LA-UR-17-20370
Approved for public release; distribution is unlimited.

Title: Accommodating Uncertainty in Prior Distributions

Author(s): Picard, Richard Roy
Vander Wiel, Scott Alan

Intended for: project report

Issued: 2017-01-19

Accommodating Uncertainty in Prior Distributions

Rick Picard and Scott Vander Wiel

Los Alamos National Laboratory

Abstract

A fundamental premise of Bayesian methodology is that a priori information is accurately summarized by a single, precisely defined prior distribution. In many cases, especially involving informative priors, this premise is false, and the (mis)application of Bayes methods produces posterior quantities whose apparent precisions are highly misleading. We examine the implications of uncertainty in prior distributions, and present graphical methods for dealing with them.

KEY WORDS: Bayesian sensitivity analysis, Imprecise probabilities, Informative priors

1. INTRODUCTION

The step-by-step strawman caricature of a complacent Bayesian analysis reduces to

1) Formulate a likelihood function suitable for the data.

2) Elicit (if in the role of a data analyst) or divine (if in the role of making a personal decision) a prior distribution for the parameters in the likelihood function.

3) Estimate via MCMC sampling or, less likely, obtain analytically, the corresponding posterior distribution.

4) Quote the posterior quantities of interest.

5) Declare victory and move on to the next analysis.

Admittedly, the above strawman is a gross oversimplification. Presentations of Bayes methods, such as a recent *The American Statistician* treatise, often contain the usual admonishments against complacency, e.g.,

> With great powers comes great responsibility, and Bayesians ... have the corresponding duty to check their predictions and abandon or extend their models as necessary. (Gelman and Robert 2013, p. 3)

Accordingly, a literature has arisen concerning related diagnostic methods.

Too frequently, applications of Bayesian methods pay little or no specific attention to uncertainty in the prior. And almost always, there are many prior distributions consistent with subjective beliefs, each prior corresponding to a different posterior. As such, Bayesian methods still encounter resistance in some quarters. As stated more positively, and prophetically (Berger 1994, p. 23), “the major objection of non-Bayesians to Bayesian analysis is uncertainty in the prior, so eliminating this concern can make Bayesian methods considerably more appealing.”

Interestingly, uncertainties in subjective beliefs not only affect analyses with informative priors, but also those with noninformative ones. Expressing the absence of knowledge in the form of a single prior remains an unsolved and vexing problem; see, e.g., Walley (1991, p. 228-229) for an amusing vignette on the choice of which noninformative prior to use in estimating an event probability. Even after a seemingly noninformative prior is chosen, complacency can still lead to poor results unless diagnostic checks ensure that the prior is truly as innocuous as casually assumed (Seaman, Seaman and Stamey 2012).

To deal with uncertainties in subjective beliefs, the field of *imprecise probabilities*, or IPs, has arisen. IPs characterize subjective beliefs more fully than is done via a single, “precise” prior distribution. Related efforts have resulted in many success stories (see, e.g., the engineering literature review of Beer et al. 2013 and its 268 references).

In the next section, we briefly review the history of IP methodology and its close cousin, Bayesian sensitivity analysis. The third section contains a worked example to illustrate basic concepts, as well as diagnostic plots and procedures useful for IPs that are greatly under-emphasized in the literature.

2. BACKGROUND

Criticisms of precisely determined priors go back to the 1800s. In place of an exact calculation, Boole discussed the "widest limits" for the probability of a compound event based on "various distinct hypotheses" placed on constituent quantities to the calculation, which in turn depended on the inherent "uncertainty in the hypotheses themselves" (Boole 1854, p. 398). Similarly, a reliance on precise probabilistic assumptions in analyses of the era led to Venn's colorful view that "it is quite true that considerable violence has to be done to some of these examples, by introducing exceedingly arbitrary suppositions into them, before they can be forced to assume a suitable form" (Venn 1888, p. 124).

The origins of IP methodology are often attributed to the economist John Maynard Keynes, who argued that, "in actual reasoning," precise subjective probabilities "occur comparatively seldom" (Keynes 1921, p. 182). He then went on to consider upper and lower probabilities in examples. Work of a similar nature by other authors followed, culminating in the classic IP textbook by Walley (1991).

Despite their intuitive appeal, early IP methods had little impact. Before the widespread advent of MCMC, pre-1990s computational tools did not exist for Bayesians to solve many real problems, which combined with other factors to limit the practicality of Bayes methods at the time (e.g., Efron 1986). Pursuit of IP methods, some of which required consideration of multiple priors, was hopeless in nontrivial applications.

Impact was further limited because many approaches put forth in the name of IP were summarily dismissed by statisticians. One example involves Dempster-Shafer belief functions, which can lead to plainly incorrect solutions, as for the classic, fully specified Monty Hall problem and other situations (e.g., Walley 1991, p. 279-281). See also the discussants of Shafer (1982) for reactions to the non-Bayesian aspects of belief functions. Most other IP methods, e.g., fuzzy set theory, are similar − despite their better characterization of subjective beliefs and successes in certain applications, their probability calculus for combining prior information with data can fail badly in special cases.

Another response to uncertainty in priors is Bayesian sensitivity analysis (BSA), sometimes called robust Bayesian analysis (e.g., Berger 1990, 1994). Formalized BSA falls under the IP umbrella, complete with coherence properties (Walley 1991, Secs. 7-8), and is examined shortly. Informal BSA, which has the analyst "try a few models and priors" (Berger 1994, p. 44), is more commonly implemented.

The first goal of formalized BSA is to characterize subjective beliefs more fully than is done via a single prior distribution. This methodology involves defining the set $\mathcal{P}$ of prior distributions and models that are consistent with available information. In economics (e.g., Weatherson 2002), the set $\mathcal{P}$ is called the representor. Numerous approaches to constructing $\mathcal{P}$ exist (Berger 1994, p. 25-27). Once $\mathcal{P}$ has been constructed, the plausible range of posterior quantities is then determined.

