

Center for Quality and Applied Statistics
Kate Gleason College of Engineering
Rochester Institute of Technology

Sequential Experimental Designs For Sensitivity Experiments

Joseph G. Voelkel
Center for Quality and Applied Statistics
Rochester Institute of Technology
jgvcqa@rit.edu

Technical Report 1999–1

August 1999

Sequential Experimental Designs For Sensitivity Experiments

by

Joseph G. Voelkel
Rochester Institute of Technology
Rochester, NY 14623

Abstract

In some experiments, the response is binary and one factor is being studied to estimate the factor setting at which probability of failure is a specified value, such as 0.10. In addition, data for such experiments are sometimes naturally collected one run at a time. The difficulty in such experiments, compared to experiments in which the response is continuous, is that the factor settings to run are not as clear. Running the settings too low or too high produces virtually no information, but even apparently reasonable values of the settings may turn out to be information-poor.

In this article we propose a Bayesian method to decide which setting to make at each successive run. This turns out to be particularly useful, even for non-Bayesians, because this method naturally incorporates the kind of vague information the experimenter has at the beginning of an experiment.

We introduce a measure of how much information is anticipated for different factor settings for the next run. We also introduce information plots as a way to gain further insight into the Bayesian strategy. We examine four practical methods to specify reasonable prior distributions, and recommend one in particular.

In addition to providing a strategy for optimal sequential design, the Bayesian method can also be used for providing estimates and their uncertainty. Several examples are presented to illustrate the strategy.

1 Introduction

ASTM D 1709-91 is “Standard Test Methods for Impact Resistance of Plastic Film by the Free-Falling Dart Method” (ASTM (1995)). Darts of various weights (made of metal, cylindrically shaped, with hemispherical heads) are dropped from a fixed height onto samples of film being tested. For the i^{th} run, a dart weight of x_i is

Key Words and Phrases: binary data; fatigue tests; logistic regression; optimal design; Bayesian design

dropped onto the i^{th} sample, and the response Y_i is recorded as either a 0 (film sample passed, or survived) or 1 (film failed). In this paper, we consider such experiments, sometimes called sensitivity experiments, and in all cases restrict ourselves to one factor such as weight.

The ASTM method provides two techniques for estimating the weight at which 50% of the samples will fail. The first technique, which we will illustrate, is a fully sequential method (i.e., no groupings of runs), called the up-and-down method. The second technique is essentially non-sequential and will not be discussed here.

The ASTM method only provides a sequential method for estimating the 50th percentile. However, other percentiles are often of interest. For example, an engineer may wish to compare several different films. If the films are made from different processes, and the engineer wishes to see which film is stronger under normal usage, it is more reasonable to compare them at a lower percentile, such as the 10th percentile. This is especially compelling when previous work has suggested that the ranking of films based on their 50th percentile do not correspond well to the rankings at lower percentiles. In fact, such a concern prompted this work.

We begin by reviewing standard models that are used in sensitivity experiments. We then review several non-Bayesian sequential methods that have been proposed for estimating a particular percentile in sensitivity experiments, and briefly mention some non-Bayesian, non-sequential results. However, because the best design for such experiments depends on the unknown parameters in the model, it is natural to incorporate initial uncertainty about the model's parameters in a Bayesian formulation. We next review Bayesian methods for the non-sequential case. We then show a Bayesian, sequential, method of constructing optimal designs.

The Bayesian formulation leads to a straightforward method for determining which setting should be run next. We gain insight into this result by illustrating from a simpler perspective why this setting is optimal. We also create a natural way to measure the anticipated information in the setting for the next run. Further insight is obtained by graphically examining the cumulative information obtained by the end of each run. We also compare this method's strategy of the selecting the next setting with other strategies.

Our emphasis is on practical utilization of such a method, so we spend some time showing how to develop prior distributions that the user can naturally specify. We examine four methods, and recommend one in particular.

The Bayesian design leads naturally to a Bayesian analysis. For our examples, we briefly compare our Bayesian analysis with other Bayesian methods and with maximum likelihood.

2 Mathematical Models

To fix ideas we refer to the film-dart example. Assume we have a large population of film samples, and that sample i has associated with it an underlying value X_i

that represents its strength. The collection of $\{X_i\}$ form a distribution $F(x)$, called the *tolerance distribution*. The most common examples of these are the normal and logistic distributions, with location and scale parameters unknown. The problem is that we cannot measure X_i directly. Rather, we drop a weight x_i on the sample, and then observe Y_i , whether the film fails or passes the test. If $X_i > x_i$, then we observe $Y_i = 0$, a pass; If $X_i \leq x_i$, then we observe $Y_i = 1$, a failure. If we are considering a generic x , we abuse notation and call the response Y_x . So, if the setting for the test is x , then

$$p_x = P(Y_x = 1) = P(X \leq x) = F(x).$$

If F is normal, with mean μ and standard deviation σ , then

$$p_x = F(x) = \Phi((x - \mu) / \sigma),$$

while if F is logistic, with location α and scale β , then

$$p_x = F(x) = 1 / (1 + \exp(-(x - \alpha) / \beta)). \quad (1)$$

Other authors have used other parametrizations, including $\alpha + \beta x$ and $\beta(x - \alpha)$. Our parametrization turns out to be more natural for the problem at hand. We work with the logistic model in this paper.

We are especially interested in estimating the setting x that corresponds to a user-specified probability p_x . We denote such a setting by δ , and the user-specified probability by γ . For example, we may wish to find the weight δ at which $\gamma = 0.10$ of the film fails. Solving for x in (1) leads to $x = \alpha + \beta \ln(p_x / (1 - p_x)) = \alpha + \beta \text{logit}(p_x)$, so

$$\delta = \alpha + \beta \text{logit}(\gamma). \quad (2)$$

Such a value for δ has many names in the literature, including $F100\gamma$, $ED100\gamma$, $LD100\gamma$, and L_γ . In this article we use either the δ or the $F100\gamma$ terminology. Also, for ease of language we refer to x as the setting, although it is sometimes more reasonable to transform the setting first, for example to $\log(\text{weight})$.

The $\{Y_i\}$ may be collected by different methods. The most common are:

1. Non-sequential. Both n and all the $\{x_i\}$ need to be specified before any results are obtained. This method might be used to find the dose δ of a chemical at which 5% of mice will develop tumors.
2. Group-sequential, e.g. two-stage. Here, using obvious notation, n_1 and $\{x_{1i}\}$ need to be specified before any results are obtained. Then the $\{Y_{1i}\}$ results are obtained. Next, n_2 and $\{x_{2i}\}$ are specified, both of which may depend on the results obtained in the first stage. Finally, the $\{Y_{2i}\}$ results are obtained. This method might be used in the mice-drug example if there is enough time for two rounds of data collection.

3. Fully sequential. We obtain $(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n)$, where the i^{th} setting x_i can be based on all results up to date. This is actually a special case of a group-sequential method that consists of n stages, and the sample size for the j^{th} stage is 1. However, it is important enough to be treated in its own right. This is the natural strategy to use in the film-dart example when only one machine is available, and this is where our interest lies.

We also need to decide on the objective of the experiment. The most common objectives, with associated reasonable strategies for achieving them, are:

1. To estimate the weight δ at which $\gamma = 0.10$ of the samples fail:

$$\delta = \alpha + \beta \logit(0.10) = \alpha - 2.2\beta.$$

So, attempt to set the $\{x_i\}$ to minimize $\text{Var}(\hat{\delta})$, the variance of a suitable estimate of δ .

2. To estimate the weight δ at which $\gamma = 0.50$ of the samples fail, so attempt to set the $\{x_i\}$ to minimize $\text{Var}(\hat{\alpha})$.
3. Obtain a good estimate of the curve “in general.” Here, attempt to set $\{x_i\}$ to minimize the generalized variance (determinant) of the variance-covariance matrix of $(\hat{\alpha}, \hat{\beta})$.
4. To estimate the scale parameter β . So, attempt to set the $\{x_i\}$ to minimize $\text{Var}(\hat{\beta})$.

Our primary interest in this article is use of the *logistic model*, in *fully sequential experiments*, to *estimate a setting δ corresponding to a given γ* . The fully sequential approach is fundamental to this article. The ideas we develop for it can be easily used for other models and for other objectives.

To begin the experiment, we need to have some idea, however vague, of reasonable values for α and β —this is the *original priors* problem. We examine how to place the $\{x_i\}$ —this is the *design* problem. We spend most of our effort on these two problems. We also examine how to estimate δ —the *analysis* problem. We assume that the probability of failure increases with x , which is equivalent to saying that the scale parameter β is positive—this is really not restrictive because the experimenter virtually always knows the sign of β . In any event, the method can be extended to the unknown-sign case.

We now review several non-Bayesian sequential methods.

3 Non-Bayesian sequential methods

3.1 The Up-and-Down Method

This method, also called the staircase method, was developed by Dixon and Mood (1948). It is used to set the $\{x_i\}$ sequentially to estimate the δ corresponding to $\gamma = 0.50$, i.e., $\delta = \alpha$.

For the up-and-down method, Dixon and Mood assumed the tolerance distribution was normal. The user begins by selecting an initial setting x_1 and a differential amount d . The first run is made at x_1 . If the sample fails, the next run is made at $x_2 = x_1 - d$; otherwise, the next run is made at $x_2 = x_1 + d$. This strategy is continued until n runs have been made. An example of this strategy, with $x_1 = 7$ and $d = 0.5$, is shown in Figure 1. This is the design solution.

To estimate δ , the average of the settings at which the smaller number of events occurred (pass or fail) is first obtained. Here, there were fewer passes, 6, and the average of the corresponding settings is $\bar{Y} = 5.17$. Then $\hat{\delta} = \bar{Y} + d/2$. ($\hat{\delta} = \bar{Y} - d/2$ if there were fewer failures). Here, $\hat{\delta} = 5.17 + 0.25 = 5.42$. An approximate s.e. for $\hat{\delta}$ can also be found—see the reference.

