Inference about constrained parameters using the elastic belief method

Duncan Ermini Leaf*, Chuanhai Liu

Department of Statistics, Purdue University, 250 N. University Street, West Lafayette, IN 47907-2067, USA

Abstract

Statistical inference about unknown parameter values that have known constraints is a challenging problem for both frequentist and Bayesian methods. As an alternative, inferential models created with the weak belief method can generate inferential results with desirable frequency properties for constrained parameter problems. To accomplish this, we propose an extension of weak belief called the elastic belief method. Compared to an existing rule for conditioning on constraint information, the elastic belief method produces more efficient probabilistic inference while maintaining desirable frequency properties. The application of this new method is demonstrated in two well-studied examples: inference about a nonnegative quantity measured with Gaussian error and inference about the signal rate of a Poisson count with a known background rate. Compared to several previous interval-forming methods for the constrained Poisson signal rate, the new method gives an interval with better coverage probability or a simpler construction. More importantly, the inferential model provides a post-data predictive measure of uncertainty about the unknown parameter value that is not inherent in other interval-forming methods.

Keywords: Bounded parameter, Confidence interval, Inferential model, Maximal belief
2000 MSC: 62F30, 62F25

*Corresponding author. Fax: +1 765 494-0558
Email addresses: `dleaf@stat.purdue.edu` (Duncan Ermini Leaf), `chuanhai@stat.purdue.edu` (Chuanhai Liu)
URL: `http://www.stat.purdue.edu/~chuanhai/` (Chuanhai Liu)

1. Introduction

The parameter space of a probability model may extend beyond what is consistent with the physical world. Currently, there is no widely accepted method for incorporating such physical constraints into statistical inference methods. A new approach to this problem, based on the theory of inferential models (IMs) [1, 2, 3], is considered here. We use two examples of particular interest to high energy physicists during the past fifteen years: inference about a nonnegative quantity measured with Gaussian error and inference about the Poisson rate from a contaminated observed count.

Suppose X is the measurement of a nonnegative quantity, μ, with Gaussian error distribution. Choosing the variance, $\sigma^2 = 1$, for simplicity, this can be represented by the probability model $X \sim N(\mu, 1)$ and the constraint $\mu \geq 0$. The Gaussian model for X allows any real-valued μ. For this unrestricted case, many inference methods have proven to be simple and produce practically the same results for μ. Somewhat surprising, when μ is known to belong to a restricted interval, the same problem becomes challenging. Bayesian inference with a flat prior on μ does not have a clear frequency interpretation [4] and frequentist procedures are difficult to construct. As discussed in [5], this problem arises when measuring particle masses, which must be non-negative and are expected to be relatively small, if nonzero.

In the Poisson example, the observed count, Y, is known to be comprised of signal and background events each coming from their own independent Poisson distributions. Suppose the background rate, b, is known, but the signal rate, λ, is unknown. Let $S \sim Poisson(\lambda)$ be the number of signal events and $B \sim Poisson(b)$ be the number of background events. Both S and B are unobserved, but the observed count, $Y = S + B$, comes from a $Poisson(\theta)$ distribution with $\theta = \lambda + b$. The Poisson model for Y only requires that θ be nonnegative or, equivalently, $\lambda \geq -b$. However, negative values of λ are not valid and so the constraint $\theta \geq b$ is required. This model is used in experiments measuring a number of events caused by neutrino oscillations. Some of the observed events are due to random background sources, but are indistinguishable from the signal events of interest. Detailed discussion and references to experimental results can be found in [5].

Much existing work on the two example problems was aimed at developing confidence intervals that involve the constraints. Methods were developed within both the Bayesian and frequentist frameworks. For a review and discussion of previous methods, see [5, with comments]. In scientific inference it is desirable that inferential results be stated with some kind of probabilistic assessment of their uncertainty, such as a confidence level. In order for such statements to be meaningful, many practitioners believe these probabilities should be calibrated to a frequency interpretation. Thus, we focus on interval constructions that provide proper coverage: for any given confidence level, γ, the unconditional probability of the interval covering the true value of the parameter over repeated experiments should be at least γ.

In the terminology of [6], the confidence level, γ, is a *pre-data* predictive probability. It describes the random coverage behavior of a confidence interval over an infinite sequence of hypothetical experiments. After data are observed and an interval is realized, the interpretation of the interval is *postdictive* [6]: a realized interval contains parameter values that would not make its realization improbable relative to the confidence level. If the true parameter value lies outside of the realized interval, then something improbable has ocurred. Although, it should not be surprising to discover that the true parameter value lies in the interval, it is incorrect to interpret the interior of a realized confidence interval to be the most likely values of the parameter. Nevertheless, deeper meaning of the parameter values inside and outside of a realized interval can be achieved. If a confidence interval is constructed by inverting the acceptance regions of hypothesis tests, then parameter values outside of a realized interval would be rejected based on the observed data while parameter values inside the interval would not be rejected. The interior of a realized interval with this construction contains parameter values that would not make the observed data improbable relative to the confidence level.

Initial approaches to the example problems in this article were motivated by the fact that traditional confidence intervals built from Neyman's method [7] can be empty for small values of X or Y. In terms of hypothesis testing, every value of the parameter in the constraint set would be rejected if X or Y is sufficiently small. Clearly, the pre-data confidence level is not a sensible measure of post-data uncertainty about whether an empty interval contains the true parameter value.

The continued development of interval-forming methods for these problems appears to be due in part to the post-data interpretation of the confidence level. It is possible to observe data that are relatively improbable given some set of parameter constraints. For the Gaussian and Poisson examples, this occurs when the experimental observation is smaller than the constraint boundary, *i.e.*, $X < 0$ or $Y < b$. Some methods will produce a shorter interval for these observations than for observations within the constraint region. If one mistakes a fixed confidence level to be the *post-data* predictive probability for the parameter lying in the realized interval, then, counterintuitively, the shorter intervals produced by improbable observations seem to provide better information about the location of the unknown parameter value. Thus, one of the motivations for developing new methods was to obtain wider intervals when improbable data are observed. However, under the postdictive interpretation, a narrower interval obtained from observations outside of the constraint region means that there is a larger range of parameter values that make the realized interval improbable with respect to the confidence level. Also, there is a smaller range of unsurprising parameter values within the interval. This postdictive interpretation makes sense when the observed data are already known to be improbable.

Gleser [8] discussed how the likelihood function can quantify uncertainty about the unknown mean in the Gaussian example. Fraser, Reid, and Wong [9] argued in favor of reporting the likelihood and one-tailed p-value as a function of hypothetical parameter values. This allows each individual to make their own judgment about the strength of evidence required for rejection. However, inference using likelihood and p-values, being postdictive in nature, lacks the predictive interpretation often sought by practitioners. As articulated in [10], care must be taken in making a probabilistic interpretation of p-values.

The elastic belief (EB) method is introduced here to create IMs that provide post-data predictive probabilistic inference about unknown parameter values with constraints. The IM framework provides inferential tools, in the form of belief functions, that measure evidence in the observed data in order to suggest which parameter values can be accepted, rejected, or neither. This naturally leads to intervals that are easier to interpret than confidence intervals. When parameter constraints are considered as in the Gaussian and Poisson examples, the IMs resulting from the EB method are calibrated to a frequency interpretation. This is convenient for decision making and constructing intervals with proper coverage. Given the interpretation of IM inferences, there is no need to create wider intervals for improbable observations in order to convey a sense of belief that the parameter falls in the interval. However, IMs give practitioners flexibility to control this behavior, if desired, while at the same time ensuring that the resulting inferences have desirable frequency properties.

In Section 2, we review IMs and weak belief (WB) methods for inference in unconstrained parameter spaces. In Section 3 we consider methods for incorporating parameter constraints into an IM. The EB method is introduced and its frequency properties are compared to those of the conditioning rule introduced in [11], known as "Dempster's rule of conditioning" [12]. Inference about the mean of a Gaussian random variable is considered as a running example throughout Sections 2 and 3. This is an extension of the example in [4]. Section 4 contains results for EB applied to the specific case where the mean is known to be nonnegative. Inference about the unknown signal rate of a Poisson count with known background is discussed with numerical results in Section 5. Finally, in Section 6, practical issues and future directions of this work are considered.

2. Fundamentals of Inferential Models

An IM for constrained parameters is built from an IM for the unconstrained parameter space. The following discussion of IMs and WB methods establishes necessary notation, motivates the use of WB, and illustrates the unconstrained problem with the simple Gaussian model. Section 3 considers how to incorporate the parameter constraints.

2.1. Background and Motivations for Inferential Models

Before proceeding with technical details, the reader may find helpful the following elaboration on the development of IMs. The goal of [1, 2] and the present work is to create a user-friendly method for probabilistic inference with desirable frequency properties. This is accomplished, in part, by working with *nonadditive* probability using the Dempster-Shafer (DS) theory of belief functions [12, 13]. What motivated the use of probability on subsets, a concept that may seem unfamiliar to many statisticians? First, the mathematics are simple when Bayesian-like posteriors are of interest for continuous-data models, such as the unconstrained Gaussian model. In this case, a basic IM is represented by familiar additive probabilities. The only difference is that these probabilities are defined on an auxiliary space rather than on the parameter space. Second, for inference with discrete data models without prior knowledge on parameters, such as the Poisson model without additional constraints on λ, the resulting lower and upper probabilities are both necessary and convenient. Lastly, applying WB methods to a basic IM produces results intended to have universal appeal. Both Bayesians and non-Bayesians have practical methods for certain kinds of probabilistic inference about assertions of interest on unknown parameters. These include frequentist rejection regions for hypothesis testing, Bayesian credible regions, and frequentist confidence intervals for parameter estimation. To some extent, the WB method incorporates all of these three concepts to produce probabilistic output for assertions about unknown parameters. Since such output has desirable frequency properties, building IMs with the WB method is a promising approach to scientific inference.

2.2. Building Inferential Models

In this section, an IM is presented for a probability model without parameter constraints. It assumed that the probability model has been chosen through a process of careful model building and considering any domain knowledge about the problem at hand. Let $\mathcal{X}$ be the sample space of all possible observations for the probability model. It is further assumed that the probability distribution of outcomes in $\mathcal{X}$ is defined by a probability measure, $\Pr_\theta$, that depends on the parameter, θ. The unconstrained parameter space, Θ, consists of all θ values for which $\Pr_\theta$ is a valid probability measure.

First, following [1] and [2], the sampling distribution of X can be characterized using an auxiliary (a)-variable, U, defined in an a-space, $\mathbb{U}$. To do so, an a-equation, $X = a(\theta, U)$, and an a-measure, π, can be defined over $\mathbb{U}$ such that $a(\theta, U)$ has the same distribution as X when $U \sim \pi$ and θ is known. Collectively, the a-equation and a-measure are known as the association model, or a-model. This first step of defining the a-model is known as the association, or a-step. The a-variable is technically the same as a pivotal variable introduced by Fisher [14] and used by Dempster [15, 6, 16]. However, for the purposes of inference, the IM framework gives it a conceptually different treatment.

The second step of building and IM is the prediction (p)-step. For fixed θ, the a-model says that the observation, $X = x$, corresponds to an unobserved realization $u^* \in \mathbb{U}$ from the π distribution. Given an observed x, inferring the unobserved value of θ can be achieved by predicting the unobserved realization of u^* using a predictive random set (PRS). The predicted u^* and the observed value x can then be mapped to values of θ by way of the a-equation. The PRS is a set, $S(u) \subseteq \mathbb{U}$, constructed for each $u \in \mathbb{U}$. When $U \sim \pi$, the PRS $S(U)$ is designed to have a large probability of covering an unobserved u^* realization from the same π distribution. Specific forms of S

will be introduced later, but assume $u \in S(u)$ for every $u \in \mathbb{U}$. A general discussion of criteria for selecting PRSs can be found in [3].

An IM is created in the third step by combining the PRS with the a-model. This is known as the combination (c)-step. Mathematically, the resulting IM contains a collection of subsets of Θ, called focal elements, that are indexed by u:

$$M_x(u, S) = \bigcup_{u' \in S(u)} \{\theta : \theta \in \Theta, x = a(\theta, u')\}, \qquad u \in \mathbb{U}. \tag{1}$$

Focal elements have a mass distribution defined by the π distribution over $\mathbb{U}$. If $S(u)$ covers the unobserved u^*, then $M_x(u, S)$ covers the unobserved value of θ corresponding to the experiment that generated x. Thus, the focal elements represent sets of parameter values allowed by the probability model (although some values may not be possible in the physical world) and π is a measure of uncertainty about these values as predictions of the true θ value given the observation, x.