In the example to follow, the quantity of interest is the probability of a set A. The maximum value of $\mathrm{Pr}_{\mathrm{posterior}}(A)$ over $\mathcal{P}$ is called the upper probability of A and is denoted $\overline{\Pr}(A)$. The minimum of $\Pr_{\text{posterior}}(A)$ over $\mathcal{P}$ is the lower probability of A and is denoted $\underline{\Pr}(A)$. By computing $\overline{\Pr}(A)$ and $\underline{\Pr}(A)$, conclusions are reached that better reflect subjective beliefs. The *imprecision* of the event A is defined as the difference

$$\Delta(A) \quad = \quad \overline{\Pr}(A) - \underline{\Pr}(A)$$

Operational differences between the IP and BSA viewpoints (e.g., Walley 1991, p. 107-108) are minor in most applications. Philosophically, the simplified BSA view postulates that a single "correct" prior exists, having an exact distributional form and parameter values expressible to several significant digits of accuracy, but − because of imperfections in soul searching and/or elicitation − the correct prior is unknown. The IP view is that there is no such thing as a single correct prior, but instead that there is a "correct" representor set $\mathcal{P}$ providing a black-and-white distinction between prior distributions that are "consistent with" subjective beliefs and those which aren't. Neither viewpoint is without its critics.

3. EXAMPLE: DETONATOR IMPACT DATA

3.1 The Data Set and Prior Information

The data set consists of 25 detonator impact tests conducted at Los Alamos National Laboratory. From various heights, a 2.5 kg anvil was dropped on detonators of a specific type. Based on review of the video and the decibel level from the audio, a "go" or "no-go" response was determined for each drop test. Data are given in Table 1.

Height (cm)	# Tested	# Go	# No-Go
45.0	1	0	1
50.5	3	1	2
57.0	3	2	1
64.0	6	2	4
71.5	7	5	2
80.5	4	3	1
90.0	1	1	0

Table 1: Detonator Impact Testing Data.

The probability of a “go” response increases monotonically from zero (at height $h = 0$) to one (when the height is sufficiently great). Of interest is the probability of a “go” response as a function of drop height. Several models have been used for such data, and we initially focus on the standard probit model. This model states that the probability of a “go” response at drop height h is

$$\Pr(\text{“go”}) \quad = \quad \Phi\left(\frac{log\ h\ -\ log\ h_{50}}{\sigma}\right)\ , \qquad (1)$$

where h_{50} is defined as the drop height whose chance of a “go” is 50%, σ is a scale factor, and $\Phi(\cdot)$ is the cumulative distribution function for a standard normal distribution.

Performance requirements for detonators are contained in Department of Energy Order O-452.1 (see www.directives.doe.gov), one of whose goals is to prevent accidents during weapons assembly and disassembly. Among the requirements (O-452.1, p. 7) is that

"... the probability of a premature nuclear explosive detonation must not exceed one in a million (1E-06) per credible nuclear weapon accident or exposure to abnormal environments."

Several accident scenarios are relevant to detonator impact testing, e.g., a worker inadvertently dropping a wrench on a detonator.

Performance requirements thus involve two types of extrapolation, one of which extrapolates the physical insult in the experiment (the drop of a 2.5 kg anvil on a detonator from a certain height) to a physical insult of interest (e.g., the drop of a wrench from a different height). For what follows, we assume that an anvil drop from 5 cm is equivalent to dropping a much lighter tool from a height of interest.

The second form of extrapolation is statistical, extrapolating results from the drop heights 45-90 cm in Table 1 to drop heights such as 5 cm. This extrapolation is necessary because a very large number of drop tests would have to be conducted at low drop heights in order to directly estimate Pr("go") with accuracy. Because the time and money involved in such an effort would be prohibitive, properties for low-probability drop heights are extrapolated using the predictive model.

Formalizing this approach, the probit model (1) is inverted to give drop height as a function of the probability of a "go" response,

$$h_{\mathrm{Pr(go)}} \;=\; h_{50} \;\times\; exp\left[\,\sigma\,\Phi^{-1}\left(\mathrm{Pr}(\text{ "go" })\right)\right]\;. \qquad (2)$$

For Pr("go") $= 10^{-6}$, the normal quantile $\Phi^{-1}(10^{-6}) = -4.75$, and the 10^{-6} drop height

$h_{-6} = h_{50} \times exp\,[\,-4.75\,\sigma\,]$ is a known function of the model parameters h_{50} and σ.

The performance requirement is that 5 cm is a safe drop height, where "safe" means that Pr("go") $< 10^{-6}$ for an anvil drop of 5 cm. Let the set

$$A_5 = \{\ h_{50}, \sigma \mid h_{-6} > 5 \text{ cm}\ \}$$

denote the safe region of parameter space. The goal is to assess the probability $\Pr(A_5)$ that the detonator type meets the performance requirement.

The scientist who was to conduct the experiment provided a prior estimate for the 50-50 drop height h_{50} based on work with detonators similar to the type examined here. That prior estimate was 70 cm, to within a relative uncertainty factor of 1.2. Equivalently, the plus-or-minus one standard deviation interval in log scale has a standard deviation of $log(1.2)$: $\;h_{50}\ \in\ (70/1.2,\ 70 \times 1.2)\ \ \Leftrightarrow\ \ log\, h_{50}\ \in\ log\, 70 \pm [log\, 1.2]$.

Prior information for the scale factor σ in (1) was elicited through drop heights besides h_{50}. Subject matter experts are more comfortable contemplating physical quantities like drop heights rather than abstract parameters like σ in a statistical model; see also Oakley and O'Hagan (2007) in this regard. Further, a single elicitation on the same physical quantities can be used in conjunction with other models besides the probit (more on this to come). Input on the 10% drop height h_{10} was obtained via the h_{50}/h_{10} ratio, which, upon solving $h_{10} = h_{50} \times exp\,[\,-1.28\,\sigma\,]$ for σ, is directly related to σ. The prior estimate for the h_{50}/h_{10} ratio was 2, to within a relative uncertainty factor of 1.5.

