THE UNIVERSITY OF BRITISH COLUMBIA

DEPARTMENT OF STATISTICS

TECHNICAL REPORT #264

On the Behaviour of Bayesian Credible Intervals in Partially Identified Models

Paul Gustafson

September 2011

On the Behaviour of Bayesian Credible Intervals in Partially Identified Models

Paul Gustafson

gustaf@stat.ubc.ca
Department of Statistics
University of British Columbia

September 21, 2011

Abstract

Partially identified models typically involve set identification rather than point identification. That is, the distribution of observables is consistent with a set of values for the target parameter, rather than a single value. Interval estimation procedures therefore behave differently than for identified models. For instance, a Bayesian credible set arising from a proper prior distribution will tend to a non-degenerate set as the sample size goes to infinity. A natural question arising is for what parameter values does the limit of the Bayesian credible interval fail to cover its target? Intuition suggests this would arise for parameter values which are not very consistent with the prior distribution. The aim of this paper is to quantify this intuition.

Keywords: Bayesian inference; credible interval; partial identification

1 Introduction

Limitations in the form of discrepancies between ideal data and available data can lead to partial identification, such that learning the distribution of observable variables only reveals a set of possible values for the target parameter, not a single value. The set of target parameter values consistent with the distributional law of the observables is often termed the *identification region*. The recent literature on interval estimation in partially identified models includes Imbens and Manski [8], Vansteelandt et al. [15], Romano and Shaikh [14], and Zhang [16]. An interesting distinction arises in this literature. In frequentist terms, interval estimators can be designed to have at least nominal probability of containing the true value of the target, or to have at least nominal probability of containing the entire identification region.

From the Bayesian perspective, starting with a proper prior distribution over all parameters one can interpret a posterior credible set as being likely to contain the true value of the target, given the combination of observed data plus prior beliefs. Work considering Bayesian interval estimators in partially identified models includes Moon and Schorfheide [11], Liao and Jiang [10], and Gustafson and Greenland [7]. As emphasized in the latter paper, the usual calibration property of Bayesian procedures is unaffected by the lack of identification. That is, the average frequentist coverage of the Bayesian credible set, taken with respect to the prior distribution over the parameter space, equals the nominal coverage. In contrast, approaches such as that of Vansteelandt et al. [15] aim to achieve nominal coverage at 'worst-case' spots in the parameter space, with better than nominal coverage, paid for via extra length, achieved elsewhere.

Another interesting distinction between Bayesian and frequentist approaches to partially identified problems is that frequentist approaches view knowledge of the identification region as the best that can be achieved in the large-sample limit. Conversely, the posterior distribution on the target concentrates asymptotically on the identification region, but its shape may purport to distinguish some values in the region as more plausible than others, in light of the data. Both Liao and Jiang [10] and Gustafson [5] tout this as a strength of the Bayesian approach in partially identified settings, though Moon and Schorfheide [11] are more circumspect.

In the partially identified context, interval estimators generally, and Bayesian credible sets specifically, will tend to a non-degenerate set as the sample size increases. Toward a full understanding of Bayesian credible sets in partially identified contexts, a natural question to ask is when does the limiting set contain its target? More particularly, the parameter space can be cleaved into parameter values under which the limiting credible set does and doesn't contain its target. Initial intuition suggests that the choice of prior distribution should play a role in this regard, with failure-to-cover occurring when the true parameter values are less consistent with the specified prior distribution. As well, the values of parameters which are less informed by the data might be suspected to play a bigger role. Through theory and examples, this paper aims to quantify these intuitions.

2 Methodology

Let $\pi(\theta, D)$ denote the joint density of a parameter vector $\theta \in \Theta$ and observable data D, as arises as the product of a proper prior density $\pi(\theta)$ and a statistical model density $\pi(D|\theta)$. Assume that θ comprises a 'scientifically intuitive' parameterization of the model, such that investigators would feel comfortable specifying a prior distribution for θ, as opposed to specifying a prior in some other parameterization. Also assume that the primary inferential interest lies in some scalar aspect of θ, denoted as the estimand $\psi = g(\theta)$.

When useful, we write D_n to emphasize observable data comprised of n observations which are independent and identically distributed given θ. To consider interval estimation, let $I_\pi^{(\alpha)}(D_n)$ denote a credible set for ψ of some type, having posterior probability $1-\alpha$ of containing ψ under prior π. The two primary examples of 'types' would be equal-tailed intervals and highest posterior density (HPD) sets.

The immutable calibration property of Bayesian interval estimation is as follows. With respect to the joint distribution π on (θ, D_n),

$$pr_\pi \left\{ \psi \in I_\pi^{(\alpha)}(D_n) \right\} \quad = \quad 1-\alpha. \tag{1}$$

This can be interpreted in the following manner. With respect to a sequence of studies arising under *different* conditions (i.e., a different 'true value' of θ each time), the $1-\alpha$ credible interval contains its target in proportion $1-\alpha$ of the studies, under a strong proviso. Particularly, (1) assumes the distribution generating the sequence of true θ values coincides with the distribution used by investigators as a prior distribution. Put another way, if 'nature's prior' and the investigators' prior match, then the investigators' interval estimation procedure is well calibrated. It is worth reinforcing that (1) is exact for any n, and does not require model identification.