The technique works best if the user can select d such that $0.5\sigma \leq d \leq 2\sigma$. The estimate $\hat{\delta}$ is an approximation to the MLE. Properties of this method for small sample sizes have also been investigated (Brownlee, et al. (1953), Dixon (1965)). This method is recommended and used in various engineering applications—see ASTM D 1709, Little (1975), Little (1990), Little and Jebe (1975), and Little and Thomas (1993) for examples.

This method also produces an estimate of β , and hence of any percentile. Because the estimate of β has poor precision, Dixon and Mood warn the reader not to use this to estimate percentiles. However, in our experience, users sometimes avoid this warning. The reason is simple—no sequential method is generally available to estimate other percentiles. We hope that this paper will solve that problem.

3.2 The Robbins-Munro Procedure

Robbins and Munro (1951) developed a general stochastic approximation method. They want to solve for δ in the equation $F(\delta) = \gamma$, where γ is given, and F is only specified to have a unique solution in the equation. Because $E(Y_x) = F(x)$, we want to find a δ such that $E(Y_\delta) = \gamma$, and this fits into their procedure. Now, Robbins and Munro did not simply want a sequence $\{x_i\}$ that could be used to *find* δ . They wanted $\{x_i\}$ to *converge* in probability to δ . For $\gamma = 0.50$, the R-M procedure looks similar to the up-and-down method:

$$\text{Up-and-Down:} \quad x_{i+1} = x_i + 2d(0.5 - Y_i)$$

$$\text{Robbins-Munro:} \quad x_{i+1} = x_i + a_i(0.5 - Y_i).$$

If the $\{a_i\}$ satisfy $c'/i < a_i < c''/i$ for two constants c' and c'' then Robbins and Munro show that the $\{x_i\}$ will converge in probability to δ . In addition, the

asymptotic variance of this estimate is minimized if $a_i = c/i$, for a certain c that depends on F . For example, $c = \beta\gamma^{-1}(1 - \gamma)^{-1}$ for the logistic case. See Hodges and Lehmann (1955) and the references in Wu (1985) for details.

However, there are problems with this procedure in practice.

1. If c is chosen too small, or if x_1 is too far from δ , the $\{a_i\}$ may shrink too quickly for the $\{x_i\}$ to converge reasonably under modest sample sizes. If x_1 is the culprit, this can be addressed in the $\gamma = 0.50$ case by not shrinking the $\{a_i\}$ until the first change of response type (fail, pass) occurs. See Figure 2 for an example.
2. For general γ , the R-M procedure is $x_{i+1} = x_i + a_i(\gamma - Y_i)$. Unfortunately, convergence is often too slow for the procedure to be of any practical value if γ is not close to 0.5. For example, say $\gamma = 0.25$, we are fortunate enough to start at the correct answer ($x_1 = \delta$), and we are able to use the optimal value of c in the case where F is logistic. If the first run is a failure (with probability equal to 0.25), then the next 31 consecutive values must be passes in order for x to return back up to δ . But the chance of achieving 31 consecutive passes, given that the first run is a failure, is 0.012. For more extreme values of γ , the problem is worsened. This example is based on one from Wetherill (1963), who investigates numerous methods for estimating δ , and suggests that methods for sequential estimation of δ when γ is outside the range (0.25, 0.75) “should be avoided at present.” Our method extends this range considerably.

3.3 Wu’s Sequential-Solving Method

Wu (1985) developed a method that was similar in spirit to the R-M procedure, but is more natural for the binary-data case. His method is illustrated with the logistic model but is asymptotically independent of that model. It is:

1. Collect initial data (Y_i, x_i) , $i = 1, \dots, n$, and use these to obtain estimates, say $(\hat{\alpha}_n, \hat{\beta}_n)$, and hence of the curve, say $\hat{F}_n(x)$.
2. Choose the next setting, x_{n+1} , to solve $\hat{F}_n(x_{n+1}) = \gamma$, that is $x_{n+1} = \hat{\delta}_n$.
3. Either stop, or collect the datum Y_{n+1} , update the estimates of α and β , and repeat step 2.

This method performs better than the R-M procedure, and much better than up-and-down method. However,

1. The performance of the method depends somewhat heavily on the initial n runs.
2. Like the R-M procedure, the method is not designed to estimate the uncertainty of $\hat{\delta}$.

3.4 Likelihood-Based Methods

Minkin (1987), following the work of Abdelbasit and Plackett (1985) and others, examined how to construct likelihood-based confidence regions of minimal area. He examined both a local (quadratic) approximation method and the more difficult method of using the actual likelihood. His results also require reasonably good initial estimates, and so he recommends a two-stage procedure. His results can naturally be extended to a fully sequential setting, but only after good initial estimates are obtained.

3.5 Some Optimal Results in a Non-Sequential Setting.

Using asymptotic variances, and local optimality results (Chernoff(1953)), it can be shown that to minimize $\text{Var}(\hat{\beta})$, data should be taken in two equal groups, at F8.5 and F91.5. Similarly, to minimize the variance of the estimates $(\hat{\alpha}, \hat{\beta})$ jointly by using their generalized variance, data should be taken in two equal groups again, but now at F17.6 and F82.4. See Wetherill (1963), and Abdelbasit and Plackett (1985). Wu (1988) showed the non-intuitive result that, for $\gamma \in (0.083, 0.917)$, the optimum design for estimating δ places all the runs at $F100\gamma$, while for other γ , data should be taken in two equal groups, at F8.3 and F91.7.

The problem with using these ideas in practice, of course, is that we don't know where these x settings are, because they depend on the unknown α and β . The user can take a best guess for such values, but if they are in error, the resulting estimates could be quite poor (Abdelbasit and Plackett (1985)). It seems more reasonable for the design problem to incorporate the natural uncertainty we have about our original guesses, and this leads naturally to Bayesian solutions.

4 Bayesian Strategies for Non-Sequential Experiments

Tsutakawa (1980) considered the problem of how to place the $\{x_i\}$ optimally in a non-sequential setting, using the logistic model to estimate δ for a given γ . To arrive at a solution he made certain assumptions, including that a constant number of runs be made at equally spaced settings, and that the user specify the total number of different settings. Tsutakawa assumed that the prior distribution on δ was normal, on $1/\beta^2$ was gamma, and that δ and β were statistically independent.

In a key paper, Chaloner and Larntz (1989), also using the logistic model, broadened Tsutakawa's results. Using some reasonable approximations, but not restricting themselves to constant number of runs or equally spaced settings, they used Whittle's (1973) version of the general equivalence theorem to generate optimal design strategies. They illustrated their results using uniform priors on α and $1/\beta$ and assuming that α and β were statistically independent. (Their examples, which suggest that the optimal settings depend more on the prior on α than on β , must be interpreted cautiously because the associated priors on β were quite tight.)

5 A Bayesian Strategy for Sequential Experiments: Two Specifications of a Prior.

Our solution to the design problem is to use a Bayesian formulation for sequential experiments. This formulation is a straightforward application of Bayesian principles, applied in a sequential manner. So, for example, the selection of the next x to run can be considered a simple example of, e.g., Chaloner and Larntz. However, we have not seen the formulation appear as such in the literature; we also provide new insights into the problem; we compare this method to other methods; and we strive to look for a workable, user-oriented, method to specify priors. In this section, we begin by specifying two of the four methods for specifying prior distributions.

5.1 A Numerical Priors Method

First consider the location parameter α , the 50th percentile. It is often reasonable to expect that the user has some idea of this value. By asking the user to specify a best guess and a 95% region around this guess, and by assuming the prior distribution of α is normal, we can specify the prior on α .

Developing a prior distribution on the scale parameter β is more naturally more difficult for the user. Because β is a scale parameter and $\beta > 0$, it is more natural to think about its uncertainty in a multiplicative sense. We do this by assuming that $\ln(\beta)$ has a normal prior. Also, instead of asking for information on β directly, we examine a δ , usually the δ of interest in the experiment, e.g. the 10th percentile, and ask the user to estimate how far this value is from α . Then we ask how good this estimate of $\alpha - \delta$ is, in a multiplicative way. For example, the user may be 95% certain that $\alpha - \delta$ ($F_{50} - F_{10}$) is in 6.6×2.0 . For $\gamma = 0.10$, this distance is 2.2β , and so can easily be translated into a prior on $\ln(\beta)$, with mean equal to $\ln(6.6/2.2) = 1.10$ and standard deviation of $\ln(2.0)/2 = 0.35$. We assume that these guesses, and hence the prior distributions, of α and β are statistically independent. We call this method the (α, δ) -numerical method.

It is fundamentally more reasonable for α and β to be independent rather than the possibly more appealing α and δ . This is so for two reasons. The first reason corresponds to why \bar{Y} and a sample quantile $\bar{Y} + zs$ are not statistically independent in the normal case. The second is that it is illogical for a user's prior beliefs on the system to depend on what particular quantile δ is to be estimated, but assuming α and δ are independent forces such a dependency. A discussion on this point is given in Bedrick, et al. (1997). For the same reason, it is illogical for priors on β and $\delta \neq \alpha$ to be independent, although some authors have done this.

Now, the prior distribution on (α, β) creates a prior distribution for δ . Before any runs are made the user can examine this prior graphically and modify his or her beliefs in (α, β) as needed.

We call these priors the *original priors* on α and β . For ease in programming, we have used a 15-point approximation to the normal prior distributions of α and

of $\ln(\beta)$. See Appendix A for details.

5.2 A Graphical Priors Method

In spite of the arguments of the previous subsection, it is easier for the user to think in terms of priors on (α, δ) , instead of $(\alpha, \alpha - \delta)$ for $\delta \neq \alpha$. It is also easier for these specifications to be done graphically. To allow the user to do this, various combinations of prior distributions of (α, δ) can be graphed in a suitable range of interest, where the combinations are created by the numerical-priors method of the previous subsection.