3.2 Complacent Bayesian Analysis

The compacent Bayesian paradigm force-fits the scientist's subjective beliefs into a single prior distribution. Even though the force-fitting does not fully capture uncertainty in those beliefs, the complacent analysis demands this force-fitting, and imposes a prior on $(h_{50}, h_{50}/h_{10})$. Lognormal priors for the drop height $h_{50} > 0$ are typically used, in this case

$$log\ h_{50} \ \sim \ N(\,log\ 70, [log\ 1.2]^2\,)\ ,$$

where the values 70 and 1.2 are the scientist's best guesses.

The prior for σ is derived from beliefs regarding the h_{50}/h_{10} ratio. Here, $h_{50}/h_{10} > 1$, or $h_{10}/h_{50} \in (0,1)$. For quantities within (0,1), the most common prior is the beta distribution. By varying its parameters, the beta density function can take on a wide variety of shapes as warranted by the situation. In IP applications (e.g., Walley 1996; Walley, Gurrin, and Burton 1996), the imprecise beta distribution is commonly used.

The beta distribution has two parameters, denoted α and β. The ratio α/β determines the mean of the distribution through the relation α/β = mean/(1 − mean). Absolute magnitudes of α and β determine the standard deviation.

The prior estimate $h_{50}\,/\,h_{10} \approx 2$ corresponds to a beta distribution with mean value $E[\,h_{10}\,/\,h_{50}\,] \approx 1/2$ and $\alpha \approx \beta$. An uncertainty factor 1.5 implies the interval

$$h_{50}/h_{10}\ \in\ (\ 2/1.5,\ 2 \times 1.5\ ) \Leftrightarrow\ h_{10}/h_{50}\ \in\ \left(\frac{1}{2 \times 1.5}, \frac{1.5}{2}\right)\ =\ (1/3, 3/4)\ .$$

Without resorting to more sophisticated elicitation (Yu, Shih, and Moore 2008), the half width of this interval, (3/4 - 1/3) / 2 $\approx$ 0.21, is equated to one standard deviation for the beta prior. Adding the constraint $\alpha \approx \beta$ gives $\alpha \approx \beta \approx 2.4$, or

$$h_{10}/h_{50} \;\sim\; \text{Be}\,(2.4, 2.4)\;.$$

Independently coupling the beta prior for h_{10}/h_{50} with the lognormal prior for h_{50} produces the "nominal" prior distribution. Combining this prior with the model (1) and data in Table 1, an MCMC sample $\{\,(h_{50}, \sigma)_j\,\}$ is simulated from the posterior. For the j-th member of the sample, its 10^{-6} drop height is $(h_{-6})_j = (h_{50})_j \times exp[\,-4.75\,\sigma_j]$, and the set $\{(h_{-6})_j\}$ is used to assess $\text{Pr}_{\text{posterior}}(A_5)$. A lengthy MCMC simulation (10^6 samples) gives the point estimate $\widehat{\text{Pr}}_{\text{posterior}}(A_5) \;=\; 0.452161$.

The use of six decimal places emphasizes that $\text{Pr}_{\text{posterior}}(A_5)$ could indeed be determined to arbitrary accuracy by running the MCMC simulation until eternity, thereby providing a misleadingly illusory sense of precision. This illusion is, unfortunately, a logical consequence of requiring a precisely defined prior distribution that leads to a precise posterior distribution, and then to precise quantities such as $\widehat{\text{Pr}}_{\text{posterior}}(A_5)$.

Pretentious accuracy aside, the probability 0.452161 provides only modest of confidence that this particular type of detonator meets the performance requirement. Were additional data obtained, the degree of confidence would improve (assuming, of course, that the additional data were consistent with safe operation).

3.3 IP Analysis of the Detonator Data

As has been noted, "a philosophy of Bayesian statistics as subjective, inductive inference can encourage a complacency" (Gelman and Shalizi 2013, p. 32). Complacent analyses as per the previous section should be resisted, and assumptions underlying such analyses should be closely examined.

In terms of "adding imprecision" to the unduly precise complacent prior, it might be tempting to construct a hierarchical model. Here, the four elicited quantities − the scientist's best guesses "70," "1.2," "2," and "1.5" − are assigned precise probability distributions based on additional elicitation. The resulting hierarchical prior is more diffuse than the complacent prior, *but it is still precisely defined*, and complacent MCMC simulation from the posterior still yields pretentiously precise $\widehat{\Pr}_{\text{posterior}}(A_5)$.

Avoiding this situation is certainly possible, through it requires a creative use of hierarchical results. A single member of the MCMC sample from the hierarchical posterior could be extracted and its hyperprior values used to simulate corresponding prior parameter values. Then a secondary MCMC simulation as in the previous section could be run to obtain $\widehat{\Pr}_{\text{posterior}}(A_5)$ conditional on those prior parameters. Repeating this process a large number of times characterizes the distribution of $\Pr_{\text{posterior}}(A_5)$ induced by the posterior on hyperprior parameters, not unlike the method alluded to in Oakley and O'Hagan (2007).

A more practical, and less computationally intensive IP approach constructs the set $\mathcal{P}$ formally identifying priors consistent with subjective beliefs. This construction is intrinsically subjective, arbitrary, and as noted above, there are several approaches to carrying it out.

