HOMELAND SECURITY
COMMUNITY OF BEST PRACTICES

# Improving the Interpretability of Bayesian Priors

January 2022

Dr. Kyle Kolsti, Acting Director


To enhance T&E science through multidisciplinary collaboration and deliver it to the DHS workforce through independent consultation and tailored resources.

About this Publication:
This work was conducted by the Homeland Security Community of Best Practices under contract FA8075-18-D-0002, Task FA8075-21-F-0074.

For more information:
Visit, https://www.afit.edu/HSCOBP/
Email, AFIT.ENS.HSCOBP@us.af.mil


Homeland Security Community of Best Practices
2950 Hobson Way
Wright-Patterson Air Force Base, Ohio



Version: 1, FY23

THEORY into PRACTICE

**Executive Summary**

Constructing a prior distribution for a Bayesian analysis with input from subject matter experts, a process called "elicitation," can be challenging. Their knowledge may not neatly fit the mathematical construct of the analysis which can confound the translation from words and thoughts to statistical distributions. This paper uses a representative scenario to illustrate two ways for analysts to help with this translation: transforming parameters to new ones that are meaningful to the experts and presenting the prior distribution in graphical products (charts and plots) that permit experts to assess the plausibility of the prior. Analysts may use either technique or both as desired.

*Keywords: Bayesian, prior, transform*

## Table of Contents

## Introduction

Bayesian statistics is garnering attention in the Department of Defense (DOD) test and evaluation (T&E) community because it provides a mechanism to aggregate understanding of a system's behavior from a variety of sources. One source to aggregate understanding, which is the focus of this paper, is from subject matter experts (SMEs) who can provide insight into what may be expected from a system. This process of collecting knowledge from SMEs is called "elicitation" (Garthwaite, 2005). Once elicited, SME insight will take the form of statistical distributions capturing both prediction and uncertainty–called the "prior" for short. This insight will later be combined mathematically with test results to provide the updated, refined "posterior" distribution. The posterior distribution is then used for inference and decision-making. An advantage of SME input is it can mitigate some of the uncertainty from a small data set; however, specialized skill and experience is needed to avoid misleading outcomes. A non-technical explanation of Bayesian statistics, including benefits and risks of employing it, can be found in *Bayesian Methods in Test and Evaluation: A Decision-Maker's Perspective* (Sieck & Kolsti, 2022).

A valid Bayesian approach depends on the appropriate construction of a prior. Unfortunately, it can be difficult to translate SME experience into the mathematical world of models, parameters, and statistical distributions. To address this challenge, this Best Practice offers two techniques to develop priors from SME inputs: (1) parameter transformation, and (2) sampling the prior. It is assumed the reader is familiar with the fundamental concepts of Bayesian statistics (Seick & Kolsti, 2022), prior distributions, and distribution sampling procedures typified by the Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithms (Hogg, 2018).

This paper offers a step-by-step scenario to demonstrate to analysts how to apply both techniques for creating a suitable prior from SME inputs. Note that either or both techniques may be used–the flow of the paper is not meant to imply that the first technique is required to perform the second one. After laying out the scenario, we will transform a model's parameters into new parameters that have intuitive meaning to SMEs. Next, we will depict the prior in a manner that is familiar to the SMEs by sampling the prior. Finally, we will pretend the first prior was deemed inadequate by the SMEs and generate a new one using the same process with updated SME inputs. This final step will illustrate how sampling the prior can create intuitive output for SMEs to assess the prior.

## Working Example

This paper will walk through a scenario to illustrate the proposed techniques. In this scenario the test team is using logistic regression to predict a sensor's probability of detection (PD) denoted by the variable $p$, where the only factor is range, $x$. The data model that describes the relationship between range and PD is

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x \tag{1}$$

Figure 1 depicts an example logistic regression prediction curve, in this case determined by the parameter values $\beta_0 = 3.296$ and $\beta_1 = -0.275$. For reference, note that the range where the PD=80% is 12, and at a range of 20 the PD is 10%.

