

On the behaviour of Bayesian credible intervals in partially identified models

Paul Gustafson

*Department of Statistics
University of British Columbia
Vancouver, B.C. CANADA V6T 1Z4
e-mail: gustaf@stat.ubc.ca*

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 and phrases: Bayesian inference, credible interval, partial identification.

Received August 2011.

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 concept of partial identification is surveyed by Manski [16], with recent literature on interval estimation in partially identified models including Imbens and Manski [12], Vansteelandt et al. [22], Romano and Shaikh [21], and Zhang [23]. 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 [17], Liao and Jiang [15], and Gustafson and Greenland [11]. One particular feature distinguishing the Bayesian approach is

that the large-sample limit of the posterior distribution will generally assign more weight to some values in the identification region than others. The extent to which learning the shape of this distribution is inferentially useful, over and above learning the identification region, is taken up in Gustafson [9]. As emphasized in Gustafson and Greenland [11], 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. [22] 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.

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. Since the posterior density will be pushed upward (downward) in regions where the prior density is high (low), we might look to parameter values in the tails of the prior distribution as values under which generated datasets would produce credible intervals missing their targets. As well, the values of those parameter components which are less informed by the data might be suspected to play a bigger role in whether or not the limiting credible interval contains its target. 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 $\psi = g(\theta)$. As a point to bear in mind for later, we presume that the proper prior for θ is elicited with regard to the scientific interpretations of the elements of θ only, and *without regard* to whether or not the model $\pi(D|\theta)$ is identified. For instance, imagine an investigator must commit to a prior on the scientific quantities at hand before finding out whether he will receive (i), a large budget for data acquisition which permits gold-standard measurements and consequently an identified model, or (ii), a small budget which necessitates cheaper proxy measurements giving rise to only a partially identified model.

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 based on prior π . The two primary examples of ‘types’ would be equal-tailed intervals and highest posterior density (HPD) sets. We do not consider ‘pathological’ choices of credible sets, such as sets which purposefully exclude values in the ‘middle’ of the posterior distribution. This proviso makes it more meaningful to discuss credible interval failure arising when true parameter values lie in the tail of the prior distribution.

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

$$pr_\pi \left\{ \psi \in I_\pi^{(\alpha)}(D_n) \right\} = 1 - \alpha. \tag{2.1}$$

This follows immediately upon writing the probability on the left-hand side as an iterated expectation, with the inner expectation taken with respect to the posterior distribution of $(\theta|D_n)$. We can interpret (2.1) 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, (2.1) 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 (2.1) 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 $n \rightarrow \infty$. Emphasizing that this limit depends on the underlying value of θ which spawns the data sequence D_n , 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 (2.1) it follows that

$$pr_\pi \left\{ \psi \in J_\pi^{(\alpha)}(\theta) \right\} = 1 - \alpha. \tag{2.2}$$

This is simply the limiting version of (2.1). Of course now the event in question involves parameters only, so that the probability is with respect to the prior distribution on θ . We can interpret (2.2) as saying that the subset of the parameter

space for which the limiting frequentist coverage of the Bayesian interval is zero must have probability mass α with respect to the prior π .

Let $B_\pi^{(\alpha)} = \{\theta : g(\theta) \notin J_\pi^{(\alpha)}(\theta)\}$ be the ‘bad’ subset of the parameter space on which the limiting credible interval fails to cover its target. That is, for given prior π and choice of α , $B_\pi^{(\alpha)}$ is the set of all θ values such that *if* data were generated under θ , then the credible interval, in the large-sample limit, would fail to contain its target. Equivalently, this is the subset of parameter values under which the limiting frequentist coverage of the credible interval is zero. As stated above, it is immediate from (2.2) that $B_\pi^{(\alpha)}$ is small in the sense of having probability α under the joint prior π across the parameter space. 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 by a number of authors (Barankin [1], Kadane [14], Dawid [4], Poirier [19], Gustafson [7]), the posterior structure in many partially identified models can be laid bare upon reparameterizing from θ to (ϕ, λ) such that (i) $D_n \perp \lambda|\phi$, and (ii) $(D_n|\phi)$ comprises a ‘regular’ model admitting root- n consistent estimation of ϕ . Barankin [1] calls ϕ a sufficient parameter, and emphasizes the need to choose a ‘minimal sufficient’ ϕ , i.e., so that the distribution of the data is not completely determined by a non-invertible function of ϕ . Gustafson [7] calls (ϕ, λ) a *transparent parameterization*. Informally, the key feature is that ϕ appears in the likelihood function, while λ does not. The limiting posterior distribution of ϕ decomposes into a point-mass at the true value of ϕ combined with the prior conditional distribution of $(\lambda|\phi)$. One can then determine the induced limiting posterior distribution on $\psi = g(\theta(\phi))$. When useful we will use $\pi^*(\cdot)$ to distinguish the prior density for (ϕ, λ) from the prior density $\pi(\cdot)$ for θ , i.e.,

$$\pi^*(\phi, \lambda) = \pi(\theta(\phi, \lambda)) \left| \frac{\partial \theta}{\partial \phi, \lambda} \right|. \quad (2.3)$$

The limiting posterior density for λ is immediately obtained, up to a constant of proportionality, by viewing (2.3) as a function of λ with ϕ fixed at the true value.