Once an IM is established, the essential tool for inference is a belief function, Bel_x [12]. It takes a subset of Θ, say $\mathcal{A}$, as an argument and outputs the mass over all focal elements that support $\mathcal{A}$, conditioned on the focal elements being nonempty:

$$\mathrm{Bel}_x(\mathcal{A}; S) = \pi\{u : M_x(u, S) \subseteq \mathcal{A} \mid M_x(u, S) \neq \emptyset\}. \tag{2}$$

The parameter space subset, $\mathcal{A}$, can be interpreted as an assertion about the true value of θ. For any observed $X = x$, the evidence about $\mathcal{A}$ is computed as:

- $\mathrm{Bel}_x(\mathcal{A}; S) = \pi\{u : M_x(u, S) \subseteq \mathcal{A} \mid M_x(u, S) \neq \emptyset\}$, which measures the evidence *for* the assertion; and
- $\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}}; S) = \pi\{u : M_x(u, S) \subseteq \mathcal{A}^{\mathrm{c}} \mid M_x(u, S) \neq \emptyset\}$, which measures the evidence *against* the assertion, with $\mathcal{A}^{\mathrm{c}}$ as the complement of $\mathcal{A}$.

It should be noted that

$$\mathrm{Bel}_x(\mathcal{A}; S) + \mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}}; S) \leq 1.$$

Any remaining probability, $1 - \mathrm{Bel}_x(\mathcal{A}; S) - \mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}}; S)$, is called (the probability of) "don't know," [13] and is neither for nor against the assertion. Another inferential tool is the plausibility function, Pl_x [12], defined as:

$$\mathrm{Pl}_x(\mathcal{A}; S) = 1 - \mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}}; S).$$

Since $\mathrm{Bel}_x(\mathcal{A}; S) \leq \mathrm{Pl}_x(\mathcal{A}; S)$, these two quantities can be thought of as lower and upper measures of evidence for the assertion, $\mathcal{A}$.

Example 2.1 (Inference about the Gaussian mean). Suppose $X \sim N(\mu, 1)$. For the a-step, an a-model can be formed as

$$X = \mu + U,$$

where U is the a-variable with the standard Gaussian distribution, $N(0,1)$. In the p-step, the PRS, $S(u) = \{u' : |u'| \leq |u|\}$ can be used. The result of the c-step is an IM for the Gaussian mean:

$$M_x(u, S) = \bigcup_{u' \in S(u)} \{\mu : x = \mu + u'\} = [x - |u|, x + |u|]. \tag{3}$$

The mass distribution over these focal elements is defined by the standard Gaussian distribution. Example 2.2 justifies the particular choice of PRS, S, used here.

Suppose we wish to make inference about whether or not $\mu > 0$. Using the IM (3) for the assertion $\mathcal{A} = \{\mu : \mu > 0\}$ we have

$$\mathrm{Bel}_x(\mathcal{A}; S) = \pi\{u : x - |u| > 0\} = \max\{0, 2\Phi(x) - 1\},$$

and

$$\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}}; S) = \pi\{u : x + |u| \leq 0\} = \max\{0, 1 - 2\Phi(x)\}.$$

Suppose $x = -1$ is observed. Then,

$$\mathrm{Bel}_{-1}(\mathcal{A}; S) = 0$$

and

$$\mathrm{Bel}_{-1}(\mathcal{A}^{\mathrm{c}}; S) = 1 - 2\Phi(-1) \approx 0.68.$$

The approach to building belief functions for statistical inference starting with a data-generating model is not unique to the IM framework. It can be found in the work of Dempster [16] and, more recently, in the theory of hints [17, 18, 19]. The problem shared by all these approaches is that observing $X = x$ gives no precise information about the value of the corresponding u^* realization. Generally, all that is known is that u^* is a realization from the π distribution, perhaps limited to some subset of $\mathbb{U}$. For any choice of u with x fixed, the a-equation, $x = a(\theta, u)$, holds true for some set of θ

values. Thus, for an assertion about θ, there is a set of u values for which the consequent sets of θ values support the assertion. In the theory of hints, the belief for the assertion would be computed using the probability on this $\mathbb{U}$ subset. This approach is called assumption-based reasoning and the resulting belief functions can be interpreted as an assertion's probability of provability [20, 21].

Because u^* is unobserved, the IM framework is not directly focused on the consequences of hypothetical values for u^*. Instead, the PRS is introduced to *predict* the value of u^* based upon the knowledge that u^* is a realization from the π distribution. If the PRS is a good predictor of an unobserved u^* realization, then one should expect the focal elements in Θ resulting from the c-step to be good predictors of the corresponding unobserved θ value. Alternatively, it would be equally valid to map the observed x and an assertion into a subset of $\mathbb{U}$ via the a-equation. Then, a belief function can be computed from the probability that the PRS supports this subset. Practitioners may find it more convenient to work with focal elements in Θ, but working directly in $\mathbb{U}$ is helpful for illustrating the goal of IM-based inference. Whether computed in Θ or $\mathbb{U}$, a belief function resulting from the IM approach represents predictive probability for an assertion, which has a different interpretation than assumption-based reasoning.

In general, the $\mathrm{Bel}_x(\mathcal{A}; S)$ and $\mathrm{Bel}_x(\mathcal{A}^c; S)$ probabilities are computed with respect to U using the π distribution and are conditioned on the set of nonempty focal elements. Empty focal elements are called conflict cases and can have undesirable consequences when constraints on θ are considered in Section 3. The $\mathrm{Bel}_x(\mathcal{A}; S)$ and $\mathrm{Bel}_x(\mathcal{A}^c; S)$ values represent the strength of evidence in the IM (1), which serve as a tool to infer the truth of $\mathcal{A}$ or $\mathcal{A}^c$. However, it may not be clear how large the values of $\mathrm{Bel}_x(\mathcal{A}; S)$ or $\mathrm{Bel}_x(\mathcal{A}^c; S)$ must be for one to believe or disbelieve $\mathcal{A}$. Their interpretation should be consistent with the distribution of X. If one interprets numerical probabilities in terms of long-run frequency, then $\mathrm{Bel}_X(\mathcal{A}; S)$ and $\mathrm{Bel}_X(\mathcal{A}^c; S)$ should behave accordingly. This long-run frequency behavior can be characterized by the concepts of validity and efficiency.

2.3. Validity and the Weak Belief Method

The following validity criteria [2, Definition 3.1] ensure that $\mathrm{Bel}_X(\mathcal{A}; S)$ and $\mathrm{Bel}_X(\mathcal{A}^c; S)$ behave in a manner consistent with the distribution of X:

Definition 2.1 (Validity). For a given assertion $\mathcal{A} \subset \Theta$, Bel_x is *valid* for $\mathcal{A}$ if

$$\mathrm{Pr}_\theta\{x : \mathrm{Bel}_x(\mathcal{A}; S) \geq 1 - \alpha\} \leq \alpha \tag{4}$$

for all $\theta \in \mathcal{A}^{\mathrm{c}}$, and

$$\mathrm{Pr}_\theta\{x : \mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}}; S) \geq 1 - \alpha\} \leq \alpha, \tag{5}$$

for all $\theta \in \mathcal{A}$ and every $\alpha \in (0, 1)$. If (4) and (5) hold for every $\mathcal{A}$, then Bel_x is valid (without reference to the assertion).

The probabilities in (4) and (5) are computed with respect to the sampling distribution of X. Suppose that we choose an α value and if either $\mathrm{Bel}_x(\mathcal{A})$ or $\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}})$ exceeds $1 - \alpha$, we will believe or disbelieve $\mathcal{A}$ accordingly. From a frequency perspective, Definition 2.1 says that the probability of making a wrong conclusion is at most α in repeated observations of X. Ideally, an IM should produce valid Bel_x for all assertions of interest. This can be achieved through the choice of the PRS.

Suppose there is a collection of PRSs, $\{S_\omega\}_{\omega \in \Omega}$. Let

$$Q_{S_\omega}(u) = \pi\{u' : S_\omega(u') \not\ni u\}, \qquad u \in \mathbb{U}.$$

Definition 1 in [1] defines the credibility of a PRS for predicting an a-variable:

Definition 2.2 (Credibility). For a given value of $\alpha \in (0, 1)$ and $\omega \in \Omega$, let

$$\varphi_\alpha(\omega) = \pi\{u : Q_{S_\omega}(u) \geq 1 - \alpha\}.$$

A PRS, S_ω, is credible at level α if $\varphi_\alpha(\omega) \leq \alpha$.

Theorem 1 in [1] shows that a credible PRS for predicting the a-variable (Definition 2.2) leads to an IM that produces valid Bel_x (Definition 2.1) with the condition that $\pi\{u : M_x(u, S_\omega) = \emptyset\} = 0$. It is shown in Section 3.2 and Appendix A.2 that this condition can be removed.

2.4. Efficiency and the Maximal Belief Method

Predicting the a-variable with a larger PRS leads to credibility. The question is: how large should the PRS be? At one extreme, $S(u) \equiv \mathbb{U}$ certainly predicts the unobserved u^* realization. When this PRS is used, each focal element in (1) becomes Θ, the entire parameter space. In this case, the IM consists of a single focal element that has unit mass and certainly contains the true unobserved parameter value. However, large focal elements (or in the extreme, one largest focal element) do not offer a great level of discernment between different possible parameter values. An ideal PRS should be large enough to meet the credibility criteria of Definition 2.2, but small enough to represent $\mathbb{U}$ (and consequently Θ) with high resolution. Smaller PRSs are more *efficient* for predicting the a-variable [1, Definition 2].

The optimality principle of the maximal belief (MB) method [1, 2] balances this tradeoff between credibility and efficiency. For a given $\alpha \in (0, 1)$ let

$$\Omega_\alpha = \{\omega \in \Omega : \varphi_\alpha(\omega) \le \alpha\}$$

be the index set for a class of credible PRSs. The MB method chooses a PRS, S_{ω^*}, from this class that satisfies

$$\varphi_\alpha(\omega^*) = \sup_{\omega \in \Omega_\alpha} \varphi_\alpha(\omega).$$

Example 2.2 (Maximal belief for standard Gaussian a-variable)*.* Suppose $U \sim N(0, 1)$. For the PRS, $S(u) = \{u' : |u'| \le |u|\}$,

$$Q_S(u) = \pi\{u' : S(u') \not\ni u\} = \pi\{u' : |u'| < |u|\} = 2\Phi(|u|) - 1.$$

Since, for any $\alpha \in (0, 1)$,

$$\begin{aligned} \pi\{u : Q_S(u) \ge 1 - \alpha\} &= \pi\{u : 2\Phi(|u|) - 1 \ge 1 - \alpha\} \\ &= \pi\{u : |u| \ge \Phi^{-1}(1 - \alpha/2)\} \\ &= \alpha, \end{aligned}$$

S is credible for predicting U as in Definition 2.2 and S also satisfies the MB criteria. Consequently, the IM (3) will be valid for any assertion.

3. Incorporating Parameter Constraints into Inferential Models

We now consider how to incorporate parameter space constraints into the IM (1). First, the EB method is introduced. Then, the existing conditioning rule [11, 12] is demonstrated and its frequency properties are compared to those of EB. Throughout this section assume θ is known to be in some constraint set $\mathcal{C} \subset \Theta$, *e.g.*, $\mathcal{C} = \{\theta : \theta \ge \theta_0\}$.

As described in Section 2.4, a PRS is designed to be credible and efficient for predicting the a-variable over the entire a-space, $\mathbb{U}$. After $X = x$ is observed, $\mathcal{C}$ can be mapped to a subset of $\mathbb{U}$ by inverting the a-equation in its second argument:

$$\mathbb{U}_{\mathcal{C},x} = \bigcup_{\theta \in \mathcal{C}} \{u : x = a(\theta, u)\}.$$

We call $\mathbb{U}_{\mathcal{C},x}$ the a-constraint set. Let θ^* be the true, unobserved value of the parameter. Then, there must exist $u^* \in \mathbb{U}_{\mathcal{C},x}$ such that $x = a(\theta^*, u^*)$. If $u^* \notin \mathbb{U}_{\mathcal{C},x}$, then the corresponding θ^* is not in the constraint set $\mathcal{C}$, which is impossible. Thus, when $S(u) \cap \mathbb{U}_{\mathcal{C},x} = \emptyset$, the focal element of the IM contains only values of θ that are not in the constraint set, *i.e.*, $M_x(u, S) \cap \mathcal{C} = \emptyset$. These focal elements are called conflict cases and are indexed by the set

$$\mathbb{U}_{\emptyset,x} = \{u : S(u) \cap \mathbb{U}_{\mathcal{C},x} = \emptyset\}.$$

The problem of incorporating parameter constraints into an IM can be framed in terms of handling conflict cases.

The probability on the set $\mathbb{U}_{\emptyset,x}$ can been seen as measuring discord between the observed value of x, its probability model, and the parameter constraint set. If the probability on conflict cases is very large, one should question whether the probability model is appropriate for the observed data or whether the constraint is correct. However, the presence of conflict alone does not justify modifying the model. In fact, rejecting a model simply because it conflicts with the observed data leads to biased inference procedures. Section 3.1 introduces a new method that modifies the PRS in a data-dependent way while preserving validity and striving for high efficiency. The result is that conflict cases become evidence for certain values of θ. In Section 3.2, the new method is compared to an existing conditioning method, which can use the probability on conflict cases to represent an additional layer of uncertainty about the model assumptions.