Of course identified and correctly specified parametric models yield nicely behaved credible intervals, under weak regularity conditions. As n increases, $I_\pi^{(\alpha)}(D_n)$ converges to the correct value of ψ, with the interval width behaving as $n^{-1/2}$. On the other hand, partially identified models typically have the feature that the posterior distribution on the target of inference does *not* shrink to a point-mass as the sample size grows. Rather, for a fixed underlying value of θ, the distribution of $(\psi|D_n)$ will converge to a non-degenerate distribution as n increases. Commensurately, $I_\pi^{(\alpha)}(D_n)$ will converge to a non-degenerate set as $n \to \infty$. Emphasizing that this limit depends on the underlying parameter values, we denote the limiting credible set as $J_\pi^{(\alpha)}(\theta)$.

Whereas model identification plus weak regularity conditions imply that Bayesian credible intervals have approximately correct frequentist coverage for large n, this cannot apply in the partially identified case. A non-degenerate $J_\pi^{(\alpha)}(\theta)$ either contains $\psi = g(\theta)$ or it doesn't. Consequently, for a given θ value the large-n limit of the frequentist coverage of the Bayesian interval estimator is either 0% or 100%. On the other hand, from (1) it follows that

$$pr_\pi \left\{ \psi \in J_\pi^{(\alpha)}(\theta) \right\} \quad = \quad 1-\alpha. \tag{2}$$

That is, the subset of the parameter space for which the limiting frequentist coverage is zero must have probability mass α with respect to the prior π.

Let $B_\pi^{(\alpha)} = \left\{ \theta : g(\theta) \notin J_\pi^{(\alpha)}(\theta) \right\}$ be the 'bad' subset of the parameter space on which the limiting credible interval fails to cover its target. Equivalently, this is the subset on which the limiting frequentist coverage of the credible interval is zero. As stated above, it is immediate that $B_\pi^{(\alpha)}$ is small in the sense of having probability α under the prior π. What this paper investigates, however, is *where* $B_\pi^{(\alpha)}$ lies in the parameter space. A first intuition is that the bad set will lie in the tails of the prior distribution, i.e., failure-to-cover will arise when the true parameter values are not very compatible with the prior distribution. A second thought might be that compatibility of prior and true values may be more important for parameters that aren't well informed by the data. The aim of this paper is to make these ideas more precise.

As emphasized in Gustafson [4], the posterior structure in many partially identified models can be laid bare upon reparameterizing from θ to $\phi = (\phi_I, \phi_N)$ such that (i) $D_n \perp \phi_N | \phi_I$, and (ii) $(D_n | \phi_I)$ comprises a 'regular' model admitting root-n consistent estimation of ϕ_I. Gustafson [4] calls this a *transparent parameterization.* The limiting posterior distribution of ϕ decomposes into a point-mass at the true value of ϕ_I combined with the prior conditional distribution of $(\phi_N | \phi_I)$. One can then determine the induced limiting posterior distribution on $\psi = g(\theta(\phi))$. In Section 3 we pursue transparent parameterizations as a route to determining limiting credible intervals, and hence the bad set $B_\pi^{(\alpha)}$, in several examples. First, however, we point out a general result.

Theorem 1: Let $\lambda = h(\theta)$ be any identified quantity, i.e., with respect to a transparent parameterization, $h(\theta(\phi_I, \phi_N))$ does not vary with ϕ_N. Then, for any value of λ_0 in the support of the induced prior $\pi(\lambda)$, we have

$$pr_\pi(\theta \in B_\pi^{(\alpha)} | \lambda = \lambda_0) \quad = \quad \alpha. \tag{3}$$

In this sense, the performance of the limiting credible interval cannot be driven by the compatibility of identified parameters and their priors, i.e., the bad sets corresponding to *a priori* likely and unlikely values of λ are of equal 'size' in the conditional probability sense of (3).

Proof. For a given λ_0, let $B_0^{(\alpha)}$ be the set of θ values for which the limiting $1 - \alpha$ credible interval arising from the prior $\pi(\theta | \lambda = \lambda_0)$ fails to cover the target. Now note that for any θ satisfying $h(\theta) = \lambda_0$, $\theta \in B_0^{(\alpha)} \iff \theta \in B_\pi^{(\alpha)}$. This happens because for data generation under such a θ, the same limiting credible interval for ψ arises under $\pi(\theta)$ as under $\pi(\theta | \lambda = \lambda_0)$. That is, since asymptotically the data already reveal the value of ϕ_I, correct *a priori* conditioning on some function of ϕ_I will have no effect whatsoever in the large-n limit. Thus

$$\begin{aligned} pr_\pi\left(\theta \in B_\pi^{(\alpha)} | \lambda = \lambda_0\right) &= pr_\pi\left(\theta \in B_0^{(\alpha)} | \lambda = \lambda_0\right) \\ &= \alpha \end{aligned}$$

as desired, where the second equality follows directly from (2) applied with the prior $\pi(\theta | \lambda = \lambda_0)$. $\square$

3 Examples

3.1 Imperfect Compliance in a Randomized Trial

Here we consider a version of the imperfect compliance model with binary variables considered by various authors, including Chickering and Pearl [1], Imbens and Rubin [9], Pearl [12, Ch. 8], and Richardson et al. [13]. Trial subjects are randomly sampled from a population comprised of never-takers, always-takers, and compliers, in unknown proportions ω_{NT}, ω_{AT}, and $\omega_{CO} = 1 - \omega_{NT} - \omega_{AT}$ respectively. Each subject is randomly assigned to either control or treatment. As the labels suggest, never-takers will not take treatment regardless of their assignment, always-takers will take treatment regardless of their assignment, and compliers will follow their assignment. We exclude the possibility of defiers in the population, though the general version of the problem allows for them.