As a motivating example in which a natural initial parameterization is already transparent, say that interest is focussed on $\lambda = E(Y)$, where $\lambda \in (0, 2)$. However, due to interval censoring, Y_L rather than Y is observed, where $Y_L \leq Y \leq Y_L + 1$, with probability one. Moreover, say the distribution of Y_L is known up to its mean $\phi = E(Y_L)$, and it is also known *a priori* that $0 < \phi < 1$. Clearly $\theta = (\phi, \lambda)$ comprises a transparent parameterization, with $\psi = g(\theta) = \lambda$ being the inferential target. An identification region of the form $\lambda \in (\phi, \phi + 1)$ results. Moreover, say any marginal prior with full support $\pi(\phi)$ is chosen, along with the conditional prior $(\lambda|\phi) \sim Unif(\phi, \phi + 1)$. Then the limit of the posterior distribution on the interest parameter λ is a uniform distribution on $(\phi_0, \phi_0 + 1)$, where ϕ_0 is the true value of ϕ . Immediately then $B_\pi^{(\alpha)}$ is the subset of the (ϕ, λ)

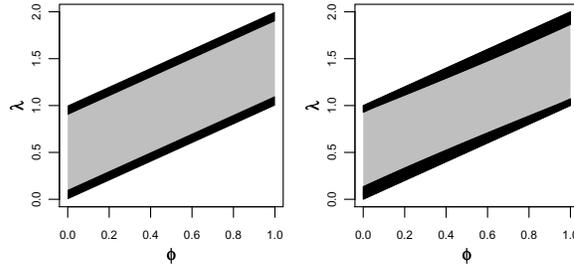


FIG 1. The failure-to-cover set $B_\pi^{(0.2)}$ in Example 1. The left panel gives the case of a marginal prior $\pi(\phi)$ with full support, along with $(\lambda|\phi) \sim \text{Unif}(\phi, \phi + 1)$. The right panel gives the case of the prior $\lambda \sim \text{Unif}(0, 2)$, $(\phi|\lambda) \sim \text{Unif}[\max(\lambda - 1, 0), \min(\lambda, 1)]$. In each case, $B_\pi^{(0.2)}$ appears in black, while its complement is in grey.

parameter space on which $\lambda \in (\phi, \phi + \alpha/2)$ or $\lambda \in (\phi + 1 - \alpha/2, \phi + 1)$. In the case of $\alpha = 0.2$, this is displayed graphically in the left panel of Figure 1.

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 2.1. *Let $h(\theta)$ be any identified quantity, i.e., with respect to a transparent parameterization, $h(\theta(\phi, \lambda))$ does not vary with λ . Then, for any value of h_0 in the support of the prior induced on $h(\theta)$, we have*

$$pr_\pi(\theta \in B_\pi^{(\alpha)} | h(\theta) = h_0) = \alpha. \tag{2.4}$$

That is, with respect to the conditional prior distribution of $(\theta|h(\theta) = h_0)$, the failure-to-cover subset of the parameter space has probability α .

Proof. For a given h_0 , let $B_0^{(\alpha)}$ be the set of θ values for which the limiting $1 - \alpha$ credible interval arising from the prior $\pi(\theta|h(\theta) = h_0)$ fails to cover the target. Now note that for any θ satisfying $h(\theta) = h_0$, $\theta \in B_0^{(\alpha)} \iff \theta \in B_\pi^{(\alpha)}$. This happens because for data generation under such a θ , the same limiting credible interval for ψ arises under $\pi(\theta)$ as under $\pi(\theta|h(\theta) = h_0)$. That is, since asymptotically the data already reveal the value of ϕ , correct *a priori* conditioning on some function of ϕ will have no effect whatsoever in the large- n limit. Thus

$$\begin{aligned} pr_\pi(\theta \in B_\pi^{(\alpha)} | h(\theta) = h_0) &= pr_\pi(\theta \in B_0^{(\alpha)} | h(\theta) = h_0) \\ &= \alpha \end{aligned}$$

as desired, where the second equality follows directly from (2.2) applied with the prior $\pi(\theta|h(\theta) = h_0)$. \square

The theorem says that the performance of the limiting credible interval cannot be driven exclusively by the compatibility of identified parameters and their priors. Regardless of whether h_0 is an *a priori* likely or unlikely value of $h(\theta)$,

(2.4) dictates that the intersection of the failure-set $B_\pi^{(\alpha)}$ with $\{\theta : h(\theta) = h_0\}$ has ‘size’ α , where size is taken as probability under the prior distribution of $\{\theta|h(\theta) = h_0\}$. A consequence of this, crudely stated, is that coverage can never be guaranteed to occur solely because an identified quantity happens to lie near its prior mode, nor can failure-to-cover be guaranteed just because an identified quantity lies in the extreme tails of its prior. This provides a sense in which the values of unidentified quantities *must* play a role in determining whether coverage occurs.

