, if I only know this sensitivity parameter eta is in H. If we can somehow directly estimate the infimum and supremum of this in this set, but then that gotta get us a way to make partial identified inference. The second method is basically to try to combine the results of point identified inference. The main idea is to sort of construct interval estimators,
for each individual sensitivity parameter
and then take a union of them.
So, these are the two broad approaches
to the partially identified inference.
And so, within the first approach
the bound estimation approach,
there are also several variety of,
there are several possible methods
depending on your problem.
So, the first approach
the first method is called separable balance.
But before that, let’s just slightly change our notation
and parameterize this range $H$ by a hyper parameter $\gamma$.
So, this is useful when we outline these methods.
And then this beta $L$ of $\gamma$, this is the lower end of the partial identify region.
So the first method is called separable bounds.
What it tries to do is to write this lower end
as a function of $\beta^*$ and $\gamma$,
where $\beta^*$ is your primary analysis estimate.
So let’s say $\theta^*_0$ is what you would do in a primary analysis
that is observationally equivalent to the true distribution.
And then, if $\beta^*$ is the corresponding causal effect,
from that model,
and if somehow can write this lower end
as a function of $\beta^*$ and $\gamma$
and the function is known,
then our life is quite easy,
because we already know how to make inference
0:31:52.54 –> 0:31:55.36 about beta star from the primary analysis.
0:31:55.36 –> 0:31:57.23 And all we need to do is just plug in
0:31:57.23 –> 0:31:59.2 that beta star in this formula,
0:31:59.2 –> 0:32:00.4 and then we’re all done.
0:32:01.81 –> 0:32:05.54 And we call this separable because it allows us
0:32:05.54 –> 0:32:09.14 to separate the primary analysis
0:32:09.14 –> 0:32:11.33 from the sensitivity analysis.
0:32:11.33 –> 0:32:14.94 And statistical inference becomes a trivial extension
0:32:14.94 –> 0:32:16.94 of the primary analysis.
0:32:16.94 –> 0:32:20.47 So, some examples of this kind of method
0:32:20.47 –> 0:32:23.65 include the classical cornfields bound
0:32:25.68 –> 0:32:27.15 and the E-value,
0:32:27.15 –> 0:32:29.32 if you have heard about them,
0:32:29.32 –> 0:32:31.34 and E-value seems quite popular
0:32:31.34 –> 0:32:33.98 these days at demonology.
0:32:36.87 –> 0:32:40.949 The second type of bound estimation
0:32:40.949 –> 0:32:44.975 is called tractable bounds.
0:32:44.975 –> 0:32:47.6 So, in these cases,
0:32:47.6 –> 0:32:51.62 we may derive this lower bound as a function
0:32:51.62 –> 0:32:54.3 of theta star and gamma.
0:32:54.3 –> 0:32:58.02 So we are not able to reduce it to just depend
0:32:58.02 –> 0:33:00.36 on beta star the causal effect
0:33:00.36 –> 0:33:04.029 under no unmeasured confounding,
0:33:04.029 –> 0:33:06.66 but we’re able to express in terms of theta star.
0:33:06.66 –> 0:33:10.72 And then the function gl is also some practical functions
0:33:10.72 –> 0:33:12.76 that we can compute.
0:33:12.76 –> 0:33:17.17 And then this also makes our lives quite a lot easier,
0:33:17.17 –> 0:33:21.149 because we can just replace this theta star,
0:33:21.149 –> 0:33:24.614 which can be nonparametric can be parametric,
0:33:28.11 –> 0:33:31.31 And, often in these cases,
we can find some central limit theorems for the corresponding sample estimator, such that the sample estimator of the bounds converges to its truth at root and rate and it follows the normal limit. And then if we can estimate this standard error, then we can use this central limit theorem to make partial identified inference because we can estimate the bounds. There’s some examples in the literature, you’re familiar with these papers. But one thing to be careful about these kind of tractable bounds is that things that get a little bit tricky with syntactic theory. This is because in a syntactic theory, the confidence intervals or the sensitivity intervals in this case, can be point wise or uniform in terms of the sample size. So it’s possible that if the convergence, if there are statistical guarantee is point wise, then you sometimes in extreme cases, even with very large sample size, they’re still exist data distributions such that your coverage is very poor. So this point is discussed very heavily in econometrics literature. And these are some references.