One intuitive approach is based on interval estimates. For example, the nominal prior's plus-or-minus one standard deviation interval is such that

$$\Pr_{m_0,uf_0}(log\, h_{50} < log\, 70 - log\, 1.2) = \Phi(-1) = 0.158655 \equiv p_0^- \qquad \text{and}$$

$$\Pr_{m_0,uf_0}(log\, h_{50} < log\, 70 + log\, 1.2) = \Phi(+1) = 0.841345 \equiv p_0^+ ,$$

where the nominal median $m_0 = 70$ and uncertainty factor $uf_0 = 1.2$. The precise normality-based values 0.158655 and 0.841345 are approximations, of course. Suppose that any probability values $p^- \in [0.05, 0.25]$ and $p^+ \in [0.75, 0.95]$ are deemed consistent with $p_0^- = 0.158655$ and $p_0^+ = 0.841345$. Then any pair of values (m, uf) such that

$$\Pr_{m,uf}(log\, h_{50} < log\, 70 - log\, 1.2) \in [0.05, 0.25] \qquad \text{and}$$

$$\Pr_{m,uf}(log\, h_{50} < log\, 70 + log\, 1.2) \in [0.75, 0.95]$$

defines a lognormal prior for h_{50} consistent with subjective beliefs. That is, the uncertainty in the prior is captured in the interval estimates $p^- \in [0.05, 0.25]$ and $p^+ \in [0.75, 0.95]$, which map to a 2-D region constraining the joint behavior of (m, uf). See Figure 1.

A plausible region for the h_{50}/h_{10} ratio follows similarly. The nominal beta prior for the h_{10}/h_{50} ratio has parameters $\alpha_0 = \beta_0 = 2.4$, which implies

$$\Pr_{\alpha_0,\beta_0}(h_{10}/h_{50} < 0.50 - 0.21) = 0.1815945 \equiv p_0^- \qquad \text{and}$$

$$\Pr_{\alpha_0,\beta_0}(h_{10}/h_{50} < 0.50 + 0.21) = 0.8184055 \equiv p_0^+ .$$

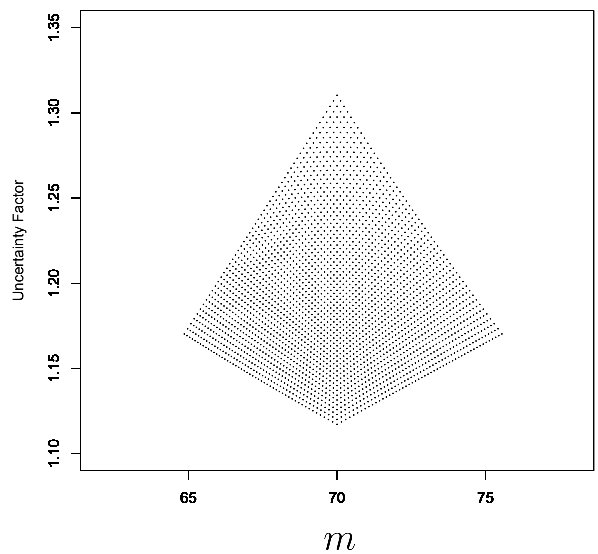

Figure 1: Plausible Region for (m, uf).

Using the same interval estimates as for h_{50}, and finding beta parameters (α, β) such that

$$\Pr_{\alpha,\beta}\left(h_{10}/h_{50} \ < \ 0.50 - 0.21\right) \ \in \ [0.05, 0.25] \qquad \text{and}$$

$$\Pr_{\alpha,\beta}\left(h_{10}/h_{50} \ < \ 0.50 + 0.21\right) \ \in \ [0.75, 0.95]$$

leads to a four-vertex region similar to that in Figure 1.

One property of $\mathcal{P}$-based formulations is that a prior deemed consistent with subjective beliefs can be virtually indistinguishable from a prior deemed inconsistent with those beliefs, e.g., as when two priors straddle the boundary of $\mathcal{P}$ in Figure 1. Such a precise boundary between plausible/implausible priors is clearly unrealistic, although IP advocates counter that, while imperfect, it is a huge improvement on the Bayesian purist's single-point set $\mathcal{P}$. The fiction of a precise boundary for $\mathcal{P}$ is essential to an IP analysis because it is needed for the maximization/minimization required to obtain $\overline{\Pr}(A_5)$ and $\underline{\Pr}(A_5)$. Explicitly dealing with ambiguity in the choice of $\mathcal{P}$ is integral to the art of applying IP methods.

As such, the most important aspect of an IP analysis is to devote careful thought to defining the set $\mathcal{P}$ of priors/models consistent with a priori information − much more thought than is typically devoted to informal BSA. If $\mathcal{P}$ is too small, such as from a subject matter expert being overconfident, IP bounds will be too narrow. If $\mathcal{P}$ is too large, as from a conservative prior-by-committee approach, IP bounds will be too wide. In examining potential representor sets, plots such as Figure 1, which map interval-estimate constraints into the elicitation frame of reference, are useful.

Similarly useful in assessing $\mathcal{P}$ are other diagnostic plots. The idea of using diagnostic plots as part of checking assumptions underlying a Bayesian analysis is by no means new (e.g., Box 1980). Further, Bayesian diagnostic plots can address more subtle issues, e.g., in the assessment of the maximum of several observed means, determining when posterior inference is (or is not) subject to selection bias (Senn 2008).

IP counterparts of Bayesian diagnostic plots do not appear to be much used. Simulating prior-predicted functionals of interest for extreme priors (e.g., priors at the vertices of Figure 1) to see if they properly span subjective beliefs provides another check on $\mathcal{P}$. For the detonator data, two important functionals are the 10^{-6} drop height h_{-6} and the curve-fit approximation (1). These functionals are shown in Figure 2, which plots the lower portion of the probit curve.