We have found it useful to create $27 = 3 \times 3 \times 3$ combinations of such priors, graphed with the associated priors on (α, δ) . For example, with $\gamma = 0.10$, the graphs are based on:

1. $F50$ best guess: 10.
2. $F50$ 95% limits: $\pm 2, 4, 6$.
3. $F50 - F10$ best guess: 1, 3, 5.
4. $F50 - F10$ 95% limits: $\times \div 2, 4, 6$.

We now illustrate how to use this method.

1. By looking at the graphs (or by other means), say the user is 95% certain that $F50$ is in 10 ± 4 . (If the best guess for $F50$ were 8, then the user would mentally offset the results by 2 or redraw the graphs.)
2. There are 9 graphs whose uncertainty in the $F50$ corresponds to these values—see Figure 3.
3. The user looks for the graph in this figure that most closely matches his or her beliefs of $F10$ and its uncertainty. If the user believed the $F10$ value is near 5, but is very unsure about this value, the graph in the last column and row would be picked. This implies that $F50 - F10 = 5 \times / \div 6$.

We call this second method of specifying priors the (α, δ) -*graphical method*. Before examining the last two methods, we next discuss how to find the settings in an optimal sequential manner.

6 Optimal Setting for the Next x

Assume we have collected data Y_1, \dots, Y_n based on settings x_1, \dots, x_n . (This includes the case of not having any data, for which $n = 0$.) Based on this, we can update the original priors on α and β to form the posterior distribution of α and β (and

hence δ). However, as we look ahead to making our next, $(n + 1)^{th}$, run, it is more natural to call this posterior distribution the *current prior* on α and β . For ease of notation we write the posterior variance of δ after n runs, $\text{Var}(\delta | Y_1, \dots, Y_n)$, simply as $\text{Var}_n(\delta)$, with similar notation for expectation and probability statements.

Our objective will be to try to select a new setting $x = x_{n+1}$ so that if we run there and obtain the result Y_x , then $\text{Var}_n(\delta | Y_x)$ ($= \text{Var}(\delta | Y_1, \dots, Y_n, Y_x)$), the new posterior variance of δ , will be as small as possible. However, we don't know what Y_x is yet, so we can't find $\text{Var}_n(\delta | Y_x)$. Instead, we choose x to minimize the result on average, or $E_n(\text{Var}_n(\delta | Y_x))$. Note that here, the expectation is with respect to Y_x , but is conditioned on the results to date. We call such a setting the *optimal setting* for the next run. So the algorithm is

1. Based on the current prior on (α, β) after n runs, find the setting x^* that minimizes $E_n(\text{Var}_n(\delta | Y_x))$.
2. Set $x_{n+1} = x^*$, make a run there, and obtain Y_{n+1} .
3. Use this information to update the current prior on (α, β) . Either stop the experiment, or return to step 1.

A common Bayesian formulation to decide when to stop an experiment requires creation of a loss function for making incorrect decisions and assumes that a cost is associated with each run. Then the strategy is to find at which point the total risk (posterior expected loss plus total cost) is minimized. For example, see DeGroot (1970, p. 275). However, we avoid this formulation because, in our experience, this is never how scientists and engineers make decisions to stop.

6.1 Insights on the Optimal Setting

Insight into x^* can be obtained by using the relationship

$$E_n(\text{Var}_n(\delta | Y_x)) = \text{Var}_n(\delta) - \text{Var}_n(E_n(\delta | Y_x)).$$

Because $\text{Var}_n(\delta)$ is already determined, we minimize $E_n(\text{Var}_n(\delta | Y_x))$ with respect to x by maximizing $\text{Var}_n(E_n(\delta | Y_x))$. But $\text{Var}_n(E_n(\delta | Y_x))$ is

$$\begin{aligned} & (E_n(\delta | Y_x = 0) - E_n(\delta))^2 P_n(Y_x = 0) \\ & + (E_n(\delta | Y_x = 1) - E_n(\delta))^2 P_n(Y_x = 1). \end{aligned} \tag{3}$$

This value is large when both of these features occur:

1. The new posterior mean of δ —which we use as a point estimate of δ —varies quite a bit depending on whether we obtain a pass or a fail. That is, $E_n(\delta | Y_x = 0)$ is quite different from $E_n(\delta | Y_x = 1)$.

2. Both events $\{Y_x = 0\}$ and $\{Y_x = 1\}$ have a reasonable probability of occurring.

So, our method is equivalent to combining these two natural features in an objective way. We provide an example in the next subsection.

6.2 Two Measures of Information

Call $1/\text{Var}_n(\delta)$ the *information* about δ after n runs, where this term is meant to suggest a similarity to the Fisher information. An equivalent way to find x^* is the x that maximizes

$$AII_\delta(x) \equiv 1/\text{E}_n(\text{Var}_n(\delta|Y_x)) - 1/\text{Var}_n(\delta),$$

where AII stands for *anticipated increase in information* and we have suppressed the dependence of AII on n .

Plotting $AII_\delta(x)$ versus x indicates how sensitive our results are to the choice of the next x . It also provides insight into what additional information we should expect to obtain by collecting one more data point. We provide examples of AII plots in the next section.

To understand the notion of information and AII better, consider the Bayesian case of sampling data from a normal distribution with a known variance σ^2 . If the original prior distribution of μ is $N(\nu, \tau^2)$, then after n data points the posterior distribution of μ has variance

$$\frac{1}{1/\tau^2 + n/\sigma^2}.$$

So, the original information is $1/\tau^2$, the anticipated (and actual) increase in information for each run is $1/\sigma^2$, and the total information is $1/\tau^2 + n/\sigma^2$. A plot of such information versus run number yields a straight line with intercept equal to the original information, slope equal to additional information per run, and endpoint the total information. So, in addition to the AII plots, we will plot the information on δ (and, e.g., α , β , ...) after n runs versus n . These *information plots* give additional insight into our experiments.

7 A Simple Example

We use simple priors on α and β here, to show more clearly how the Bayesian strategy works. Our objective will be to find the δ corresponding to $\gamma = 0.10$. For reference on the calculations, see Appendix B.

Suppose our original prior on α has values (8, 9, 10) with corresponding probabilities (0.25, 0.50, 0.25) and our original prior on β has values (1, 2, 3) with probabilities (0.25, 0.50, 0.25). This generates a prior for $\delta = \alpha - 2.2\beta$, along with an initial estimate, $E_0(\delta)$, of 4.6—See Table 1.

To find the best setting at which to make the first run, the *AII* for a variety of x settings is graphed in Figure 4 and partly tabled in Table 2. From the Figure, we can easily see that

1. The optimal setting for the next run is approximately $x = 6$.
2. Running at $x = 10$ would be useless.
3. Running near $x = 2$ or $x = 8$ would provide only half as much *AII* as running at $x = 6$.
4. Some information can be obtained by running near 13 or 14, but not enough to consider.

Now, turning our attention to Table 2, and recalling (3), we see that

1. At $x = 6$, our new estimate of δ would change quite a bit, from 4.79 to 3.75, depending on whether the run there yields a pass or a fail, and that there is a reasonable chance, 0.82 and 0.18, for either of these events to happen.
2. At $x = 10$, our new estimate of δ would barely change at all, from 4.58 (pass) to 4.62 (fail), depending on the next run's results. Such a run is information-poor.
3. At $x = 2$, say, our new estimate of δ would change very much, from 4.66 (pass) to 3.14 (fail). However, it is so unlikely to produce a fail (0.04) that this run is also information-poor.

On the other hand, if all priors were the same but we were interested in finding the δ for which $\gamma = 0.50$, the *AII* curve would be very different—see Figure 5. Hence, the user needs to give careful consideration to the choice of γ , or more generally the objective function, in the experiment.

To see the strategy in action, we simulated some data. We selected $(\alpha, \beta) = (8.00, 1.82)$, which makes $\delta = 4.0$, and examined x values from 0 to 20 in increments of 1. A simulation at $x_1 = 6$ produced $Y_1 = 0$, a pass. Using Bayes' Theorem to update the priors of α and β , and hence δ , yields the results in column (2) of Table 3. These now become the current priors of α and β .

Repeating the process for 13 runs produced the results shown in columns (3), (4), and (5) of Table 3. Continuing for 40 runs produced results that are graphed in Figure 6. A close examination of this figure indicates that failures tend to make the routine search at lower x values, similar to Figures 1 and 2.

To give the reader a sense of the variability of these results, we ran a second simulation—see Figure 7. Note that at runs 14, 15, and 16, we obtained 3 consecutive failures, all at $x = 4$, for which the probability of each failure is only 0.10. As a result, our estimates of δ were quite low. Also note that at run 23, the routine

lowered the value of x at which to run to 3, even though passes occurred in the immediately preceding runs. This indicates that the method employs a different strategy than either the R-M or the Wu procedure.

8 A Realistic Example—Flatter Priors.

To illustrate the technique in a more realistic setting, say the user's estimates of uncertainty are $F50 = 10 \pm 4$ and $F50 - F10 = 5 \frac{\times}{\div} 6$. This corresponds to the initial prior on δ in the lower right-hand graph of Figure 3. We simulated data with $F50 = 8$, $F50 - F10 = 4$. The results of 60 runs of this simulation are shown in Figures 8 and 9. Here,

1. The flatter prior causes the routine to search around more for interesting features—it is more data-driven than our previous example.
2. Each fail causes the routine to select a lower value of x to run next. Initially, these jumps in x were quite large, but decreased as more information was obtained. Search increments of size 1 for x were used here.
3. Each pass causes the routine to stay at the current x or move up, but the movements up are more conservative than the movements down. This is analogous to the R-M procedure.
4. Usually, more information could be gained by running the next x at a slightly lower value than the current estimate of δ . This should be compared to Wu's method, for which this next x equals the current estimate of δ . But this result does not always hold. For example, if we had used the original prior distribution shown in the upper left-hand corner of Figure 3, the initial estimate of δ would be 8.9, but the best first run would be at $x = 10$.
5. The *AII* curve starts off quite flat—with such a poor initial idea of where δ is, many settings will help to narrow down the range of reasonable δ values. However, as information is accumulated the optimal settings become more well-defined. Also, the optimal *AII* at these later settings is larger.

