

Bayesian Modeling and Inference for One-Shot Experiments

Jonathan Rougier & Andrew Duncan

To cite this article: Jonathan Rougier & Andrew Duncan (24 Jul 2023): Bayesian Modeling and Inference for One-Shot Experiments, Technometrics, DOI: [10.1080/00401706.2023.2224524](https://doi.org/10.1080/00401706.2023.2224524)

To link to this article: <https://doi.org/10.1080/00401706.2023.2224524>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



View supplementary material [↗](#)



Published online: 24 Jul 2023.



Submit your article to this journal [↗](#)



Article views: 579



View related articles [↗](#)



View Crossmark data [↗](#)

Bayesian Modeling and Inference for One-Shot Experiments

Jonathan Rougier ^{a,b} and Andrew Duncan^c

^aSchool of Mathematics, University of Bristol, Bristol, UK; ^bAWE plc, Aldermaston, UK; ^cDepartment of Mathematics, Imperial College London, London, UK

ABSTRACT

In one-shot experiments, units are subjected to varying levels of stimulus and their binary response (go/no-go) is recorded. Experimental data is used to estimate the “sensitivity function”, which characterizes the probability of a “go” for a given level of stimulus. We review the current GLM approaches to modeling and inference, and identify some deficiencies. To address these, we propose a novel Bayesian approach using an adjustable number of cubic splines, with physically-plausible smoothness, monotonicity, and tail constraints introduced through the prior distribution on the coefficients. Our approach runs “out of the box,” and in roughly the same time as the GLM approaches. We illustrate with two contrasting datasets, and show that our more flexible Bayesian approach gives different inferences to the GLM approaches for both the sensitivity function and its inverse.

ARTICLE HISTORY

Received November 2022
Accepted May 2023

KEYWORDS

go/no-go; No-Fire Threshold (NFT); Quantal function; Sensitivity testing

1. Introduction

A detonator is given an energy stimulus, and the probability that the detonator fires (a “go”) is a function of the amount of energy. This is an example of a “single event go/no-go” test, also known as a “one-shot” test or a “sensitivity” test. We will use “one-shot” test, and “experiment” for a set of tests at specified values of the stimulus. “One-shot” and “single event” indicate that a unit can only be tested once, regardless of the outcome, because after the stimulus the unit is no longer factory-fresh. For explosive devices, there is also the cost of disposing of a “no-go” unit, which still has the potential to “go”. Thus, one-shot experiments tend to be expensive, and there is a preference not to do tests at low stimulus values.

There is a large literature on one-shot experiments; see, for example, Langlie (1963), Wetherill (1963), Young and Easterling (1994), Neyer (1994), Dror and Steinberg (2008), and Wu and Tian (2014). Although much of the literature concerns explosive devices, we prefer to write “stimulus” and “go” rather than, say, “energy” and “fire,” to reflect that there are also plenty of other applications. In ecotoxicology, for example, the stimulus is the concentration of the chemical, and “go” is the survival of an organism, or some other endpoint such as growth or reproduction (see, e.g., Hickey and Hart 2013, sec. 14.2.1).



Denote the probability of a “go” as a function of the stimulus the “sensitivity function.” Broadly, the purpose of one-shot experiments is to characterize the sensitivity function. Sometimes the focus is the center of the sensitivity function. For example, in ecotoxicology one quantity of interest is the EC_{50} , the concentration at which 50% of the organisms survive. But sometimes the focus is deep in the tails. For example, evaluation of the No-Fire Threshold (NFT) of a detonator requires an uncertainty assessment of the inverse sensitivity function at $p = 0.001$, which is a stimulus far lower than those used in the tests.


Section 2 reviews current approaches, which fall into three overlapping categories. We show that these approaches have deficiencies, and we claim that a Bayesian approach can address these, but at a potentially higher cost, both in terms of code complexity and compute time. Section 3 presents our novel Bayesian approach. Section 4 is a short discussion about uncertainty assessment. Section 5 illustrates with two contrasting datasets, and Section 6 concludes with a brief summary and an outline of current research directions. The Markov chain Monte Carlo (MCMC) sampler for the Bayesian posterior distribution is described in the supplementary material.

2. Review

Let $x \mapsto p(x)$ be the sensitivity function of a specified type of unit, where x is the stimulus, and $p(x)$ is the probability of “go”. In an ideal world the sensitivity function would be a step-function, with a transition at some critical stimulus, but in reality it is an increasing function, where the difference between $p(x)$ and 0 or 1 represents the influence of many uncontrolled factors. These include miscalibration of the stimulus, the profile of the delivered stimulus (which may ramp up over a short time-interval), manufacturing variations in the unit, exposure to contaminants during emplacement, environmental conditions such as temperature and humidity, “shake, rattle and roll,” and so on.

We will use a parametric model for the sensitivity function. Interest in one-shot experiments often focuses on the inverse of the sensitivity function, which Wetherill (1963) in an early review refers to as the “quantal response function,” $q(\cdot) := p^{-1}(\cdot)$, which we shorten to “quantal function.” The quantal function comes for free, with a parametric model for the sensitivity function, and therefore this section and the next focus on the sensitivity function.

CONTACT Jonathan Rougier  j.c.rougier@bristol.ac.uk  School of Mathematics, University of Bristol, Fry Building, Woodland Road, Bristol, BS8 1UG, UK.

 Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/TECH.

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

2.1. Current GLM-based Approaches

Consider this parametric model for the sensitivity function:

$$p(x) = \Phi(x; \mu, \sigma^2), \quad (1a)$$

where $\Phi(\cdot)$ is the Gaussian distribution function with mean and variance arguments. Another way to express this model is

$$p(x) = \Phi\left(\frac{x - \mu}{\sigma}; 0, 1\right) = \Phi_S(\alpha_0 + \alpha_1 x) \quad (1b)$$

where $\Phi_S(\cdot)$ is the standard Gaussian distribution function, and $\alpha_0 := -\mu/\sigma$, $\alpha_1 := 1/\sigma$. When attached to a Binomial likelihood function, this is a Binomial Generalized Linear Model (GLM) with a Probit link function; see (6). This is the standard model for one-shot experiments, with either a Probit or a Logit link function (see, e.g., Teller et al. 2016).

Unfortunately, (1b) has the feature that $p(x) > 0$ for $x \leq 0$, which is nonphysical. This would be inappropriate for any analysis of the sensitivity function at low values of the stimulus, and hard to defend in the presence of alternatives. The obvious alternative to (1) is

$$p(x) = \Phi_S(\alpha_0 + \alpha_1 \log(x)), \quad (2a)$$

for which $\lim_{x \downarrow 0} p(x) = 0$. Reversing the argument in (1),

$$p(x) = \Phi(\log(x); \mu, \sigma^2) = L(x; \mu, \sigma^2), \quad (2b)$$

where “L” is the distribution function of the Lognormal distribution. So the problem with (1) can be fixed simply by using $\log(x)$ instead of x , and the result is equivalent to using a Lognormal link function in x , and fitting over (μ, σ) .

So, to summarize, there are three approaches for parametric models of the sensitivity function, which overlap:

1. *GLM-lin*: Binomial GLM in x , (1b);
2. *GLM-log*: Binomial GLM in $\log(x)$, (2a);
3. *Distribution approach*: the link function is the quantile function for a nonnegative random variable, like (2b).