Displayed are 80 such curves, consisting of 5 samples from each of the 16 priors formed

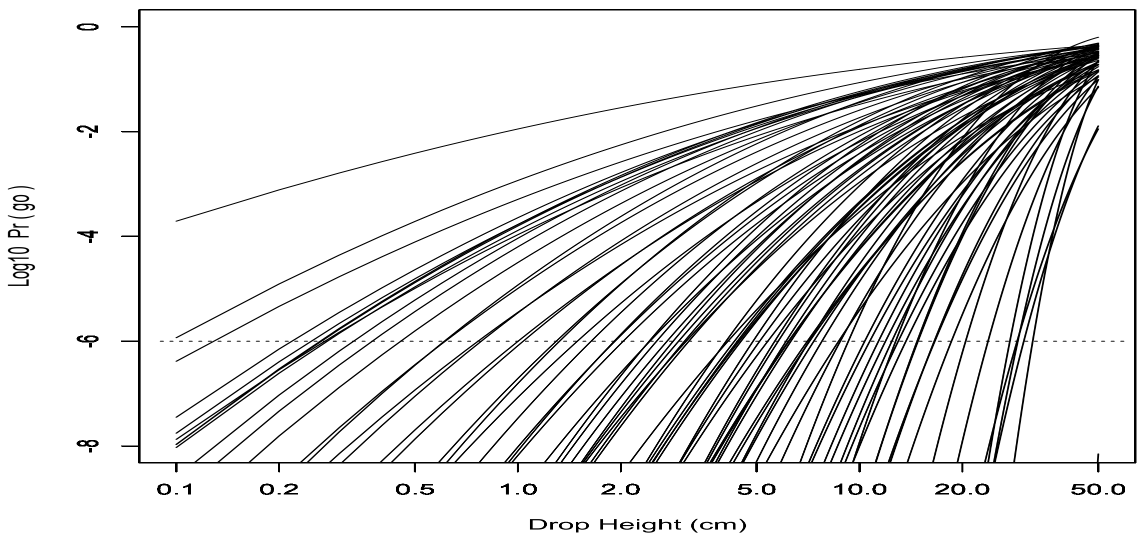

Figure 2: $\mathcal{P}$-Based Spectrum of Prior-Predicted 10^{-6} Drop Heights.

from the combinations of the four h_{50} and four h_{50}/h_{10} vertices in $\mathcal{P}$. Intersections of the curves with the dotted 10^{-6} intercept define the corresponding drop heights, and confirm that the scientist's prior intuition on small (10^{-6}) tail probabilities is limited.

In isolated cases, prior-predicted data may be inconsistent with subsequently observed data. This is not the case for the detonator example, but in general, observed data should lie within the envelope of simulated prior-predicted data over extreme priors in $\mathcal{P}$. For i.i.d. samples, overlay plots of observed data versus prior-predicted data are obviously useful. When observed data do not fall within the $\mathcal{P}$-based envelope, an analyst has a dilemma of producing IP intervals with poor frequentist properties or re-analyzing with a new set $\mathcal{P}$ that is incoherently obtained only after having observed the data. The dilemma here is an old one for Bayesians; see, e.g., Dawid (1982) and discussants for lively reactions to it.

Table 2 summarizes posterior probabilities $\widehat{\Pr}_{\text{posterior}}(A_5)$ for the 16 combinations of the h_{50} and h_{50}/h_{10} vertices in $\mathcal{P}$ (100K MCMC samples per case). Recall that the complacent analysis gave $\widehat{\Pr}_{\text{posterior}}(A_5) = 0.452161$. Maximum and minimum $\widehat{\Pr}_{\text{posterior}}(A_5)$ over the 16 priors in Table 2 are $\overline{\Pr}(A_5) = 0.65$ and $\underline{\Pr}(A_5) = 0.30$. The non-probabilistic IP bounds [0.30, 0.65] span a factor-of-2 range and provide a much more realistic interpretation of the data than the precise value 0.452161 alone.

$(1\ /\ E\,[h_{10}/h_{50}]\,, uf)$	(m, uf)			
	(64.9, 1.17)	(70, 1.12)	(70, 1.31)	(75.5, 1.17)
(1.72, 1.33)	.63	**.65**	.59	.62
(2, 1.28)	.49	.53	.46	.50
(2, 1.63)	.43	.45	.38	.40
(2.39, 1.48)	.33	.36	**.30**	.32

Table 2. Probit Posterior Probabilities $\widehat{\Pr}_{\text{posterior}}(A_5)$

Sample sizes for Table 2, 100K post burn-in samples per case, are large enough that $\overline{\Pr}(A_5)$ and $\underline{\Pr}(A_5)$ are significantly different. A good diagnostic check when sample sizes and imprecisions are smaller is to assess how much of $\overline{\Pr}(A_5) - \underline{\Pr}(A_5)$ is reasonably ascribed MCMC error alone versus how much is attributable to intrinsic uncertainty in the prior. Complications (in a general sense) to this assessment are that $\overline{\Pr}(A)$ and $\underline{\Pr}(A)$ are not binomial proportions because of the correlated MCMC sampling, and $\overline{\Pr}(A)$ and $\underline{\Pr}(A)$ can be subject to selection bias.

Despite work on characterizing maxima/minima of certain posterior quantities over certain types of sets $\mathcal{P}$ (e.g., Sivaganesan and Berger 1989, Wasserman and Kadane 1992, Abraham and Daures 2000), theoretical results are limited. Moreover, any formal optimization should be taken with a grain of salt because the precise boundary of $\mathcal{P}$ shouldn't be viewed literally. In obtaining results, relevant computational tricks include:

1) Additional MCMC runs carried out on a space filling design over $\mathcal{P}$ can help quantify extreme posterior quantities that do not occur on the boundary of $\mathcal{P}$.

2) Posterior quantities for priors local to vertices and space filling design points can be estimated using importance sampling, re-weighting results for priors near the locations, as opposed to running additional importance samples from scratch.

3) For extrapolated quantities like rare event probabilities, specialized importance sampling techniques are much more effective than MCMC sampling from a posterior (Picard and Williams 2013).

Prior-posterior comparisons are also useful. The detonator example is typical in that the data aid in reducing uncertainty. Posterior-predicted 10^{-6} drop heights are shown in Figure 3. The plot is analogous to Figure 2, overlaying 80 curves (5 posterior samples from each of the 16 priors in Table 2). A prior-to-posterior shrinkage is apparent, more for large drop heights (23 of the 80 prior-predicted 10^{-6} drop heights exceed 10 cm, but only 4 of the 80 posterior-predicted drop heights do) than for smaller ones.