This simulation produced the information plot shown in Figure 10. Our choice of these three plotted features is based on the relation $\delta = \alpha - 2.2\beta$. From this plot, it is apparent that the strategy used in this experiment was to estimate δ better by learning more about the parameter β , but not α . We discuss the importance of this point in the next section.

The decrease in information seen in the region between runs 28 and 35 occurred because there were 3 failures in 8 runs—see Figure 8—that occurred at x values that were expected to produce passes. This “surprised” the routine, in the same way that

collecting an additional data point to estimate the mean μ in the one-sample normal-distribution case with unknown variance can produce a larger standard error if the addition data point is very inconsistent with previous results.

The information about δ at run 0 is very small compared to the total information after 60 runs. So, the estimate of δ is based virtually entirely on the data collected. Of course, the original priors on α and β needed to be broad enough to be consistent with the process that is actually generating the data.

9 Two Other Methods of Specifying Priors

9.1 Equal-Contribution Prior Method

We return to the information plot in Figure 10. For this problem, $\delta = \alpha - 2.2\beta$, and because our original priors on α and β are independent, the original variance of δ is

$$\begin{aligned} \text{Var}_0(\delta) &= \text{Var}_0(\alpha) + 4.84 \text{Var}_0(\beta) \\ &= 3.92 + 4.84(12.18) = 3.92 + 58.96 \\ &= 62.88. \end{aligned}$$

This indicates why the routine tried to increase its information on δ through β . Although this decomposition of variances cannot be used later on (for example, $\text{corr}_n(\alpha, \beta) = 0.80$ after $n = 60$ runs for this simulation), it suggests a third method to produce an initial prior. The method is to *construct the initial prior such that the contributions of variance due to α and $\text{logit}(\gamma)\beta$ are equal*. We call this the *equal-contribution priors method*. In this case, if we wanted to keep $\text{Var}_0(\delta) = 62.88$, we would need $\text{Var}_0(\alpha) = 31.44$, $\text{Var}_0(\beta) = 6.50$. To specify such priors, the user would need to specify only three pieces of information, e.g. best guess for α , 95% region around α , and best guess for δ . Graphs analogous to Figure 3 could be used here.

To make a direct comparison with the first simulation of the flatter prior, we next examined an original prior on α of 10 ± 11.21 , an initial guess of δ of 5, and an original prior on $\ln(\beta)$ such that $\text{Var}_0(\beta) \cong 6.50$. The information plot of another simulation of 60 runs, again using $(\alpha, \beta) = (8, 1.28)$, is shown in Figure 11(a). This shows that our design gains information on δ , but at the expense of other features of the response curve. For more resolution, we have replotted these results for only α and 2.2β in Figure 11(b). This strongly suggests that our revised strategy was successful, as now information is sought on both α and β . A third simulation, not shown, shows a similar information plot.

However, the equal-contribution-priors strategy, although illuminating, has flaws that render it unacceptable in practice. First, it forces the prior distribution to depend on the γ of interest, which we have previously argued is illogical. Second, it ignores the fact that the γ of interest *should* affect the relative interest in learning more about α and β . For example, at $\gamma \cong 0.50$, little information needs to be

obtained on β —in fact, the equal-contribution-priors strategy fails at $\gamma = 0.50$. So we consider a fourth method for generating priors.

9.2 Relative Priors Method

To understand the method we are about to propose, suppose for a moment that the tolerance distribution is normal. Imagine that the user has (mentally) collected m X_i 's from the tolerance distribution. From this, we know that $\text{Var}(\bar{X}) = \sigma^2/m$ and (Evans, et al.(1993), p. 49)

$$\text{Var}(s) = \frac{\sigma^2/2}{m-1} \left[2m - 2 - 4 \left(\frac{\Gamma(m/2)}{\Gamma((m-1)/2)} \right)^2 \right].$$

Now, the term in the brackets converges quickly to 1 as m increases. For example, at $m = 10$, the value in brackets is 0.97. So $\text{Var}(s) \cong \sigma^2(1/2(m-1))$, the Taylor series approximation based on $\text{Var}(s^2) = \sigma^4(2/(m-1))$. Similarly, for reasonable sized m , we replace $m-1$ by m and find that

$$\frac{\text{sd}(s)}{\text{sd}(\bar{X})} \cong 1/\sqrt{2} \cong 0.7.$$

For $m = 10$, the actual values of this ratio are $1.04/\sqrt{2}$. Based on this, we suggest for the normal tolerance distribution that the user only specify best guesses for μ and δ and an estimate of uncertainty for μ . Then the best guess for σ is obtained directly and the estimate of uncertainty for σ is simply 0.7 times the uncertainty estimate of μ . We call such a method of specifying priors the *relative-priors method*. Finally, we can translate this into a symmetric prior on $\ln(\sigma)$. Let $\lambda = (m-1)/2$. Then $\text{Var}(\ln(s)) = (1/4)d^2 \ln(\Gamma(\lambda))/d\lambda^2$, where “ $\Gamma(\cdot)$ ” is the gamma function (based on Johnson and Kotz (1970), p. 196). The Taylor series approximation is $\text{Var}(\ln(s)) \cong 1/(2(m-1))$. The ratio of the Taylor approximation to the actual value at $m = 10$ is 1.12, so again the approximation is good for reasonable sized m .

Returning to the logistic tolerance distribution, assume the method of moments has been used to summarize the (mentally) collected set of m X_i 's. Then, $\text{Var}(\bar{X}) = \text{Var}(X_1)/m$ and, letting γ_2 be the coefficient of kurtosis,

$$\begin{aligned} \text{Var}(s^2) &= \text{Var}(X_1)^2 \left(\frac{2}{m-1} + \frac{\gamma_2}{m} \right) \\ &= \text{Var}(X_1)^2 \left(\frac{2}{m-1} + \frac{1.2}{m} \right), \end{aligned} \tag{4}$$

where the first equality is from Scheffé (1959, p. 83), and $\gamma_2 = 1.2$ is from Evans, et al. (1993, p. 99). Then, to a Taylor series approximation, and replacing $m-1$

by m in the equation again,

$$\frac{\text{sd}(s)}{\text{sd}(\bar{X})} \cong \sqrt{0.8} \cong 0.9.$$

In fact, for any location-scale tolerance distribution whose fourth moment exists, (4) indicates that the relative magnitudes of $\text{sd}(s)$ and $\text{sd}(\bar{X})$ depend on the tolerance distribution but have little dependence on m as long as m is not too small. In addition, it is easy to show that s^2 and \bar{X} are uncorrelated as long as third moments exist and the tolerance distribution is symmetric. So, in this sense, it is reasonable to consider our priors on the location and scale parameters to be independent.

Using this result for our problem, we again we will force $\text{Var}_0(\delta) = 62.88$ for comparison purposes. Then the original-prior variances are $\text{Var}_0(\alpha) = 12.91$ and $\text{Var}_0(\beta) = 10.33$. So we now use an original prior α of 10 ± 7.19 , an initial guess of δ of 5, and an original prior on $\ln(\beta)$ such that $\text{Var}_0(\beta) \cong 10.33$. Results of the corresponding fourth simulation are shown in Figures 12 and 13.

We believe that the relative-priors strategy has an *a priori* justification, should lead to reasonable design strategies, and *makes it as simple as possible for the user to specify reasonable priors*. For these reasons, *this is the strategy we recommend for constructing original priors*. Finally, from a practical point of view, we point out to the user that it is far better to overstate the original uncertainty in the priors, because this allows the design strategy to be more data driven.

10 Analysis Strategies

The main purpose of this article has been to show an optimal strategy in the design stage. We now briefly discuss the analysis of such designs. It is clearly reasonable that prior knowledge be incorporated into the design of sensitivity experiments using Bayesian methods. However, it can be argued that the analysis of such experiments should not depend on these priors. So, the likelihood function could be used both to provide an estimate of δ and to provide an interval of likely values, using the profile likelihood function.

Even in the Bayesian framework, we have other choices for estimates besides $E_n(\delta) \pm 1.96\sqrt{\text{Var}_n(\delta)}$. We could calculate the 2.5%, 50%, and 97.5% points of the posterior distribution of δ . We could also estimate δ by calculating the x for which $P_n(Y_x = 1) = \gamma$.

To give a sense of the how these methods compare, we used the four simulations with flatter priors—see Table 4. For point estimation, all methods agree well. For interval estimation, the two Bayesian methods compare well, but the likelihood method yields shorter intervals. In fact for second simulation, this interval misses the actual value of δ . This is not too surprising, because likelihood intervals are often a somewhat narrow in their coverage probabilities. Other examples, not presented here, lead to the same conclusions.

11 Recommendations

A Bayesian method of designing sequential sensitivity tests is attractive from both a theoretical and practical position. We recommend that the method of relative-priors specification, along with ensuring that the user has not understated the uncertainty, be used to obtain the original priors. The *AII* and information plots can be used to provide insight into the experiment as data are collected.

In the analysis stage, either Bayesian methods or likelihood methods can be used. This would depend in part upon the audience to whom the results must be presented.

These methods have been successfully used in the proprietary application for which they were originally developed.