The GLM approaches can use a standard GLM algorithm such as `glm` in the R statistical programming environment (R Core Team 2020). The Distribution approach requires a numerical optimization to fit the distribution parameters by Maximum Likelihood, unless it happens to overlap with the GLM-log approach, like the Lognormal overlaps with GLM-log and a Probit link function. Apart from the Lognormal, other obvious choices of two-parameter distribution would be the Gamma, Weibull, and Gompertz.

The second and third approaches respect the condition $p(x) = 0$ for $x \leq 0$. For completeness, note that this condition can be generalized to $p(x) = 0$ for $x \leq x_0$ simply by replacing x with $x - x_0$.

Finally, it is important to recognize that there is one situation where GLM-lin and GLM-log will give effectively the same fit. Let x_{mid} be some central value for the test stimuli, and write $x = x_{\text{mid}} + \delta$. If $|\delta| \ll x_{\text{mid}}$, then

$$\begin{aligned} \alpha_0 + \alpha_1 \log(x) &= \alpha_0 + \alpha_1 \log\{x_{\text{mid}}(1 + \delta/x_{\text{mid}})\} \\ &\approx \alpha_0 + \alpha_1 \log(x_{\text{mid}}) + \alpha_1 (\delta/x_{\text{mid}}) \\ &= \alpha'_0 + \alpha'_1 x \end{aligned} \quad (3)$$

where $\alpha'_0 := \alpha_0 + \alpha_1 \{\log(x_{\text{mid}}) - 1\}$ and $\alpha'_1 := \alpha_1/x_{\text{mid}}$. It is useful to operationalize this property for the whole dataset, albeit heuristically.

Let x_u be the largest “no-go” stimulus, and x_ℓ be the smallest “go” stimulus, and let x_{mid} be their midpoint. The interval $[x_\ell, x_u]$ is the “overlap,” assuming, as is usual in a designed experiment, that $x_\ell < x_u$, so that the overlap is the interval of stimulus values containing both “no-go’s” and “go’s”. A fitted GLM will tend to set $p(x) \approx 0$ for $x < x_\ell$, and $p(x) \approx 1$ for $x > x_u$, these being the maximum likelihood values of a fully flexible model: crudely, the overlap is where the fitting occurs. The approximation in (3) relies on $\log(1 + v) \approx v$, which holds to two significant figures for $|v| < 1/8$. Therefore, the twin conditions

$$\frac{x_{\text{mid}} - x_\ell}{x_{\text{mid}}} < \frac{1}{8}, \quad \frac{x_u - x_{\text{mid}}}{x_{\text{mid}}} < \frac{1}{8}$$

place the overlap inside the region where the approximation in (3) applies. Equivalently,

$$\frac{x_u - x_\ell}{x_u + x_\ell} < \frac{1}{8}, \quad (4)$$

which we call the “overlap condition,” with the term on the left being the “overlap ratio.” If the overlap condition holds, then the fitted sensitivity functions in GLM-lin and GLM-log will be effectively the same.

The overlap condition often holds in practice, which probably explains why GLM-lin persists, despite being nonphysical: it is “effectively physical.” However, being nonphysical matters whenever very small values of the stimulus are of interest, as will be shown in Section 5. Since it costs nothing to replace x with $\log(x)$ in a GLM, we strongly deprecate the use of GLM-lin in practice. After all, if the overlap condition holds then the fitted GLM-lin and GLM-log are effectively the same, and if the overlap condition does not hold then the fitted GLM-lin is potentially nonphysical.

2.2. Limitations in these Approaches

These different approaches highlight an important distinction when it comes to generalizations. We do not believe that two parameters are sufficient to capture the complexity of the sensitivity function, over the full range of the stimulus. However, uncontrolled factors blur the sensitivity function in experiments, and affect the unit in usage. What is estimated from an experiment is that part of the sensitivity function which emerges from the blur, and this might be quite low-dimensional, especially in small experiments. But to insist that it is two-dimensional is very prescriptive. In general it would be better to allow the number of parameters to be larger and adjustable. This in turn requires approaches that can be generalized to allow for an adjustable number of parameters.

The GLM approaches can easily be generalized to more than two parameters simply by including more regressors, for example

$$p(x) = \Phi_S(\alpha_0 + \alpha_1 \log(x) + \alpha_2 \{\log(x)\}^2) \quad (5)$$

for GLM-log. However, the resulting estimated sensitivity function is not necessarily increasing. This could be enforced with

a careful choice of regressors and sign constraints on the coefficients, but this would no longer be a standard GLM, and would require more complicated methods with less-well-understood properties (Py and Wood 2015).

The Distribution approach will ensure that the sensitivity function is increasing, but the pool of distribution functions with an adjustable number of parameters is poorly-defined. For example, there are three-parameter generalizations of the Gamma and Weibull distributions, but no obvious further extensions. Of course we could construct such distributions, but they would be largely arbitrary, and intractable.

So both approaches are deficient, if we want a parametric sensitivity function which is both increasing and has an adjustable number of parameters.

These approaches have another deficiency too, which is uncertainty assessment. Formal characterizations of the sensitivity function are expressed with a specified level of confidence. For example, the NATO definition of the No-Fire Threshold (NFT) is the value x_L for which (x_L, ∞) is a 95% confidence interval for $q(0.001)$, known as a lower one-sided 95% confidence interval for $q(0.001)$; see STANAG 4560 (2016, p. 4). The NFT requires confidence intervals for stimuli which are far lower than those used in the experiment, which are typically around $p = 0.5$.

Unfortunately, there are no practical algorithms for producing accurate confidence intervals at very low values of the stimulus. Thus, the coverage of practical algorithms which produce a lower one-sided nominal 95% confidence interval for $q(0.001)$ cannot be guaranteed to be at least 95%. The accuracy of practical algorithms such as the delta method (Casella and Berger 2002, p. 240), likelihood ratios (Neyer 1992), or the bootstrap (Efron and Hastie 2016, chap. 10 and 11) is only guaranteed under regularity conditions and as the number of tests goes to infinity. Even where the regularity conditions hold, convergence in the tails of the quantal function is likely to be slow. A typical one-shot experiment might involve tens of tests, which seems far too low to trust an asymptotic result in the tail of the quantal function. So a value x_L can be computed, for the NFT, but the true coverage of the algorithm which produces x_L is poorly known.

This pessimistic assessment is supported by the empirical study of Young and Easterling (1994), who find that both point estimates and standard asymptotic confidence intervals perform poorly for $q(0.99)$ and $q(0.999)$, for sample sizes up to $n = 50$.

The need for accurate uncertainty assessment in the tails of the sensitivity function favors a Bayesian approach to estimation and prediction. This will produce an effectively exact credible interval for the sensitivity function at any stimulus. Steinberg, Dror, and Villa (2014) also advocate a Bayesian parametric approach, for a variety of reasons including uncertainty assessment. It is a moot point whether a user such as a regulator would prefer an approximate confidence interval or an exact credible interval; we are content to provide an approach for the latter, to widen the pool of options.

