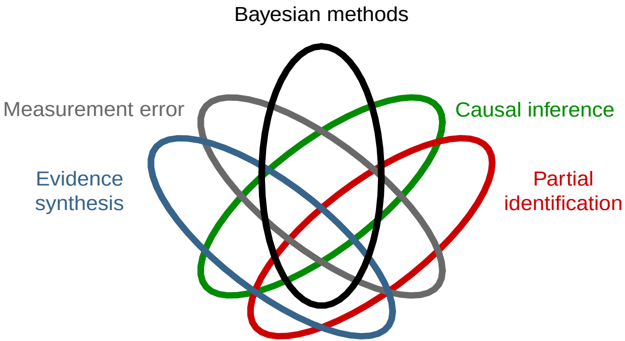
Handling Partial Identification: Sensitivity Analysis, Inference, or Stuck in the Middle with Bayes?

Paul Gustafson
Department of Statistics
University of British Columbia

ISBA World Meeting
Montréal, QC
June 29, 2022

A rough schematic of my research persona



. . . all with an eye to biostatistics, epidemiology, public health . . .

Abbreviated outline

- Two motivating examples
- Three rabbit holes

Ex. #1: Inferring HIV prevalence in a target population

Verstraeten et. al. (1998).

Y = HIV status.

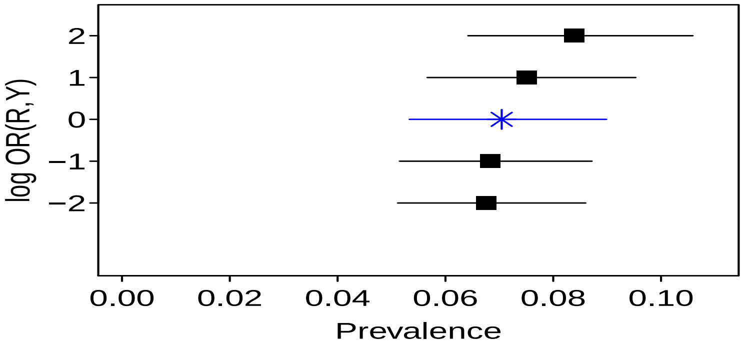
R = Response indicator.

	$Y=0$	$Y=1$	Total
$R=1$	699	52	751
$R=0$	???	???	36
			787

$$\begin{aligned} \text{Target parameter:} \quad & \psi = Pr(Y=1) \\ \text{"Sensitivity" parameter:} \quad & \lambda = \log OR(R, Y) \end{aligned}$$

$\lambda = 0$: Missing completely at random (MCAR).

Main analysis (MCAR) + (essentially tabular) sensitivity



E.g. Posterior distribution of $(\psi|\lambda, \text{Data})$.

Low-hanging fruit - Bayesianize this

Bayesian sensitivity analysis (BSA)

Assign a prior distribution to the sensitivity parameter λ.

Report out the posterior marginal of $(\psi|\text{Data})$.

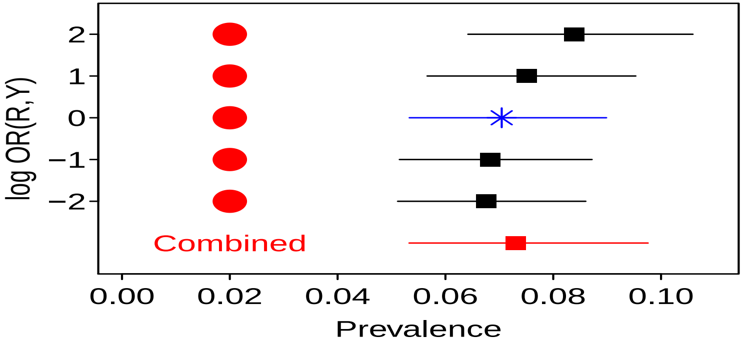
Seemless uncertainty management, e.g., think:

$$\text{Var}(\psi|\text{Data}) = E\left\{\text{Var}(\psi|\lambda, \text{Data})|\text{Data}\right\} + \text{Var}\left\{E(\psi|\lambda, \text{Data})|\text{Data}\right\}.$$

Appropriately combining:

- Uncertainty because $n < \infty$ (if λ were known).
- Uncertainty because λ isn't known.