3.1. The Elastic Belief Method

The PRS, $S(u)$, and the π distribution on $\mathbb{U}$ represent a set of predictions and a measure of uncertainty about those predictions. Intuitively, a conflict case results from $S(u)$ being too small. If the probability model for the observed x and the parameter constraint are not in doubt, then $S(u)$ should be enlarged in an adaptive fashion. The EB method eliminates conflict cases by allowing the PRS to stretch until it includes at least one member of $\mathbb{U}_{\mathcal{C},x}$ while retaining the same π distribution. Technically, the EB method equips the PRS with an elasticity parameter, $e \in [0,1]$, thus forming a PRS collection, $\mathcal{S} = \{S_e : e \in [0,1]\}$, called an *elastic* PRS (EPRS).

Definition 3.1 (Elastic predictive random set). A collection of PRSs, $\mathcal{S}$, indexed by $e \in [0,1]$, is called *elastic* if,

(a) for any $e \in [0,1]$, S_e satisfies Definition 2.2;

(b) for any $e_1 \le e_2$, $S_{e_1}(U) \subseteq S_{e_2}(U)$ with probability one; and

(c) for any $u \in \mathbb{U}$ and any $(x,\theta) \in \mathcal{X} \times \Theta$, there exists an $e \in [0,1]$ and $u' \in S_e(u)$ such that $x = a(\theta, u')$.

We call these three properties, (a) *credibility*, (b) *monotonicity*, and (c) *completeness*.

Example 3.1 (An EPRS for the Gaussian problem). Consider the PRS from Examples 2.1 and 2.2: $S(u) = \{u' : |u'| \le |u|\}$. One way to make S elastic is

$$S_e(u) = \begin{cases} \{u' : |u'| \le \frac{1}{1-e}|u|\}, & \text{if } e \in [0,1); \\ \mathbb{R}, & \text{if } e = 1. \end{cases}$$

Using the Gaussian a-equation, $x = \theta + u$, it is easy to verify that $\mathcal{S} = \{S_e : e \in [0,1]\}$ satisfies Definition 3.1.

The existence of an EPRS is ensured by the nature of the a-equation. Let S_0 be a PRS satisfying Definition 2.2 and let let $S_1(u) \equiv \mathbb{U}$. For $e \in (0,1)$, an arbitrary S_e increasing in e from S_0 to $\mathbb{U}$ will satisfy (a) and (b) in Definition 3.1. Since $a(\theta, U)$ has the same distribution as X for fixed $\theta \in \Theta$, then $\mathcal{X} = \bigcup_{u \in \mathbb{U}} a(\theta, u)$. Thus, for all $(x,\theta) \in \mathcal{X} \times \Theta$, there must exist $u' \in S_1(u)$ such that $x = a(\theta, u')$, which satisfies (c).

To use the EB method, each focal element in the IM (1) is simply replaced with

$$M_x^{\mathrm{EB}}(u, \mathcal{S}) = \mathcal{C} \cap M_x(u, S_{\hat{e}}),$$

where

$$\hat{e} = \min\{e : S_e(u) \cap \mathbb{U}_{\mathcal{C},x} \neq \emptyset\} = \min\{e : \mathcal{C} \cap M_x(u, S_e) \neq \emptyset\}.$$

In effect, the EB method stretches the IM focal element until it is just large enough to intersect with $\mathcal{C}$. The amount of stretching is characterized by $\hat{e}$. The completeness property (c) in Definition 3.1 ensures that $\{e : S_e(u) \cap \mathbb{U}_{\mathcal{C},x} \neq \emptyset\}$ is not empty for any u. Therefore, if $\hat{e}$ exists, then $M_x^{\mathrm{EB}}(u, \mathcal{S})$ is also not empty. Finally, if applying the EB method results in any duplicate focal elements (*i.e.*, $M_x^{\mathrm{EB}}(u,\mathcal{S}) = M_x^{\mathrm{EB}}(u',\mathcal{S})$ for $u \neq u'$), they can be considered as a single element with mass aggregated from the duplicate elements. After building an IM with the EB method, the belief for any assertion, $\mathcal{A} \subseteq \mathcal{C}$, can be computed as:

$$\mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A};\mathcal{S}) = \pi\{u : M_x^{\mathrm{EB}}(u,\mathcal{S}) \subseteq \mathcal{A}\}. \tag{6}$$

When using an EPRS, all the IM focal elements are non-empty. The conditioning on non-empty focal elements in (2) is omitted in (6) because $\pi\{u : M_x^{\mathrm{EB}}(u,\mathcal{S}) \neq \emptyset\} = 1$.

Example 3.2 (EB method for constrained Gaussian mean). Suppose it is known that the Gaussian mean, μ, lies in the range $[a,b]$ for some known constants $a < b$. The EB method can be applied with the a-model from Example 2.1 and the EPRS from Example 3.1. This gives

$$M_x(u,S_e) = \left[x - \tfrac{1}{1-e}|u|, x + \tfrac{1}{1-e}|u|\right].$$

To handle conflict cases, the EPRS is expanded with

$$\begin{aligned}\hat{e} &= \min\{e : [a,b] \cap M_x(u,S_e) \neq \emptyset\} \\ &= \begin{cases} 1 + \frac{|u|}{x-a} & \text{if } x + |u| < a; \\ 1 - \frac{|u|}{x-b} & \text{if } x - |u| > b; \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

This yields an IM with the following focal elements,

$$M_x^{\mathrm{EB}}(u,\mathcal{S}) = \begin{cases} \{a\}, & \text{if } |u| < a - x; \\ [\max\{a, x - |u|\}, \min\{b, x + |u|\}], & \text{if } |u| \geq \max\{a - x, x - b\}; \\ \{b\}, & \text{if } |u| < x - b. \end{cases}$$

Now, suppose $a = -1/4$, $b = 1/4$, and $x = -1$ is observed. Then, for the assertion $\mathcal{A} = \{\mu : \mu \geq 0\}$, the EB method gives

$$\mathrm{Bel}_{-1}^{\mathrm{EB}}(\mathcal{A};\mathcal{S}) = \pi\{u : -1 - |u| \geq 0\} = 0$$

and

$$\mathrm{Bel}_{-1}^{\mathrm{EB}}(\mathcal{A}^{\mathrm{c}};\mathcal{S}) = \pi\{u : -1 + |u| < 0\} = 1 - 2\Phi(-1) \approx 0.68,$$

which is the same result as in Example 2.1 where the constraint was not part of the IM.

The following theorem shows how $M_x^{\mathrm{EB}}(U, \mathcal{S})$ can be used for valid inference in the sense of Definition 2.1. A proof can be found in Appendix A.1.

Theorem 3.1. *Let $\mathcal{S} = \{S_e : e \in [0,1]\}$ satisfy properties (b) and (c) of Definition 3.1. Let $\mathrm{Bel}_x(\mathcal{A}; S)$ and $\mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A}; \mathcal{S})$ be defined as in (2) and (6), respectively. Then, for any $x \in \mathcal{X}$, the following are true:*

(i) For any assertion $\mathcal{A} \subset \mathcal{C}$,

$$\mathrm{Bel}_x(\mathcal{A}; S_0) \leq \mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A}; \mathcal{S}) \leq \mathrm{Bel}_x(\mathcal{A} \cup \mathcal{C}^{\mathrm{c}}; S_0).$$

(ii) If $\mathrm{Bel}_x(\mathcal{A} \cup \mathcal{C}^{\mathrm{c}}; S_0)$ and $\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}} \cup \mathcal{C}^{\mathrm{c}}; S_0)$ satisfy (4) for some assertion $\mathcal{A} \subset \mathcal{C}$, then $\mathrm{Bel}_x^{\mathrm{EB}}$ is valid for inference about $\mathcal{A}$.

(iii) If $\mathcal{S}$ also satisfies property (a) of Definition 3.1, then $\mathrm{Bel}_x^{\mathrm{EB}}$ is valid for inference about every assertion $\mathcal{A} \subset \mathcal{C}$.

An important application of Theorem 3.1 is when an IM has been created using a PRS satisfying the efficiency criteria in Section 2.4 without considering constraints. That PRS can be used for S_0 when creating an EPRS. Thus, when constraints are incorporated using the EB method, $\mathrm{Bel}_x^{\mathrm{EB}}$ will be valid for inference about any $\mathcal{A} \subset \mathcal{C}$.

3.2. Elastic Belief Compared to Conditioning Rule

Another method for incorporating parameter constraints into an IM is the conditioning rule described in [11], known as "Dempster's conditioning rule," [12]. In effect, this method uses $S(u) \cap \mathbb{U}_{\mathcal{C},x}$ as the PRS and conditions the distribution, π, on the event, $\mathbb{U}_{\emptyset,x}^{\mathrm{c}}$. Define $K_x = \pi\{u : M_x(u, S) \cap \mathcal{C} = \emptyset\}$ as a function of the observed data, x. For any practical assertion, $\mathcal{A} \subset \mathcal{C}$, we have

$$\mathrm{Bel}_x(\mathcal{A} \mid \mathcal{C}; S) = \frac{\pi\{u : M_x(u, S) \cap \mathcal{C} \subseteq \mathcal{A},\, M_x(u, S) \cap \mathcal{C} \neq \emptyset\}}{1 - K_x}$$

and

$$\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}} \mid \mathcal{C}; S) = \frac{\pi\{u : M_x(u, S) \cap \mathcal{C} \subseteq \mathcal{A}^{\mathrm{c}},\, M_x(u, S) \cap \mathcal{C} \neq \emptyset\}}{1 - K_x}.$$

The following theorem, an extension of Theorem 1 in [1], states that over repeated observations of X, the conditional $\mathrm{Bel}_x(\mathcal{A} \mid \mathcal{C}; S)$ and $\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}} \mid \mathcal{C}; S)$ will be valid for $\mathcal{A}$. A proof can be found in Appendix A.2.

Theorem 3.2. *For a given value of $\alpha \in (0,1)$, suppose S is credible by Definition 2.2 and that $\pi\{u : M_x(u,S) = \emptyset\} > 0$ for some $x \in \mathcal{X}$. Then,* $\mathrm{Bel}_x(\mathcal{A} \mid \mathcal{C}; S)$ *and* $\mathrm{Bel}_x(\mathcal{A}^c \mid \mathcal{C}; S)$ *will satisfy Definition 2.1 for any $\mathcal{A} \subset \mathcal{C}$.*

Although beliefs resulting from conditioning are valid, the following example illustrates that they may be less likely to suggest the truth of $\mathcal{A}$ or $\mathcal{A}^c$ than beliefs computed with the unconstrained IM (1).

Example 3.3 (Conditioning method for constrained Gaussian mean). For the constraint set $\mathcal{C} = [a,b]$, the unconstrained Gaussian IM (3) will have conflict cases for focal elements indexed by $\{u : |u| < \max(x-b, a-x)\}$. If $a = -1/4$, $b = 1/4$, and $x = -1$, as in Example 3.2, then for $\mathcal{A} = \{\mu : \mu \geq 0\}$, the conditioning rule gives

$$\mathrm{Bel}_{-1}(\mathcal{A} \mid \mathcal{C}; S) = 0,$$

which is the same as Example 2.1, but

$$\mathrm{Bel}_{-1}(\mathcal{A}^c \mid \mathcal{C}; S) = \frac{\Phi(1) - \Phi(3/4)}{1 - \Phi(3/4)} \approx 0.30,$$

which is less than half of what was found in Example 2.1 where no constraint on μ was known.

As shown in Examples 2.1 and 3.3, introducing the constraint on μ values leads to weaker indications of whether or not $\mu \geq 0$, given the same evidence: $x = -1$. This is due to the large mass of conflict cases, $2\Phi(3/4) - 1 \approx 0.55$. The conditioning rule effectively ignores all these cases and distributes their mass over the non-conflict set. While both $\mathrm{Bel}_{-1}(\mathcal{A}^c; S)$ in Example 2.1 and $\mathrm{Bel}_{-1}(\mathcal{A}^c \mid \mathcal{C}; S)$ in Example 3.3 represent uncertainty about whether or not $\mathcal{A}^c$ is true, the reduction in $\mathrm{Bel}_{-1}(\mathcal{A}^c \mid \mathcal{C}; S)$ could be attributed to additional uncertainty about the data and model assumptions that led to conflict. However, for any $\mu \in [-1/4, 1/4]$ the probability of observing $x \leq -1$ is greater than 0.10. Under the model assumptions, it is not a rare event to observe $x = -1$ or something more extreme. If there is no other reason to doubt the validity of the data or model, then it seems paradoxical that introducing more information about possible μ values in the form of a constraint leads to weaker indications of whether or not $\mu \geq 0$. Conflict cases become subsets of $\mathcal{C}$ when using EB and therefore may become evidence for an assertion, $\mathcal{A} \subseteq \mathcal{C}$. Thus, when more information is known about a parameter via constraints, and there is no reason to doubt the data and model assumptions, the EB method may find stronger evidence for an assertion where the conditioning rule would find weaker evidence.