A Bayesian approach can also meet the other needs discussed above: (i) $p(x) = 0$ for $x \leq 0$, (ii) $p(\cdot)$ increasing, and (iii) $p(\cdot)$ parameterized by an adjustable number of parameters. In fact, as we show in Section 3.2, a Bayesian approach can generalize the condition that $p(x) = 0$ for $x \leq 0$. The GLM-log and

Distribution approaches have $p(x) > 0$ for $x > 0$. That is, they “lift off” at $x = 0$. The Bayesian approach allows liftoff to happen at some unknown value which may be substantially greater than zero. The additional requirement to specify a prior distribution for the parameters should not diminish the appeal of a Bayesian approach, because a flexible model with a flexible prior distribution is far less prescriptive than the two-parameter models that are currently used.

The issue with a Bayesian approach is that it can be time-consuming if a sampling approach like Markov chain Monte Carlo (MCMC) is required. Overall, there is no clear winner, but we believe the arguments above favor a Bayesian approach if the technical and computational cost can be managed, or if the application justifies a high cost. The next section presents a Bayesian approach, to widen the set of options available to the analyst.

Finally, if we were required to provide just one justification for favoring a Bayesian approach over the GLM approaches, it would be the flexibility of the Bayesian approach to depart from symmetry. GLM-lin is symmetric by construction, since the Gaussian distribution is symmetric. Then GLM-log is often nearly symmetric, because, as shown in Section 2.1, the predictions of GLM-log and GLM-lin will be effectively the same when the overlap condition holds. But there is seldom anything in the physics, chemistry, biology, engineering, or other aspects of the application, to justify this symmetry. The same argument applies to the Distribution approach, which also imposes a very tight relationship between the lower and the upper half of the sensitivity function.

3. Bayesian Approach

This section presents our novel Bayesian approach for analyzing datasets from one-shot experiments. We are not the first authors to propose a Bayesian approach, per se. For example, Dror and Steinberg (2008), Steinberg, Dror, and Villa (2014), and Teller et al. (2016) pursue an approach to experimental design in which the GLM-lin model with a Probit link function is treated in a Bayesian fashion with a prior distribution on (α_0, α_1) , rather than estimated through the usual GLM machinery. This is advantageous in experimental design, where expert judgement can provide an informative prior distribution. The critical feature of our approach is its functional flexibility, as summarized in the final paragraph of the previous section. We show below that this imposes a high level of complexity on the parameter space, which requires a Bayesian inference to manage the uncertainty assessment. So when we write “Bayesian approach,” it is this flexibility that is the crucial feature, not simply the presence of a prior distribution on the parameters.

Furthermore, there are other approaches to Bayesian isotonic regression, such as Neelon and Dunson (2004). Our approach is different from these approaches because of the need to respect physically-plausible features of sensitivity functions. Additionally, we are sensitive to two user needs. First, to provide defensible assessments of uncertainty in the tails of the sensitivity function and the quantal function, despite possibly a small number of tests. Second, not to substantially increase user effort over the current GLM approaches.

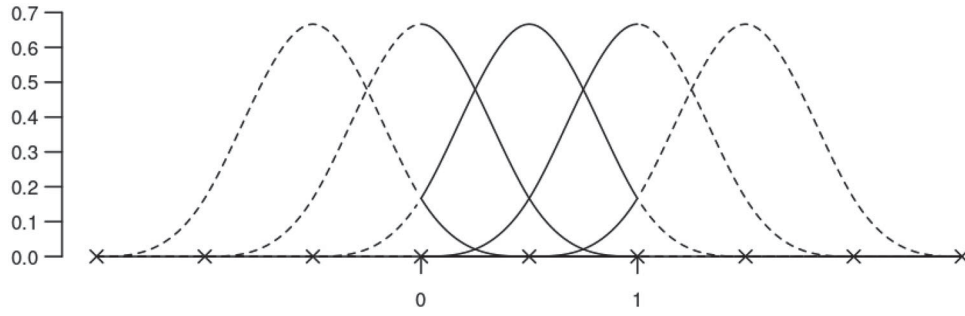


Figure 1. Cubic spline basis functions on $[0, 1]$ with $k = 5$. The 9 knots are shown as crosses on the horizontal axis; the knot spacing is $\Delta = 0.5$. The five basis functions on $[0, 1]$ are shown as solid lines, while their continuations outside $[0, 1]$ are shown as dashed lines. At most values of $x \in [0, 1]$ there are four nonzero basis functions, but at the knots, including the two endpoints, there are only three.

3.1. The Model

Start with the GLM-lin approach described in the previous section. The statistical model is

$$\begin{aligned} Y_i &\sim \text{Bin}(n_i, p(x_i)), \quad i = 1, \dots, m \\ g(p(x)) &= \alpha_0 + \alpha_1 x, \end{aligned} \quad (6)$$

where “Bin” is a Binomial distribution with size and probability arguments, and $g(\cdot)$ is a link function, like the quantile function of the standard Gaussian distribution (Probit). We have chosen, without loss of generality, to group the tests within m stimulus bins, taken to be closed on the left. In this case, n_i is the number of tests which fall within the i th stimulus bin, Y_i is the number of “go’s”, and x_i is the midpoint of the bin.

Denote the zero-point of the stimulus as x_0 : the value at which it is practically impossible for a “go” to occur, possibly the physical lower limit of the stimulus values. Usually this would be $x_0 = 0$, but we write x_0 rather than 0 for clarity and generality.

In our model we require the link function $g(\cdot)$ to satisfy

$$g^{-1}(v) = 0 \quad \text{for all } v \leq 0,$$

that is, to have a hard lower bound, and we replace $\alpha_0 + \alpha_1 x$ with a more flexible function, but still smooth, continuous, and increasing:

$$h(x; \alpha) := \sum_{j=1}^k \alpha_j b_j(x) = \alpha^T \mathbf{b}(x) \quad (7)$$

where the $b_j(\cdot)$ are cubic B-spline basis functions on an evenly-spaced set of knot locations (see, e.g., Wood 2017, sec. 5.3.3). The use of splines requires specified lower and upper bounds for x , denoted \underline{x} and \bar{x} . These bounds delimit the range of stimulus values of interest, where $x_i \in [\underline{x}, \bar{x}]$ for each i , and $x_0 \leq \underline{x}$. The condition

$$\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_k \quad (8)$$

is sufficient for $h(\cdot)$ in (7) to be increasing; see (Wood 2017, sec. 5.3.6) or (17). The sufficiency of this condition holds whether or not the knots are evenly-spaced, and it would be an option to use an uneven knot spacing in $[\underline{x}, \bar{x}]$. However, it is simpler to have evenly-spaced knots which are implicitly defined by the range $[\underline{x}, \bar{x}]$ and the number k .

For later reference, cubic B-splines have $k + 4$ knots in total, here denoted $x^{(j)}$, for which $\underline{x} = x^{(4)}$, $\bar{x} = x^{(k+1)}$, so that the knot spacing is

$$\Delta := \frac{\bar{x} - \underline{x}}{k - 3}. \quad (9)$$

The j th cubic spline $b_j(\cdot)$ is zero for $x \leq x^{(j)}$ and for $x \geq x^{(j+4)}$, and strictly positive and symmetric between $x^{(j)}$ and $x^{(j+4)}$. Therefore, for any x , either three or four basis functions are strictly positive: four in general, but three at the knots, including \underline{x} and \bar{x} . See (Wood 2017, sec. 5.3.3) for a summary, or de Boor (2001) for full details. Figure 1 shows the basis functions on $[0, 1]$ with $k = 5$.