Assume that a patient's binary response is Y_0 if treatment is not taken, and Y_1 if treatment is taken, regardless of treatment assignment. Then a subject's outcome is $Y = (1 - X)Y_0 + XY_1$, where X indicates reception of treatment, whereas Z indicates assignment to treatment. For compliance type $C \in \{NT, AT, CO\}$, let $\gamma_{C,i}$ by the mean of Y_i amongst the sub-population of that type. The observed data consist of (Z, X, Y) for sampled subjects. Assuming that Z is based only on randomization (i.e., is independent of X, Y_0, Y_1 and compliance type), the observed distribution of $(X, Y|Z)$, in terms of parameters $\theta = (\omega_{NT}, \omega_{AT}, \gamma_{NT,0}, \gamma_{NT,1}, \gamma_{AT,0}, \gamma_{AT,1}, \gamma_{CO,0}, \gamma_{CO,1})$, is characterized by:

$$
\begin{aligned}
pr(X = 1|Z = 0) &= \omega_{AT} \\
pr(X = 1, Y = 1|Z = 0) &= \omega_{AT}\gamma_{AT,1} \\
pr(X = 0, Y = 1|Z = 0) &= \omega_{CO}\gamma_{CO,0} + \omega_{NT}\gamma_{NT,0} \\
pr(X = 0|Z = 1) &= \omega_{NT} \\
pr(X = 0, Y = 1|Z = 1) &= \omega_{NT}\gamma_{NT,0} \\
pr(X = 1, Y = 1|Z = 1) &= \omega_{CO}\gamma_{CO,1} + \omega_{AT}\gamma_{AT,1}.
\end{aligned}
$$

Note that a transparent parameterization is obtained by simply cleaving θ as

$$
\begin{aligned}
\phi_I &= (\omega_{NT}, \omega_{AT}, \gamma_{NT,0}, \gamma_{AT,1}, \gamma_{CO,0}, \gamma_{CO,1}), \\
\phi_N &= (\gamma_{NT,1}, \gamma_{AT,0}).
\end{aligned}
$$

Particularly, it is easy to verify that the map from ϕ_I to the $(X, Y|Z)$ distribution is invertible. Thus a first thought is that the important consideration for interval coverage might be compatibility between the prior and the true values for $(\gamma_{NT,1}, \gamma_{AT,0})$, since these parameters are absent from the likelihood function. This is backed up with the intuition that these are precisely the counterfactual quantities that are inherently uninformed by the data, i.e., the mean response for never-takers if they take, and the mean response for always-takers if they don't take.

Consider taking the prior $\pi(\theta)$ to be a uniform distribution, i.e., a Dirichlet(1,1,1) prior for $\omega = (\omega_{NT}, \omega_{AT}, \omega_{CO})$ and a uniform prior on $(0, 1)^6$ for the elements of γ, with *a priori* independence between ω and γ. Now, say the target of inference is the (global) average causal effect (ACE), given as:

$$
\begin{aligned}
ACE &= (1 - \omega_{NT} - \omega_{AT})(\gamma_{CO,1} - \gamma_{CO,0}) + \omega_{NT}(\gamma_{NT,1} - \gamma_{NT,0}) + \\
&\quad \omega_{AT}(\gamma_{AT,1} - \gamma_{AT,0}).
\end{aligned}
$$

Gustafson [6] shows that for n independent and identically distributed realizations of (Y, X, Z), as $n \to \infty$, the posterior distribution of ACE converges to a symmetric, trapezoidal density. Particularly, let

$$\begin{aligned}
a(\phi_I) &= (1-\omega_{NT}-\omega_{AT})(\gamma_{CO,1}-\gamma_{CO,0}) + \omega_{NT}(1/2-\gamma_{NT,0}) \; + \\
&\quad\; \omega_{AT}(\gamma_{AT,1}-1/2), \\
b(\phi_I) &= (\omega_{NT}+\omega_{AT})/2, \\
b^*(\phi_I) &= |\omega_{NT}-\omega_{AT}|/2.
\end{aligned}$$

The support of the limiting trapezoidal density is $a(\phi_I) \pm b(\phi_I)$ and the 'top' is $a(\phi_I) \pm b^*(\phi_I)$, so consequently the height of the trapezoid is $\{b(\phi_I)+b^*(\phi_I)\}^{-1}$.

Since the limiting posterior is symmetric, equal-tailed and HPD credible intervals agree (with HPD suitably interpreted in light of the 'flat-topped' density). Thus the limiting level $1-\alpha$ credible interval for the target will have the form

$$a(\phi_I) \quad \pm \quad k_\alpha(\omega_{AT}/\omega_{NT})(\omega_{NT}+\omega_{AT})/2,$$

where

$$k_\alpha(r) \quad = \quad \begin{cases} \{(1-\alpha)/2\}\{1+g(r)\} & \text{if } g(r) > (1-\alpha)/(1+\alpha), \\ 1-\sqrt{\alpha\{1-g(r)^2\}} & \text{otherwise,} \end{cases}$$

with $g(r) = |1-r|/(1+r)$. Note that the limiting interval is narrowest relative to the support of the limiting posterior when $\omega_{AT} = \omega_{NT}$, with the limiting density becoming triangular and $k_\alpha(1) = 1-\sqrt{\alpha}$. The interval is widest when $\omega_{AT} = 0$ or $\omega_{NT} = 0$, with the limiting density becoming rectangular and $k_\alpha(r) \to 1-\alpha$ as $r \to 0$ or $r \to \infty$.