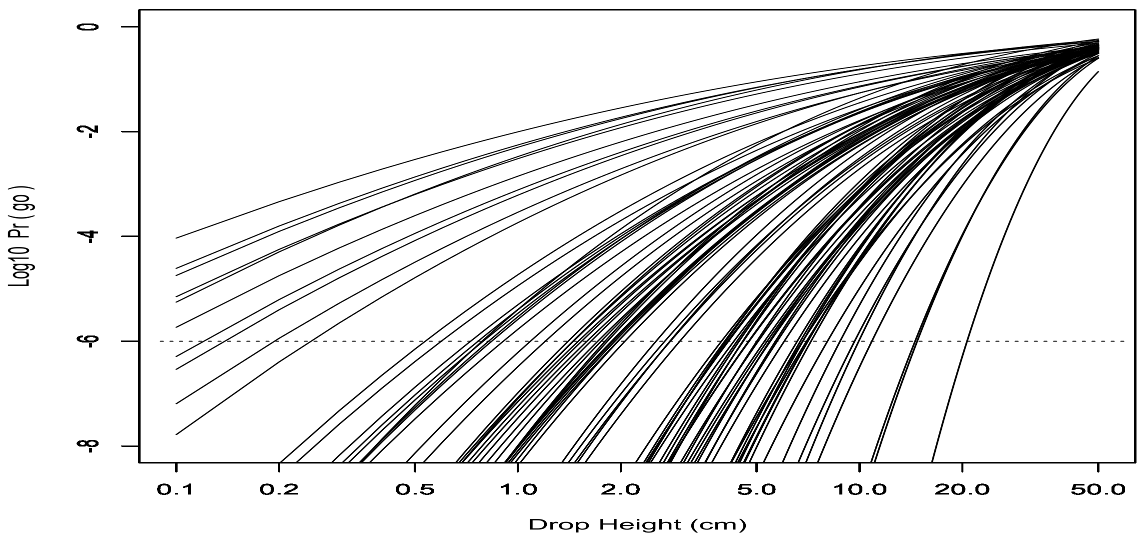

Figure 3: $\mathcal{P}$-Based Spectrum of Posterior-Predicted 10^{-6} Drop Heights.

Prior-to-posterior shrinkage is not guaranteed. The term *dilation* (Wasserman and Seidenfeld 1994) refers to situations where posterior imprecision exceeds prior imprecision. Imprecision can increase, for example, when unexpected data arise that leave an expert feeling less certain after seeing the data than he thought he was beforehand (e.g., Walley 1991, p. 225), or when $\Pr(A)$ is a full-system reliability and $\mathcal{P}$ contains multiple priors on subsystem performance reflecting disagreement among experts, one or more of whom is miscalibrated. When imprecisions increase, it is important to understand the reason(s) why and determine whether further action is needed.

Not yet mentioned is uncertainty in the model. There is no first-principles, detonator-physics-based justification for the probit function (1), and alternatives should be considered.

Other sigmoidal forms, such as logistic regression and the Arrenhius model for chemical kinetics, are also used in conjunction with monotonically increasing phenomena in (0,1).

For present purposes, it suffices to consider the two-parameter Weibull model (Meeker and Escobar 1998, Eq. 4.7), which postulates that the probability of a "go" at height h is

$$\Pr(\text{ “go” }) \;=\; 1 \;-\; exp\left\{ -\, exp\left[\frac{log\, h \;-\; \mu}{\sigma_{\rm W}} \right] \right\} \,, \qquad (3)$$

where $(\mu, \sigma_{\rm W})$ are model parameters. Similar to the probit model, prior information on the 50-50 drop height h_{50} and the h_{50}/h_{10} ratio can be translated directly into a representor set $\mathcal{P}$ for Weibull model parameters.

Using the nominal prior on (h_{50}, h_{50}/h_{10}), posterior means for μ and $\sigma_{\rm W}$ can be substituted into the Weibull model (3) to provide a curve-fit approximation to Pr("go") as a function of drop height. The same can be done for the probit model (1), and overlaying the two nominal curves in semi-log scale (which highlights the differences at low drop heights at the cost of obscuring the sigmoidal shapes) gives Figure 4.

There is no practical difference between the nominal probit and Weibull model curve fits over the 45-90 cm range of the drop tests. When the models are extrapolated, the Weibull curve extrapolates substantially above the probit curve for low drop heights, which greatly affects the estimated 10^{-6} drop height. A 1M MCMC sample yields the complacent Weibull posterior probability $\widehat{\Pr}^{\rm W}_{\rm posterior}(A_5) = 0.035125$ (six decimals to re-emphasize the misleading accuracy in lengthy MCMC samples).

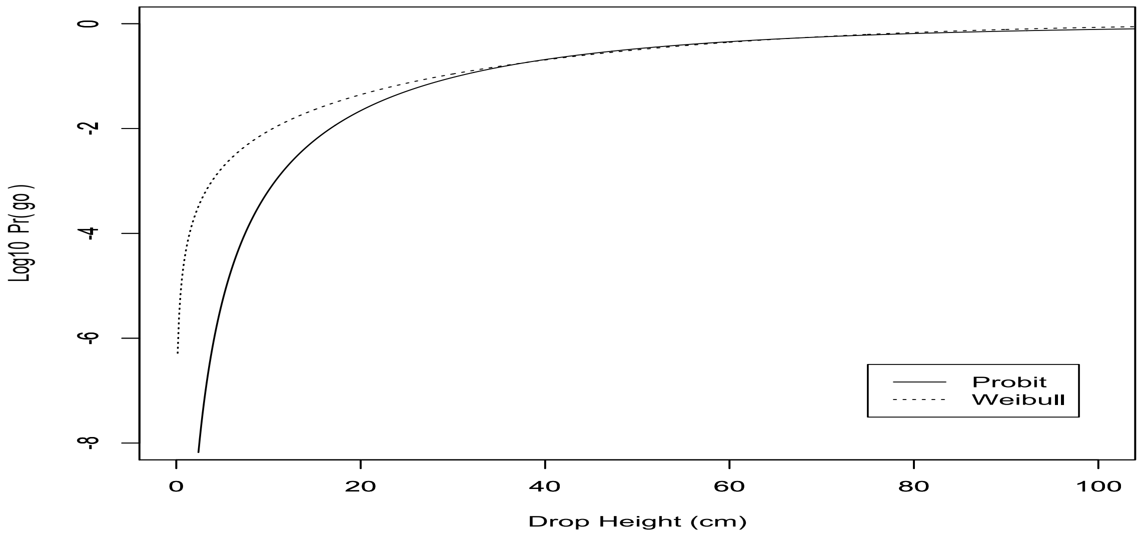

Figure 4: Overlay of Probit and Weibull Data Fits.