Finally in this section, we suggest default choices for $[\underline{x}, \bar{x}]$, k , and $g(\cdot)$. The range $[\underline{x}, \bar{x}]$ is usually determined by the application. Well-known experimental design approaches, like Langlie (1963), take such a range as their starting-point. But where the user is reluctant to provide a range, a “pretty” range is easy to construct based on the test stimulus values, possibly extending down to x_0 .

For the number of basis functions we reason as follows. Current practice is satisfied with two parameters: α_0 and α_1 in (6). As discussed in Section 2, this is too restrictive. On the other hand, we do not want make k huge, because there is only a limited amount of information in a one-shot experiment. We have experimented with using a variety of k values and selecting one using an information criterion (WAIC, see Gelman et al. 2014, sec. 7.2), but the results were not compelling across the wide range of experimental datasets encountered in practice. Therefore, our starting-point is $k \leftarrow 5$, which is modestly larger than current practice.

However, if x_0 is not too far below \underline{x} , then it seems sensible to set $\underline{x} \leftarrow x_0$ so that the model “sees” x_0 , and increase k to keep the density of knots about the same. Specifically, compute Δ from (9) with $k = 5$. If x_0 is closer to \underline{x} than \underline{x} is to \bar{x} , then set $\underline{x} \leftarrow x_0$, and increase k to keep the knot spacing about the same:

$$k \leftarrow \left\lceil \frac{\bar{x} - \underline{x}}{\Delta} \right\rceil + 3. \quad (10)$$

where $\lceil v \rceil$ is the smallest integer not smaller than v . The user may want to select a smaller or larger k , depending on the number of tests in the experimental dataset, and also examine sensitivity to k for quantities of interest, such as low quantals.

For the link function we reason as follows. Current practice is satisfied with a symmetric sigmoidal link function (Probit).

Therefore, we will also use a symmetric sigmoidal link function, although the resulting sensitivity function will not be symmetric, because $h(x; \alpha)$ is not an affine function of x . We use the quantile function of a Beta(2, 2) distribution,

$$g^{-1}(v) := \Pr(V \leq v) \quad V \sim \text{Beta}(2, 2), \quad (11)$$

which satisfies our requirement that $g^{-1}(v) = 0$ for $v \leq 0$. This choice implies that $g^{-1}(v) = 1$ for $v \geq 1$; that is, sufficiently high stimulus values are certain to cause a “go,” although these values may be well above \bar{x} , the top of the range of interest. This particular Beta was chosen because of its width, but the result will not be sensitive to the choice because the spline coefficients can adjust.

3.2. Prior Distribution

This section explains how to specify a prior distribution for the basis coefficients α , which adjusts automatically to the modeling choices of the previous section.

Putting aside all shape constraints on the sensitivity function, the natural prior distribution for α is a multivariate Gaussian distribution of the general form

$$\alpha \mid \sigma \sim N_k(\bar{\alpha} \mathbf{1}, \sigma^2 W). \quad (12)$$

Both $\bar{\alpha}$ and W can be given simple values. For $\bar{\alpha}$, note that

$$\mathbb{E}(h(x; \alpha) \mid \sigma) = (\bar{\alpha} \mathbf{1})^T \mathbf{b}(x) = \bar{\alpha},$$

because $\mathbf{1}^T \mathbf{b}(x) = 1$ for all $x \in [\underline{x}, \bar{x}]$ (de Boor 2001, p. 89). Therefore, $\bar{\alpha}$ is the vertical offset, and the natural choice is

$$\bar{\alpha} \leftarrow g(0.5). \quad (13)$$

For the link function in (11), $\bar{\alpha} = 0.5$.

For W we reason as follows. As outlined above, cubic B-splines are localized. Therefore, consecutive smoothness in the components of α promotes smoothness in the sensitivity function. This feature is exploited in spline-based smoothing; for example, Eilers and Marx (1996) proposed a penalty on α which amounted to a random walk prior (Wood 2017, sec. 5.3.3). However, the monotonic constraint (8) already imposes a high degree of smoothness on α , and we have found that it is unnecessary to add more. Therefore, we make the simplest choice, that

$$W \leftarrow I,$$

the identity matrix.

Next, we need a prior distribution for σ . Letting $\bar{\sigma} := \mathbb{E}(\sigma)$, we tentatively propose that $\bar{\sigma}$ satisfies

$$\text{sd}(h(x_{\text{mid}}) \mid \bar{\sigma}) = g(0.75) - g(0.25)$$

for some representative value $x_{\text{mid}} \in [\underline{x}, \bar{x}]$. In other words, one standard deviation covers 50% of the range. This would imply

$$\bar{\sigma} \leftarrow \frac{g(0.75) - g(0.25)}{\|\mathbf{b}(x_{\text{mid}})\|_2} \quad (14)$$

after substituting $W = I$, where $\|\cdot\|_2$ is the Euclidean norm. However, the denominator can vary according to where x_{mid} lies relative to the knots, whose location is determined by $[\underline{x}, \bar{x}]$ and k . It is largest when x_{mid} lies at a knot, and smallest when x_{mid}

lies equidistant between knots. So we define x_{mid} to be a stimulus value equidistant between knots. For the link function in (11) with $k = 5$ basis functions, $\bar{\sigma} = 0.512$.

We choose an Exponential prior distribution for σ , so that $\bar{\sigma}$ is the only argument required. The Exponential distribution is the Shannon entropy maximizing distribution on the positive real line, subject to an expectation constraint. As will be seen below (Figures 2 and 4) this large $\bar{\sigma}$ and Exponential distribution, in combination with the other choices, generates a wide and fairly neutral set of prior samples for the sensitivity function.

Finally, we come back to the shape constraints, that $p(\cdot)$ is increasing, and that $p(x) = 0$ for $x \leq x_0$. The prior distribution on (α, σ) defines a stochastic process for $p(\cdot)$, the sensitivity function. In this stochastic process every sample path is a function which maps a stimulus value x into $[0, 1]$.

First, consider the case where $x_0 = \underline{x}$. The sample paths do not all pass through $(x_0, 0)$, since there will be values of α for which $v = \alpha^T \mathbf{b}(x_0) > 0$, and these sample paths pass through $(x_0, g^{-1}(v))$, where $g^{-1}(v) > 0$. As the link function satisfies $g^{-1}(v) = 0$ for $v \leq 0$, the conjunction

$$x_0 = \underline{x} : \alpha_1 \leq \dots \leq \alpha_k \text{ and } \alpha^T \mathbf{b}(x_0) \leq 0 \quad (15)$$

is sufficient for $p(\cdot)$ increasing and $p(x) = 0$ for $x \leq x_0$. The second condition ensures that $p(x_0) = 0$, and the first condition ensures that $p(\cdot)$ is increasing, which then implies that $p(x) = 0$ for all $x \leq x_0$. This conjunction does not rule out the possibility that $p(x) = 0$ for $x > x_0$. This is in contrast with GLM-log, discussed in Section 2, where $p(x) > 0$ for all $x > x_0$.