In some sense, the conditioning rule can also be understood as a different way of stretching $S(u)$ by replacing it with a larger one, especially when the PRS $\{S(u)\}_{u\in\mathbb{U}}$ forms a nested sequence. In that case, only those $S(u)$ large enough to intersect with the a-constraint set, $\mathbb{U}_{\mathcal{C},x}$, will be considered. For $u \in \mathbb{U}_{\emptyset,x}$, the set, $S(u)$, is too small and will be thrown away. Compared to EB, the conditioning rule stretches stochastically more than necessary. This explains intuitively why the conditioning rule is valid but sometimes inefficient.

4. Gaussian Observation with Bounded Mean

Consider computing $\mathrm{Bel}_x^{\mathrm{EB}}$ for the nonnegative mean example using the IM obtained with the EB method in Example 3.2. In this case $a = 0$ and $b \to \infty$. Let $\mathcal{A} = \{\mu : \mu = \mu_0\}$ be the assertion of interest. Then,

$$\mathrm{Bel}_x^{\mathrm{EB}}(\{\mu_0\}; \mathcal{S}) = \begin{cases} 1 - 2\Phi(x), & \text{if } \mu_0 = 0 \text{ and } x < 0; \\ 0, & \text{otherwise;} \end{cases}$$

$$\mathrm{Bel}_x^{\mathrm{EB}}(\{\mu_0\}^{\mathrm{c}}; \mathcal{S}) = \begin{cases} 2\Phi(|x-\mu_0|) - 1, & \text{if } \mu_0 > 0; \\ 2\Phi(x) - 1, & \text{if } \mu_0 = 0 \text{ and } x > 0; \\ 0, & \text{otherwise;} \end{cases}$$

for $\mu_0 \in [0, \infty)$. In many situations, the goal is to infer the presence or absence of a signal and so $\mu = 0$ is the assertion of interest. Fig. 1 illustrates $\mathrm{Bel}_x^{\mathrm{EB}}$ and $\mathrm{Pl}_x^{\mathrm{EB}}$ for $\mathcal{A}_0 = \{\mu : \mu = 0\}$. For negative values of x, $\mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A}_0; \mathcal{S})$ is large. For $x \leq -2$, $\mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A}_0; \mathcal{S}) > 0.95$, suggesting that $\mathcal{A}_0$ is true. For positive x values, $\mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A}_0; \mathcal{S})$ drops to zero and $\mathrm{Pl}_x^{\mathrm{EB}}(\mathcal{A}_0; \mathcal{S})$ becomes small. For $x \geq 2$, $\mathrm{Pl}_x^{\mathrm{EB}}(\mathcal{A}_0; \mathcal{S}) < 0.05$, suggesting that $\mathcal{A}_0$ is false. When x is close to zero, $\mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A}_0; \mathcal{S})$ is small while $\mathrm{Pl}_x^{\mathrm{EB}}(\mathcal{A}_0; \mathcal{S})$ is large. In these cases, it may be difficult to make any conclusion about $\mathcal{A}_0$.

Applying the conditioning rule to this problem, one obtains:

$$\mathrm{Bel}_x(\{\mu_0\} \mid \mathcal{C}; S) = 0$$

for any μ_0 and

$$\mathrm{Bel}_x(\{\mu_0\}^{\mathrm{c}} \mid \mathcal{C}; S) = \begin{cases} \frac{\Phi(\mu_0 - x) - \Phi(-x)}{1 - \Phi(-x)}, & \text{if } x < 0; \\ 2\Phi(|x-\mu_0|) - 1, & \text{if } x \geq 0. \end{cases}$$

For the assertion $\mathcal{A}_0 = \{\mu : \mu = 0\}$, both the conditioning rule and EB give the same plausibility. No matter what value of x is observed, $\mathrm{Pl}_x^{\mathrm{EB}}(\mathcal{A}_0; \mathcal{S}) = \mathrm{Pl}_x(\mathcal{A}_0 \mid \mathcal{C}; S)$. However, $\mathrm{Bel}_x(\mathcal{A}_0 \mid \mathcal{C}; S) = 0$, and so, unlike with EB, no x observation ever supports the assertion.

The IM obtained with the EB method in Example 3.2 can also be used to create a *plausibility interval* for μ based on the observed x. For a level $\gamma \in (0, 1)$, let $z_\gamma = \Phi^{-1}(\frac{1+\gamma}{2})$. Then,

$$\{\mu_0 : \mathrm{Pl}_x^{\mathrm{EB}}(\{\mu_0\}; \mathcal{S}) \geq 1 - \gamma\} = [\max\{0, x - z_\gamma\}, \max\{0, x + z_\gamma\}].$$

A level γ plausibility interval has coverage probability of at least γ over repeated experiments [22]. Similarly, a level γ plausibility interval for μ can be created using the IM formed with the conditioning rule:

$$[\max\{0, x - z_\gamma\}, x + z_\gamma\}],$$

when $x \geq 0$, and

$$[0, x + \Phi^{-1}(\gamma + (1 - \gamma)\Phi(-x))]$$

when $x < 0$. Both of these intervals are illustrated in Fig. 2 for $\gamma = 0.9$. The shaded region is the plausibility interval obtained by the EB method while the dashed line marks the boundary of the plausibility interval found with the conditioning rule. Their lower boundaries coincide for every x and their upper boundaries coincide when x is non-negative. When $x \in (-z_\gamma, \infty)$, for any specific μ_0 in the EB interval interior $\mathrm{Bel}_x^{\mathrm{EB}}(\{\mu_0\}; \mathcal{S}) < \gamma$ and $\mathrm{Bel}_x^{\mathrm{EB}}(\{\mu_0\}^{\mathrm{c}}; \mathcal{S}) < \gamma$. So there is not enough evidence to either support or deny μ_0 at level γ. When $x \in (-\infty, -z_\gamma]$, the EB interval collapses to a single point where one concludes that $\mu = 0$ with $\mathrm{Bel}_x^{\mathrm{EB}}(\{0\}; \mathcal{S}) \geq \gamma$. If there is no reason to doubt the data and model assumptions, then the EB method says that these improbable observations are consistent with $\mu = 0$ far more than any other value of μ, and in fact these improbable x values support the hypothesis that $\mu = 0$. As expressed in [23], an interval construction that collapses to a point for improbable observations is a reflection of the strength of evidence. Using the EB method, this is explicitly quantified by computing $\mathrm{Bel}_x^{\mathrm{EB}}(\{\mu_0\}; \mathcal{S})$ and $\mathrm{Pl}_x^{\mathrm{EB}}(\{\mu_0\}; \mathcal{S})$ for hypothetical μ_0 values. The interval found with the conditioning rule has strictly positive length for any observed x. All μ_0 values within the interval are plausible with respect to level γ. However, since $\mathrm{Bel}_x(\{\mu_0\} \mid \mathcal{C}; \mathcal{S}) < \gamma$ in the interval, no specific μ_0 value is supported, no matter how improbable the observed x is.

5. Poisson Count with Known Background Rate

Now inference is considered for the signal rate, λ, from a Poisson count, Y, when there is a known background rate, b. With b known, the overall rate is $\theta = \lambda + b$. For inference about λ, it is sufficient to perform inference about θ with the constraint set, $\mathcal{C} = \{\theta : \theta \geq b\}$.

5.1. Inferential Model

The a-step relies on the the following relationship between the Poisson and Gamma distributions. Let G_y be the cdf for the Gamma distribution with shape y and scale 1. Also, let F_θ be the cdf for the Poisson distribution with rate θ. Then, if y is a nonnegative integer,

$$G_{y+1}(\theta) = \int_0^\theta \frac{t^y e^{-t}}{y!}\,dt = 1 - \int_\theta^\infty \frac{t^y e^{-t}}{y!}\,dt = 1 - \sum_{k=0}^{y} \frac{e^{-\theta}\theta^k}{k!} = 1 - F_\theta(y).$$

Let $G_0(\theta) = 1 - F_\theta(-1) = 1$. The a-variable, U, has a uniform mass distribution over $\mathbb{U} = [0, 1]$. Because Y is discrete, the a-equation is a many-to-one mapping:

$$\begin{aligned} a(\theta, u) &= \{Y : F_\theta(Y-1) \leq u \leq F_\theta(Y)\} \\ &= \{Y : G_{Y+1}(\theta) \leq 1 - u \leq G_Y(\theta)\} \\ &= \{Y : G_Y^{-1}(1-u) \leq \theta \leq G_{Y+1}^{-1}(1-u)\}, \end{aligned} \tag{7}$$

where G_y^{-1} is the quantile function for the Gamma distribution with shape y and scale 1, and $G_0^{-1}(\theta) \equiv 0$. An alternative a-model can be derived from the waiting times in a Poisson process [24, 13]. That a-model introduces additional challenges because the a-variable has more than one dimension and its distribution depends on x. The Poison process a-model is compared to the present a-model in Appendix B.

For an unconstrained model,

$$S(u) = \left[\tfrac{1}{2} - |u - \tfrac{1}{2}|,\ \tfrac{1}{2} + |u + \tfrac{1}{2}|\right] \tag{8}$$

satisfies the criteria in Section 2.4 to be an efficient PRS for predicting U from a uniform distribution on $[0, 1]$. This can be adapted for an EPRS in the p-step:

$$S_e(u) = \left[(1-e)\left(\tfrac{1}{2} - \left|u - \tfrac{1}{2}\right|\right),\ (1-e)\left(\tfrac{1}{2} + \left|u - \tfrac{1}{2}\right|\right) + e\right], \qquad e \in [0, 1].$$

which is (8) when $e = 0$ and increases to $\mathbb{U} = [0, 1]$ as $e \to 1$. This gives

$$M_y(u, S_e) = \left[G_y^{-1}\left((1-e)(\tfrac{1}{2} - \left|u - \tfrac{1}{2}\right|)\right), G_{y+1}^{-1}\left((1-e)(\tfrac{1}{2} + \left|u - \tfrac{1}{2}\right|) + e\right)\right].$$

To handle conflict cases, the EPRS is expanded with

$$\begin{aligned}\hat{e} &= \min\{e : [b, \infty) \cap M_y(u, S_e) \neq \emptyset\} \\ &= \begin{cases} \frac{G_{y+1}(b) - \frac{1}{2} - \left|u - \frac{1}{2}\right|}{\frac{1}{2} - \left|u - \frac{1}{2}\right|} & \text{if } F_b(y) < \frac{1}{2} - |u - \frac{1}{2}|; \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

The resulting IM focal elements are:

$$M_y^{\text{EB}}(u, \mathcal{S}) = \left[\max\left\{b, G_y^{-1}\left(\tfrac{1}{2} - \left|u - \tfrac{1}{2}\right|\right)\right\}, \max\left\{b, G_{y+1}^{-1}\left(\tfrac{1}{2} + \left|u - \tfrac{1}{2}\right|\right)\right\}\right].$$

For point assertions of the form $\mathcal{A} = \{\theta : \theta = \theta_0\}$ we have the following Bel_y^{EB} when $\theta_0 = b$,

$$\text{Bel}_y^{\text{EB}}(\{b\}; \mathcal{S}) = \begin{cases} 2G_{y+1}(b) - 1, & \text{if } F_b(y) \leq 1/2; \\ 0, & \text{otherwise;} \end{cases}$$

$$\text{Bel}_y^{\text{EB}}(\{b\}^{\text{c}}; \mathcal{S}) = \begin{cases} 1 - 2G_y(b), & \text{if } F_b(y-1) > 1/2; \\ 0, & \text{otherwise;} \end{cases}$$

and for $\theta_0 > b$:

$$\text{Bel}_y^{\text{EB}}(\{\theta_0\}; \mathcal{S}) = 0$$

$$\text{Bel}_y^{\text{EB}}(\{\theta_0\}^{\text{c}}; \mathcal{S}) = \begin{cases} 2G_{y+1}(\theta_0) - 1, & \text{if } F_{\theta_0}(y) < 1/2; \\ 1 - 2G_y(\theta_0), & \text{if } F_{\theta_0}(y-1) > 1/2; \\ 0, & \text{otherwise.} \end{cases}$$

Just as in the constrained Gaussian example, we can test for the absence of a signal. This is represented by the assertion $\mathcal{A}_b = \{\theta : \theta = b\}$. Fig. 3 illustrates Bel_y^{EB} and Pl_y^{EB} for this assertion when $b = 15$.