Weibull IP bounds over the 16 priors in Table 2 are [.01, .07], and do not even overlap with those of the probit model, [.30, .65]. The probit IP bounds [0.30, 0.65] and Weibull IP bounds [0.01, 0.07] are displayed at drop height $h = 5$ cm in Figure 5, which plots $\widehat{\Pr}_{\text{posterior}}(A_h)$ against drop height h, where $A_h = \{\ h_{50}, \sigma \mid h_{-6} > h \text{ cm } \}$ is the region of parameter space where the 10^{-6} drop height exceeds h.

To the right side of the plot, there is no practical difference, in that both models agree that there is essentially no chance that the 10^{-6} drop height exceeds 30 cm. As is common with extrapolated quantities, the width of IP bounds increases with the degree of extrapolation for each model, a phenomenon that could easily be overlooked in an examination of complacently-determined precise quantities.

Extrapolation in tail probability is also important. Consider Figure 6 and the 10% drop

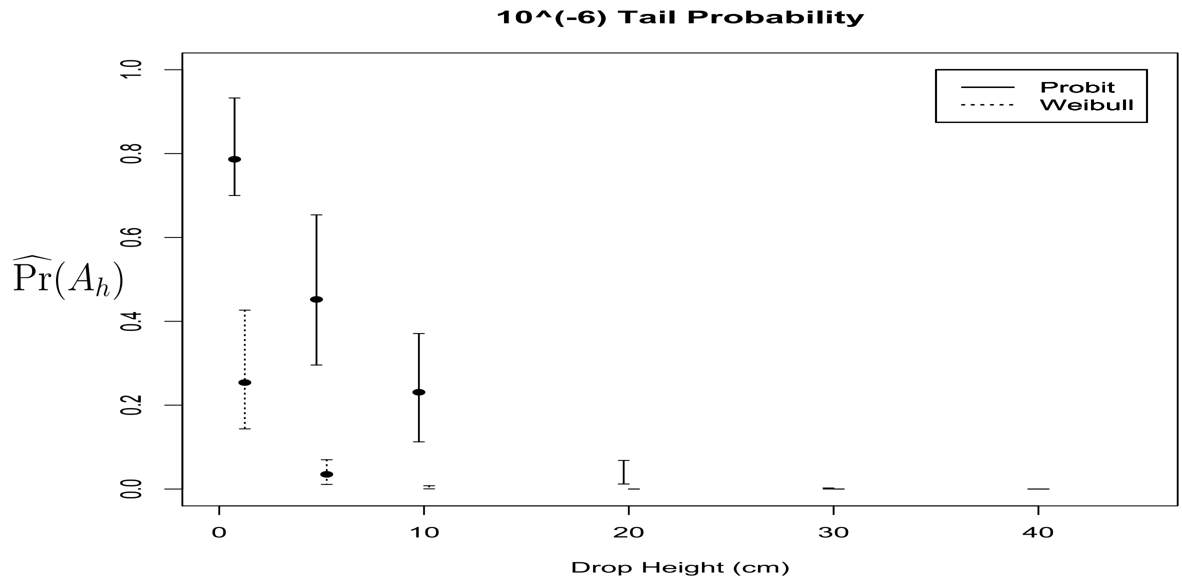

Figure 5: IP Bounds.

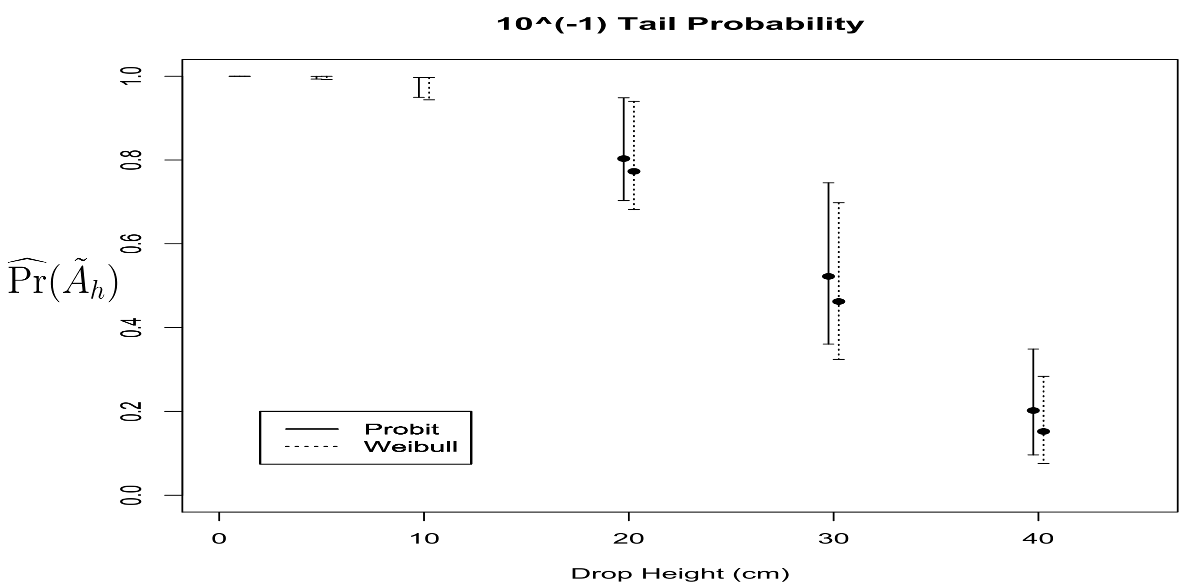

Figure 6: IP Bounds.

height h_{10}, a much less severe extrapolation than the 10^{-6} drop height. The models agree that the 10% drop height exceeds 5 cm. Moreover, comparing IP bounds for both models, the minimal extrapolation in tail probability yields model-to-model differences for h_{10} that are comparatively minor. As illustrated in these plots, whether the width of the IP bounds and/or the model-to-model differences are large enough to matter depends on the specific posterior quantity of interest.