A Discrete Approximation to the Normal Distribution

Our approximation was arbitrary but reasonable. We found the 16 z_j from the $N(0, 1)$ distribution that solved $\Phi(z_j) = 0, 1/100, 2/100, 1/11, 2/11, \dots, 10/11, 98/100, 99/100, 1$. Then our 15 discrete values were $E_j = E(Z | z_j < Z < z_{j+1})$, where $Z \sim N(0, 1)$. For example, the fourth value was $E(Z | \Phi^{-1}(z_4) < Z < \Phi^{-1}(z_5)) = E(Z | -1.33518 < Z < -0.90846) = -1.10191$. We attached weights to these values corresponding to $\Phi(z_{j+1}) - \Phi(z_j)$, i.e., $1/100, 1/100, 1/11-2/100, 1/11, \dots, 1/11, 1/11-2/100, 1/100, 1/100$. The resulting symmetric probability distribution has a mean of 0 and a variance of 0.98.

More sophisticated methods could have been used—see the discussion in D’Errico and Zaino (1988), for example, or direct numerical integration, or Markov Chain Monte Carlo (Tanner (1996), e.g.)—but the vagueness in setting the prior and the little information it adds to the final results suggest that such methods would not change the basic results of this article.

An issue here is how to modify the discretized distribution after a large number of runs have been collected. For example, in the relative-priors simulation, the original prior on δ had 95% of its probability placed on 152 of the 225 combinations. But after 60 runs, 95% of the probability was placed only on 29 of these combinations. It would be useful to smooth this distribution out over a finer grid in this new region of interest to make it more continuous-like. However, this must be done carefully—for example, any new grid of values should at least preserve the first and second moments of α and $\ln(\beta)$ and their covariance structure. We have not investigated how this might easily be done.

B Summary of Calculations

Running the routine requires 4 steps:

1. Create initial information, such as the δ of interest and the original priors on α and β .
2. Find the x setting that maximizes $AI I_\delta(x)$.
3. Collect a datum there, either a pass or fail.
4. Use this to update the priors on α and β . Summarize the information on δ . Either stop collecting data, or return to step 2.

B.1 Create Initial Information

1. Create an N_α -point original-prior distribution for α . Place the values of this distribution into the vector $\tilde{\alpha}$ and the probabilities into \mathbf{p}_α . Do the same for the original-prior distribution for β , yielding N_β , $\tilde{\beta}$, and \mathbf{p}_β . These may be done using the method in Appendix A, or by other means. Note that as these distributions are updated, the values for α and for β will remain the same, but their probabilities will change. (Actually, it is the probabilities of the joint distribution that will concern us.) Because β is a scale parameter, we choose the values in $\tilde{\beta}$ to be all positive in line with comments from Section 2 as well as Item 5 below.
2. Set $\boldsymbol{\alpha} \leftarrow \mathbf{1}_{N_\beta \times 1} \otimes \tilde{\alpha}$, $\boldsymbol{\beta} \leftarrow \tilde{\beta} \otimes \mathbf{1}_{N_\alpha \times 1}$, and $\mathbf{p} \leftarrow \mathbf{p}_\beta \otimes \mathbf{p}_\alpha$ where “ \otimes ” is the Kronecker product, and set $N \leftarrow N_\beta N_\alpha$. This stores the joint distribution of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in \mathbf{p} . (The original priors of α and β need not be independent as we have assumed here—any joint distribution of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ could be entered into \mathbf{p} .) The values in \mathbf{p} will be updated as data are collected, so \mathbf{p} is associated with the joint distribution $P_n(\alpha, \beta)$.
3. Define the value γ of interest, and set $c \leftarrow \ln(\gamma/(1-\gamma))$. Set $\boldsymbol{\delta} \leftarrow \boldsymbol{\alpha} + c\boldsymbol{\beta}$. The distribution of δ is given by the values of $\boldsymbol{\delta}$ and \mathbf{p} .
4. Find the mean and variance of δ , as $E\delta \leftarrow \boldsymbol{\delta} \cdot \mathbf{p}$, where “ \cdot ” is the dot product, and $V\delta \leftarrow (\boldsymbol{\delta} - E\delta)^2 \cdot \mathbf{p}$.
5. Decide whether higher or lower x values will correspond to a greater chance of producing a failure. If higher x values, then in an abusive but obvious notation associate the events “ $Y_x = 1$ ” with “ $Y_x = F$ ” and “ $Y_x = 0$ ” with “ $Y_x = P$ ”. Otherwise, do the opposite. Note in both cases that $P(Y_x = 1)$ increases with x . This is done only to increase the readability of the results.

B.2 Find the x setting that maximizes $AI I_\delta(x)$

We usually maximize $AI I_\delta(x)$ in our work by defining it as a function of x , calculating it for a range of x values (e.g., $x = 2$ to 20 in steps of 0.5). Occasionally,

we have instead used an algorithm to find the optimal x more precisely. Here, we simply define $AII_\delta(x)$, and let the user decide the method of finding x^* .

1. Define the function $P(Y_x = 1 | \alpha, \beta)$ as

$$P1ab(x, \alpha, \beta) \equiv 1 / (1 + \exp(-(x - \alpha) / \beta))$$

Let $P1ab(x, \alpha, \beta)$ be the $N \times 1$ function vector of these values.

2. Define the function associated with $P_n(Y_x = 1)$ as $P1(x) \equiv P1ab(x, \alpha, \beta) \cdot \mathbf{p}$. Define $P0(x) \equiv 1 - P1(x)$
3. Define the function associated with $E_n(\delta | Y_x = 1)$ as $Ed1(x) \equiv (\boldsymbol{\delta} * P1(x, \alpha, \beta)) \cdot \mathbf{p} / P1(x)$, where “ $*$ ” is the direct product. Similarly, define the function associated with $E_n(\delta | Y_x = 0)$ as

$$Ed0(x) \equiv (\boldsymbol{\delta} * (1 - P1(x, \alpha, \beta))) \cdot \mathbf{p} / P0(x).$$

4. Define the function associated with $E_n(\text{Var}_n(\delta | Y_x))$ as

$$EVd(x) \equiv Vd - (Ed1(x) - Ed)^2 P1(x) - (Ed0(x) - Ed)^2 P0(x).$$

5. Define the function associated with $AII_\delta(x)$ as $AII(x) \equiv 1 / EVd(x) - 1 / Vd$.
6. Find, through a search of discrete values of x , or by an optimization method, the value of x that maximizes $AII(x)$. Denote this value by $xstar$. Note that the $AII(x)$ function may have more than one local maximum, as Figure 9 indicates, so any optimization method should take this into consideration. (Based on the nature of location-scale tolerance distributions, and upon our experience, we believe that $AII_\delta(x)$ will have at most two local maxima.)

B.3 Update the priors on α and β and summarize δ .

We assume that the new datum Y_x , here denoted as Y , has been collected at $xstar$. (If not, just modify the following formulas in the obvious way). So we need to update \mathbf{p} .

If $Y = 1$, then

$$\mathbf{p} \leftarrow P1(xstar, \alpha, \beta) * \mathbf{p} / P1(xstar).$$

If $Y = 0$, then

$$\mathbf{p} \leftarrow (1 - P1(xstar, \alpha, \beta)) * \mathbf{p} / P0(xstar).$$

If the Bayes method used in the simulations will be used for analysis purposes, find $Ed \leftarrow \boldsymbol{\delta} \cdot \mathbf{p}$, and $Vd \leftarrow (\boldsymbol{\delta} - Ed)^2 \cdot \mathbf{p}$.

References

- Abdelbasit, K. M., and Plackett, R. L. (1983), “Experimental Designs for Binary Data,” *Journal of the American Statistical Association*, 78, 90–98.
- ASTM (1995), “Standard Test Methods for Impact Resistance of Plastic Film by the Free-Falling Dart Method,” Annual Book of ASTM Standards, Vol. 08.01 Plastics (I), 378–384, ASTM, Philadelphia.
- Bedrick, E. J., Christensen, R., and Johnson, W. (1997), “Bayesian Binomial Regression: Predicting Survival at a Trauma Center,” *American Statistician*, 51, 211–218.
- Brownlee, K. A., Hodges, Jr., J. L., and Rosenblatt, M. (1953), “The Up-and-Down Method with Small Samples,” *Journal of the American Statistical Association*, 48, 262–277.
- Chaloner, K., and Larntz, K. (1989), “Optimal Bayesian Design Applied to Logistic Regression Experiments,” *Journal of Statistical Planning and Inference*, 21, 191–208.
- Chernoff, H. (1953), “Locally Optimal Designs for Estimating Parameters,” *Journal of the American Statistical Association*, 24, 586–602.
- DeGroot, M. (1970), *Optimal Statistical Decisions*, McGraw-Hill, New York.
- D’Errico, J. R., and Zaino, N. A., Jr. (1988), “Statistical Tolerancing Using a Modification of Taguchi’s method,” *Technometrics*, 30, 397–405.
- Dixon, W. J., and Mood, A. M. (1948), “A Method for Obtaining and Analyzing Sensitivity Data,” *Journal of the American Statistical Association*, 43, 109–126.
- Dixon, W. J. (1965), “The Up-and-Down Method for Small Samples,” *Journal of the American Statistical Association*, 60, 967–978.
- Evans, M., Hastings, N., and Peacock, B. (1993). *Statistical Distributions*, 2nd edition, Wiley, New York.
- Hodges, J. L., and Lehmann, E. L. (1955), “Two Approximations to the Robbins-Munro Procedure,” *Proceedings of the 3rd Berkeley Symposium*, 1, 95–104.
- Johnson, N., and Kotz, S. (1970), *Continuous Univariate Distributions-1*, Wiley, New York.
- Little, R. E. (1975), *Tables for Estimating Median Fatigue Limits*, ASTM STP 731, American Society for Testing and Materials, Philadelphia.