A plausibility interval can also be created for the unknown θ. For $\gamma \in (0, 1)$,

$$\{\theta_0 : \text{Pl}_y^{\text{EB}}(\{\theta_0\}; \mathcal{S}) \geq 1 - \gamma\} = [\max\{b, G_y^{-1}(\tfrac{1-\gamma}{2})\}, \max\{b, G_{y+1}^{-1}(\tfrac{1+\gamma}{2})\}].$$

The interval behaves similarly to the Gaussian interval: when $F_b(y) \leq \frac{1-\gamma}{2}$, then $\text{Bel}_y^{\text{EB}}(\{b\}; \mathcal{S}) \geq \gamma$, but for $F_b(y) > \frac{1-\gamma}{2}$, any θ_0 on the interval interior has $\text{Bel}_y^{\text{EB}}(\{\theta_0\}; \mathcal{S}) < \gamma$ and $\text{Bel}_y^{\text{EB}}(\{\theta_0\}^{\text{c}}; \mathcal{S}) < \gamma$. Fig. 4 illustrates the level 0.9 plausibility interval for $b = 15$.

5.2. Numerical Comparison

The level γ plausibility interval coverage probability is at least γ in repeated experiments. The following methods were also designed to achieve proper coverage probability. Numerical results illustrate the relative performance of the new Poisson plausibility interval compared to the existing methods.

Feldman and Cousins [25] constructed confidence bounds with proper coverage by filling acceptance intervals with points ordered according to a likelihood ratio. Giunti [26] argued that it is undesirable for the upper confidence bound to decrease in b when small values of Y are observed and proposed a modification to the ranking method that lessens the rate of decrease.

Roe and Woodroofe [27] noted that observing $Y = 0$ is equivalent to observing $S = 0$ and $B = 0$. When the number of signal events is known, the interval bounds for λ should not depend on b. This issue is addressed in [27] by forming an interval conditioned on the fact that $B \leq y$ when $Y = y$ is observed. This method may undercover over all repeated experiments. Mandelkern and Schultz [28] provided an "*ad hoc*" [5] remedy by shifting the upper bound of each acceptance interval until proper unconditional coverage was achieved.

The conditional probability used to form intervals in [27] has the same form as the posterior density for λ when given a uniform prior over $[0, \infty)$. Roe and Woodroofe [29] developed this into a procedure for constructing a Bayesian credible interval. While this method has appropriate conditional coverage probability, Roe and Woodroofe [29] employed an "*ad hoc*" adjustment of the bounds to obtain appropriate unconditional coverage.

Confidence intervals derived from maximum likelihood estimators [30] differ from other methods in that the interval bounds remain constant for all observations outside of the constrained parameter space. Constructing the interval from the sampling distribution of the estimator ensures proper coverage.

For $\gamma = 0.9$, $b = 3$, and λ ranging from 0 to 4, Fig. 5 shows the plausibility interval coverage probability compared to the existing methods. The Feldman and Cousins [25] and Roe and Woodroofe [29] methods had coverage probability at least as large the plausibility interval for most values of λ. The "*ad hoc*" adjustment of Mandelkern and Schultz [28] to the Roe and Woodroofe [27] conditional intervals tended to have coverage closer to

0.9 than the plausibility interval. However, the construction of the plausibility interval guarantees proper coverage so that *ad hoc* adjustments are not necessary. Furthermore, to our knowledge there is no analytical expression for the Roe and Woodroofe [27] interval nor an expression that includes the Mandelkern and Schultz [28] adjustment. For this Poisson example, the plausibility interval expression requires less computation to produce numerical values. The intervals of Giunti [26] and Mandelkern and Schultz [30] provided coverage closer to 0.9 than the plausibility interval for most values of λ, with the Giunti [26] method providing the best coverage of all.

Table 1 lists the level 0.9 interval bounds obtained from the EB plausibility interval and the other methods for several values of y when $b = 3$. The interval widths for the different methods are plotted in Fig. 6. For $y < b$, the plausibility interval is narrower than those produced by most of the other methods. When $y \geq b$, the plausibility interval becomes wider than the others. This greater width causes the peaks in coverage probability seen in Fig. 5. For example, λ values in [1.70, 1.74] are covered by the plausibility interval when $y \in [1, 9]$. Hence, the coverage probability is near 0.97. Most of the other methods cover λ values in this range when $y \in [0, 8]$, which gives coverage probabilities closer to 0.95. Fig. 7 shows the maximum and minimum coverage probabilities of the level γ plausibility interval when $\gamma \in [0.5, 1]$ and $\lambda \in [0, 100]$. Within this range of λ values the minimum coverage probability is close to γ. As the λ range is narrowed, the minimum coverage probability becomes larger for many values of γ due to discreteness. It may be possible to obtain a specific minimum coverage probability for a given λ range by choosing a smaller γ value.

6. Concluding Remarks

The theory of IMs allows direct probabilistic inference from data to parameters without introducing priors or relying on asymptotic arguments. The EB method presented here extends the IM theory to situations where conflict cases can arise from parameter constraints. As an alternative to the conditioning rule, it achieves higher efficiency by using conflict cases as evidence for specific parameter values. This is a reasonable choice when one holds the constraint and model assumptions to be valid and hence cannot attribute conflict to uncertainty about these assumptions. The probability represented by $\mathrm{Bel}_x^{\mathrm{EB}}$, the belief function obtained from the EB method, is calibrated to a frequency interpretation for any assertion. As functions of an

Table 1: Level 0.9 interval bounds for λ when $b = 3$.

	EB Plaus. Int.		FC98 [25]		Giunti99 [26]		RW99+MS00a [28]		RW00 [29]		MS00b [30]	
y	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
0	0	0	0	1.08	0	1.82	0	2.44	0	2.53	0	4.69
1	0	1.74	0	1.88	0	2.42	0	2.95	0	3.09	0	4.69
2	0	3.30	0	3.04	0	3.52	0	3.75	0	3.82	0	4.69
3	0	4.75	0	4.42	0	4.76	0	4.80	0	4.71	0	4.69
4	0	6.15	0	5.60	0	5.69	0	6.01	0	5.74	0	5.60
5	0	7.51	0	6.99	0	7.10	0	7.28	0	6.85	0	7.04
6	0	8.84	0.15	8.47	0.15	8.54	0.16	8.42	0	8.07	0.16	8.64
7	0.29	10.15	0.89	9.53	0.90	9.56	0.9	9.58	0.55	9.29	0.90	9.54
8	0.98	11.43	1.51	10.99	1.66	11.03	1.66	11.02	1.21	10.62	1.66	11.08
9	1.70	12.71	1.88	12.30	2.38	12.30	2.44	12.23	1.90	11.91	2.44	12.30
10	2.43	13.96	2.63	13.50	2.98	13.53	2.98	13.51	2.64	13.24	3.23	13.55
11	3.17	15.21	3.04	14.81	3.52	14.81	3.75	14.77	3.37	14.47	4.03	14.81
12	3.92	16.44	4.01	16.00	4.36	16.03	4.52	16.01	4.14	15.69	4.69	16.05

assertion, likelihood and p-value functions [8, 9] are also available as inferential tools in the constrained Gaussian and Poisson examples, but $\mathrm{Bel}_x^{\mathrm{EB}}$ has the advantage of a predictive probability interpretation.

From $\mathrm{Bel}_x^{\mathrm{EB}}$ it is easy to construct plausibility intervals containing hypothetical parameter values that are supported by, or at least consistent with, the evidence presented in the data. The two-sided PRSs considered here resulted in two-sided plausibility intervals (Figs. 2 and 4). Although efficient and mathematically convenient, the symmetrical PRS (8) used in the Poisson example is sometimes larger than necessary. This caused the EB plausibility interval to be slightly wider for certain y values than intervals created with other methods. A more efficient IM and narrower plausibility interval may be obtained by considering an assertion-specific PRS for each $\{\theta_0\}$ assertion. The authors are currently investigating this approach. One-sided plausibility intervals may be obtained using one-sided PRSs. In the Poisson example, a one-sided plausibility interval is expected to have better performance than other methods because the skewness of the Poisson distribution will no longer create the difficulties that arise when using interval length as a criterion.

In the presentation of the EB method, it was assumed that there existed a minimum intersection of the EPRS and the a-constraint set, $\mathbb{U}_{\mathcal{C},x}$. The EPRS may be designed so that it is always a closed set (except, possibly, the a-space itself). However, in some situations the constraint set may be problematic. For example, the Gaussian mean could be strictly positive: $\mu \in \mathcal{C} = (0, \infty)$. In this case one could build an IM with $\mathcal{C} = [0, \infty)$ instead. Any mass placed on $\mu = 0$ could be logically interpreted as evidence for 0^+, a point infinitesimally larger than zero.

Finally, the EB method can be used for more general, data-dependent conflict cases. The EB approaches demonstrated here can be extended to situations with nuisance parameters as in [31] and [32]. Before applying the EB method, however, a problem may be simplified by handling nuisance parameters with the marginalization methods of [33]. When there are multiple observations, the conditioning methods described in [22] can reduce the data dimensionality in a manner similar to sufficient statistics.

7. Acknowledgments

The authors thank Ryan Martin and the anonymous reviewers for their insightful and constructive comments during the preparation of this article.

Chuanhai Liu is supported by the National Science Foundation, Grant DMS-1007678.

Appendix A. Technical Results

Appendix A.1. Validity of the Elastic Belief Method

The following is a proof of Theorem 3.1.

(i) The relationships between $\mathrm{Bel}_x^{\mathrm{EB}}$ and Bel_x can be shown by expanding the definition of $\mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A};\mathcal{S})$:

$$\begin{aligned}\mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A};\mathcal{S}) &= \pi\{u : M_x^{\mathrm{EB}}(u,\mathcal{S}) \subseteq \mathcal{A}\}\\ &= \pi\{u : M_x(u,S_0)\cap\mathcal{C} \subseteq \mathcal{A}, M_x(u,S_0)\cap\mathcal{C} \neq \emptyset\}\\ &+ \pi\{u : M_x(u,S_{\hat{e}})\cap\mathcal{C} \subseteq \mathcal{A}, M_x(u,S_0)\cap\mathcal{C} = \emptyset\}.\end{aligned}$$

The inequality,

$$\mathrm{Bel}_x(\mathcal{A};S_0) \leq \mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A};\mathcal{S}),$$

follows from the fact that

$$\begin{aligned}\pi\{u : M_x(u,S_0)\cap\mathcal{C} \subseteq \mathcal{A}, M_x(u,S_0)\cap\mathcal{C} \neq \emptyset\} &\geq \pi\{u : M_x(u,S_0) \subseteq \mathcal{A}\}\\ &= \mathrm{Bel}_x(\mathcal{A};S_0).\end{aligned}$$

The inequality,

$$\mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A};\mathcal{S}) \leq \mathrm{Bel}_x(\mathcal{A}\cup\mathcal{C}^{\mathrm{c}};S_0), \tag{A.1}$$

is determined by the mass on conflict cases that support $\mathcal{A}$ in the IM resulting from the EB method:

$$\begin{aligned}\pi\{u : M_x(u,S_{\hat{e}})\cap\mathcal{C} \subseteq \mathcal{A}, M_x(u,S_0)\cap\mathcal{C} = \emptyset\} &\leq \pi\{u : M_x(u,S_0)\cap\mathcal{C} = \emptyset\}\\ &= K_x,\end{aligned}$$

with equality when all of the conflict cases support $\mathcal{A}$ after using the EB method. It follows that

$$\begin{aligned}\mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A};\mathcal{S}) &\leq \pi\{u : M_x(u,S_0)\cap\mathcal{C} \subseteq \mathcal{A}, M_x(u,S_0)\cap\mathcal{C} \neq \emptyset\} + K_x\\ &= \pi\{u : M_x(u,S_0) \subseteq \mathcal{A}\cup\mathcal{C}^{\mathrm{c}}\} = \mathrm{Bel}_x(\mathcal{A}\cup\mathcal{C}^{\mathrm{c}};S_0).\end{aligned}$$

(ii) The validity of $\mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A};\mathcal{S})$ follows by considering the random variables, $\mathrm{Bel}_X^{\mathrm{EB}}(\mathcal{A};\mathcal{S})$ and $\mathrm{Bel}_X(\mathcal{A}\cup\mathcal{C}^{\mathrm{c}};S_0)$ as functions of the random variable, X. By satisfying (4),

$$\mathrm{Pr}_\theta\{x:\mathrm{Bel}_x(\mathcal{A}\cup\mathcal{C}^{\mathrm{c}};S_0)\geq 1-\alpha\}\leq\alpha$$

for any $\theta\in(\mathcal{A}\cup\mathcal{C}^{\mathrm{c}})^{\mathrm{c}}=\mathcal{A}^{\mathrm{c}}\cap\mathcal{C}$. The inequality (A.1) implies $\mathrm{Bel}_X^{\mathrm{EB}}(\mathcal{A};\mathcal{S})$ is stochastically smaller than $\mathrm{Bel}_X(\mathcal{A}\cup\mathcal{C}^{\mathrm{c}};S_0)$. Thus,

$$\mathrm{Pr}_\theta\{x:\mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A};\mathcal{S})\geq 1-\alpha\}\leq\mathrm{Pr}_\theta\{x:\mathrm{Bel}_x(\mathcal{A}\cup\mathcal{C}^{\mathrm{c}};S_0)\geq 1-\alpha\}\leq\alpha.$$

for any $\theta\in\mathcal{A}^{\mathrm{c}}\cap\mathcal{C}$. Also, by the same argument when $\mathrm{Bel}_X(\mathcal{A}^{\mathrm{c}}\cup\mathcal{C}^{\mathrm{c}};S_0)$ satisfies (4),

$$\mathrm{Pr}_\theta\{x:\mathrm{Bel}_x^{\mathrm{EB}}(\mathcal{A}^{\mathrm{c}};\mathcal{S})\geq 1-\alpha\}\leq\mathrm{Pr}_\theta\{x:\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}}\cup\mathcal{C}^{\mathrm{c}};S_0)\geq 1-\alpha\}\leq\alpha.$$

for any $\theta\in\mathcal{A}\cap\mathcal{C}=\mathcal{A}$.