So that’s the second type of method in the first broad approach.
The third kind of method is called stochastic programming. And this applies when the model is separable. So and we can write this parameter we’re interested in as some expectation of some function of the theta and the sensitivity parameter eta. Okay, so in this case, the bound becomes the optimal value for an optimization problem, which you want to minimize expectation of some function. And the parameter in this function is in some set as defined by U. So, this is known as stochastic programming. So, this type of problem is known as stochastic programming in the optimization literature. And what people do there is they sample from the distribution, and then they try to use it to solve the empirical version and try to use that as approximate solution to this population optimization problem, which we can’t directly evaluate. And the method is called sample average approximation in the optimization literature. So, what is shown there. And Alex Shapiro did a lot of great work on this, that nice problems with compact set age, and everything is euclidean. So it’s finite dimensional. Then you actually have a central limit theorem.
0:37:03.73 → 0:37:05.693 for the sample optimal value.
0:37:07.15 → 0:37:11.82 And this link, is a link between sensitivity analysis
0:37:11.82 → 0:37:15.753 and stochastic programming is made in this paper
0:37:15.753 → 0:37:17.263 by Tudball et al.
0:37:20.33 → 0:37:22.89 Okay, so that’s the first broad approach
0:37:22.89 → 0:37:25.003 with doing bounds estimation.
0:37:26.29 → 0:37:29.33 The second broad approach is to combine the results
0:37:29.33 → 0:37:31.423 of points identified inference.
0:37:32.37 → 0:37:36.93 So, the first possibility is to take a union
0:37:36.93 → 0:37:40.02 of the individual confidence intervals.
0:37:40.02 → 0:37:43.332 Suppose these are the confidence intervals
0:37:43.332 → 0:37:45.282 when the sensitivity from eta is given.
0:37:46.51 → 0:37:51.134 Then, it is very simple to just apply a union bound
0:37:51.134 → 0:37:54.06 and to show that if you take a union
0:37:54.06 → 0:37:57.46 of these individual confidence intervals,
0:37:57.46 → 0:38:01.1 then they should satisfy the criteria
0:38:01.1 → 0:38:03.35 for sensitivity interval.
0:38:03.35 → 0:38:06.994 So now, if you take a union this interval only depends
0:38:06.994 → 0:38:07.96 on the range H,
0:38:07.96 → 0:38:11.511 and then you just apply the union bound
0:38:11.511 → 0:38:13.933 and get this formula from the first.
0:38:13.933 → 0:38:19.61 And this can be slightly improved
0:38:19.61 → 0:38:23.27 to cover not just these parameters,
0:38:23.27 → 0:38:27.21 but also the entire partial identified region
0:38:27.21 → 0:38:29.91 if the intervals if the confidence intervals
0:38:29.91 → 0:38:32.653 have the same tail probabilities.
0:38:35.05 → 0:38:36.923 So we discussed this in our paper.
0:38:38.653 → 0:38:43.35 And here, so, all we need to do
0:38:43.35 → 0:38:45.113 is to compute this union.
0:38:45.97 → 0:38:49.23 So, which essentially is an optimization problem
0:38:49.23 → 0:38:52.48 we’d like to minimize the lower bound,
that the lower confidence point $C_l$ of $\eta$ over $\eta$ in $H$ and similarly for the upper bound. And usually using of syntactic theory, we can get some normal base confidence intervals for each fixed $\eta$. And then we just need to optimize this thing this confidence interval over $\eta$. But for many problems this can be computationally challenging because the standard errors are usually quite complicated and it has some very nonlinear dependence on the parameter $\eta$. So optimizing this can be tricky. This is where another method of percentile bootstrap method can greatly simplify the problem. It’s proposed by this paper that we wrote, and what it does is instead of using the syntactic confidence interval for fixed $\eta$, we use the percentile bootstrap interval. Where we take theta samples, and then you estimate the causal effect $\beta$ in each resample and then take quantiles. Okay, so if you use this confidence interval, then there is a general, generalized minimax inequality that allows us to construct this percentile bootstrap sensitivity interval. So what it does is this thing in the inside is just the union of these percentile construct intervals for fixed $\eta$, taken over $\eta$ in $H$. 
And then this generalized minimax inequality allows us to interchange the infimum with quanto and the supremum of a quanto.