- Little, R. E. (1990), “Optimal Stress Amplitude Selection in Estimating Median Fatigue Limits Using Small Samples,” *Journal of Testing and Evaluation*, 18, 115–122.
- Little, R. E., and Jebe, E. H. (1975), *Statistical Design of Fatigue Experiments*, Halstead Press, New York.
- Little, R. E., and Thomas, J. J. (1993), “Up-and-Down Methodology Applied to Statistically Planned Experiments,” *Journal of Testing and Evaluation*, 21, 14–20.
- Minkin, S. (1987), “Optimal Designs for Binary Data,” *Journal of the American Statistical Association*, 82, 1098–1103.
- Robbins, H., and Munro, S. (1951), “A Stochastic Approximation Method,” *Annals of Mathematical Statistics*, 22, 400–407.
- Sitter, R. A., and Wu, C. F. J. (1999), “Two-Stage Design of Quantal Response Studies,” *Biometrics*, 55, 396–402.
- Tanner, M. A. (1996), *Tools for Statistical Inference. Methods for the Exploration of Posterior Distributions and Likelihood Functions*, 3rd edition, Springer, New York.
- Tsutakawa, R. K. (1980), “Selection of Dose Levels for Estimating a Percentage Point of a Logistic Quantal Response Curve,” *Applied Statistics*, 29, 25–33.
- Wetherill, G. B. (1963), “Sequential Estimation of Quantal Response Curves” (with discussion), *Journal of the Royal Statistical Society, Series B*, 25, 1–48.
- Whittle, P. (1973), “Some General Points in the Theory of Optimal Experimental Design,” *Journal of the Royal Statistical Society, Series B*, 35, 123–130.
- Wu, C. F. J. (1985), “Efficient Sequential Designs with Binary Data,” *Journal of the American Statistical Association*, 80, 974–984.
- Wu, C. F. J. (1988). “Optimal Design for Percentile Estimation of a Quantal Response Curve,” in *Optimal Design and Analysis of Experiments* (eds. Y. Dodge, V. Federov, H. P. Wynn), 213–223, Elsevier, Amsterdam.
- Zocchi, S. S., and Atkinson, A. C. (1999), “Optimum Experimental Designs for Multinomial Logistic Models,” *Biometrics*, 55, 437–444.

Table 1. Example 1. Original Prior of α and β with Corresponding δ Values

α	β	δ	Prob
8	1.0	5.8	0.0625
8	2.0	3.6	0.1250
8	3.0	1.4	0.0625
9	1.0	6.8	0.1250
9	2.0	4.6	0.2500
9	3.0	2.4	0.1250
10	1.0	7.8	0.0625
10	2.0	5.6	0.1250
10	3.0	3.4	0.0625

Table 2. Example 1. Calculating the AII versus x after 0 Runs.

$E_0(\delta) = 4.606$, $\text{Var}_0(\delta) = 2.914$. (Pass \equiv 0, Fail \equiv 1)

x	$P_0(Y_x = 1)$	$E_0(\delta Y_x = 0)$	$E_0(\delta Y_x = 1)$	$E_0(\text{Var}_0(\delta Y_x))$	$AII_\delta(x)$
0	0.02	4.64	2.94	2.863	0.006
2	0.04	4.66	3.14	2.828	0.010
4	0.08	4.71	3.40	2.783	0.016
6	0.18	4.79	3.75	2.758	0.019
8	0.37	4.82	4.24	2.837	0.009
10	0.63	4.58	4.62	2.914	0.000
12	0.82	4.13	4.71	2.865	0.006
14	0.92	3.77	4.68	2.852	0.007
16	0.96	3.52	4.65	2.867	0.006
18	0.98	3.32	4.63	2.883	0.004

Table 3. Distribution of α , β , δ after 0, 1, 10, 12, and 13 Runs.

α	β	δ	(1) Prob after 0 runs	(2) Prob after 1 run at X=6	(3) Prob after 9 more runs at X=6	(4) Prob after 2 more runs at X=6	(5) Prob after 1 more run, at X=5
				P	PPPPP PPPP	FF	F
8	1	5.8	0.0625	0.0668	0.0754	0.0777	0.0287
9	1	6.8	0.1250	0.1446	0.3300	0.0538	0.0076
10	1	7.8	0.0625	0.0745	0.2237	0.0052	0.0003
8	2	3.6	0.1250	0.1110	0.0234	0.1227	0.1747
9	2	4.6	0.2500	0.2482	0.1432	0.3455	0.3214
10	2	5.6	0.1250	0.1337	0.1508	0.1553	0.0920
8	3	1.4	0.0625	0.0501	0.0043	0.0355	0.0745
9	3	2.4	0.1250	0.1110	0.0234	0.1227	0.1997
10	3	3.4	0.0625	0.0601	0.0258	0.0816	0.1011
$E(\delta) \pm 2\sigma_\delta$			4.6 \pm 3.4	4.8 \pm 3.4	6.2 \pm 2.8	4.4 \pm 2.6	3.8 \pm 2.5

Table 4. Flatter Prior, Simulation 1, 2, 3, and 4.

Comparison of Three Methods for Point and 95% Interval Estimation of δ after 60 Runs. Actual Value of $\delta=4.0$.

Simulation		$E(\delta) \pm 1.96\sigma^1$	MLE/profile ²	Posterior of δ^3	$P(Y_x=1)^4$
1	Estimate	3.8	3.7	3.8	3.8
	Lower	1.4	1.7	1.1	
	Upper	6.1	5.8	5.9	
2	Estimate	2.1	2.2	2.2	2.1
	Lower	-0.5	0.1	-0.1	
	Upper	4.7	3.8	4.5	
3	Estimate	4.3	4.4	4.4	4.3
	Lower	2.2	2.8	2.0	
	Upper	6.4	5.6	6.3	
4	Estimate	4.5	4.6	4.5	4.5
	Lower	2.7	3.4	3.1	
	Upper	6.3	5.6	5.8	

¹Posterior mean, and ± 1.96 posterior standard deviations.

²MLE, and profile likelihood values 1.92 units from maximum likelihood.

³Posterior median, and 2.5% and 97.5% posterior percentiles.

⁴Solution x to equation $P(Y_x=1)=0.10$.

Figure 1. Example of the Up-Down Method.

○ Indicates a Pass, ● indicates a Fail. Increment $d=0.5$

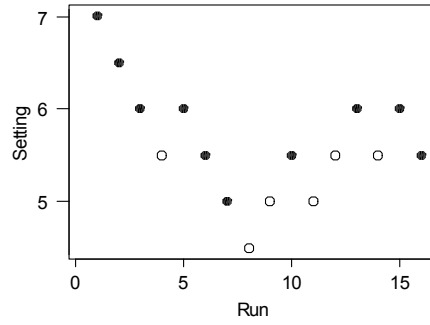


Figure 2. Example of modified R-M Procedure.

No shrinking is done until first change of response type.

○ Indicates a Pass, ● indicates a Fail.

First 4 values correspond to Up-and-Down method data set, $d=0.5$.

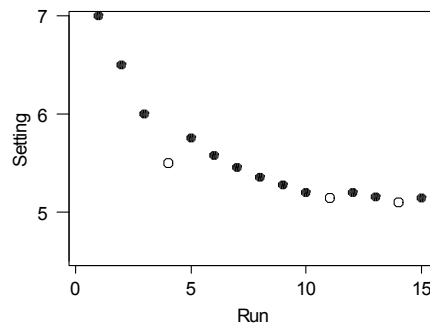
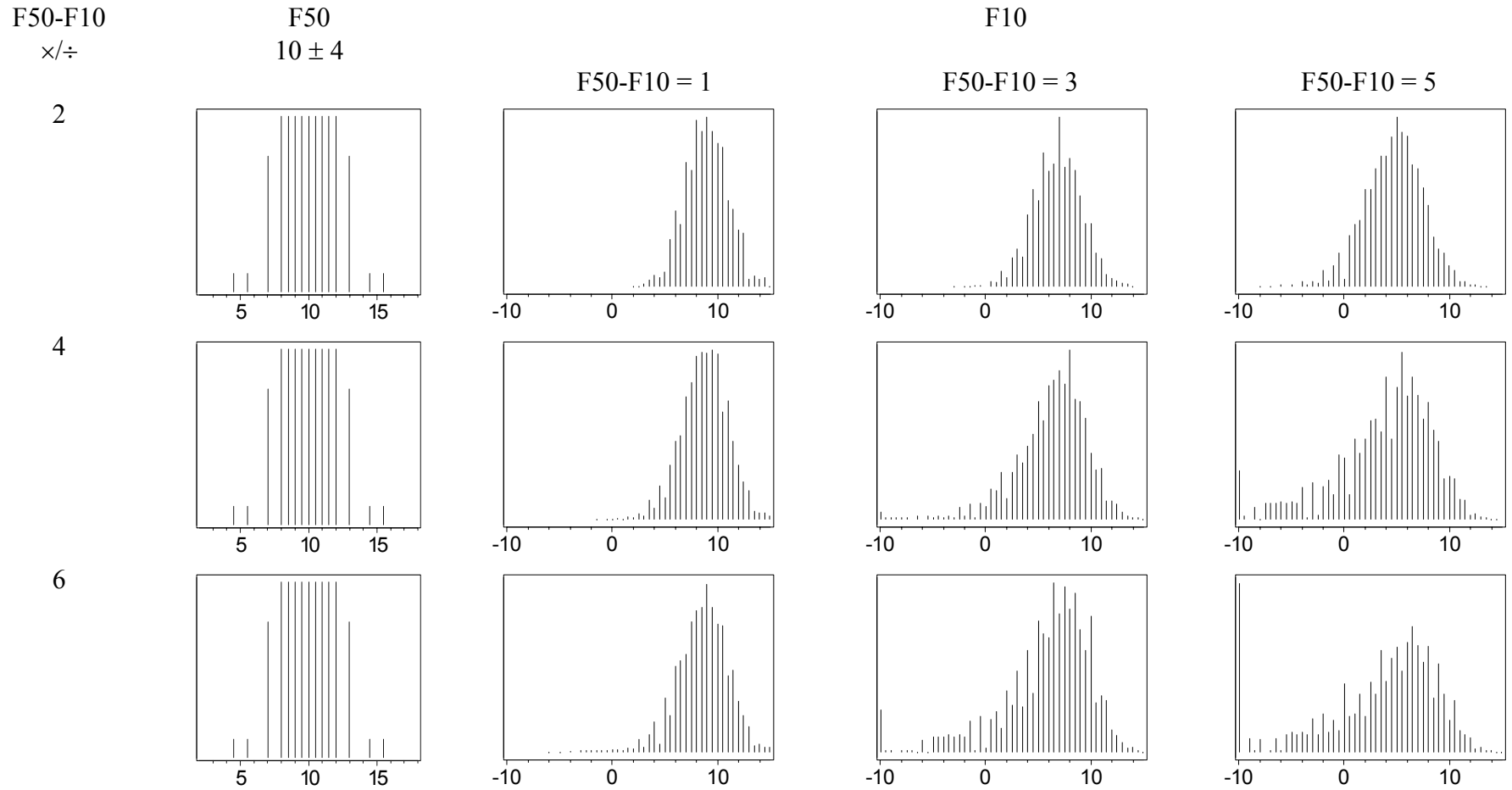


Figure 3. Example of Prior Distributions of F10.