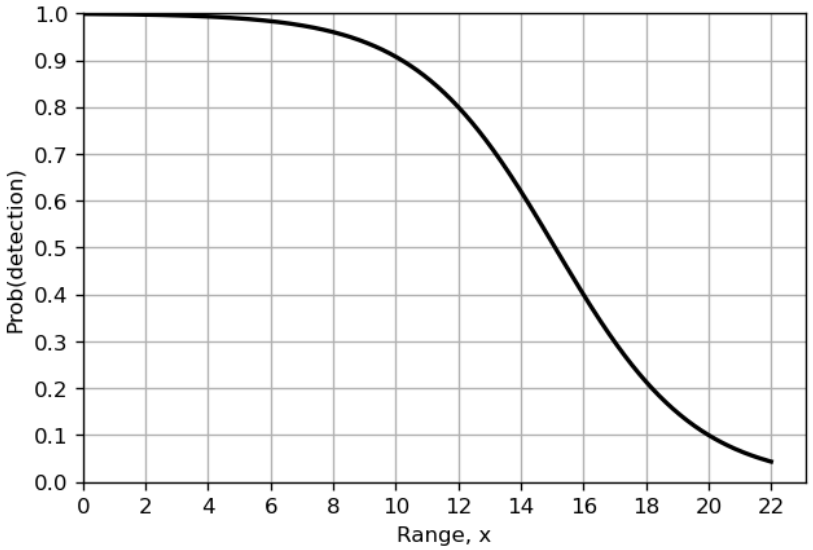



**Figure 1**
*Example of a Curve Created by the Logistic Function*

To perform a Bayesian analysis, we must build a prior over the 2-dimensional (2-D) space $\beta = \{\beta_0, \beta_1\}^T$ , which we will call $f_\beta(\boldsymbol{\beta})$. Asking SMEs "what range of values should we expect for the parameters $\beta_0$ and $\beta_1$?" will probably not be productive because these parameters are not inherently relatable to engineering specifications or a system's behavior. Instead, suppose you elicit this informative feedback from SMEs who are highly experienced with this type of system and application:

1. "The 80% detection range is probably between 3 and 11."

2. "At a range of 20 or greater, detection is highly unlikely."

How do we turn this information into the prior which we have labeled $f_\beta(\boldsymbol{\beta})$? The next section will describe parameter transformation (Technique #1) as a way to incorporate these statements into the model. Later in this paper, a graphical approach (Technique #2) will be illustrated that can help SMEs assess the realism of their inputs.

**Technique #1: Parameter Transformation**

***Step 1: Defining the New Parameters***
Let's define two new parameters that are more intuitive for the SME and that relate directly to the requirement and the SME inputs:

| Parameter | Definition |
|---|---|
| $x_q$ | The range $x$ at which the probability of detection equals $q$, which is chosen in advance. Given the SME inputs, $q = 0.8$. |
| $p_r$ | The probability $p$ at the range $r$, which is chosen in advance. Given the SME inputs, $r = 20$. |

For notation convenience we define a vector that contains the two new parameters, $\boldsymbol{\theta} = \{x_q, p_r\}^T$. Since there are two parameters being transformed between $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, two equations are needed. These two formulas can be obtained from the data model of Equation 1, with each formula corresponding to the definition of one of our new parameters. Recall that $q$ and $r$ are both constants with values selected by the SMEs.

$$\log\left(\frac{q}{1-q}\right) = \beta_0 + \beta_1 x_q$$
$$\log\left(\frac{p_r}{1-p_r}\right) = \beta_0 + \beta_1 r \tag{2}$$

Solving this system of equations provides the transformation formulas for the original parameters $\boldsymbol{\beta}$ in terms of the two new parameters $\boldsymbol{\theta}$. For notation convenience we will use $\boldsymbol{G}$ as the transformation operator, so the transformation is denoted as $\boldsymbol{\beta} = \boldsymbol{G}(\boldsymbol{\theta})$. Note the condition $x_q \neq r$ means the SMEs must provide two mathematically distinct descriptions of system behavior; in other words, the two formulas cannot be built using only one piece of information. The formulas that make up the operator $\boldsymbol{G}$ are

$$\beta_0(x_q, p_r) = \log\left(\tfrac{q}{1-q}\right) - \left(\tfrac{x_q}{x_q - r}\right)\left[\log\left(\tfrac{q}{1-q}\right) - \log\left(\tfrac{p_r}{1-p_r}\right)\right], \quad x_q \neq r$$
$$\beta_1(x_q, p_r) = \frac{\log\left(\frac{q}{1-q}\right) - \log\left(\frac{p_r}{1-p_r}\right)}{x_q - r}, \quad x_q \neq r \tag{3}$$

The inverse transform $G^{-1}$ can similarly be derived through some algebra to obtain $\boldsymbol{\theta} = G^{-1}(\boldsymbol{\beta})$. Note that the new parameters do not permit a zero-slope solution, which occurs when $\beta_1 = 0$; this means when using the new parameters the range must have some non-zero influence on PD, however small. This restriction should have no impact on the analysis because the selection of the data model already indicated a belief that range matters. The formulas that make up the inverse operator $G^{-1}$ are

$$x_q(\beta_0, \beta_1) = \frac{\log\left(\frac{q}{1-q}\right) - \beta_0}{\beta_1}, \quad \beta_1 \neq 0$$
$$p_r(\beta_0, \beta_1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 r)}} \tag{4}$$