Okay, so the infimum of a quanto is greater than equal to the quanto of infimum and that it’s always true.

So it’s just a generalization of the familia minimax inequality.

Now, if you look at this order interval, this is much easier to compute, because all it needs to do is you gather data resample, then you just need to repeat method 1.3. So just get the infimum of the point estimate for that resample and the supremum for that resample.

Then you do this over many, many resamples and then you take the quantiles of the infimum, lower of the infimum and upper quantile of the supremum, and because this union sensitivity interval is always valid, if the individual confidence intervals are valid. So you almost got a very you got a free lunch in some sense, you don’t need to show any heavy theory. All you need to show is that these percentile bootstrap intervals are valid for each fixed eta, which are much easier to establish in real problems.

And this is sort of selfish,
where I’d like to compare this idea with Efron’s bootstrap, where what was found there is that you’ve got a point estimator, you resample your data, and then many times and then use bootstrap to get the confidence interval.

For partially identified inference, you need to do a bit more. So for each resample you need to get extrema optimal estimator. Then the minimax inequality allows you just sort of transfer the intuition from the bootstrap, for bootstrap from point identification to partial identification.

So the third approach in this, is a third method in this general approach is to take the supremum of key value. And this is used in Rosenbaum sensitivity analysis. If you’re familiar with that.

Essentially it’s a hypothesis testing analog of the Union confidence interval method. What it does is that if you have individually valid P values for a fixed \( \eta \), then you just take the supremum of the P values over all the \( \eta \)s in this range. And that can be used for partially identified inference.

So what Rosenbaum did, and Rosenbaum is really a pioneer in this area in the partially identify sensitivity analysis. So what he did was use randomization tests.
0:43:59.41 –> 0:44:01.073 to construct these key values.

0:44:02.54 –> 0:44:06.57 So, this is usually done for matched observational studies

0:44:06.57 –> 0:44:11.57 and the inside of this line of work

0:44:11.79 –> 0:44:16.044 is that you can use these inequalities

0:44:16.044 –> 0:44:18.94 particularly Holley’s inequality

0:44:18.94 –> 0:44:21.5 in probabilistic combinatorics

0:44:21.5 –> 0:44:25.113 to efficiently compute these supremum of the P values.

0:44:26.44 –> 0:44:29.74 So, usually what is done there is that

0:44:29.74 –> 0:44:32.47 the Holley’s inequality gives you a way

0:44:32.47 –> 0:44:36.983 to upper bound the distribution of a that,

0:44:38.68 –> 0:44:40.92 to upper bound family of distributions

0:44:42.08 –> 0:44:45.424 in the stochastic dominance sense.

0:44:45.424 –> 0:44:49.793 So, that is used to get these supremum of the P values.

0:44:51.07 –> 0:44:56.07 And so, basically the idea is to use some theoretical tool

0:44:58.52 –> 0:45:02.023 to simplify the computation.

0:45:05.366 –> 0:45:08.14 Okay, so that’s the statistical inference.

0:45:08.14 –> 0:45:10.27 The third part, the third component

0:45:10.27 –> 0:45:13.19 is interpretation of sensitivity analysis.

0:45:13.19 –> 0:45:16.95 And this is the area that we actually really need

0:45:16.95 –> 0:45:19.293 a lot of good work at the moment.

0:45:20.46 –> 0:45:25.46 So, overall, there are two good ideas that seem to work,

0:45:25.77 –> 0:45:27.56 that seem to improve the interpretation

0:45:27.56 –> 0:45:28.873 of sensitivity analysis.

0:45:29.99 –> 0:45:31.69 The first is sensitivity value,

0:45:31.69 –> 0:45:35.203 the second is the calibration using measured confounders.