For resolution, F10 values less than -10 (greater than 15) were graphed as -10 (15).

Figure 4. Simple Prior, Simulation 1. All versus x after 0 Runs, for $\gamma=0.10$.

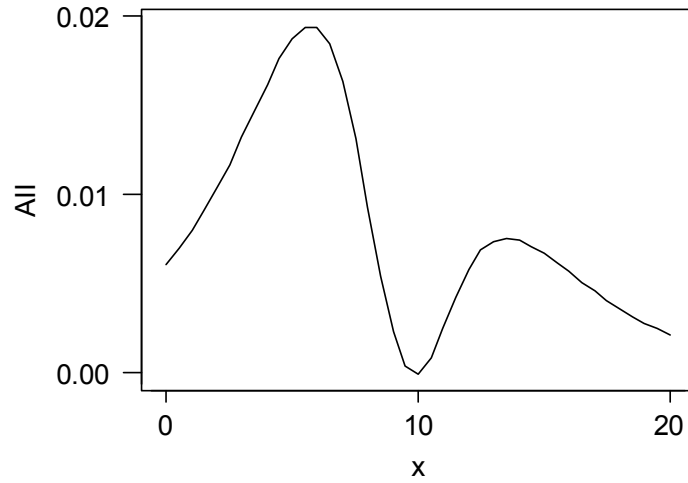


Figure 5. Simple Prior, Simulation 1. All versus x after 0 Runs if Estimation was for $\gamma=0.50$

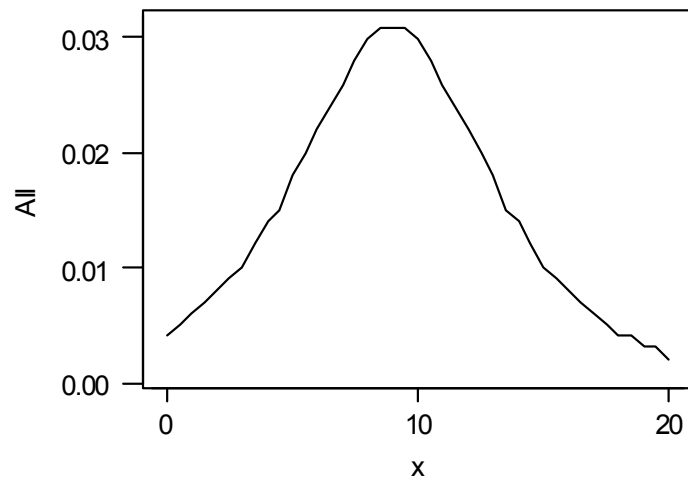


Figure 6. Simple Prior, Simulation 1. Results of 40 Runs.

The x's indicate settings. Circles are $E(\delta)$. Dashed line is actual value of δ .
 ○ indicates a Pass, ● indicates a Fail, ⊕ indicates initial value.

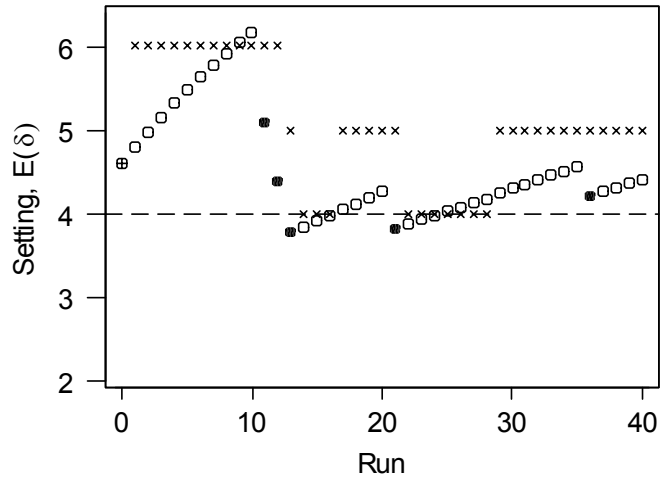


Figure 7. Simple Prior, Simulation 2. Results of 40 Runs.

The x's indicate settings. Circles are $E(\delta)$. Dashed line is actual value of δ .
 ○ indicates a Pass, ● indicates a Fail, ⊕ indicates initial value.

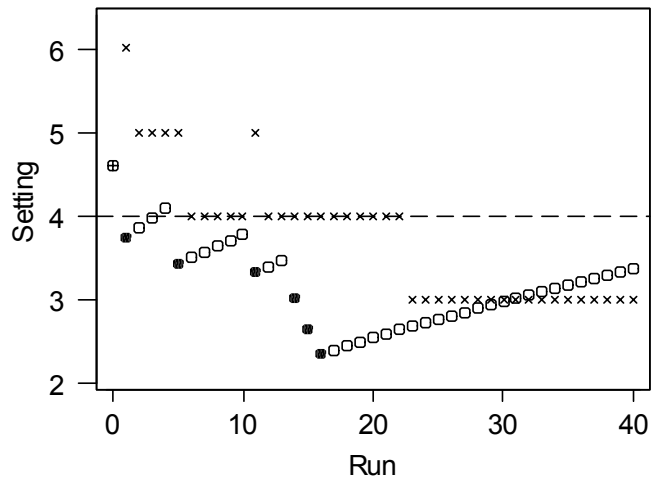


Figure 8. Flatter Prior, Simulation 1. Settings and $E(\delta)$ for 60 Runs.

The x's indicate settings. Circles are $E(\delta)$. Dashed line is actual value of δ . \circ indicates a Pass, \bullet indicates a Fail, \oplus indicates initial value.

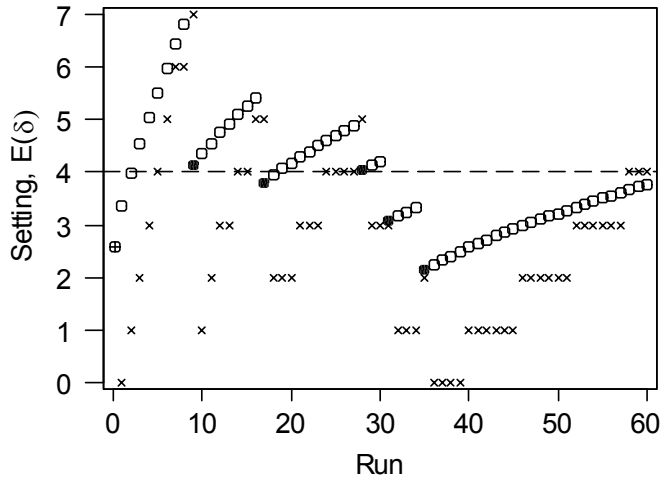


Figure 9. Flatter Prior, Simulation 1. AII versus Possible Settings, after 0, 2, 10, 20, and 60 Runs.

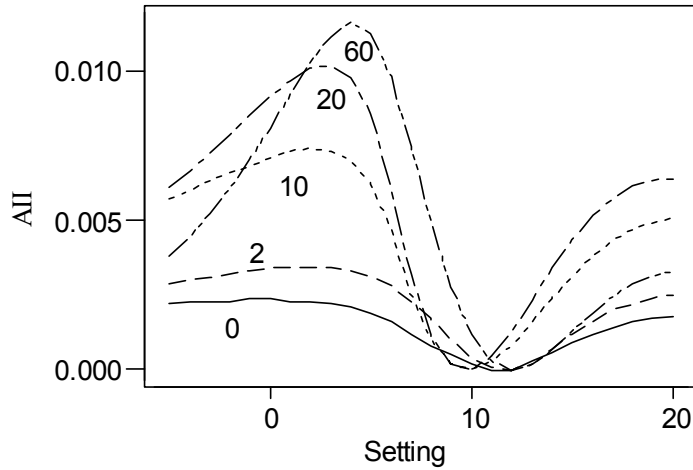


Figure 10. Flatter Prior, Simulation 1.
Information on α , 2.2β , and $\delta=\alpha-2.2\beta$, versus Run Number.

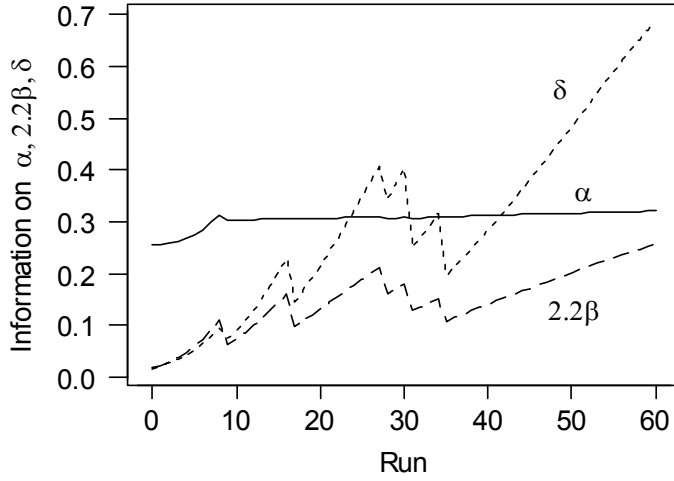


Figure 11. Flatter Prior, Simulation 2 (Equal-Contribution Priors on α and 2.2β).