To gain some intuition, we simulate draws from the prior $\pi(\theta)$ and ascertain which realizations fall in $B_\pi^{(\alpha)}$, for $\alpha = 0.05$. Theory tells us that the proportion of draws falling in this 'bad' set is α, but the results in Figure 1 show that the location of $B_\pi^{(\alpha)}$ in the parameter space is not understood trivially. First, membership in $B_\pi^{(\alpha)}$ is not driven exclusively by the values of $\phi_N = (\gamma_{AT,0}, \gamma_{NT,1})$. Plots show that a necessary, but not sufficient, condition for failure-to-cover is that these two parameters take on extreme values in opposite directions. That is, 'failures' are found to have ϕ_N values in the bottom-right or upper-left corners of the unit-square. However, 'successes' are found in these regions as well, with no smooth boundary in ϕ_N-space to separate $B_\pi^{(\alpha)}$ from its complement.

The other feature evident from Figure 1 is that failure-to-cover is not strongly driven by the value of the target. The prior distribution of $(\psi|\theta \in B_\pi^{(\alpha)})$ is more dispersed than that of $(\psi|\theta \notin B_\pi^{(\alpha)})$, but not to a great extent. Put more succinctly, many θ values for which failure occurs have $\psi = g(\theta)$ values near the middle of the prior for ψ, and many of the θ values yielding ψ values in the tails of the prior correspond to successful coverage. Thus there is not a direct explanation for failure-to-cover in terms of the unidentified parameters alone, nor in terms of the target parameter alone.

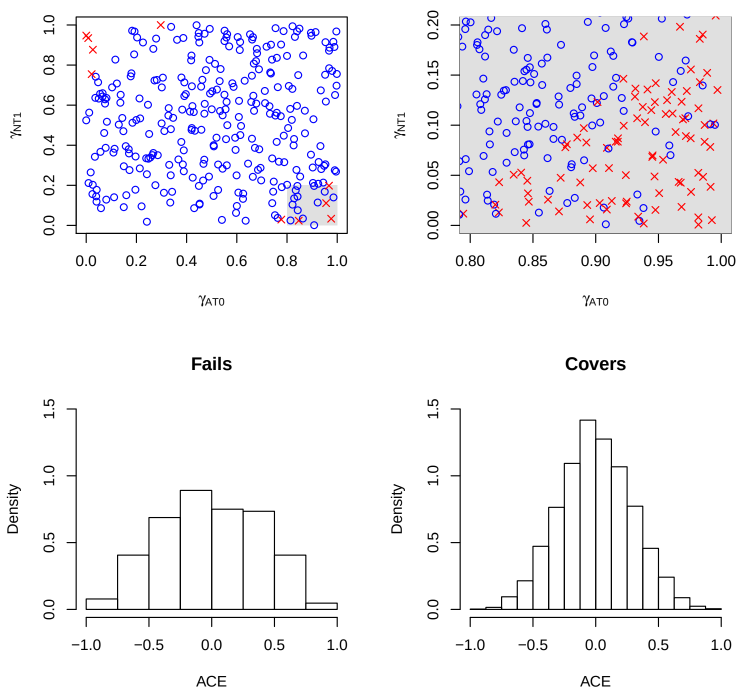

Figure 1: Exploration of coverage in Example 1, via simulated draws from the prior. The upper-left panel gives the $(\gamma_{AT,0}, \gamma_{NT,1})$ values of the draws, with success/failure to cover indicated by 'o'/'x'. The upper-right panel focuses only on the shaded corner region of the upper-left panel. The lower panels depict simulated values of the target ACE, stratified by failure/success to cover.

To explore in more depth, from the form of the ACE and the form of the limiting credible interval, it is apparent that failure to cover occurs when

$$|(\gamma_{NT,1} - 1/2) - (\omega_{AT}/\omega_{NT})(\gamma_{AT,0} - 1/2)| \quad > \quad k_\alpha \left(\frac{\omega_{AT}}{\omega_{NT})} \right) \frac{1 + \omega_{AT}/\omega_{NT}}{2}.$$

This underscores that failure is not determined exclusively by the values of of nonidentified parameters $\phi_N = (\gamma_{NT,1}, \gamma_{AT,0})$, since the failure region depends on identified parameters via the ratio $r = \omega_{AT}/\omega_{NT}$. Particularly, the failure region intersected with a fixed value of $r \in (0, 1)$ corresponds to equal-sized triangles in the lower-right and upper-left corners of the unit-square describing $(\gamma_{AT,0}, \gamma_{NT,1}) \in [0, 1]^2$. More particularly, these triangles are formed via the parallel lines of slope r,

$$\gamma_{NT,1} - 1/2 \quad = \quad r(\gamma_{AT,0} - 1/2) \pm k_\alpha(r)(1 + r)/2, \tag{4}$$

passing through the unit square. While the size of the triangles could be deduced from (4), direct appeal to Theorem 1 immediately gives the areas as $\alpha/2$ each.

Figure 2 gives a pictorial representation of how the failure-to-cover region for ϕ_N varies with $r = \omega_{AT}/\omega_{NT}$. In the $r = 0$ limit the region devolves to rectangles defined by $\gamma_{NT,1} < \alpha/2$ and $\gamma_{NT,1} > 1 - \alpha/2$. This is intuitively sensible: when there are fewer (or no) always-takers, success/failure of the interval will be determined more (or exclusively) by the value of $\gamma_{NT,1}$. Congruently, $\gamma_{AT,0}$ plays a larger role as r becomes large. This behaviour explains the lack of separation between $B_\pi^{(\alpha)}$ in terms of ϕ_N alone, i.e., how it is possible to have successes lying 'south-east' of failures in the top right panel of Figure 1.