0:45:36.08 –> 0:45:38.46 So the sensitivity value is basically

0:45:38.46 –> 0:45:41.062 the value of the sensitivity parameter

0:45:41.062 –> 0:45:42.17 or the hyper parameter,
where some qualitative conclusions about your study change.

And in our motivating example, this is where the estimated average treatment effect is reduced by half an Rosenbaum sensitivity analysis if you are familiar with that.

This is where, this is the value of the gamma in his model, where we can no longer reject the causal null hypothesis.

So, this is can be seen as kind of an extension of the idea of a P value.

So P value is used for primary analysis, so assuming no unmeasure confounding, and then for sensitivity analysis, you can use the sensitivity value to sort of sorry, that’s the P value it basically measures how likely your results are due to random chance.

But then what a sensitivity value does is measures how much sort of how sensitive your resources is in some sense, so, how much deviation from the unmeasured confounding it takes to alter your conclusion. And for sensitivity value, there often exists a phase transition phenomenon for partially identified inference.

This is because if you take your hyper parameter gamma to be very large, then essentially your partially identify region
0:47:15.21 -> 0:47:17.06 already covered in null.
0:47:17.06 -> 0:47:20.33 So, no matter how large your sample size is
0:47:23.24 -> 0:47:26.31 So, this is sort of an interesting phenomenon
0:47:27.632 -> 0:47:32.07 and explained first discovered by Rosenbaum
0:47:32.07 -> 0:47:37.07 in this paper I wrote also clarified some problems
0:47:37.61 -> 0:47:42.153 some issues in both the phase transition.
0:47:44.08 -> 0:47:46.37 So, the second idea is the calibration
0:47:48.65 -> 0:47:50.69 So, you have already seen an example
0:47:50.69 -> 0:47:54.3 in a motivating study.
0:47:54.3 -> 0:47:59.18 It’s really a very necessary and practical solution
0:47:59.18 -> 0:48:01.086 to quantify the sensitivity,
0:48:01.086 -> 0:48:05.066 because it’s not really very useful if you tell people,
0:48:05.066 -> 0:48:08.23 we are sensitive at gamma equals to two,
0:48:08.23 -> 0:48:09.4 what does that really mean?
0:48:09.4 -> 0:48:12.568 That depends on some mathematical model.
0:48:12.568 -> 0:48:14.95 But if we can somehow compare that
0:48:14.95 -> 0:48:17.483 with what we do observe,
0:48:18.39 -> 0:48:19.51 and we have,
0:48:19.51 -> 0:48:22.95 often the practitioners have some good sense
0:48:22.95 -> 0:48:26.8 about what are the important confounders and what are not.
0:48:26.8 -> 0:48:30.61 Then this really gives us a way to calibrate
0:48:30.61 -> 0:48:35.15 and strengthen the conclusions of a sensitivity analysis.
0:48:35.15 -> 0:48:37.99 But unfortunately, although there are some good heuristics
0:48:37.99 -> 0:48:39.373 about the calibration,
0:48:40.366 -> 0:48:43.89 they’re often suffer from some subtle issues,
0:48:43.89 -> 0:48:46.1 like the ones that I described
0:48:46.1 -> 0:48:47.55 in the beginning of the talk.
0:48:48.627 -> 0:48:51.113 If you carefully parameterize your models
this can become easier. And this recent paper sort of explored this in terms of linear models. But really there’s not a unifying framework then you can cover more general cases and lots of work are needed. And when I was writing the slides, I thought maybe what we really need is to somehow build this calibration into the sensitivity model. Because currently our workflow is that we assume a sensitivity model, and we see where things get changed, and then we try to interpret those values where things get changed. But suppose if we somehow build that, if we left the range H eta to be defined in terms of this calibration. Perhaps gamma directly means some kind of comparisons that measured confounders this would solve some a lot of the issues. This is just a thought I came up when I was preparing for this talk. Okay, so to summarize, so there is number of messages, which I hope you can take home. There are three components of a sensitivity analysis.