3. Examples

3.1. Interval censoring, continued

Consider the interval censoring problem described in the previous section. We have already seen the form of the failure-to-cover set in this problem, for a prior under which ϕ has full support while $(\lambda|\phi)$ is uniformly distributed. We now examine the failure-to-cover set under a different prior specification, namely a uniform marginal prior for λ and a uniform conditional prior for $(\phi|\lambda)$. More specifically, $\lambda \sim \text{Unif}(0, 2)$, $(\phi|\lambda) \sim \text{Unif}[\max(\lambda - 1, 0), \min(\lambda, 1)]$, results in a joint prior with full support on the (ϕ, λ) parameter space. We then have the limiting posterior distribution governed by the prior conditional density

$$\begin{aligned} \pi(\lambda|\phi) &\propto \frac{1}{\min(\lambda, 1) - \max(\lambda - 1, 0)} I_{(\phi, \phi+1)}(\lambda) \\ &\propto \left\{ \frac{I_{(0,1)}(\lambda)}{2 - \lambda} + \frac{I_{(1,2)}(\lambda)}{\lambda} \right\} I_{(\phi, \phi+1)}(\lambda). \end{aligned}$$

Since this is readily integrated analytically, the $\alpha/2$ and $1 - \alpha/2$ quantiles of $\pi(\lambda|\phi)$ are easily obtained for a given ϕ , and they constitute the limit of the $1 - \alpha$ equal-tailed credible interval. Values of (ϕ, λ) for which λ falls outside the interval constitute $B_\pi^{(\alpha)}$. This ‘bad set,’ as illustrated in the right panel of Figure 1, is seen to differ from that arising from the earlier choice of prior. Particularly the top (bottom) failure band widens (narrows) as ϕ increases. We also note that the Figure is consistent with Theorem 2.1, with different ϕ values giving rise to the same chance of failure to cover.

3.2. Prevalence estimation with nonignorable missingness

As a second example where the failure-to-cover set is readily characterized, consider estimating the prevalence of a binary trait in the face of nonignorable missingness. We seek to infer $p = \text{Pr}(X = 1)$ based on *iid* observations of (R, XR) , i.e., R is a binary variable with $R = 0$ indicating missingness of X . The unknown parameters are $\theta = (p, q_0, q_1)$, where $q_i = \text{Pr}(R = 1|X = i)$, for $i = 0, 1$, while $\psi = g(\theta) = p$ is the inferential target. A transparent parameterization is obtained by taking $\phi = (s, t)$, $\lambda = p$, where $s = \text{Pr}(R = 1) = (1 - p)q_0 + pq_1$

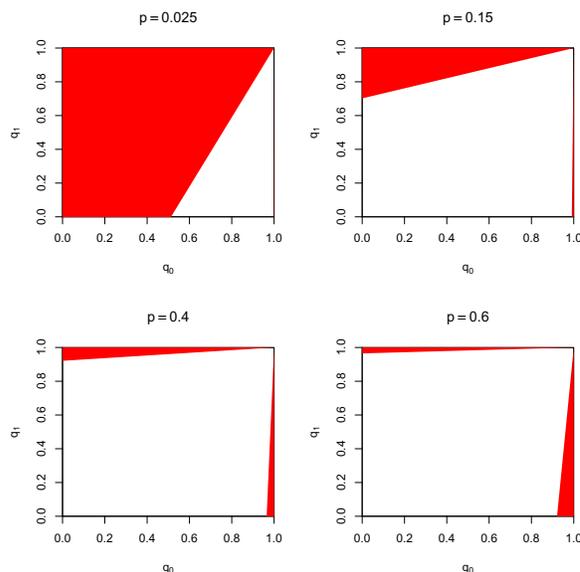


FIG 2. The failure-to-cover set $B_{\pi}^{(0.1)}$ in Example 2. For select values of the target prevalence p , the values of (q_0, q_1) for which failure-to-cover occurs are shaded.

and $t = Pr(X = 1|R = 1) = pq_1/\{(1 - p)q_0 + pq_1\}$. Immediately we have

$$\pi^*(p|s, t) \propto \pi(\theta(p, s, t)) \frac{1}{p(1 - p)} I\{st < p < st + (1 - s)\}. \quad (3.1)$$