Roughly put, the situation with this example can be summarized as follows. Failure-to-cover arises when the nonidentified parameters lie in 'a tail' of their prior distribution, but 'which tail' depends on a particular aspect of the identified parameters. This interaction between nonidentified and identified components is more complicated than simply seeing failure-to-cover when the value of the inferential target is extreme compared to its prior. However the interaction is quite understandable, in terms of which nonidentified parameter is a bigger determinant of failure-to-cover when the relevant identified quantity is smaller or larger.

3.2 Example: Prevalence of Misclassified Trait

Say that X is a binary trait, with interest lying in its population prevalence $r = Pr(X = 1)$. However, the observable binary variable is $\tilde{X}$, which is subject to misclassification, i.e., $Pr(\tilde{X} = 1) = r\gamma_N + (1-r)(1-\gamma_P)$, where $\gamma_N = Pr(\tilde{X} = 1|X = 1)$ and $\gamma_P = Pr(\tilde{X} = 0|X = 0)$ are the sensitivity and specificity of the classification scheme respectively.

Consider the situation where the investigator commits to lower bounds on sensitivity and specificity but applies uniform prior distributions above the bounds and also applies a uniform prior to the target parameter r. That is

$$\pi(r, \gamma_N, \gamma_P) \quad = \quad (1 - a)^{-1}(1 - b)^{-1} I_{(0,1)}(r) I_{(a,1)}(\gamma_N) I_{(b,1)}(\gamma_P).$$

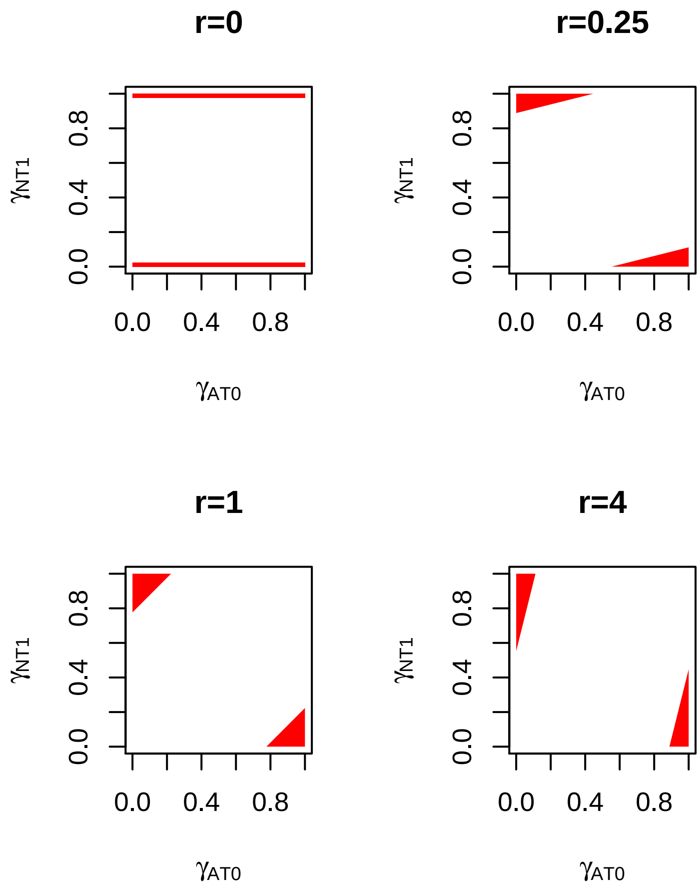

Figure 2: The failure-to-cover region intersected with different values of $r = \omega_{AT}/\omega_{NT}$ in Example 1, with $\alpha = 0.05$.

Clearly in the large-sample limit of observing independent and identically distributed realizations of $\tilde{X}$, conditioning on the observed data is equivalent to conditioning on the true value of $\tilde{r} = r\gamma_N + (1-r)(1-\gamma_P)$. More particularly, a transparent parameterization is obtained by taking $\phi_I = \tilde{r}$ and $\phi_N = (\gamma_N, \gamma_P)$, and the large-sample limit of the posterior distribution over the target r is identically the prior conditional distribution $\pi(r|\tilde{r})$ (where the conditioning is on the true value of $\tilde{r}$). This distribution is given in Gustafson [5] as having support

$$r \quad \in \left(1 - \tfrac{1-\tilde{r}}{b^*}, \tfrac{\tilde{r}}{a^*}\right), \tag{5}$$

where $a^* = \max\{a, \tilde{r}\}$, $b^* = \max\{b, 1-\tilde{r}\}$. The interval (5) is then the identification region for the target parameter in this partially identified model. Moreover, the conditional prior density (and equivalently the limiting posterior density) over this support is given by:

$$\pi(r|\tilde{r}) \quad \propto \quad r^{-1}\left[\min\left\{\frac{\tilde{r}-a^*r}{1-r}, 1-b^*\right\} - \max\left\{\frac{\tilde{r}-r}{1-r}, 0\right\}\right]. \tag{6}$$

For the hyperparameters $a = 0.7$ and $b = 0.85$, examples of the limiting posterior density (6), for different values of $\tilde{r}$, appear in Figure 3.