Simple Bayesianization: $\lambda \sim \text{Unif}\{-2, -1, 0, 1, 2\}$



Full disclosure: Have used a prior specification which guarantees the marginal prior and posterior distributions of λ to be the same (more soon).

Ex. #2: Infer (X, Y) association from case-control data

Disease status Y. Measured exposure status X^*.

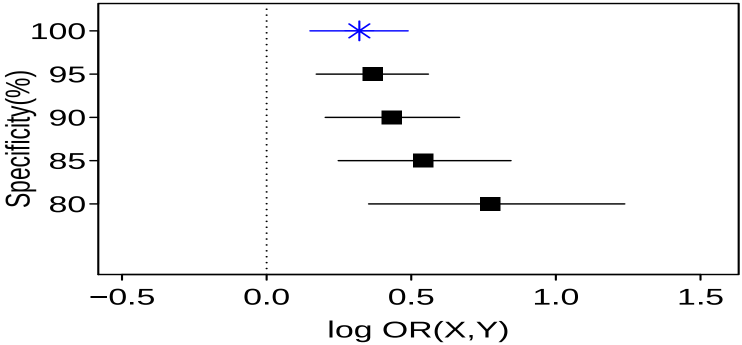
Concern that perhaps for some subjects $X^* \neq$ *actual* exposure X.

	Dataset #1		Dataset #2	
	$Y=1$	$Y=0$	$Y=1$	$Y=0$
$X^*=1$	300	240	188	168
$X^*=0$	580	640	692	712
Total	880	880	880	880

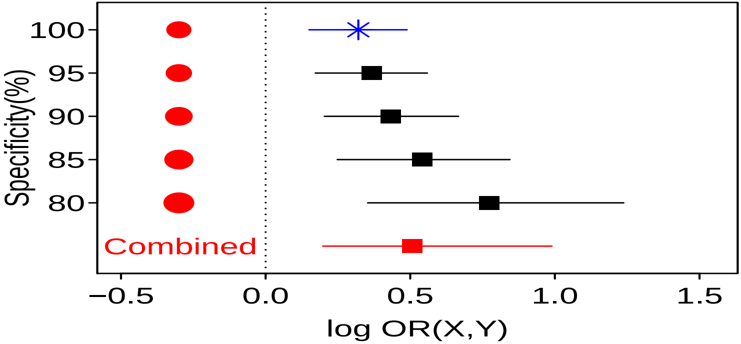
Say $Pr(X^*=1|X=1)=1$ known. But concern that specificity, $\lambda = Pr(X^*=0|X=0)$, may fall below one.

E.g., think interviewer-assisted survey, $X=1$ is virtuous (veggies, exercise, flossing ...)