For the sake of a very tractable illustration, say that the prior π on θ is chosen such that (p, q_0, q_1) are mutually independent, with $p \sim \text{Beta}(2, 2)$ and $q_i \sim \text{Unif}(0, 1)$. This can be interpreted as a ‘flat’ prior on the parameters governing the (R, X) association, along with a prior on the target prevalence which slightly downweights lower/higher values. Under this specification, (3.1) specializes to a limiting posterior distribution on p which is uniform on the identification interval $(st, st + (1 - s))$. Thus for an equal-tailed $1 - \alpha$ credible interval, failure-to-cover occurs if $p < st + (\alpha/2)(1 - s)$ or $p > st + (1 - \alpha/2)(1 - s)$. Reexpressed in terms of the original parameterization θ , we have

$$B_{\pi}^{(\alpha)} = \{\theta : q_1 > 1 - \{\alpha/(2 - \alpha)\}\{(1 - p)/p\}(1 - q_0)\} \cup \{\theta : q_1 < 1 - \{(2 - \alpha)/\alpha\}\{(1 - p)/p\}(1 - q_0)\}.$$

For fixed p then, the failure-to-cover region is described by linear boundaries in the (q_0, q_1) plane. This is depicted in Figure 2 for a few values of p . We see that an extreme value of p can lead to failure-to-cover for a ‘majority’ of q values. For a mid-range value of p , however, either q_0 or q_1 must be near one for failure to arise, with the requirement for nearness weaker if the other q is close to zero. For this problem then, we have a fairly simple and full understanding of the way in which $\theta = (p, q_0, q_1)$ must be extreme in order for the credible interval to miss the target.

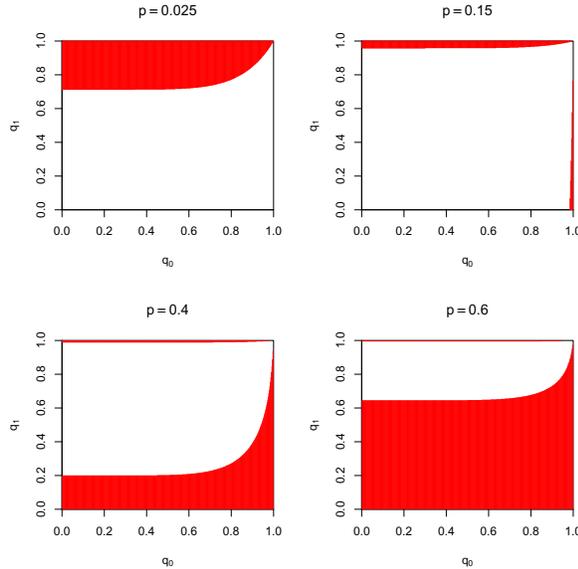


FIG 3. The failure-to-cover set $B_{\pi}^{(0.1)}$ when the prior distribution on the target is changed to $p \sim \text{Beta}(2, 8)$.

This example also lends itself to investigation of how the failure-to-cover set $B_{\pi}^{(\alpha)}$ depends on the choice of prior π . Say we retain uniform prior distributions on q_0, q_1 but generalize from $p \sim \text{Beta}(2, 2)$ to $p \sim \text{Beta}(a, b)$, so that the limiting posterior distribution on p follows a $\text{Beta}(a - 1, b - 1)$ distribution truncated to the identification interval. Then

$$B_{\pi}^{(\alpha)} = \left\{ (p, q_0, q_1) : \frac{F(p) - F(pq_1)}{F(p + (1 - p)(1 - q_0)) - F(pq_1)} \notin \left(\frac{\alpha}{2}, 1 - \frac{\alpha}{2} \right) \right\},$$

where $F()$ is the $\text{Beta}(a - 1, b - 1)$ distribution function. For instance, say that background knowledge speaks to high prevalence being unlikely, so that investigators settle on hyperparameters $(a, b) = (2, 8)$. Figure 3 indicates that the resulting failure-to-cover set is quite different than for $(a, b) = (2, 2)$, with highly nonlinear boundaries. There is a temptation to make direct comparisons between Figures 2 and 3. For instance, one notes that the intersection of $B_{\pi}^{(0.1)}$ with $\{p = 0.4\}$ is smaller under the first prior than under the second, whereas the reverse is true when we intersect with $\{p = 0.15\}$ instead. However, since such discrepancies must indeed ‘even out,’ in the sense that (2.2) applies to both priors, such observations do not seem particularly useful.

3.3. Imperfect compliance in a randomized trial

Next we consider a version of the imperfect compliance model with binary variables considered by various authors, including Chickering and Pearl [3], Imbens

and Rubin [13], Pearl [18, Ch. 8], and Richardson, Evans and Robins [20]. Clinical 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 subject’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 $\gamma_{C,i}$ be 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 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

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

Particularly, it is easy to verify that the map from ϕ 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 $\psi = g(\theta)$ is the (global) average causal effect (ACE), given as:

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

which depends on components of ϕ and components of λ .