Now consider the case where $x_0 < \underline{x}$. This situation is more complicated because x_0 lies outside the domain of the basis functions. In this case, we adapt the second condition in (15) by linearly extrapolating from \underline{x} to x_0 , which gives

$$\alpha^T \mathbf{b}(\underline{x}) + (x_0 - \underline{x})h'(\underline{x}; \alpha) \leq 0 \quad (16)$$

which generalizes (15). de Boor (2001, p. 116, eq. 12') states the expression for the derivative of $h(\cdot)$:

$$h'(\underline{x}; \alpha) = 3 \sum_{j=1}^{k+1} \frac{\alpha_j - \alpha_{j-1}}{x^{(j+3)} - x^{(j)}} q_j(\underline{x}), \quad \alpha_0 = \alpha_{k+1} = 0, \quad (17)$$

where $x^{(j)}$ are the knot locations and $q_j(\cdot)$ are quadratic splines (unfortunately the expression in Wood 2017, p. 209, is wrong). It is straightforward to check that (17) implies

$$h'(\underline{x}; \alpha) = \frac{-\alpha_1 + \alpha_3}{2\Delta}.$$

Furthermore, $\mathbf{b}(\underline{x})$ is only nonzero for its first three components (see Figure 1). Therefore, (16) can be written as

$$\mathcal{R} := \left\{ \alpha \in \mathbb{R}^k : \alpha_1 \leq \dots \leq \alpha_k \text{ and } \alpha^T \tilde{\mathbf{b}} \leq 0 \right\} \quad (18a)$$

where $\tilde{\mathbf{b}} = (\tilde{b}_1, \tilde{b}_2, \tilde{b}_3, 0, \dots, 0)$, and

$$\tilde{b}_1 := b_1(\underline{x}) - c, \quad \tilde{b}_2 := b_2(\underline{x}), \quad \tilde{b}_3 := b_3(\underline{x}) + c, \quad (18b)$$

where $c := (x_0 - \underline{x})/(2\Delta)$.

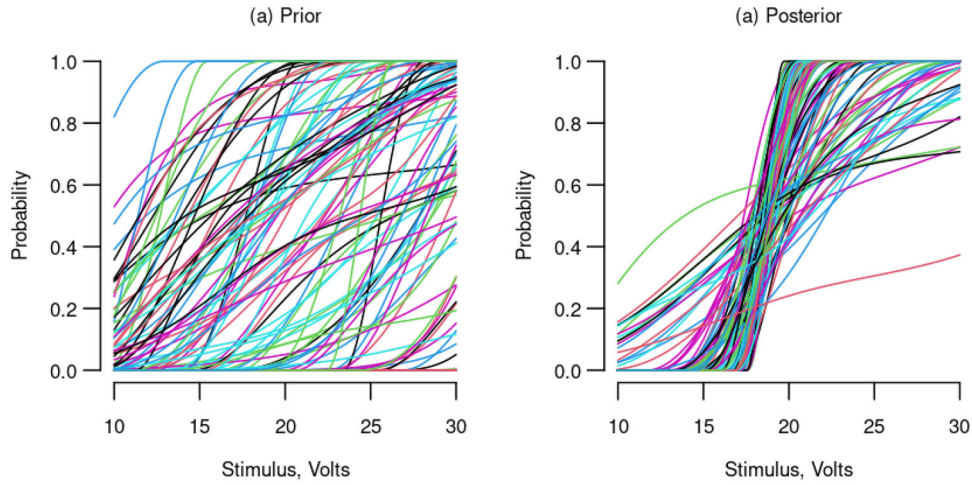


Figure 2. Dror dataset, 100 samples from the prior and the posterior sensitivity functions.

Constraining $\alpha \mid \sigma$ to \mathcal{R} gives the final model:

$$\begin{aligned} Y_i \mid \alpha &\sim \text{Bin}(n_i, p(x_i; \alpha)), \quad i = 1, \dots, m \\ g(p(x; \alpha)) &= \alpha^T \mathbf{b}(x) \\ \alpha \mid \sigma &\sim \text{TN}_k(\bar{\alpha} \mathbf{1}, \sigma^2 I, \mathcal{R}) \\ \sigma &\sim \text{Exp}(1/\bar{\sigma}), \end{aligned} \quad (19)$$

where $\text{TN}_k(\mu, \Sigma, \mathcal{R})$ is the k -variate Normal distribution with mean μ and variance Σ , restricted to the set \mathcal{R} , and $\bar{\alpha}$ and $\bar{\sigma}$ are specified in (13) and (14), respectively. The Supplementary Material describes our MCMC approach for targeting the posterior distribution of α .

As a reviewer has noted, the quantification of $\bar{\sigma}$ is made before the restriction of α to \mathcal{R} . The concern is that this restriction will filter out many sample paths for the sensitivity function, and therefore $\bar{\sigma}$ as quantified might be too low to generate a wide variety of sample paths in the prior distribution. However, the prior distribution for σ is Exponential, and so many of the σ values in the prior will be larger than $\bar{\sigma}$. Moreover, we inspect samples from the prior distribution, to ensure that they are not overly prescriptive; see Figures 2 and 4.

4. Uncertainty Assessment

As described in Section 2, some applications require specific uncertainty assessments, such as a lower one-sided 95% confidence interval for $q(0.001)$, for the no-fire threshold (NFT). Under a Bayesian approach, these can be straightforwardly extracted from the output of the MCMC sampler. Each sample has a posterior α , each α implies a sensitivity function, each sensitivity function can be inverted to a quantal function, and each quantal function yields a value for $q(0.001)$. The posterior distribution of $q(0.001)$ is then summarized as an interval. The lower one-sided 95% credible interval would start at the 5th percentile of the resulting sample.

Things are not so straightforward for the GLM approaches, bearing in mind that confidence intervals for nonlinear models must be based on some form of approximation: the delta method is the default for GLM algorithms, like `glm` in R. However, the value $q(0.001)$ is itself a nonlinear function of the parameters

(α_0, α_1) :

$$q(0.001) = \exp \left\{ \frac{g(0.001) - \alpha_0}{\alpha_1} \right\} \quad (20)$$

for GLM-log, and so uncertainty about $q(0.001)$ is doubly hard to assess accurately.

For the GLM approaches we suggest using a simulation-based approach based on treating (α_0, α_1) as Gaussian, using the asymptotic approximation to the covariance based on inverting the observed information matrix. This approach is commonly used in practice, and has a “quasi-Bayesian” justification (Wood 2017, sec. 2.2, 6.10, Appendix A.4). It is a unified approach which can be applied to approximating pointwise (vertical) confidence intervals for both the sensitivity function and the quantal function. For example, ten thousand samples of (α_0, α_1) would be generated from the fitted GLM-log model, $q(0.001)$ would be computed in each case using (20), and the lower one-sided 95% credible interval would start at the 5th percentile of the resulting sample. For the sensitivity function, the simulation-based approach is the same as using the delta method to construct a confidence interval on the link scale, and then transforming by the inverse link function, which is quicker and does not incur Monte Carlo error.

5. Illustrations

This section illustrates the performance of the different approaches, using two contrasting experimental datasets, both given below. The three approaches are GLM-lin and GLM-log, presented in Section 2.1, and the Bayesian approach presented in Section 3. To summarize their similarities and differences:

1. GLM-lin does not respect the physical constraint $p(x) = 0$ for $x \leq x_0$, where in both illustrations $x_0 = 0$. GLM-log and the Bayesian approach both respect this constraint.
2. GLM-lin has a symmetric sensitivity function, and GLM-log has an effectively symmetric sensitivity function when the overlap condition holds, as it will do in many applications. The Bayesian approach is not necessarily symmetric.