(a) Information on α , 2.2β , and $\delta=\alpha-2.2\beta$, versus run number. (b) Information on α and 2.2β versus run number.

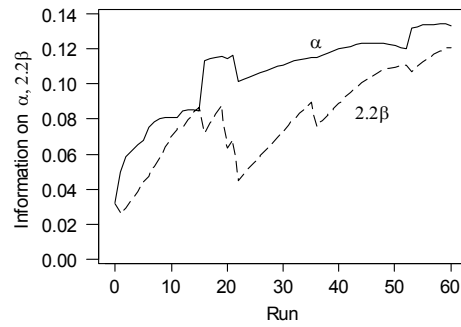
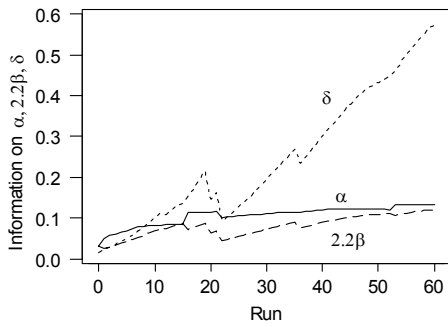


Figure 12. Flatter Prior, Simulation 4—Relative-Prior Strategy on α and β . Settings and $E(\delta)$ for 60 Runs.

The x's indicate settings. Circles are $E(\delta)$. Dashed line is actual value of δ . \circ indicates a Pass, \bullet indicates a Fail, \oplus indicates initial value.

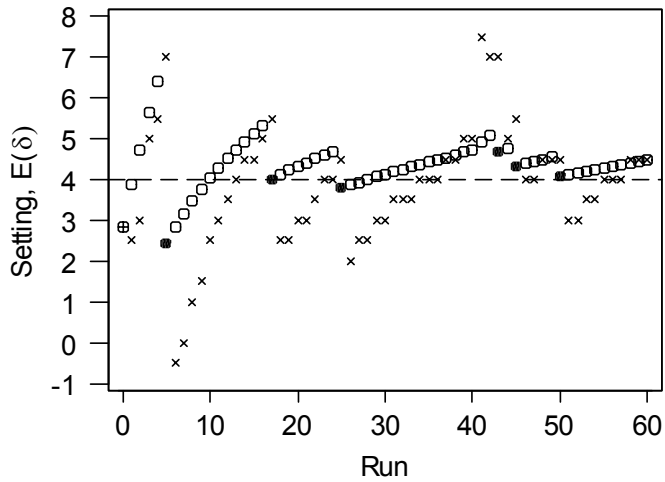
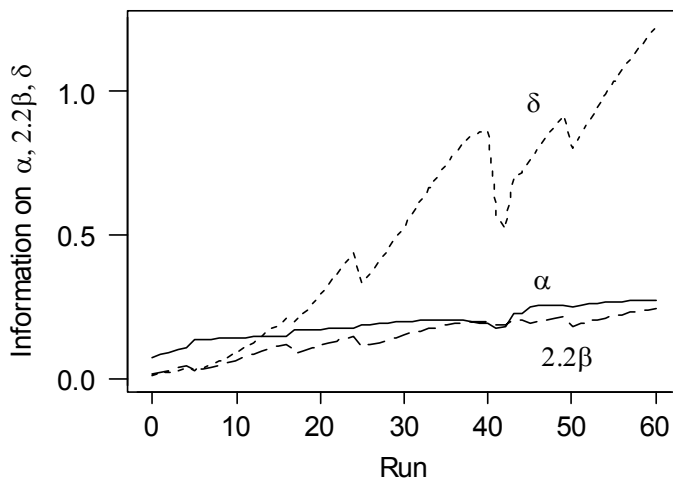
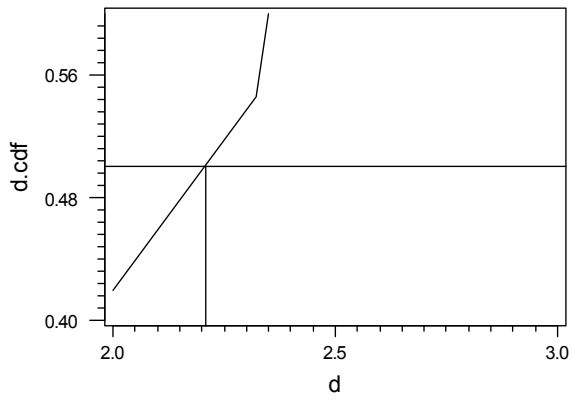
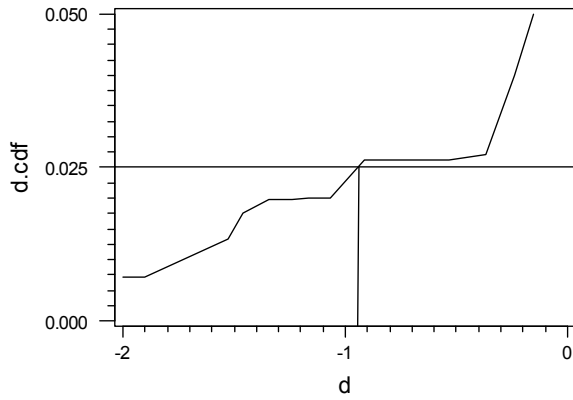


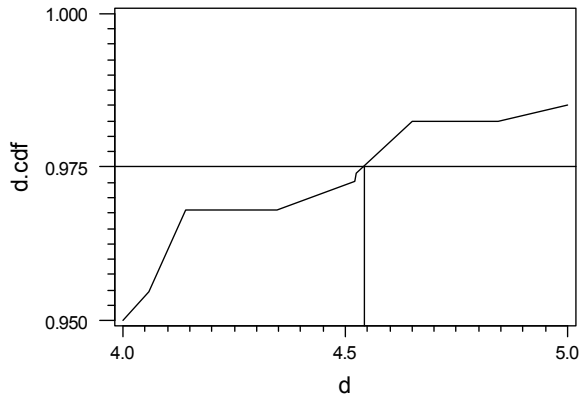
Figure 13. Flatter Prior, Simulation 4—Relative Prior Strategy on α and β . Information on α , 2.2β , and $\delta = \alpha - 2.2\beta$, versus Run Number.



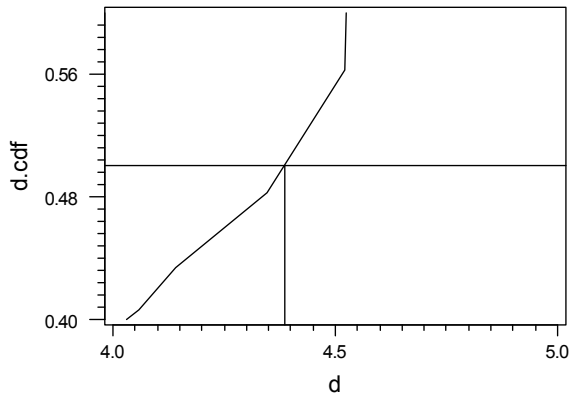
Not used in report itself

Eqinfa.mtw (equal information, first run). So lower .025 point is -0.093, 0.5 is 2.21, .975 is 4.54





eqinfb



Flat prior, first simulation

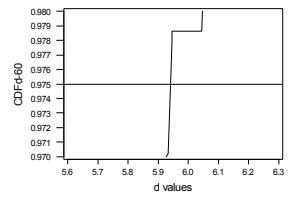
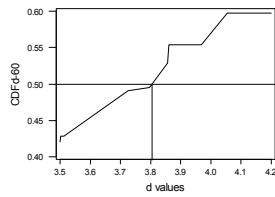
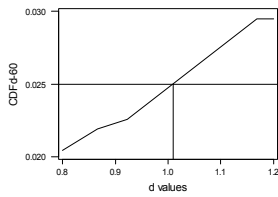


Figure 14. Flatter Prior, simulation 1. Distribution of (α, β) after 60 runs.

Area of symbol is proportional to $P(\alpha, \beta)$

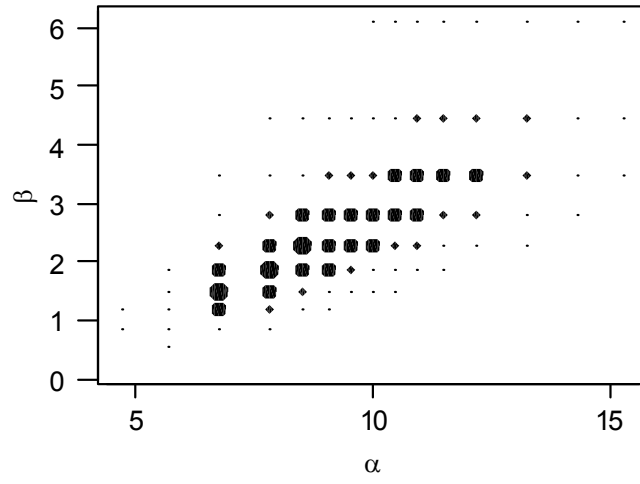
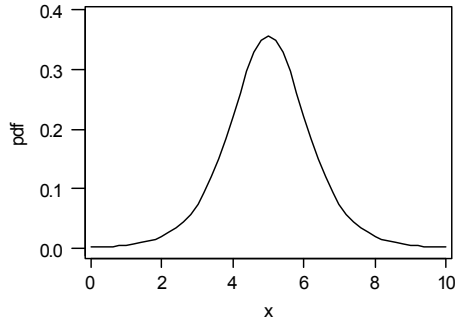


Figure 15. Example of a logistic distribution ($\alpha=5.0, \beta=0.7$)

(a) Tolerance distribution



(b) Quantal Response Curve: $P(Y_x = 1)$