(iii) If $\mathcal{S}$ satisfies property (a) of Definition 3.1, then S_0 satisfies Definition 2.2. By [1, Theorem 1], for any $\mathcal{A}\subset\mathcal{C}$, Bel_x is valid for inference about $\mathcal{A}\cup\mathcal{C}^{\mathrm{c}}$ and $\mathcal{A}^{\mathrm{c}}\cup\mathcal{C}^{\mathrm{c}}$. The result follows from applying part (ii) of Theorem 3.1.

∎

Appendix A.2. Validity of the Conditioning Rule

The following is a proof of Theorem 3.2. For the sample space, $\mathcal{X}$, and parameter space, Θ, let U be defined in the a-space, $\mathbb{U}$ with distribution π. Then, let $S(U)\subseteq\mathbb{U}$ be the PRS and obtain the IM (1) for Θ with focal elements $\{M_x(u,S)\}_{u\in\mathbb{U}}$. Assume $M_x(u,S)$ was designed so that $\pi\{u: M_x(u,S)=\emptyset\}=0$ for every $x\in\mathcal{X}$. This gives

$$\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}};S)=\pi\{u:M_x(u,S)\subseteq\mathcal{A}^{\mathrm{c}}\}\tag{A.2}$$

as the evidence *against* the assertion, $\mathcal{A}$. Further, suppose $M_x(u,S)$ is valid for inference about $\mathcal{A}$ as in Definition 2.1 so that,

$$\mathrm{Pr}_\theta\{x:\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}};S)\geq 1-\alpha\}\leq\alpha,$$

for $\alpha\in(0,1)$ and every $\theta\in\mathcal{A}$.

Next, suppose a constraint on the parameter space, $\mathcal{C} \subset \Theta$, is introduced such that θ is known to lie inside $\mathcal{C}$. The evidence against an assertion, $\mathcal{A} \subseteq \mathcal{C}$, is defined by the conditioning rule as the conditional probability,

$$\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}} \mid \mathcal{C}; S) = \pi\{u : M_x(u, S) \cap \mathcal{C} \subseteq \mathcal{A}^{\mathrm{c}} \mid M_x(u, S) \cap \mathcal{C} \neq \emptyset\}.$$

This can be written as:

$$\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}} \mid \mathcal{C}; S) = \frac{\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}} \cup \mathcal{C}^{\mathrm{c}}; S) - \mathrm{Bel}_x(\mathcal{C}^{\mathrm{c}}; S)}{1 - \mathrm{Bel}_x(\mathcal{C}^{\mathrm{c}}; S)}.$$

The proof follows from the fact that $\mathcal{A} \subseteq \mathcal{C}$ implies $\mathcal{C}^{\mathrm{c}} \subseteq \mathcal{A}^{\mathrm{c}}$. Thus,

$$\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}} \mid \mathcal{C}; S) = \frac{\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}}; S) - \mathrm{Bel}_x(\mathcal{C}^{\mathrm{c}}; S)}{1 - \mathrm{Bel}_x(\mathcal{C}^{\mathrm{c}}; S)}.$$

The proof is completed by noting that $\frac{a-b}{1-b} \leq a$ when $a, b \in [0, 1)$ and $a \geq b$. Here,

$$a = \mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}}; S)$$

and

$$b = \mathrm{Bel}_x(\mathcal{C}^{\mathrm{c}}; S).$$

Thus, for every $x \in \mathcal{X}$, we have $\mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}} \mid \mathcal{C}; S) \leq \mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}}; S)$, with equality when $\pi\{u : M_x(u, S) \cap \mathcal{C} = \emptyset\} = 0$. Therefore,

$$\mathrm{Pr}_\theta\{x : \mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}} \mid \mathcal{C}; S) \geq 1 - \alpha\} \leq \mathrm{Pr}_\theta\{x : \mathrm{Bel}_x(\mathcal{A}^{\mathrm{c}}; S) \geq 1 - \alpha\} \leq \alpha$$

for every $\theta \in \mathcal{A}$. The same argument can be repeated with $\mathrm{Bel}_X(\mathcal{A} \mid \mathcal{C}; S) = \mathrm{Bel}_X((\mathcal{A}^{\mathrm{c}})^{\mathrm{c}} \mid \mathcal{C}; S)$ to show that

$$\mathrm{Pr}_\theta\{x : \mathrm{Bel}_x(\mathcal{A} \mid \mathcal{C}; S) \geq 1 - \alpha\} \leq \mathrm{Pr}_\theta\{x : \mathrm{Bel}_x(\mathcal{A}, S) \geq 1 - \alpha\} \leq \alpha$$

for every $\theta \in \mathcal{A}^{\mathrm{c}}$. This shows that the conditioning rule preserves the IM's validity for an assertion in the presence of a constraint on Θ. ∎

Although not required for the proof, it is worthwhile to consider conditions under which $\pi\{u : M_x(u, S) \cap \mathcal{C} = \emptyset\} = 0$. If S has a neutral point, u_0, such that $u_0 \in S(u)$ for every $u \in \mathbb{U}$, then we can partition $\mathcal{X}$ by the impossibility of conflict cases. Let $M_x(u)$ be the basic IM obtained with the singleton PRS, $S(u) = \{u\}$ and define

$$\mathcal{X}_{\mathrm{NC}} = \{x : M_x(u_0) \cap \mathcal{C} \neq \emptyset\}.$$

Then, for any $x \in \mathcal{X}_{\rm NC}$ and any $u \in \mathbb{U}$, we have $M_x(u_0) \subseteq M_x(u, S)$ and so $M_x(u, S) \cap \mathcal{C} \neq \emptyset$. Therefore, on $\mathcal{X}_{\rm NC}$,

$$\{u : M_x(u, S) \cap \mathcal{C} = \emptyset\} = \emptyset,$$

which implies $\mathrm{Bel}_x(\mathcal{A}^{\rm c} \mid \mathcal{C}; S) = \mathrm{Bel}_x(\mathcal{A}^{\rm c}; S)$ for any assertion, $\mathcal{A}$.

Appendix B. Poisson process a-model

A data-generating model for a Poisson random variable can be built from the waiting times of a Poisson process. In [24] and [13], this model was used to build belief functions for inference about the parameter in the Poisson model, θ. Compared to the a-model in Section 5.1, the Poisson process model poses additional challenges to building an efficient IM. These challenges are present even in situations without parameter constraints.

Suppose there is an infinite sequence, $T_1, T_2, \ldots$, of independent, exponentially distributed random variables with unit rate. Let $T_0 = 0$ and let

$$S_i = \sum_{j=0}^{i} T_j.$$

If $Y = \max\{i : S_i \leq \theta\}$, then Y follows the Poisson distribution with rate θ. For fixed θ, a realization of Y can be simulated by generating successive T_i from the exponential distribution until

$$S_{i-1} \leq \theta < S_{i-1} + T_i \tag{B.1}$$

and then taking $y = i - 1$ as the realization. When $Y = y$ is observed and θ is unknown, this implies a random interval for θ:

$$S_y \leq \theta < S_y + T_{y+1}.$$

that can be used to build an a-model with the infinite-dimensional a-variable, $\mathbf{T} = (T_0, T_1, T_2, \ldots)$. A valid IM can be created using this a-model by finding a PRS for some function of $\mathbf{T}$. The infinite dimensionality makes it difficult to find an efficient PRS for $\mathbf{T}$ directly. However, a conditional IM [22] can be created using the finite-dimensional a-variable $(T_1, T_2, \ldots, T_{y+1})$. The dimensionality can be reduced further to the two-dimensional a-variable: (S_y, T_{y+1}).

In general, higher dimensional a-variables are more difficult to predict efficiently. Ultimately, efficient inference is performed by reducing the dimensionality of the a-variable to that of the parameter using the conditioning methods of [22] whenever possible. This is facilitated by choosing an initial a-model with an a-variable of lowest possible dimension. If an a-model is built from a data generating model, it is usually possible to find an a-variable that has the same dimensionality as the number of independent data observations. For inference about the parameter of a Poisson model from a single observation, the a-model in Section 5.1 has a one-dimensional a-variable while the Poisson process leads to an a-model with a two-dimensional a-variable. One expects that predicting a two-dimensional a-variable will be less efficient for inference about a scalar parameter than predicting a one-dimensional a-variable. Consequently, the resulting plausibility intervals will be wider for an IM based on the Poisson process a-model.

In order to predict (S_y, T_{y+1}), let:

$$U_1 = G_y(S_y)$$

and

$$U_2 = 1 - \exp\{-T_{y+1}\},$$

where $\mathbf{U} = (U_1, U_2)$ has a uniform mass distribution over $[0, 1]^2$. An efficient PRS for $\mathbf{U}$ is:

$$S(\mathbf{u}) = \{\mathbf{u}' : ||\mathbf{u}' - \mathbf{h}||_\infty \leq ||\mathbf{u} - \mathbf{h}||_\infty\},$$

where $||\mathbf{t}||_\infty = \max\{|t_1|, |t_2|\}$ and $\mathbf{h} = (\frac{1}{2}, \frac{1}{2})$. Applying this PRS to the a-model (B.1) gives focal elements of the form:

$$M_y^{(2)}(\mathbf{u}, S) = \\ \left[G_y^{-1}(\tfrac{1}{2} - ||\mathbf{u} - \mathbf{h}||_\infty),\, G_y^{-1}(\tfrac{1}{2} + ||\mathbf{u} - \mathbf{h}||_\infty) + \exp\{\tfrac{1}{2} + ||\mathbf{u} - \mathbf{h}||_\infty\}\right]. \quad \text{(B.2)}$$

Using the a-model in Section 5.1 and PRS (8) gives the focal elements:

$$M_y^{(1)}(u, S) = \left[G_y^{-1}(\tfrac{1}{2} - |u - \tfrac{1}{2}|),\, G_{y+1}^{-1}(\tfrac{1}{2} + |u - \tfrac{1}{2}|)\right]. \quad \text{(B.3)}$$

Fig. 8 shows that the plausibility interval based upon the Poisson process IM (B.2) is indeed wider than the plausibility interval from the a-model of Section 5.1 (B.3) for $y \in [0, 50]$.

References

[1] R. G. Martin, J. Zhang, C. Liu, Dempster-Shafer theory and statistical inference with weak beliefs, Statist. Sci. 25 (1) (2010) 72–87.

[2] J. Zhang, C. Liu, Dempster-Shafer inference with weak beliefs, Statist. Sinica 21 (2) (2011) 475–494.

[3] R. G. Martin, C. Liu, Inferential models: A framework for prior-free posterior probabilistic inference, unpublished manuscript, available at http://www.stat.purdue.edu/∼chuanhai/docs/imbasics.pdf (2011).

[4] D. Ermini Leaf, J. Hui, C. Liu, Statistical inference with a single observation of $N(\theta, 1)$, Pakistan J. Statist. 25 (4) (2009) 571–586.

[5] M. Mandelkern, Setting confidence intervals for bounded parameters, Statist. Sci. 17 (2) (2002) 149–172.

[6] A. P. Dempster, On the difficulities inherent in Fisher's fiducial argument, J. Amer. Statist. Assoc. 59 (1964) 56–66.

[7] J. Neyman, Outline of a theory of statistical estimation based on the classical theory of probability, Philos. Trans. Roy. Soc. London Ser. A 236 (1937) 333–380.

[8] L. J. Gleser, Comment on "Setting confidence intervals for bounded parameters", Statist. Sci. 17 (2) (2002) 161–163.

[9] D. A. S. Fraser, N. Reid, A. C. M. Wong, Inference for bounded parameters, Phys. Rev. D 69 (3) (2004) 033002.