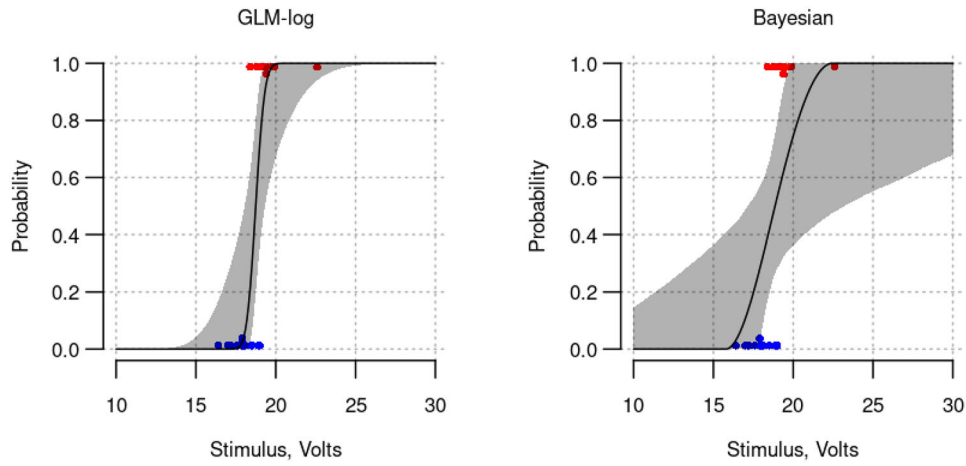


Figure 3. Dror dataset, sensitivity function under the GLM and Bayesian approaches. The solid line is the fitted function, and the shaded polygon shows pointwise 95% prediction intervals. The dataset is shown as dots: at $p = 0$ for “no-go”, and $p = 1$ for “go”. The overlap condition in (4) holds, and therefore the two GLM approaches are effectively identical; only GLM-log is shown.

Table 1. The Dror dataset, hand-digitized from Figure 5, Panel (b) of Dror and Steinberg (2008), where the stimulus is voltage.

No-Go	16.4, 17, 17.2, 17.6, 17.9, 17.9, 18, 18.1, 18.5, 18.9, 19
Go	18.4, 18.8, 19.1, 19.2, 19.4, 19.4, 19.8, 19.9, 22.6

NOTE: These test results are not aggregated into bins.

3. More generally, the flexibility the GLM approaches is restricted by having only two parameters, while the Bayesian approach has more: we use five and seven below.
4. The GLM approaches require approximations to compute pointwise confidence intervals for the sensitivity and quantal functions; therefore, their actual coverage will differ from their nominal coverage. The Bayesian approach provides effectively exact coverage in its pointwise credible intervals.

The calculations in this section were all carried out using the R package *oneshot*. This is available in the Supplementary Material, which also contains more details about the MCMC results below. The run-time of the entire set of calculations was just a few seconds on a standard laptop, using 10,000 samples in each approach, plus a spin-up of one thousand samples in the Bayesian approach.

5.1. The Dror Dataset

The first illustration uses the small detonator dataset from Figure 5, Panel (b) of Dror and Steinberg (2008), with 20 tests; hereafter “Dror”. This dataset is given in Table 1. The overlap ratio is 0.016, and therefore the overlap condition in (4) holds, so that there is no practical difference between GLM-lin and GLM-log. We will only fit the latter, but note the implication, that the GLM approaches are necessarily symmetric for this dataset.

As stated in Dror and Steinberg (2008, sec. 4), the objectives of the experiment “were to verify the manufacturer’s statement that the explosives will not detonate at 12 V (being a safe voltage) and will detonate (“all fire”) at 25 V.” Using these two values as indicative of lower and upper limits, there is a case for replacing the lower limit with $x_0 = 0$ and increasing the knots from the default value of $k = 5$, as described around (10). For this

illustration, though, we will stick with $[x, \bar{x}] = [10, 30]$ and $k = 5$.

Figure 2 shows samples from the prior and posterior sensitivity functions for the Dror dataset. The samples from the prior show that there are a wide variety of sensitivity functions in the prior distribution, so that the choice of prior described in Section 3.2 is not overly prescriptive. Comparing these two samples shows that conditioning on the dataset has substantially reduced uncertainty about the sensitivity function, but apparently these 20 tests are not able to eliminate uncertainty about the sensitivity function in the left-hand tail, despite the monotonicity constraint.

Figure 3 shows the sensitivity function for the GLM and Bayesian approaches, along with pointwise 95% prediction intervals. When presented with these two estimates, a subject matter expert, or regulator, might agree that they are clearly different, but neither is obviously wrong. To return to the question posed by Dror and Steinberg (2008), above, it would be prudent to conclude from Figure 3 that more testing is required, for both the lower limit of 12 V and the upper limit of 25 V.

5.2. The ESD Dataset

The second illustration uses a large detonator dataset from an experiment carried out at the Atomic Weapons Establishment (AWE), at Aldermaston in the United Kingdom; hereafter “ESD,” as it was used as part of an electrostatic discharge safety assessment. This dataset is given in Table 2. It is very different from the Dror dataset, both in its size (79 tests), and in its overlap ratio, which is 0.34, so that the overlap condition in (4) does NOT hold. Therefore, the two GLM approaches may give different results, and GLM-log may not be symmetric. In order to investigate the low quantals, we use $[x, \bar{x}] = (0, 300)$ and $k = 7$.

Figures 4 and 5 show the prior and posterior samples, and the estimated sensitivity functions, respectively. The effect of the constraint set \mathcal{R} in (19) is apparent in the left-hand panel of Figure 4, where every prior sensitivity function satisfies $p(0) = 0$. With this much larger dataset there is greater similarity between the three fits, although also some differences. In Figure 5 the

Table 2. The ESD dataset, from an experiment carried out at the Atomic Weapons Establishment (AWE), Aldermaston.

No-Go	44.1, 58.3, 59, 59.3, 59.3, 59.4, 60.3, 60.3, 60.4, 60.8, 61.3, 61.3, 61.3, 61.9, 62.5, 62.9, 63.3, 66, 67.3, 69.4, 70.8, 72.4, 72.8, 73.2, 74.3, 74.5, 76.2, 77.1, 77.8, 78.2, 78.5, 78.7, 79.1, 79.5, 79.9, 80, 80.1, 80.2, 82.3, 83.3, 84.4, 85.7, 92.6, 95.6, 99.5, 101, 105, 114, 120, 123, 130, 133, 140
Go	68.8, 78.2, 78.4, 78.9, 83.9, 84.3, 88.8, 89.5, 92.8, 103, 111, 111, 112, 112, 114, 127, 130, 132, 135, 137, 145, 149, 150, 151, 156, 196