The form of (6) is such that limiting credible intervals are readily computed, but closed-form expressions for their endpoints would be cumbersome. We focus on 90% HPD credible sets. In the case of hyperparameters $(a, b) = (0.7, 0.85)$, we draw values of $\theta = (r, \gamma_N, \gamma_P)$ from the prior distribution, and check membership in $B_\pi^{(\alpha)}$, for $\alpha = 0.05$. The top panels of Figure 4 show that failure-to-cover can occur when one of sensitivity and specificity is large but the other is small. Echoing findings in the previous example, however, failure-to-cover is not completely determined by the extremity of $\phi_N = (\gamma_N, \gamma_P)$, i.e., $B_\pi^{(\alpha)}$ is not determined by a smooth boundary in ϕ_N-space.

Again we appeal to Theorem 1. Drawing values of θ by sampling from $\pi(\gamma_N, \gamma_P|\tilde{r} = \tilde{r}_0)$ for selected values of $\tilde{r}_0$, we examine the intersection of $B_\pi^{(\alpha)}$ with $\{\theta : \tilde{r}(\theta) = \tilde{r}_0\}$. Theorem 1 guarantees this intersection to have probability α with respect to the conditional prior, while Figure 4 illustrates that this set is determined by a smooth boundary in ϕ_N-space. Moreover, this boundary is seen to vary with $\tilde{r}_0$. Thus again we see that failure-to-cover arises when ϕ_N lies in a tail of the prior distribution, but *which* tail depends on an identified quantity. As an aside, Figure 4 also illustrates the 'indirect learning' that can arise because the support of $\pi(\gamma_N, \gamma_P|\tilde{r} = \tilde{r}_0)$ can, for some values of $\tilde{r}_0$, be smaller than the support of $\pi(\gamma_N, \gamma_P)$.

To touch briefly on finite-sample performance, using hyperparameters $(a, b) = (0.7, 0.9)$, we consider five successively more extreme values of θ given by $\tilde{r} = 0.25$ along with (a) $(\gamma_N, \gamma_P) = (0.75, 0.95)$, (b) $(\gamma_N, \gamma_P) = (0.74, 0.96)$, (c) $(\gamma_N, \gamma_P) = (0.73, 0.97)$, (d) $(\gamma_N, \gamma_P) = (0.72, 0.98)$, (e) $(\gamma_N, \gamma_P) = (0.71, 0.99)$. For each value we numerically evaluate the frequentist coverage of the 90% equal-tailed credible interval for r, for a variety of sample sizes. While the asymptotic results in Figure 4 pertain to HPD intervals, it is somewhat easier to control

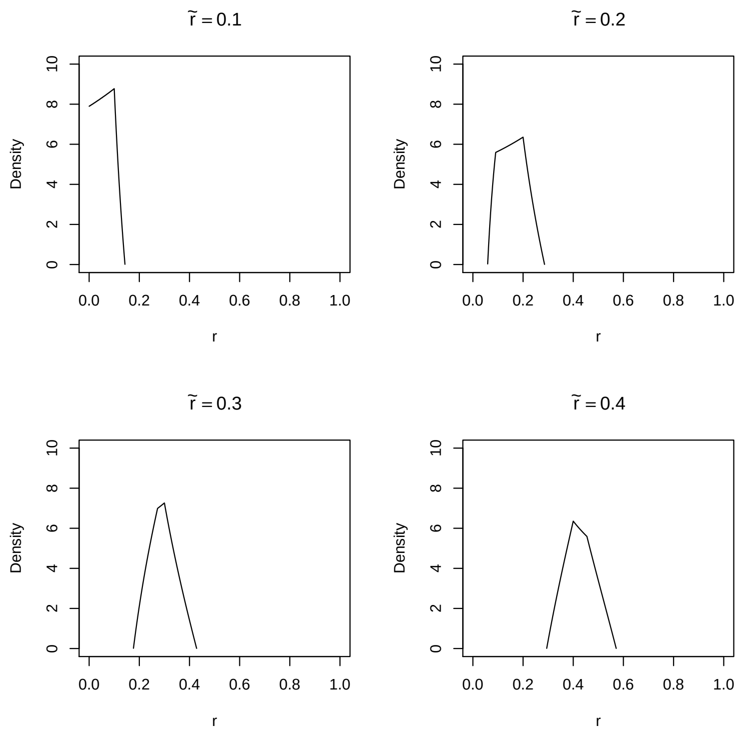

Figure 3: The limiting posterior density on r in Example 2, in the cases that $\tilde{r} = 0.1, 0.2, 0.3, 0.4$. The hyperparameters, in the form of lower bounds on sensitivity γ_N and specificity γ_P, are $a = 0.7$ and $b = 0.85$.