[10] R. A. Fisher, Statistical Methods and Scientific Inference, Hafner Press [A Division of Macmillan Publishing Co., Inc.], New York, 1973.

[11] A. P. Dempster, Upper and lower probabilities induced by a multivalued mapping, Ann. Math. Statist. 38 (1967) 325–339.

[12] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, NJ, 1976.

[13] A. P. Dempster, Dempster-Shafer calculus for statisticians, Internat. J. of Approx. Reason. 48 (2008) 365–377.

[14] R. A. Fisher, Inverse probability, Mathematical Proceedings of the Cambridge Philosophical Society 26 (04) (1930) 528–535.

[15] A. P. Dempster, On direct probabilites, J. Roy. Statist. Soc. Ser. B 25 (1963) 100–110.

[16] A. P. Dempster, New methods for reasoning towards posterior distributions based on sample data, Ann. Math. Statist. 37 (1966) 355–374.

[17] J. Kohlas, P.-A. Monney, A Mathematical Theory of Hints: An Approach to the Dempster-Shafer Theory of Evidence, Vol. 425 of Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, 1995.

[18] P.-A. Monney, A Mathematical Theory of Arguments for Statistical Evidence, Contributions to Statistics, Physica-Verlag, Heidelberg, 2003.

[19] J. Kohlas, P.-A. Monney, An algebraic theory for statistical information based on the theory of hints, Internat. J. Approx. Reason. 48 (2) (2008) 378–398.

[20] J. Pearl, On probability intervals, Internat. J. Approx. Reason. 2 (3) (1988) 211–216.

[21] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, The Morgan Kaufmann Series in Representation and Reasoning, Morgan Kaufmann, San Mateo, CA, 1988.

[22] R. G. Martin, J. S. Hwang, C. Liu, Conditional inferential models, unpublished manuscript, available at http://www.stat.purdue.edu/∼chuanhai/docs/imcond.pdf (2011).

[23] G. Casella, Comment on "Setting confidence intervals for bounded parameters", Statist. Sci. 17 (2) (2002) 159–160.

[24] R. G. Almond, Graphical Belief Modelling, Chapman and Hall, London, 1995.

[25] G. J. Feldman, R. D. Cousins, Unified approach to the classical statistical analysis of small signals, Phys. Rev. D 57 (7) (1998) 3873–3889.

[26] C. Giunti, New ordering principle for the classical statistical analysis of Poisson processes with background, Phys. Rev. D 59 (5) (1999) 053001.

[27] B. P. Roe, M. B. Woodroofe, Improved probability method for estimating signal in the presence of background, Phys. Rev. D 60 (5) (1999) 053009.

[28] M. Mandelkern, J. Schultz, Coverage of confidence intervals based on conditional probability, J. High Energy Phys. (11) (2000) Paper 36, 9 pp.

[29] B. P. Roe, M. B. Woodroofe, Setting confidence belts, Phys. Rev. D 63 (1) (2000) 013009.

[30] M. Mandelkern, J. Schultz, The statistical analysis of Gaussian and Poisson signals near physical boundaries, J. Math. Phys. 41 (8) (2000) 5701–5709.

[31] T. Zhang, M. B. Woodroofe, Credible and confidence sets for restricted parameter spaces, J. Statist. Plann. Inference 115 (2) (2003) 479–490.

[32] J. Heinrich, Review of the Banff challenge on upper limits, in: H. B. Prosper, L. Lyons, A. De Roeck (Eds.), Proceedings of the 2007 PHYSTAT LHC Workshop on Statistical Issues for LHC Physics, CERN, Geneva, Switzerland, 2008, pp. 125–134.

[33] R. G. Martin, J. S. Hwang, C. Liu, Marginal inferential models, unpublished manuscript, available at http://www.stat.purdue.edu/∼chuanhai/docs/immarg.pdf (2011).

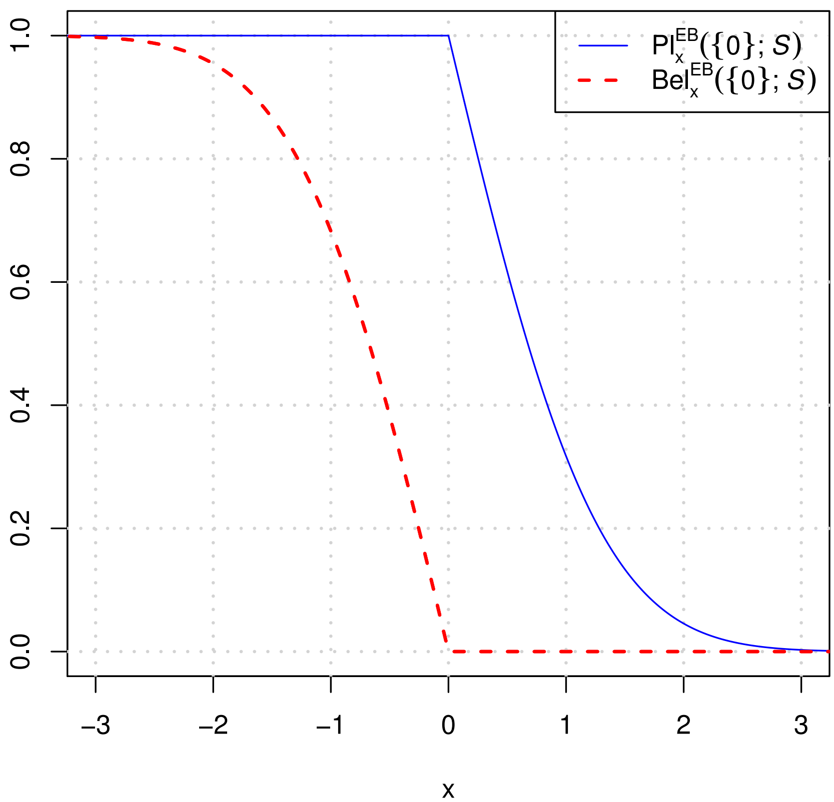

Figure 1: $\mathrm{Bel}_x^{\mathrm{EB}}$ and $\mathrm{Pl}_x^{\mathrm{EB}}$ for the assertion $\mathcal{A}_0 = \{\mu : \mu = 0\}$ with $\sigma^2 = 1$ in the Gaussian example of Section 4.

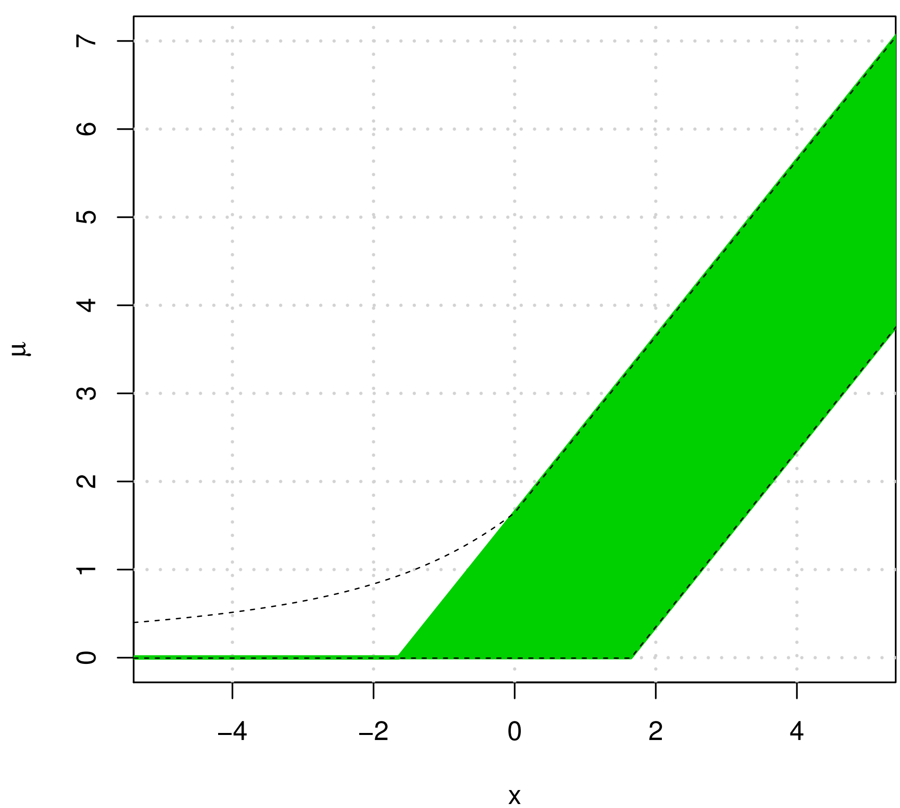

Figure 2: Level $\gamma = 0.9$ plausibility intervals for μ with $\sigma^2 = 1$ in the Gaussian example of Section 4. The shaded region is the plausibility interval found with the EB method, which collapses to the point $\mu = 0$ for $x < \Phi^{-1}(0.05)$. The dotted lines marks the boundary of the plausibility interval obtained from the conditioning rule. Both methods have the same lower boundary.

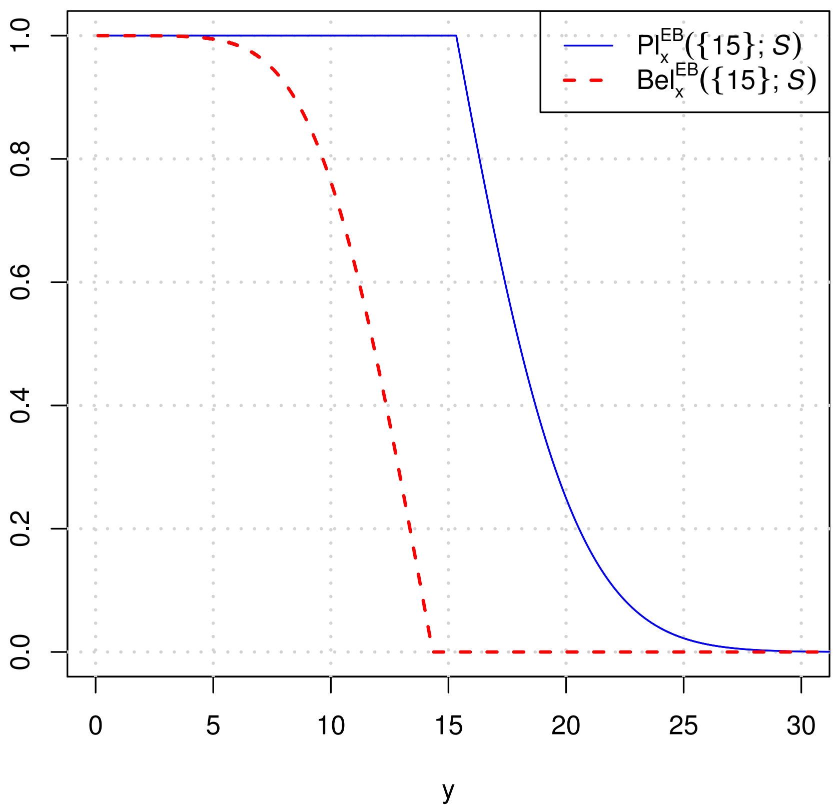

Figure 3: $\mathrm{Bel}_y^{\mathrm{EB}}$ and $\mathrm{Pl}_y^{\mathrm{EB}}$ for the assertion $\mathcal{A}_b = \{\theta : \theta = b\}$ with $b = 15$ in the Poisson example of Section 5.

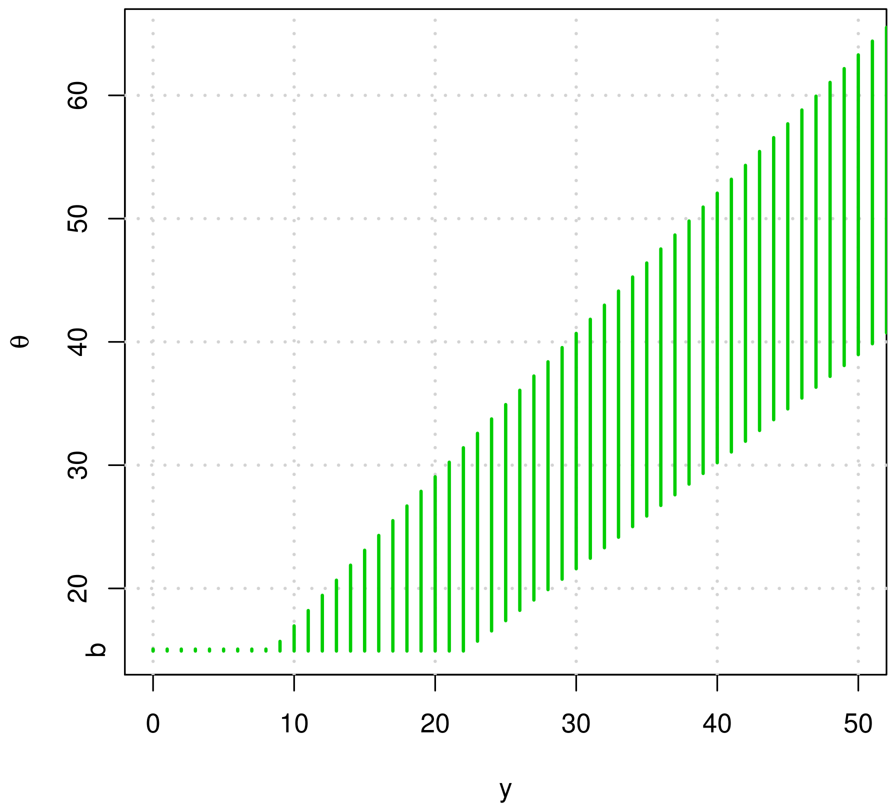

Figure 4: Level $\gamma = 0.9$ plausibility interval for θ with $b = 15$ in the Poisson example of Section 5.