Model augmentations, statistical inference and the interpretation of sensitivity analysis. So sensitivity model is about parameterizing, the full data distribution.
And that’s basically about over parameterizing the observed data distribution. And you can understand these models by factorizing the distribution differently and specify different models for those that are on identifiable. And there’s a difference between point identified inference and partially identified inference, and partially identified inference is usually much harder. And there are two general approaches for partially identified inference, bound estimation and combining point identified inference. For interpretation of sensitivity analysis, there seem to be two good ideas so far, to use the sensitivity value, and to calibrate that sensitivity value using measured confounders. But overall, I’d say this is still a very open area for this prototypical example that people have studied for decades, it seems there’s still a lot of questions that are unresolved. And there are methods that need to be developed for this sensitivity analysis.
0:51:44.86 –> 0:51:48.03 to be regularly used in practice.
0:51:48.03 –> 0:51:50.85 And then there are many other related problems
0:51:50.85 –> 0:51:52.81 in missing data in causal inference
0:51:53.73 –> 0:51:57.703 that need to see more developments of sensitivity analysis.
0:51:58.81 –> 0:52:00.82 So that’s the end of my talk.
0:52:00.82 –> 0:52:03.423 And there are some references that are used.
0:52:04.63 –> 0:52:08.14 I’m happy to take any questions.
0:52:08.14 –> 0:52:11.167 Still have about four minutes left.
0:52:14.307 –> 0:52:17.18 Thank you, I’m sorry I couldn’t introduce you earlier,
0:52:17.18 –> 0:52:20.627 but my connection but it did not to work.
0:52:20.627 –> 0:52:23.153 So we have time for a couple of questions.
0:52:25.56 –> 0:52:28.82 You can write the question in the chat box,
0:52:28.82 –> 0:52:30.663 or just unmute yourselves.
0:52:43.482 –> 0:52:44.315 Any questions?
0:52:53.67 –> 0:52:55.962 I guess I’ll start with a question.
0:52:55.962 –> 0:53:00.12 Yeah I guess I’ll start with a question.
0:53:00.12 –> 0:53:03.89 This was a great connection between I think,
0:53:03.89 –> 0:53:05.77 sensitivity analysis literature
0:53:05.77 –> 0:53:07.493 and the missing data literature.
0:53:08.6 –> 0:53:11.982 Which I think it’s kind of overlooked.
0:53:11.982 –> 0:53:16.982 Even when you when you run a prometric sensitivity analysis,
0:53:17.36 –> 0:53:20.27 it’s really something, like most of the times
0:53:20.27 –> 0:53:22.06 people really don’t understand
0:53:22.06 –> 0:53:24.023 how much information is given.
0:53:24.86 –> 0:53:28.56 Like, how much information the model actually gives
0:53:29.4 –> 0:53:31.66 on the sensitivity parameters.
0:53:31.66 –> 0:53:34.08 And as you said,
0:53:34.08 –> 0:53:35.53 like it’s kind of inconsistent
to set the sensitivity parameters when sensitivity parameters are actually identified by the model.

So I think like my I guess a question of like, clarifying question is, you mentioned there is this testable models, essentially are wherein the sensitivity model is such that the sensitivity barometer are actually point identified. Right?

So it re, so you said, you reshooting use the sensitivity analysis to actually set the parameters if the sensitivity parameters are actually identified model.

Yeah.

Is that what you’re trying?

All right, so and. - Yes, yeah.

Basically what happened there is the model is too specific, and it wasn’t constructed carefully.

So it’s possible to construct parametric models that are not testable that are perfectly fine.

But sometimes, if you just sort of write down the most natural model,

if it just extend what the parametric model you used for observed data to also model full data,

then you don’t do it carefully,

then the entire full data distribution becomes identifiable.