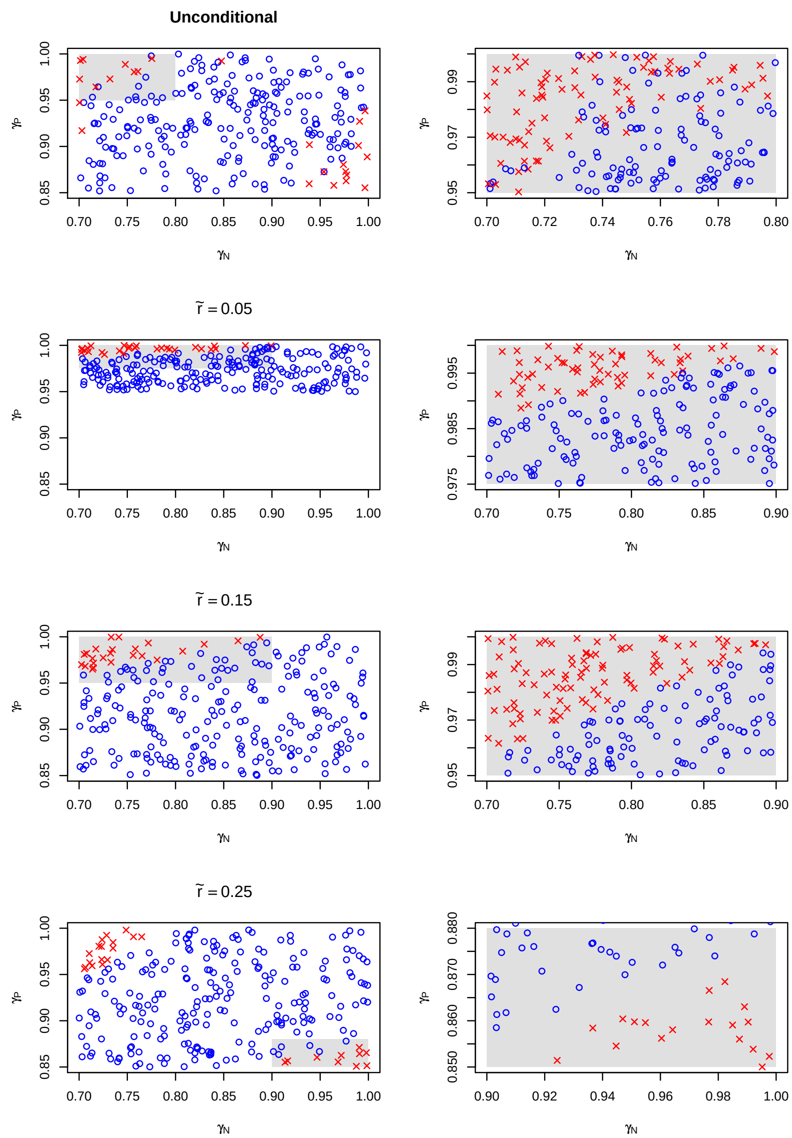

Figure 4: Failure to cover in Example 2. The top panels report asymptotic success ('o') or failure ('x') to cover, for a sample of θ values drawn from the unconditional prior $\pi(\theta)$. The remaining panels report asymptotic success/failure to cover for a sample of θ values obtained via draws from the conditional prior $\pi(\gamma_N, \gamma_p|\tilde{r})$, for $\tilde{r} = 0.05$ (second row), $\tilde{r} = 0.15$ (third row), and $\tilde{r} = 0.25$ (fourth row). In each case, the right-hand plot is an enlargement of the shaded region in the left-hand plot, along with an addition of more sampled points.

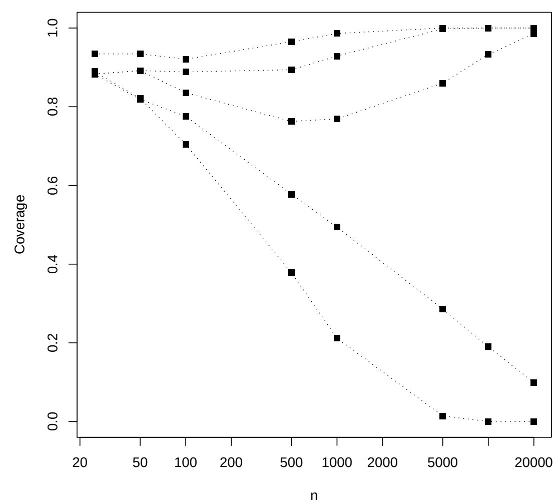

Figure 5: Frequentist coverage of the 90% equal-tailed credible interval as a function of n, for the five selected values of θ described in the text. The coverage decreases as we move from values (a) through (e), i.e., further toward the corner of the prior support for (γ_N, γ_P). At each selected value of n and θ, coverage is approximated via 4000 realized data values, implying a Monte Carlo simulation standard error for coverage evaluation of 0.008 or less. Note the use of a logarithmic axis for n.

the numerical error in computing finite-sample coverage in the equal-tailed case. Details of the numerical evaluation are given in the Appendix.

For each θ value the frequentist coverage is plotted against sample size n in Figure 5. As is consistent with theory, the coverage is seen to tend to zero or one as n increases. Particularly, the limit is one for the less extreme cases (a) through (c), and zero for the more extreme cases (d) and (e). In these latter instances convergence is slow in practical terms.

4 Discussion

As illustrated in our examples, as the sample size increases we learn both the support and the shape of the limiting posterior distribution of the target, with the limiting support corresponding to the identification region. In the first example, different identified quantities control the support $(a(\phi_I), b(\phi_I))$ and the shape $(b^*(\phi_I))$, such that two different distributions of observables could produce the same identification region but different limiting posterior distributions over this region. In the second example, the distribution of observables is governed by a single parameter ($\tilde{r}$) which gives rise to both the support and shape of the limiting posterior. In either case, the limiting credible interval is not a 'pre-ordained' subset of the identification region - the data have something to say.