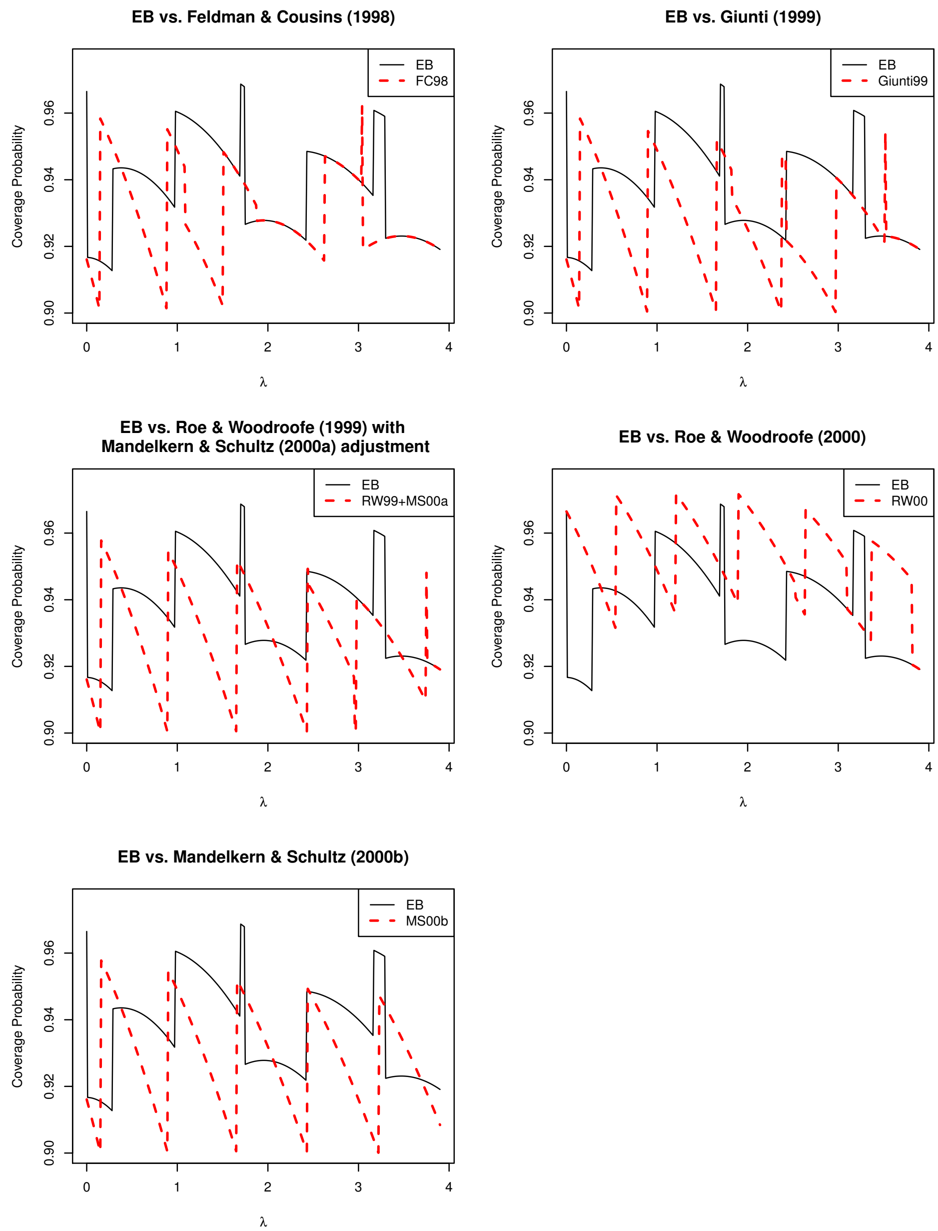

Figure 5: Coverage probability of plausibility interval for Poisson signal rate, λ, compared to the intervals of Feldman and Cousins [25] (top left), Giunti [26] (top right), Roe and Woodroofe [27] with Mandelkern and Schultz [28] adjustment (middle left), Roe and Woodroofe [29] (middle right), and Mandelkern and Schultz [30] (bottom left), when $\gamma = 0.9$ and $b = 3$.

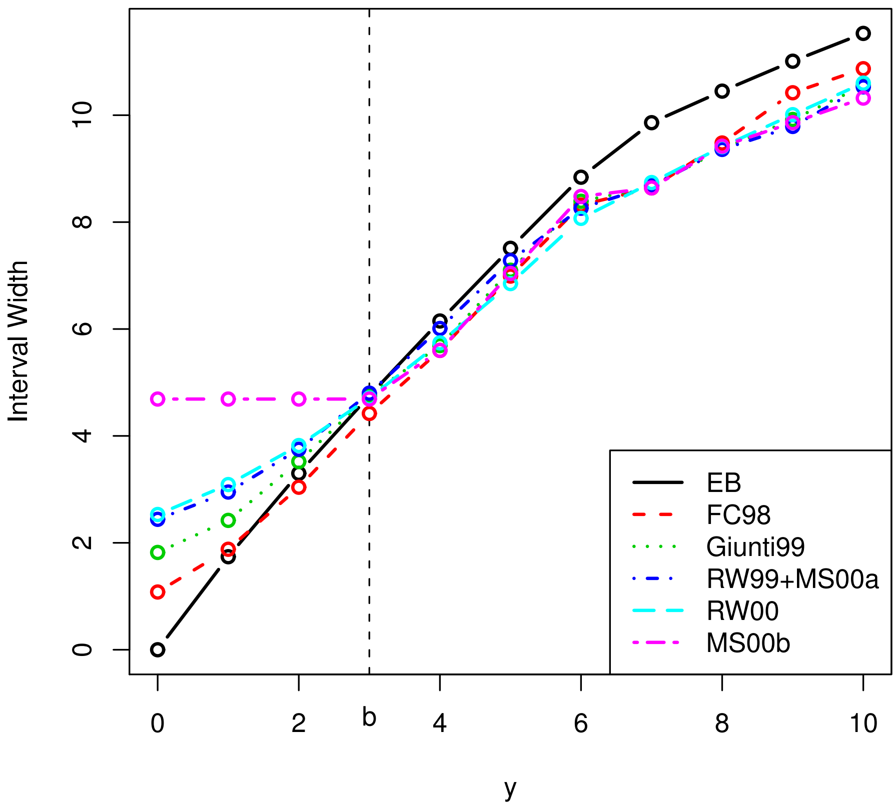

Figure 6: Width of level 0.9 plausibility interval (EB) for Poisson signal rate, λ, compared to the intervals of Feldman and Cousins [25] (FC98), Giunti [26] (Giunti99), Roe and Woodroofe [27] with Mandelkern and Schultz [28] adjustment (RW99+MS00a), Roe and Woodroofe [29] (RW00), and Mandelkern and Schultz [30] (MS00b) for $y = 0, \ldots, 10$ and $b = 3$.

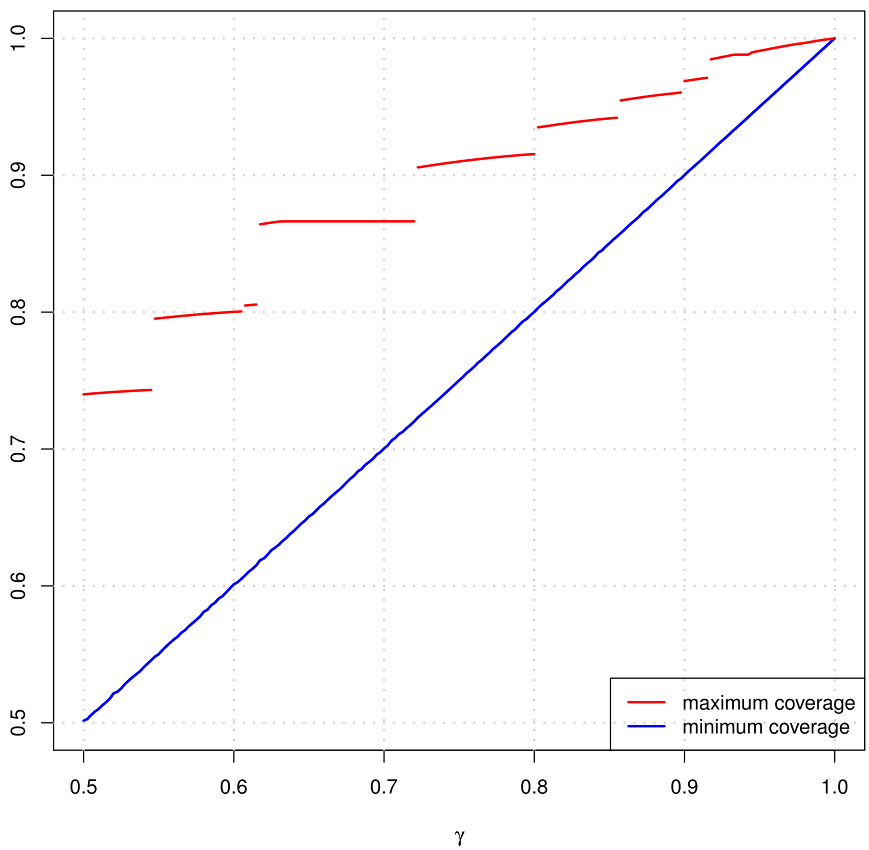

Figure 7: Maximum and minimum coverage probabilities for level γ plausibility interval over $\lambda \in [0,\, 100]$ with $b = 3$.

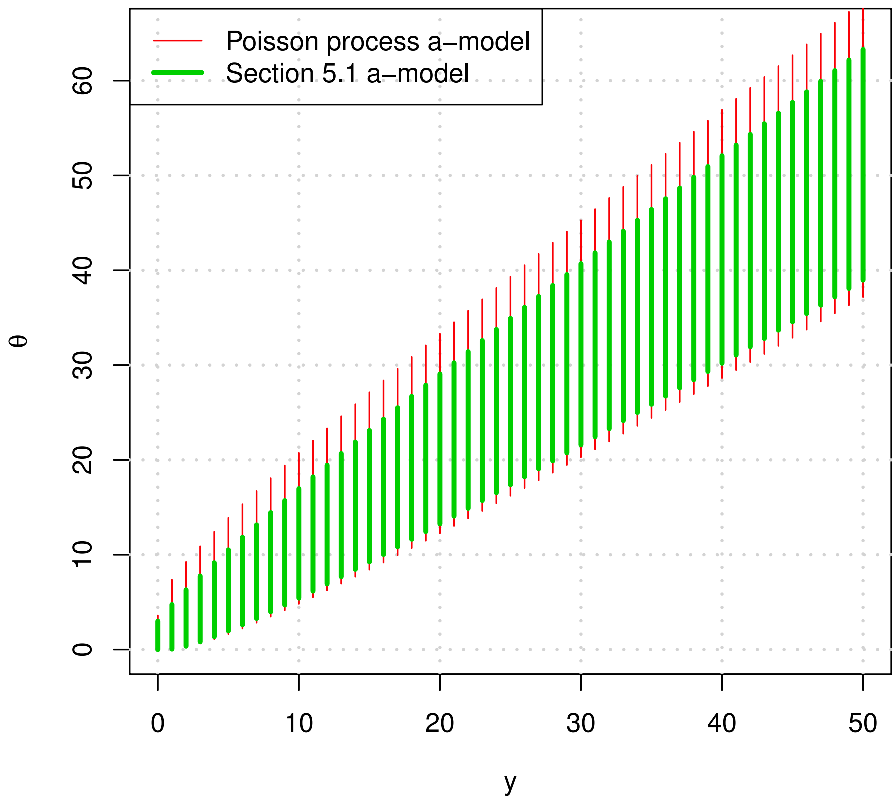

Figure 8: Level 0.9 plausibility intervals based upon the Poisson process IM (B.2) and the IM built from the a-model in Section 5.1 (B.3). The plausibility interval created with (B.3) lies within the plausibility interval from (B.2) for every $y \in [0, 50]$.