***Step 2: Constructing the Prior with the New Parameters***

Now that we have defined parameters that match the elicited information about the system, we can build the prior based off the information we elicited from the SME. Since the parameters are independent of each other, we can construct a simple one-dimensional (1-D) prior for each parameter, denoted as $f_{xq}(x_q)$ and $f_{pr}(p_r)$. The analyst will engage with the SMEs to elicit the degree to which values or ranges of values for a parameter may be expected, plausible, and physically realistic. Previously existing data may also be incorporated. The details of this elicitation process are beyond the scope of this paper. Suppose for this scenario that the

outcome of this collaborative process is the selection of a normal distribution and a beta distribution for the priors of $x_q$ and $p_r$, respectively.

$$f_{xq}(x_q) = \text{Normal}(\mu = 7, \sigma = 2)$$

$$f_{pr}(p_r) = \text{Beta}(2, 48)$$

The 1-D probability density functions for these two priors are shown in Figure 2.

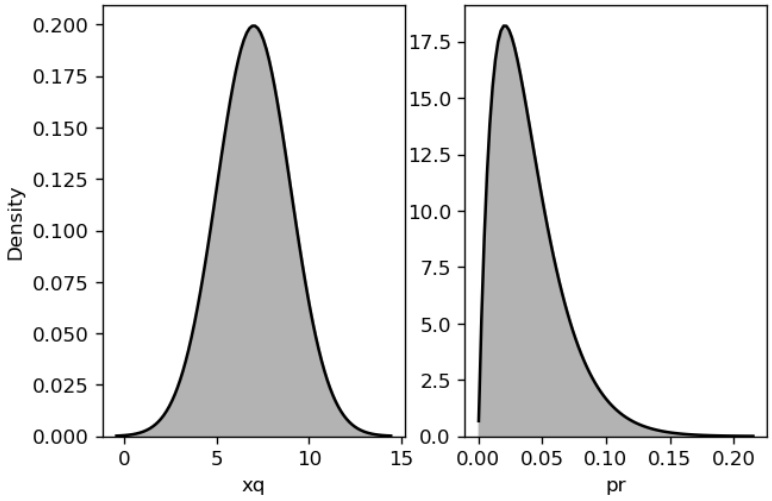
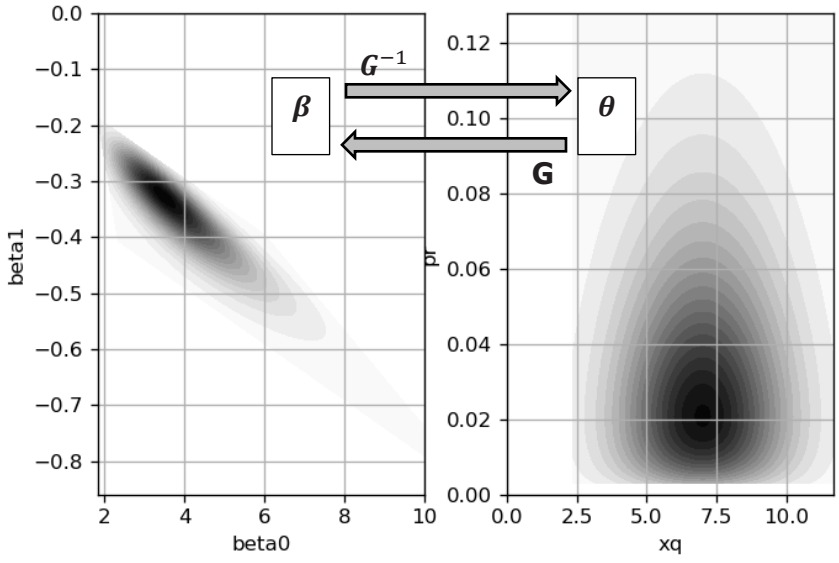



**Figure 2**

*Prior Distributions Chosen for $x_q$ and $p_r$*

These two independent 1-D priors can be multiplied by each other to create the 2-D prior over the transformed parameter space $\boldsymbol{\theta}$, $f_\theta(x_q, p_r) = f_{xq}(x_q) f_{pr}(p_r)$. This joint prior $f_\theta$ is depicted in Figure 3b.