Gustafson [10] shows that for n independent and identically distributed realizations of (Y, X, Z) , as $n \rightarrow \infty$, the posterior distribution of ψ converges to a symmetric, trapezoidal density. Particularly, let

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

The support of the limiting trapezoidal density is $a(\phi) \pm b(\phi)$, while the ‘top’ of the trapezoid extends along $a(\phi) \pm b^*(\phi)$, and consequently the height of the trapezoid is $\{b(\phi) + b^*(\phi)\}^{-1}$.

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

$$a(\phi) \pm k_\alpha(\omega_{AT}/\omega_{NT})b(\phi),$$

where

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

with $v(r) = |1 - r|/(1 + r)$. 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 $k_\alpha(r) \rightarrow 1 - \alpha$ as $r \rightarrow 0$ or $r \rightarrow \infty$.

To gain some intuition, we simulate draws from the joint prior density $\pi(\theta)$ and ascertain which realizations fall in $B_\pi^{(\alpha)}$, for $\alpha = 0.05$. We know from (2.2) that the proportion of draws falling in this ‘bad’ set is α , but the results in Figure 4 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 $\lambda = (\gamma_{AT,0}, \gamma_{NT,1})$. The 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 λ 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 λ -space to separate $B_\pi^{(\alpha)}$ from its complement.

The other feature evident from Figure 4 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.

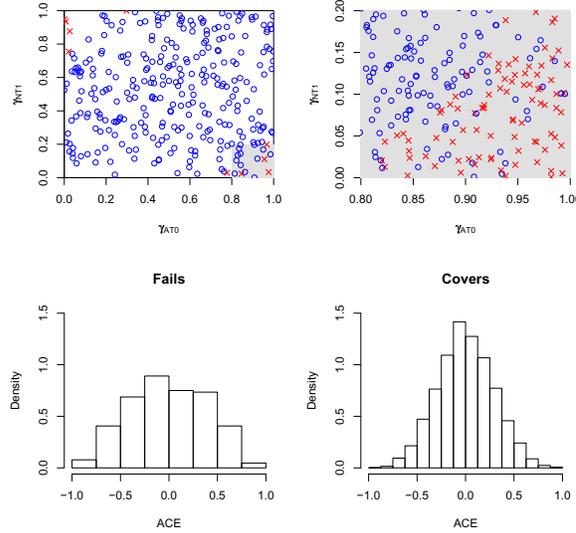


FIG 4. Exploration of coverage in Example 3, via simulated draws from the prior. The upper-left panel gives the $(\gamma_{AT,0}, \gamma_{NT,1})$ values of the draws, with success/failure to cover indicated by ‘o’/‘x’. The upper-right panel focuses only on the shaded corner region of the upper-left panel. The lower panels depict simulated values of the target ACE, stratified by failure/success to cover.

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

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

This underscores that failure is not determined exclusively by the values of nonidentified parameters $\lambda = (\gamma_{NT,1}, \gamma_{AT,0})$, 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 ,

$$\gamma_{NT,1} - 1/2 = r(\gamma_{AT,0} - 1/2) \pm k_\alpha(r)(1 + r)/2, \tag{3.2}$$

passing through the unit square. While the size of the triangles could be deduced from (3.2), direct appeal to Theorem 2.1 immediately gives the areas as $\alpha/2$ each.

Figure 5 gives a pictorial representation of how the failure-to-cover region for λ 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

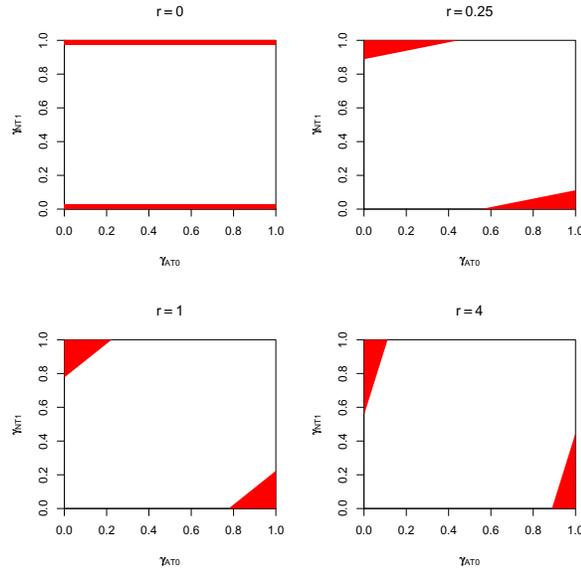


FIG 5. The failure-to-cover region intersected with different values of $r = \omega_{AT}/\omega_{NT}$ in Example 3, with $\alpha = 0.05$.

a larger role as r becomes large. This behavior explains the lack of separation between $B_{\pi}^{(\alpha)}$ and its complement in terms of λ alone, i.e., how it is possible to have successes lying ‘south-east’ of failures in the top right panel of Figure 4.

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.