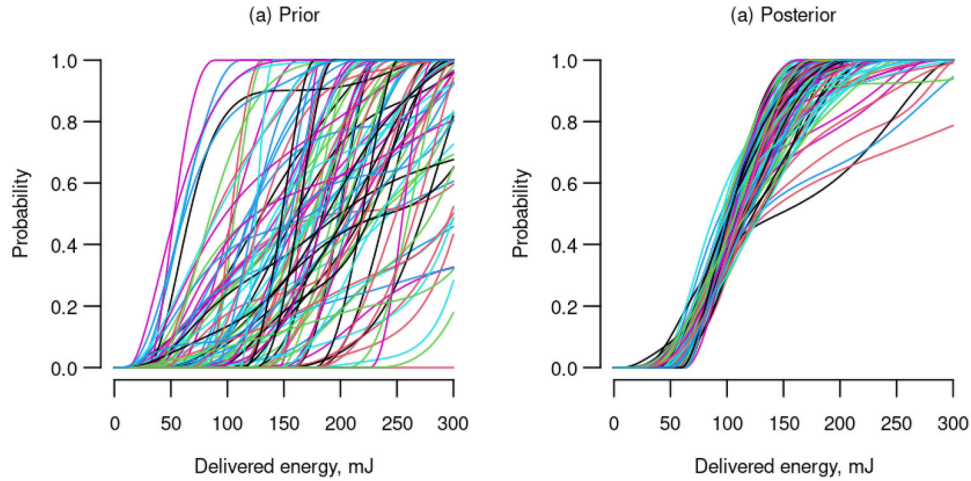
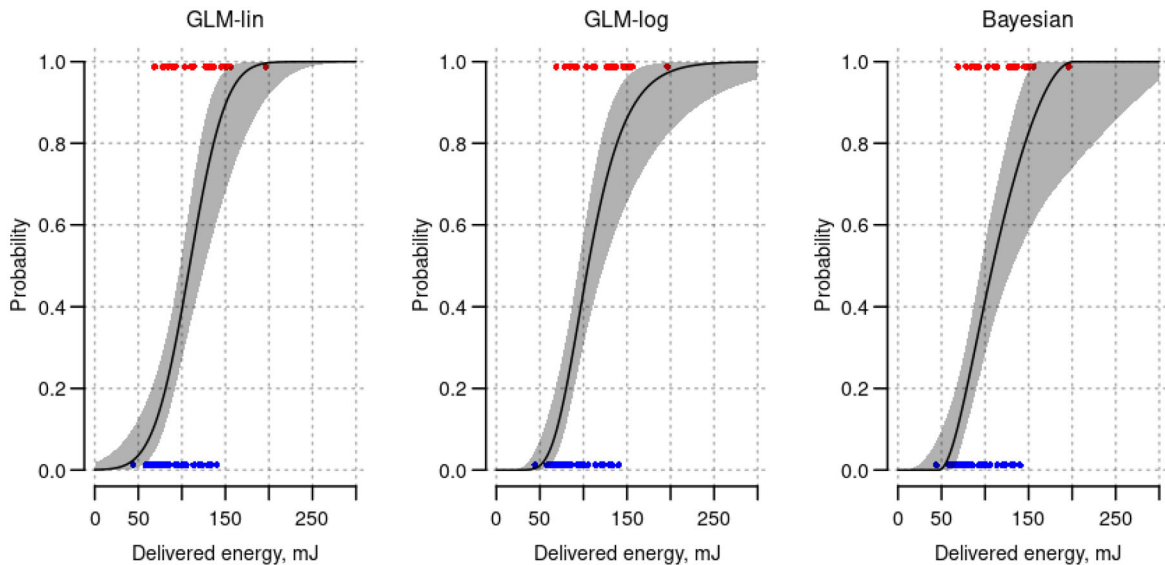
NOTE: The values are delivered energy in mJ. These test results are not aggregated into bins.

GLM-lin prediction interval has not gone to zero at $x = 0$, which will have a catastrophic effect on the low-probability quantals (see below). Comparing GLM-log and the Bayesian approach, the fits are similar around $p = 0.5$, but the Bayesian approach is more uncertain in the tails.

Figure 6 shows the fitted sensitivity functions along with the same functions rotated around their midpoints. This figure shows asymmetry in the fitted sensitivity functions. Due to the

shape of the Lognormal distribution, which has a long right-hand tail, the only asymmetry that GLM-log can manifest is a sensitivity function which is steeper taking off from $p = 0$ than it is arriving at $p = 1$. The Bayesian approach, which has no such constraint, finds that this asymmetry is present in the ESD dataset, and GLM-log has found it too. GLM-log appears to be more asymmetric than the Bayesian approach, but this is tenuous, because the asymmetry of GLM-log is determined entirely by the fitted values of μ and σ —the asymmetry may be a side-effect.

Figure 7 shows some low quantals for the three approaches, using two-tailed 95% prediction intervals. GLM-lin has given a nonsensical outcome for the $p = 0.001$ and $p = 0.01$ quantals, owing to not respecting the constraint $p(x) = 0$ for $x \leq 0$. The quantals for GLM-log and the Bayesian approach are roughly comparable, although the prediction intervals for the latter are wider. This could be due to differences in the two models, or it could be coverage error in GLM-log.

**Figure 4.** ESD dataset, 100 samples from the prior and the posterior sensitivity functions.**Figure 5.** ESD dataset, sensitivity function under the GLM and Bayesian approaches. The solid line is the fitted function, and the shaded polygon shows pointwise 95% prediction intervals.

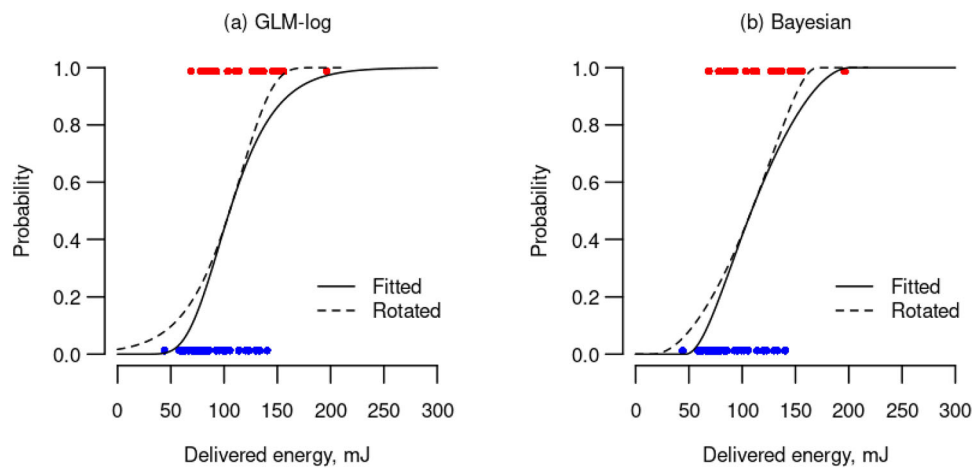


Figure 6. ESD dataset, rotating the fitted sensitivity functions around their midpoints, $(q(0.5), 0.5)$. (a) GLM-log, (b) Bayesian. If the fit is symmetric, the solid line and the dashed line will coincide.

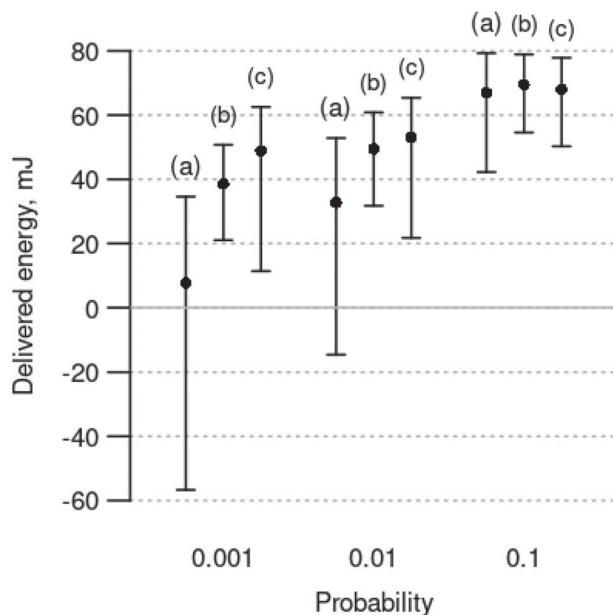


Figure 7. ESD dataset, low quantals. The dot is the fitted value and the intervals are two-tailed 95%. (a) GLM-lin, (b) GLM-log, (c) Bayesian.