Both examples possess parameters which intuitively aren't informed by the data. In the first example these are the counterfactual average responses of never-takers who take and always-takers who don't take. In the second example these are the sensitivity and specificity of the classification scheme. Intuition suggests we can only do sensitivity analysis with respect to such parameters, and indeed applying prior distributions to them and proceeding with Bayesian inference is often regarded as a probabilistic form of sensitivity analysis (see, for instance, Greenland [2, 3]). Thus we might only expect a 'wrong' answer to arise if the true values of the uninformed-by-data parameters are extreme with respect to the chosen priors. We have shown this to be correct in the main, but with some devil lurking in the details. Notably, what constitutes extreme can vary with identified parameters. While Theorem 1 precludes the size of the extreme set varying with an identified quantity, our examples illustrate that the location of the extreme set can vary with an identified quantity.

Appendix

The numerical evaluation of finite-sample frequentist coverage in Example 2 proceeds as follows. Assume the hyperparameters satisfy $a + b > 1$, so that the prior ensures $1 - \gamma_P < \gamma_N$. Upon reparameterizing to $(\tilde{r}, \gamma_N, \gamma_P)$ we have the prior density

$$\pi(\tilde{r}, \gamma_N, \gamma_P) \quad \propto \quad (\gamma_N + \gamma_P - 1)^{-1} I_{(1-\gamma_P, \gamma_N)}(\tilde{r}) I_{(a,1)}(\gamma_N) I_{(b,1)}(\gamma_P).$$

Then observing y out of n units to have $\tilde{X} = 1$, yields updating as

$$\pi(\tilde{r}, \gamma_N, \gamma_P | y) \quad \propto \quad \tilde{r}^y (1 - \tilde{r})^{n-y} \pi(\tilde{r}, \gamma_N, \gamma_P). \tag{7}$$

Note that the event $r < r_0$ can be reexpressed as $\tilde{r} < (1 - r_0)(1 - \gamma_P) + r_0 \gamma_N$, and that integration of (7) with respect to $\tilde{r}$ can be expressed via the Beta$(y + 1, n - y + 1)$ distribution function, which we denote as $G_y()$. Thus

$$pr(r < r_0 | y) \quad = \quad \frac{E\left[\frac{G_y\{(1-r_0)(1-\gamma_P)+r_0\gamma_N\} - G_y(1-\gamma_P)}{\gamma_N+\gamma_P-1}\right]}{E\left[\frac{G_y(\gamma_N) - G_y(1-\gamma_P)}{\gamma_N+\gamma_P-1}\right]}, \tag{8}$$

where the expectations are with respect to the prior distribution of (γ_N, γ_P). Thus a Monte Carlo sample from this distribution can be used to numerically approximate both expectations in (8). Moreover, since this takes the form of a ratio estimator, a Monte Carlo standard error is readily obtained to quantify the approximation error. The results appearing in Figure 5 use 20000 realizations of (γ_N, γ_P).

Thus for a given θ_0 value and given sample size n, the frequentist coverage of the $1-\alpha$ equal-tailed credible interval is obtained by repeatedly simulating a value $Y = y$ and reporting the proportion of times for which $pr(r < r_0|Y = y)$ lies between $\alpha/2$ and $1-\alpha/2$. Note that for a given y this only requires numerical evaluation of (8) at a single value of r_0, whereas checking coverage for an HPD interval would require evaluation over a fine grid of values.

References

[1] Chickering, D. and Pearl, J. (1996). A clinician's tool for analyzing non-compliance. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96),* Portland, OR, volume 2, pages 1269–1276.

[2] Greenland, S. (2003). The impact of prior distributions for uncontrolled confounding and response bias: A case study of the relation of wire codes and magnetic fields to childhood leukemia. *Journal of the American Statistical Association* **98,** 47–55.

[3] Greenland, S. (2005). Multiple-bias modelling for analysis of observational data. *Journal of the Royal Statistical Society, Series A* **168,** 267–306.

[4] Gustafson, P. (2005). On model expansion, model contraction, identifiability, and prior information: two illustrative scenarios involving mismeasured variables (with discussion). *Statistical Science* **20,** 111–140.

[5] Gustafson, P. (2010). Bayesian inference for partially identified models. *International Journal of Biostatistics* **6,** issue 2 article 17.

[6] Gustafson, P. (2011). Comment on 'Transparent parameterizations of models for potential outcomes,' by Richardson, Evans, and Robins. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., editors, *Bayesian Statistics 9*, page in press. Oxford University Press.

[7] Gustafson, P. and Greenland, S. (2009). Interval estimation for messy observational data. *Statistical Science* **24,** 328–342.

[8] Imbens, G. W. and Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica* **72,** 1845–1857.

[9] Imbens, G. W. and Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* **25,** 305–327.

[10] Liao, Y. and Jiang, W. (2010). Bayesian analysis in moment inequality models. *Annals of Statistics* **38,** 275–316.

[11] Moon, H. R. and Schorfheide, F. (2011). Bayesian and frequentist inference in partially identified models. *Econometrica, to appear* .

[12] Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

[13] Richardson, T. S., Evans, R. J., and Robins, J. M. (2011). Transparent parameterizations of models for potential outcomes. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., editors, *Bayesian Statistics 9*, page in press. Oxford University Press.

[14] Romano, J. P. and Shaikh, A. M. (2008). Inference for identifiable parameters in partially identified econometric models. *Journal of Statistical Planning and Inference* **138,** 2786–2807.

[15] Vansteelandt, S., Goetghebeur, E., Kenward, M. G., and Molenberghs, G. (2006). Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica* **16,** 953–979.

[16] Zhang, Z. (2009). Likelihood-based confidence sets for partially identified parameters. *Journal of Statistical Planning and Inference* **139,** 696–710.