3.4. Example: Prevalence of misclassified trait

Say that X is a binary trait, with interest lying in its population prevalence $p = Pr(X = 1)$. However, the observable binary variable is \tilde{X} , which is subject to misclassification, i.e., if we let $\tilde{p} = Pr(\tilde{X} = 1)$, then $\tilde{p} = r\gamma_N + (1 - r)(1 - \gamma_P)$, 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. Thus the original scientific parameterization is $\theta = (p, \gamma_N, \gamma_P)$ and the inferential target is $\psi = g(\theta) = p$.

Consider the situation where the investigator commits to lower bounds on sensitivity and specificity but applies uniform prior distributions above the

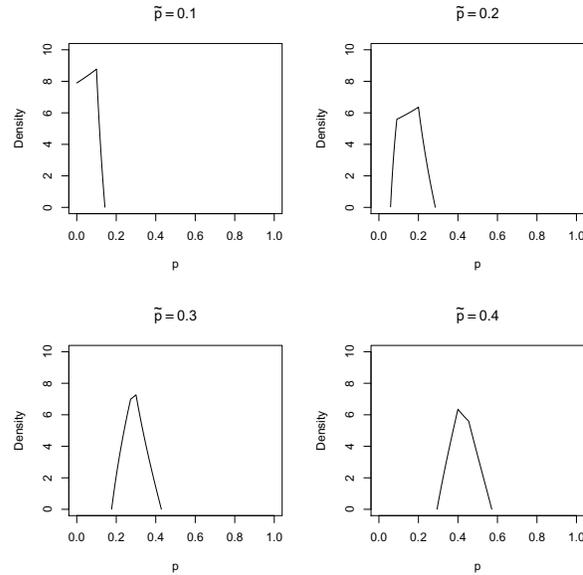


FIG 6. The limiting posterior density on p in Example 4, in the cases that $\tilde{p} = 0.1, 0.2, 0.3, 0.4$. The hyperparameters, in the form of lower bounds on sensitivity γ_N and specificity γ_P , are $a = 0.7$ and $b = 0.85$.

bounds and also applies a uniform prior to the target parameter p . That is

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

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{p} . More particularly, a transparent parameterization is obtained by taking $\phi = \tilde{p}$ and $\lambda = (\gamma_N, \gamma_P)$, and the large-sample limit of the posterior distribution over the target p is identically the prior conditional distribution of $(p|\tilde{p})$ (where the conditioning is on the true value of \tilde{p}). This distribution is given in Gustafson [8] as having support

$$p \in \left(1 - \frac{1 - \tilde{p}}{b^*}, \frac{\tilde{p}}{a^*}\right), \tag{3.3}$$

where $a^* = \max\{a, \tilde{p}\}$, $b^* = \max\{b, 1 - \tilde{p}\}$. The interval (3.3) 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:

$$\pi^*(p|\tilde{p}) \propto p^{-1} \left[\min \left\{ \frac{\tilde{p} - a^*p}{1 - p}, 1 - b^* \right\} - \max \left\{ \frac{\tilde{p} - p}{1 - p}, 0 \right\} \right]. \tag{3.4}$$

For the hyperparameters $a = 0.7$ and $b = 0.85$, examples of the limiting posterior density (3.4), for different values of \tilde{p} , appear in Figure 6. Bollinger and van

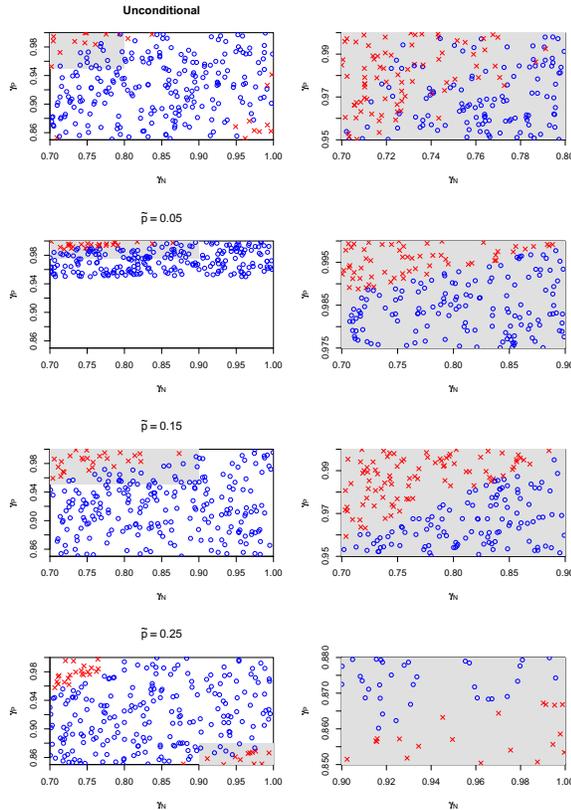


FIG 7. Failure to cover in Example 4. The top panels report asymptotic success ('o') or failure ('x') to cover, for a sample of θ values drawn from the unconditional prior $\pi(\theta)$. The remaining panels report asymptotic success/failure to cover for a sample of θ values obtained via draws from the conditional prior $\pi^*(\gamma_N, \gamma_P|\tilde{p})$, for $\tilde{p} = 0.05$ (second row), $\tilde{p} = 0.15$ (third row), and $\tilde{p} = 0.25$ (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.