First dataset - tabular

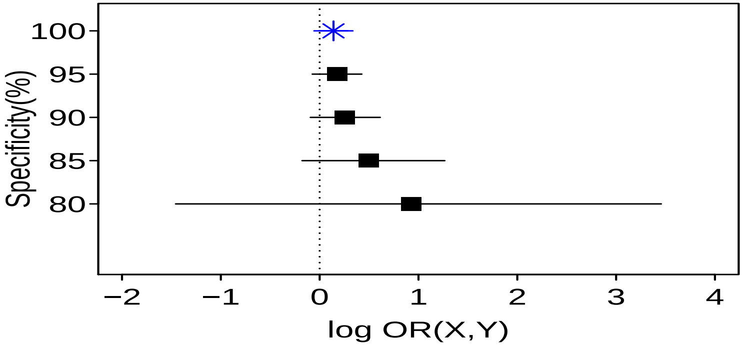


First dataset - BSA, $\lambda \sim \text{Unif}\{0.80, 0.85, 0.90, 0.95, 1\}$

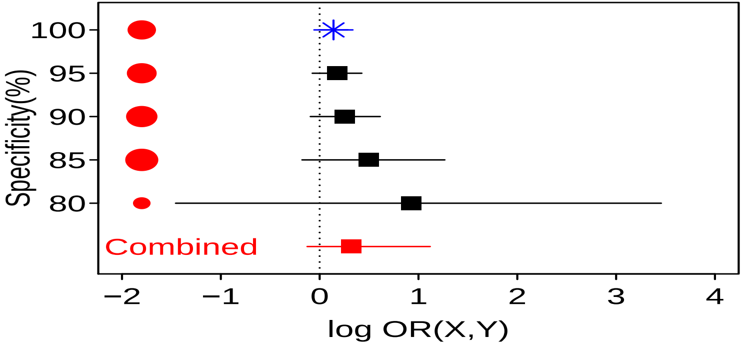


Marginal posterior for the sensitivity parameter closely resembles prior.

Second dataset - tabular



Second dataset - BSA



Marginal posterior for the sensitivity parameter less closely resembles prior.

For some datasets we do inference for λ, others not so much?

Stuck in between with Bayes.

OK, now for three of these

Rabbit Hole #1: Prior dependence structure

(Re)parameterizations:

$$\begin{array}{ccc} \text{“scientific”} & & \text{“transparent”} \\ \theta & \leftrightarrow & (\phi, \lambda) \end{array}$$

θ: initial, interpretable, intuitive

ϕ: governs distribution of observables

λ: sensitivity parameter

Issue: *A priori* independence of ϕ and λ?

If yes, then

- *a posteriori* independence
- marginal posterior for $\lambda =$ prior

Issue: *A priori* independence of ϕ and λ?

$$\begin{array}{ccc} \text{scientific} & & \text{transparent} \\ \theta & \leftrightarrow & (\phi, \lambda) \end{array}$$

Sometimes $\phi \perp \lambda$ argued for as **something to impose**.
E.g., specify (elicit?) on right, left is a consequence.

- makes the Bayesian sensitivity analysis clean/simple
- dovetails with the sizeable epidemiologic literature on *probabilistic sensitivity analysis*

But

- left-specification may be more interpretable
- left-specification often won't yield independence on right
- sometimes independence on right is (nearly?) impossible

Example 1 via this lens

Results shown used *a priori* (and therefore *a posteriori*) independence of $\phi = \{Pr(R=1), Pr(Y=1|R=1)\}$ and $\lambda = \log OR(R, Y)$.

But FYI, had we started with θ being (R, Y) cell probs, then a specification such as $\theta \sim$ Dirichlet would not yield $\phi \perp \lambda$.

Example 2 via this lens

Results shown based on left-specification: λ independent of dist$(X|Y)$.

This Induces dependence on right.

In fact, this problem really can't admit *a priori* independence of ϕ and λ, even if we wanted:

Constrained parameter space:

$$\lambda \quad > \quad 1 - \min_{y=0,1} Pr(X^* = 1 | Y = y)$$

- Support of $(\lambda|\phi)$ prior must depend on ϕ.
- Low apparent exposure rate rules out high FP rate.

Example 2 via this lens, continued

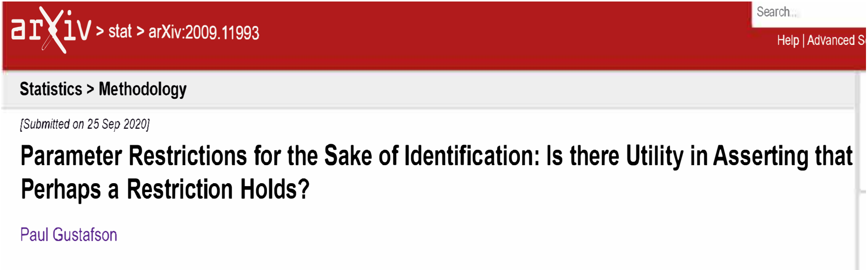
So the (intuitive/interpretable) sensitivity parameter is bounded by an estimable quantity.