6. Conclusion

Current practice for analysing data from one-shot experiments, for example for detonator safety and reliability assessment, is based on a Binomial Generalized Linear Model (GLM). We show that this approach has several deficiencies which can be addressed within a Bayesian approach in which the prior distribution represents the physical constraints of the application. We develop a MCMC sampling scheme which allows the Bayesian approach to run “out of the box,” in about the same time as the current GLM-based approaches. We show that our Bayesian approach gives different inferences to those of the GLM-based approaches.

Along the way, we explain why GLM-log should always be favored over GLM-lin (Section 2.1), and give a simple condition for when they will provide effectively identical fits (the overlap condition, (4)). When this condition holds, GLM-log will have an effectively symmetric sensitivity function, a restriction which

is unlikely to be justified by the underlying application. When this condition does not hold, GLM-log can only manifest one type of asymmetry, in which the sensitivity function is steeper taking off from $p = 0$ than it is arriving at $p = 1$. Our Bayesian approach can manifest both types of asymmetry, although in Section 5.2 we find that the direction of asymmetry in the ESD dataset is the same as that of GLM-log.

Finally, we would like to highlight two directions of ongoing research. First, we are planning a detonator experiment with several hundred tests, designed to investigate asymmetry in the sensitivity function. With such a large number of tests we ought to be able to use a large number of basis functions in the Bayesian approach, and—we hope—see whatever asymmetry exists, in high resolution. This will inform future inference for one-shot experiments to test detonators.

Second, we are examining experimental design to target a particular tail of the quantal function, such as the lower tail to assess the No-Fire Threshold (NFT). Experimental design for one-shot experiments has been well-studied (see, e.g., Wu and Tian 2014), but usually with either a GLM approach, or with a nonparametric approach. The nonparametric approach is flexible and effective around $p = 0.5$, but not practical for very low quantals. The rigidity of the two-parameter GLM approach has the unfortunate feature that tests at medium and high stimulus values will be informative for the sensitivity at very low stimulus values, a viewpoint which is rejected by subject matter experts. We have found that the Bayesian spline model combined with minimizing the expected Shannon entropy of the quantity of interest (such as the NFT) is effective at identifying stimulus values for further tests. The expected Shannon entropy can be estimated directly from the MCMC sample.

Supplementary Materials

The MCMC sampler is described in the Supplementary Material, along with diagnostics for ESD dataset. The code and datasets are also available, including the oneshot package for R.

Acknowledgments

We would like to thank the following for their very helpful comments on previous versions of this article: Rod Drake, David Steinberg, Simon Wood, Andrew Zammit Mangion, and Matt Stapleton. The comments of

the Editor, an Associate Editor, and three reviewers substantially improved the paper's focus and clarity.

Disclosure Statement

The authors report there are no competing interests to declare.

ORCID

Jonathan Rougier  <https://orcid.org/0000-0003-3072-7043>

References

- Casella, G., and Berger, R. L. (2002), *Statistical Inference* (2nd ed.), Pacific Grove, CA: Duxbury. [3]
- de Boor, C. (2001), *A Practical Guide to Splines* (revised ed.), New York: Springer. [4,5]
- Dror, H. A., and Steinberg, D. M. (2008), "Sequential Experimental Designs for Generalized Linear Models," *Journal of the American Statistical Association*, 103, 288–297. [1,3,7]
- Efron, B., and Hastie, T. (2016), *Computer Age Statistical Inference*, New York, NY: Cambridge University Press. Available at <https://hastie.su.domains/CASI/>. [3]
- Eilers, P. H. C., and Marx, B. D. (1996), "Flexible Smoothing with *b*-splines and Penalties," *Statistical Science*, 11, 89–102. With discussion and rejoinder, 102–121. [5]
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014), *Bayesian Data Analysis* (3rd ed.), Boca Raton, FL: Chapman and Hall/CRC. Available at <http://www.stat.columbia.edu/~gelman/book/>. [4]
- Hickey, G., and Hart, A. (2013), "Statistical Aspects of Risk Characterisation in Ecotoxicology," in *Risk and Uncertainty Assessment for Natural Hazards*, eds. J. C. Rougier, R. S. J. Sparks, and L. J. Hill, Cambridge, UK: Cambridge University Press. [1]
- Langlie, H. J. (1963), "A Reliability Test Method for 'One-Shot' Items," Technical Report ADP014612, Aeronutronic Division, Ford Motor Company. Available at <https://apps.dtic.mil/sti/citations/ADP014612>. [1,4]
- Neelon, B., and Dunson, D. B. (2004), "Bayesian Isotonic Regression and Trend Analysis," *Biometrics*, 60, 398–406. [3]
- Neyer, B. T. (1992), "An Analysis of Sensitivity Tests," Technical Report, Mound, Miamisburg, OH, US. Available at <https://www.osti.gov/biblio/5654343>, OSTI identifier 5654343. [3]
- (1994), "A D-optimality-based Sensitivity Test," *Technometrics*, 36, 61–70. [1]
- Pya, N., and Wood, S. N. (2015), "Shape Constrained Additive Models," *Statistics and Computing*, 25, 543–559. [3]
- R Core Team. (2020), *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at <https://www.R-project.org/>. [2]
- STANAG 4560. (2016), *Electro-Explosive Devices, Assessment and Test Methods for Characterization* (3rd ed.). Full reference number NS0/1418(2016)SGA/4560, 21 November 2016. [3]
- Steinberg, D. M., Dror, H., and Villa, L. (2014), "Discussion of 'Three-phase Optimal Design of Sensitivity Experiments' by Wu and Tian," *Journal of Statistical Planning and Inference*, 149, 26–29. [3]
- Teller, A., Steinberg, D. M., Teper, L., Rozenblum, R., Mendel, L., and Jaeger, M. (2016), "New Methods for Sensitivity Tests of Explosive Devices," in *13th International Conference on Probabilistic Safety Assessment and Management (PSAM 13)*, Seoul, Korea. [2,3]
- Wetherill, G. B. (1963), "Sequential Estimation of Quantal Response Curves," *Journal of the Royal Statistical Society, Series B*, 25, 1–48. [1]
- Wood, S. N. (2017), *Generalized Linear Models: An Introduction with R* (2nd ed.), Boca Raton FL: CRC Press. [4,5,6]
- Wu, C. F. J., and Tian, Y. (2014), "Three-Phase Optimal Design of Sensitivity Experiments," *Journal of Statistical Planning and Inference*, 149, 1–15. With comments and rejoinder, pp. 16–32. [1,9]
- Young, L. J., and Easterling, R. G. (1994), "Estimation of Extreme Quantiles based on Sensitivity Tests: A Comparative Study," *Technometrics*, 36, 48–60. [1,3]