So it does makes sense to treat those parameters
as sensitivity parameters.
So this kind of is a reminiscent of the discussion in the 80s about the Hackmann selection model.
Because in that case,
there was also sir Hackmann has this great selection model for reducing or getting rid of selection bias,
but it’s based on very heavy parametric assumptions.
And you can adapt certainly identify the selection effect directly from the model where you actually have no data
to support that identification.
Which led to some criticisms in the 80s.
But I think we are seeing this things repeatedly again and again in different areas.
And it’s, I think it’s fine to use the power metric models that are testable, actually,
if you really believe in those models,
but it doesn’t seem that they should be used this sensitivity analysis,
because just logically,
it’s a bit strange.
It’s hard to interpret those models.
And but sometimes I’ve also seen people who use the sort of parameterize the model in a way that you include enough terms.
So the sensitivity parameters are weakly identified in a practical example.
So with a practical data set of maybe the likelihood test,
Likelihood Ratio Test rejection region,
that acceptance region is very, very large. So there are a suggestions like that, that kind of it’s a sort of a compromise for good practice.

Right in that case you gave it either set the parameters and drag the causal effects, or kind of treat that as a partial identification problem and just write use bounds or the methods you were mentioning, I guess.

Yeah.

Well I guess you can read the question? It’s a question from Kiel Sint. Sorry if I didn’t pronounce your name correctly. “In the applications of observational studies ideally, what confounders should be collected for sensitivity analysis for unmeasured confounding?” Thank you.

So if I understand your question correctly, basically what sensitivity analysis does is you have observational study, where you for already collected confounders that you believe are important or relevant that really that are real confounders, that change the causal unchanged the treatment and the outcome. But often that’s not enough. And what sensitivity analysis does is it tries to say, “based on what the components already
you have already collected, what if there is still something missing that we didn’t collect?
And then if those things behave in a certain way, does that change our results?
So, I guess sensitivity analysis is always relative to a primary analysis.
So I think you should use the same set of confounders that the primary analysis uses.
I don’t see a lot of reasons to vary to say use a primary analysis with more confounders, but a sensitivity analysis with fewer confounders. Sensitivity analysis is really a supplement to what you have in the primary analysis. 
- Just one more question if we have? There not. Yes.
- So from Ching Hou Soo, “How to specify the setup sensitivity parameter gamma in the real life question?
When gamma is too large the inference results will always be non informative?” Yes, this is always a tricky problem any, so it tries to directly look at the value of this sensitivity parameter that changes your conclusion.
So in some sense, you don’t need to specify a parameter a priori. But obviously, in the end of the day,
1:00:24.76 –> 1:00:29.76 we need some clue about what value of sensitivity parameter
1:00:30.02 –> 1:00:31.36 is considered large.
1:00:31.36 –> 1:00:35.5 In a practical sense, in this application.
1:00:35.5 –> 1:00:39.156 That’s something this calibration clause
1:00:39.156 –> 1:00:43.62 this calibration analysis is trying to address.
1:00:43.62 –> 1:00:44.49 But as I said,
1:00:44.49 –> 1:00:47.31 they’re not perfect at the moment.
1:00:47.31 –> 1:00:52.31 So for some time, now, at the least,
1:00:52.36 –> 1:00:55.83 we’ll have to sort of live through this and
1:00:55.83 –> 1:01:00.6 or will either need to understand really
1:01:00.6 –> 1:01:02.18 what the sensitivity model means,
1:01:02.18 –> 1:01:06.01 and then use your domain knowledge
1:01:06.01 –> 1:01:10.46 to set the sensitivity parameter,
1:01:10.46 –> 1:01:15.46 or we have to use these rely on these
1:01:15.471 –> 1:01:19.323 imperfect visualization tools to calibrate analysis.
1:01:27.689 –> 1:01:28.62 - Yeah, all right.
1:01:30.372 –> 1:01:33.209 I think we need to wrap up we’ve run over time.
1:01:33.209 –> 1:01:36.04 So thank you again Qingyuan,
1:01:36.04 –> 1:01:37.85 for sharing your work with us.
1:01:37.85 –> 1:01:40.672 And thank you, everyone for joining.
1:01:40.672 –> 1:01:41.78 Thank you.
1:01:41.78 –> 1:01:42.96 Bye bye.
1:01:42.96 –> 1:01:44.222 See you next week.
1:01:44.222 –> 1:01:45.325 - It’s a great pleasure.
1:01:45.325 –> 1:01:46.158 Thank you.