4. SUMMARY

Avoiding complacency requires care and is time consuming. Running MCMC simulations for multiple priors/models in $\mathcal{P}$ can be computationally intensive. Still more effort is required to generate and examine diagnostic plots (IP-related and otherwise). Practical considerations allow only a finite effort to be devoted to an analysis, thus imposing tradeoffs: the representor set $\mathcal{P}$ should be rich but not too rich, and the maximization/minimization of posterior quantities over $\mathcal{P}$ should be reasonably accurate but not amount to overkill.

Relative to this tradeoff spectrum, the complacent Bayes analysis is one endpoint. It provides a desired result − namely, a probabilistic answer − with a minimum investment of time, and appeals to Bayesians that "tend to be aggressive and optimistic with their modeling assumptions" (Efron 2005, p. 1). Unfortunately, the fundamental basis of complacency requires that a single prior distribution accurately reflect subjective beliefs, an assumption recognized for more than 100 years to be false. Costs of the complacent approach can sometimes be minimal, such as for truly noninformative priors or data sets with large sample sizes, coupled with non-extrapolated posterior quantities of interest, but are serious in applications such as the detonator impact testing data.

IP analyses quantify, in a non-probabilistic way, effects of uncertainty in priors and models. Though not emphasized in the literature, numerous diagnostic plots are useful in IP analyses, including those aimed at

1) assessing the adequacy of a candidate set $\mathcal{P}$ by relating the boundary of $\mathcal{P}$ to quantities directly elicited, such as interval estimates,

2) identifying prior-data mismatches through overlay plots of observed and simulated prior-predicted data across extreme priors in $\mathcal{P}$,

3) understanding prior-to-posterior shrinkage or dilation effects beyond a simple comparison of nominal-prior-versus-nominal-posterior standard deviations, and

4) understanding which posterior quantities are more robust to prior/model uncertainty than others.

In many Bayesian analyses, the only quoted variation in posterior quantities involves MCMC sampling error, with no allowance for uncertainty in the prior or model, thus misleading data analysts and their clients. IP methods aid in quantifying these effects.

References

Abraham, C., and Daures, J. P. (2000), "Global Robustness with Respect to the Loss Function and the Prior," *Theory and Decision*, 48, 359–381.

Beer, M., Ferson, S., and Kreinovich, V. (2013), "Imprecise Probabilities in Engineering Analyses," *Mechanical Systems and Signal Processing*, 37, 4–29.

Berger, J. O. (1990), "Robust Bayesian Analysis: Sensitivity to the Prior," *Journal of Statistical Planning and Inference*, 25, 303–328.

Berger, J. O. (1994), "An Overview of Robust Bayesian Analysis," *Test*, 3, 5–59.

Boole, G. (1854), *An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities*, London: Macmillan.

Box, G. E. P. (1980), "Sampling and Bayes Inference in Scientific Modelling and Robustness," *Journal of the Royal Statistical Society, A*, 143, 383–404.

Dawid, A. P. (1982), "The Well-Calibrated Bayesian," *Journal of the American Statistical Association*, 77, 605–610.

Efron, B. (1986), "Why Isn't Everyone a Bayesian?," *The American Statistician*, 40, 1–11.

Efron, B. (2005), "Bayesians, Frequentists, and Scientists," *Journal of the American Statistical Association*, 100, 1–5.

Gelman, A., and Robert, C. P. (2013), "Not Only Defended but Applied: The Perceived Absurdity of Bayesian Inference," *The American Statistician*, 67, 1–5.

Gelman, A., and Shalizi, C. R. (2013), "Philosophy and the Practice of Bayesian Statistics," *British Journal of Mathematical and Statistical Psychology*, 66, 8–38.

Keynes, J. M. (1921), *A Treatise on Probability*, London: Macmillan.

Meeker, W. Q., and Escobar, L. A. (1998), *Statistical Methods for Reliability Data*, New York: Wiley.

Oakley, J. E., and O'Hagan, A. (2007), "Uncertainty in Prior Elicitations: A Nonparametric Approach," *Biometrika*, 94, 427–441.

Picard, R., and Williams, B. (2013), "Rare Event Estimation for Computer Models," *The American Statistician*, 67, 22–32.

Seaman, J. W., Seaman, J. W., and Stamey, J. D. (2012), "Hidden Dangers of Specifying Noninformative Priors," *The American Statistician*, 66, 77–84.

Senn, S. (2008), "A Note Concerning a 'Paradox' of Dawid's," *The American Statistician*, 62, 206–210.

Shafer, G. (1982), "Lindley's Paradox," *Journal of the American Statistical Association*, 77, 325–334.

Sivaganesan, S., and Berger, J. O. (1989), "Ranges of Posterior Measures for Priors with Unimodal Contaminations," *Annals of Statistics*, 17, 868–889.

Venn, J. (1888), *The Logic of Chance*, New York: Macmillan.

Walley, P. (1991), *Statistical Reasoning with Imprecise Probabilities*, London: Chapman Hall.

Wasserman, L., and Kadane, J. B. (1992), "Computing Bounds on Expectations," *Journal of the American Statistical Association*, 87, 516–522.

Wasserman, L., and Seidenfeld, T. (1994), "The Dilation Phenomenon in Robust Bayesian Inference," *Journal of Statistical Planning and Inference*, 40, 345–356.