Hasselst [2] give related expressions for limiting posterior distributions in this problem under different choices of prior distributions.

The form of (3.4) 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 $\theta = (p, \gamma_N, \gamma_P)$ from the joint prior distribution, and check membership in $B_\pi^{(\alpha)}$. The top panels of Figure 7 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 $\lambda = (\gamma_N, \gamma_P)$, i.e., $B_\pi^{(\alpha)}$ is not determined by a smooth boundary in λ -space.

Again we appeal to Theorem 2.1. Drawing values of θ by sampling from $\pi^*(\gamma_N, \gamma_P|\tilde{p} = \tilde{p}_0)$ for selected values of \tilde{p}_0 , we examine the intersection of

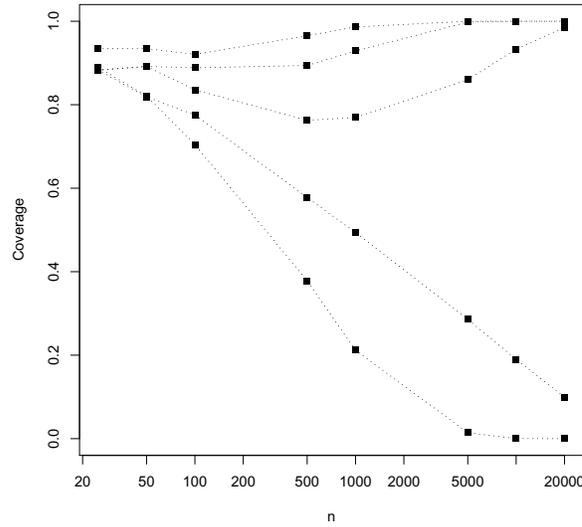


FIG 8. Frequentist coverage of the 90% equal-tailed credible interval for p in Example 4. Coverage is given as a function of n for the five selected values of θ described in the text. The coverage decreases as we move from values (a) through (e), i.e., further toward the corner of the prior support for (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 .

$B_\pi^{(\alpha)}$ with $\{\theta : \tilde{p}(\theta) = \tilde{p}_0\}$. Theorem 2.1 guarantees this intersection to have probability α with respect to the conditional prior, while Figure 7 illustrates that this set is determined by a smooth boundary in λ -space. Moreover, this boundary is seen to vary with \tilde{p}_0 . Thus again we see that failure-to-cover arises when λ lies in a tail of the prior distribution, but *which* tail depends on an identified quantity. As an aside, Figure 7 also illustrates the ‘indirect learning’ that can arise because the support of $\pi^*(\gamma_N, \gamma_P | \tilde{p} = \tilde{p}_0)$ can, for some values of \tilde{p}_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 $\tilde{p} = 0.25$ 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), $(\gamma_N, \gamma_P) = (0.72, 0.98)$, and (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 p , for a variety of sample sizes. While the asymptotic results in Figure 7 pertain to HPD intervals, it is somewhat easier to control 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 of the five θ values the frequentist coverage is plotted against sample size n in Figure 8. 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

All four of our examples possess parameters which intuitively aren't informed by the data. In the first example this is the location of the target mean inside the interval induced by censoring. In the second example these are the probabilities of missingness for the two levels of the outcome. In the third example these are the counterfactual average responses of never-takers if they take treatment and always-takers if they do not. In the fourth 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 [5, 6]). 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 2.1 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.

It should also be noted that failure-to-cover is inevitable, and cannot be made to go away via a particular choice of prior. As emphasized by Moon and Schorfheide [17], the limits of credible sets must lie *inside* the identification region, hence failure-to-cover can arise. Of course, this is not a bad thing, in that $1 - \alpha$ credible intervals are *supposed* to miss their target sometimes (unless $\alpha = 0$). The difficulty in thinking about this stems from the fact that in the identified case, 'sometimes' refers to some data realizations. In the nonidentified case, however, 'sometimes' refers to some values of the parameters.