Short of making the sensitivity parameter wildly uninterpretable...

$$\tilde{\lambda} \;=\; \frac{\lambda - LB(\theta)}{1 - LB(\theta)} \quad ???$$

... prior independence between dist(observables) and the sensitivity parameter ain't happening.

Rabbit Hole #2


arXiv > stat > arXiv:2009.11993

Search...

Help | Advanced S

Statistics > Methodology

[Submitted on 25 Sep 2020]

Parameter Restrictions for the Sake of Identification: Is there Utility in Asserting that Perhaps a Restriction Holds?

Paul Gustafson

BSA may be particularly appealing in settings where an assumption that results in an identified model is **plausible, but not incontrovertible**.

E.g. $\lambda = 0$ is plausible but not incontrovertible

E.g., $\lambda = 0$ represents MAR, or perfect exposure classification, or no unobserved confounding, or ...

Spike-and-slab (i.e., mixture) prior on $\theta \in \Theta$.

Slab: continuous, full support $[Pr(\lambda = 0) = 0]$

Spike: semi-degenerate, $\lambda \equiv 0$

Literal representation of: maybe the identifying restriction holds.

Identifying restriction $\lambda = 0$: Prior says maybe

Some problems yield the simple behaviour seen earlier:

$$Pr(\lambda = 0|\text{Data}) \quad = \quad Pr(\lambda = 0)$$

But we can exhibit (relatively simple and realistic) problems where the partial identification structure yields curious behavior of $\lim_{n\to\infty} Pr(\lambda = 0|\text{Data})$.

Curious behaviour - teaser

Inferring a risk diference under *possible* outcome misclassification.

Let $p^*(\theta) = \lim_{n \to \infty} Pr(\lambda = 0|\text{Data})$.

Image of Θ under $p^*()$ is $(c, 1)$, where $c > 0$.

Data never definitive.

There are select truths under which we get arbitrarily close to proving perfect classification.

But can't get similary close to refuting this.

Rabbit hole # 3: Performance Evaluation

Simulation studies can be a bit predictable: Compared to the "main analysis" inference, BSA will win/lose if the assumption being interrogated is false/true.

Spice this up a bit by considering Bayes risk.

E.g. Consider the Root Average Mean Squared Error (RAMSE) of the Bayesian point estimator built using the **investigator's prior**.

Inner expectation: Mean wrt data given truth (θ).

Outer expectation: Average across different truths (θ's) weighted by **Nature's prior**.

RAMSE to contrast pro/con of asserting that maybe an identifying restriction holds

$$\begin{aligned} \pi_0 : &\quad \text{distribution for } \theta \in \Theta \text{ giving no weight to } \lambda = 0. \\ \pi_{MIX} : &\quad \text{mixture of } \pi_0 \text{ and a distribution asserting } \lambda = 0. \end{aligned}$$

RAMSE		Investigator	
		π_0	π_{MIX}
Nature	π_0	a_0	$a_0(1 + \delta_{tax})$
	π_{MIX}	a_1	$a_1(1 - \delta_{div})$

Standard guarantees: $\delta_{div} \geq 0$, $\delta_{tax} \geq 0$.

Dividend/tax teaser

Inferring a risk difference under *possible* outcome misclassification.

Monte Carlo computation: $\delta_{div} = 0.25$, $\delta_{tax} = 0.45$.

Asserting that the identifying restriction **might** hold yields an **aggregate 25% reduction** in RMSE, across a mix of scenarios, some where the assumption holds and others where not.

Saying maybe is worthwhile, if you mean it!

Asserting that the identifying restriction **might** hold yields an **aggregate 45% increase** in RMSE, across a mix of scenarios, none of which have the assumption holding.

No free lunch. Don't say maybe unless you mean it!

Parting thoughts

Bayesian sensitivity analysis is conceptually straightforward and interpretable.

But many problems occupy a murky middle ground between:

- λ is completely uninformed by data
- λ is consistently estimable

In such problems, we can understand what kinds of prior assertions engender what kinds of posterior distributions, with what kind of performance.

So stuck in the (murky) middle with Bayes - but this is a feature, not a bug!