Appendix

The numerical evaluation of finite-sample frequentist coverage in Example 4 proceeds as follows. Assume the hyperparameters satisfy $a + b > 1$, so that the prior ensures $1 - \gamma_P < \gamma_N$. Upon reparameterizing to $(\tilde{p}, \gamma_N, \gamma_P)$ we have the prior density

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

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

$$\pi^*(\tilde{p}, \gamma_N, \gamma_P | y) \propto \tilde{p}^y (1 - \tilde{p})^{n-y} \pi(\tilde{p}, \gamma_N, \gamma_P). \quad (4.1)$$

Note that the event $p < p_0$ can be reexpressed as $\tilde{p} < (1 - p_0)(1 - \gamma_P) + p_0 \gamma_N$, and that integration of (4.1) with respect to \tilde{p} can be expressed via the Beta($y + 1, n - y + 1$) distribution function, which we denote as $G_y(\cdot)$. Thus

$$pr(p < p_0 | y) = \frac{E \left[\frac{G_y\{(1-p_0)(1-\gamma_P)+p_0\gamma_N\}-G_y(1-\gamma_P)}}{\gamma_N+\gamma_P-1} \right]}{E \left[\frac{G_y(\gamma_N)-G_y(1-\gamma_P)}{\gamma_N+\gamma_P-1} \right]}, \quad (4.2)$$

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 (4.2). 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 8 use 20,000 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 $pr(p < p_0 | Y = y)$ lies between $\alpha/2$ and $1 - \alpha/2$. Note that for a given y this only requires numerical evaluation of (4.2) at a single value of p_0 , whereas checking coverage for an HPD interval would require evaluation over a fine grid of values.

References

- [1] BARANKIN, E. W. (1960). Sufficient parameters: Solution of the minimal dimensionality problem. *Annals of the Institute of Mathematical Statistics* **12** 91-118. [MR0126308](#)
- [2] BOLLINGER, C. and VAN HASSELT, M. (2009). A Bayesian analysis of binary misclassification: inference in partially Identified models. *Preprint*.
- [3] CHICKERING, D. M. 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 **2** 1269-1276.
- [4] DAWID, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B* **41** 1-31. [MR0535541](#)
- [5] 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. [MR1977199](#)
- [6] GREENLAND, S. (2005). Multiple-bias modelling for analysis of observational data. *Journal of the Royal Statistical Society, Series A* **168** 267-306. [MR2119402](#)
- [7] GUSTAFSON, P. (2005). On model expansion, model contraction, identifiability, and prior information: two illustrative scenarios involving mis-measured variables (with discussion). *Statistical Science* **20** 111-140. [MR2183445](#)
- [8] GUSTAFSON, P. (2010). Bayesian inference for partially identified models. *International Journal of Biostatistics* **6** issue 2 article 17. [MR2602560](#)
- [9] GUSTAFSON, P. (2011a). Bayesian inference in partially identified models: Is the shape of the posterior density useful? *Technical Report #265, Department of Statistics, University of British Columbia*.
- [10] GUSTAFSON, P. (2011b). Comment on 'Transparent parameterizations of models for potential outcomes,' by Richardson, Evans, and Robins. In *Bayesian Statistics 9: Proceedings of the Ninth Valencia International Meeting* (J. M. BERNARDO, M. J. BAYARRI, J. O. BERGER, A. P. DAWID,

- D. HECKERMAN, A. F. M. SMITH and M. WEST, eds.). Oxford University Press.
- [11] GUSTAFSON, P. and GREENLAND, S. (2009). Interval estimation for messy observational data. *Statistical Science* **24** 328-342. [MR2757434](#)
 - [12] IMBENS, G. W. and MANSKI, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica* **72** 1845-1857. [MR2095534](#)
 - [13] IMBENS, G. W. and RUBIN, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics* **25** 305-327. [MR1429927](#)
 - [14] KADANE, J. B. (1974). The role of identification in Bayesian theory. In *Studies in Bayesian Econometrics and Statistics, In Honor of Leonard J. Savage* (S. E. FIENBERG and A. ZELLNER, eds.) 175. [MR0483124](#)
 - [15] LIAO, Y. and JIANG, W. (2010). Bayesian analysis in moment inequality models. *Annals of Statistics* **38** 275-316. [MR2589323](#)
 - [16] MANSKI, C. F. (2003). *Partial Identification of Probability Distributions*. Springer. [MR2151380](#)
 - [17] MOON, H. R. and SCHORFHEIDE, F. (2012). Bayesian and frequentist inference in partially identified models. *Econometrica* **80** 755-782. [MR2951948](#)
 - [18] PEARL, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press. [MR1744773](#)
 - [19] POIRIER, D. J. (1998). Revising beliefs in nonidentified models. *Econometric Theory* **14** 483-509. [MR1650041](#)
 - [20] RICHARDSON, T. S., EVANS, R. J. and ROBINS, J. M. (2011). Transparent parameterizations of models for potential outcomes. In *Bayesian Statistics 9: Proceedings of the Ninth Valencia International Meeting* (J. M. BERNARDO, M. J. BAYARRI, J. O. BERGER, A. P. DAWID, D. HECKERMAN, A. F. M. SMITH and M. WEST, eds.) 569-610. Oxford University Press.
 - [21] 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. [MR2422399](#)
 - [22] VANSTEEELANDT, 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. [MR2281311](#)
 - [23] ZHANG, Z. (2009). Likelihood-based confidence sets for partially identified parameters. *Journal of Statistical Planning and Inference* **139** 696-710. [MR2479821](#)