**(a)** **(b)**

**Figure 3**

*Joint Prior Distributions for the Original Parameter Space, $f_\beta(\beta_0, \beta_1)$ (left) and the Transformed Parameter Space, $f_\theta(x_q, p_r)$ (right)*

***Step 3: Transforming the Prior***

We have so far obtained the prior distribution over the transformed parameter space $\boldsymbol{\theta}$, as shown in Figure 3b. However, the prior distribution we are seeking is in the original parameter space $\boldsymbol{\beta}$, as shown in Figure 3a. Fortunately, there are well-known formulas available for this purpose. The formula which transforms the known prior distribution $f_\theta$ in Figure 3b to the desired prior distribution $f_\beta$ in Figure 3a is

$$f_\beta(\boldsymbol{\beta}) = f_\theta(\boldsymbol{\theta}) \ \left|\det\big(\boldsymbol{J}(\boldsymbol{\beta})\big)\right| \tag{5}$$

where $\left|\det\big(\boldsymbol{J}(\boldsymbol{\beta})\big)\right|$ is a scalar value calculated as the absolute value of the determinant of the Jacobian matrix $\boldsymbol{J}$ (Christensen, 2011). The Jacobian matrix contains the partial derivatives of the transformation $G^{-1}$, and is numerically evaluated at the point $\boldsymbol{\beta}$. In this paper's working example, the transformation involves two parameters so the Jacobian matrix is a $2 \times 2$ matrix defined as

$$\boldsymbol{J}(\boldsymbol{\beta}) = \begin{bmatrix} \dfrac{\partial x_q}{\partial \beta_0}(\boldsymbol{\beta}) & \dfrac{\partial x_q}{\partial \beta_1}(\boldsymbol{\beta}) \\ \dfrac{\partial p_r}{\partial \beta_0}(\boldsymbol{\beta}) & \dfrac{\partial p_r}{\partial \beta_1}(\boldsymbol{\beta}) \end{bmatrix} \tag{6}$$

Numerical approximation of these partial derivatives may be used if necessary, but in this working example the analytical derivatives are readily obtainable:

$$\begin{aligned} \frac{\partial x_q}{\partial \beta_0} &= -\frac{1}{\beta_1} \\ \frac{\partial x_q}{\partial \beta_1} &= -\frac{\log\left(\frac{q}{1-q}\right) - \beta_0}{{\beta_1}^2} \\ \frac{\partial p_r}{\partial \beta_0} &= \left(\frac{1}{1+e^{-(\beta_0+\beta_1 r)}}\right)\left(1-\frac{1}{1+e^{-(\beta_0+\beta_1 r)}}\right) \\ \frac{\partial p_r}{\partial \beta_1} &= \left(\frac{1}{1+e^{-(\beta_0+\beta_1 r)}}\right)\left(1-\frac{1}{1+e^{-(\beta_0+\beta_1 r)}}\right) r = r\,\frac{\partial p_r}{\partial \beta_0} \end{aligned} \tag{7}$$

Using the procedure of Equation 5 for parameter transformation, it can be verified using numerical integration that the volume under both priors in Figure 3 equals 1.0 as required by the definition of a statistical distribution. At this point, we have successfully invented new parameters which are more meaningful to the SMEs and constructed the 2-D prior distribution over both the original and new parameters.

**Technique #2: Sampling the Prior**

This technique may be used regardless of whether the first technique of parameter transformation was employed. For the flow of this paper, we will continue with the working example.

At this point, the prior distribution shown in Figure 3a has been constructed. It may be difficult for a SME to visually inspect Figure 3a and assess its validity. One effective way to assess this prior is to sample it, and then use the sample to create analysis products that are familiar to the SMEs. For example, if the SMEs are accustomed to seeing plots of detection probability versus range like Figure 1 then you should create that product from the sample. Specifically, you may consider creating the graphical and tabular products that will go into the final report. If the results in these familiar data products look right to the SMEs, the prior is more likely to be suitable. (Note: the procedure described in this section may be applied to any statistical distribution; here we apply it to the prior of Figure 3a, but later after data are collected, this same sampling and plotting procedure may be used on the posterior distribution to obtain the figures for the final report.)

To demonstrate this technique, a sample of 5,000 points was drawn from the prior in the original parameter space, $f_\beta(\beta_0, \beta_1)$ using the Metropolis-Hastings Markov Chain Monte Carlo (MCMC) method. These randomly sampled points are shown in Figure 4. By inspection, the point cloud corresponds to the contour plot of Figure 3a as it should. The sample size was selected somewhat arbitrarily as sufficient for this demonstration; in practice, you should follow published guidance to ensure adequate sample size (Hogg, 2018).

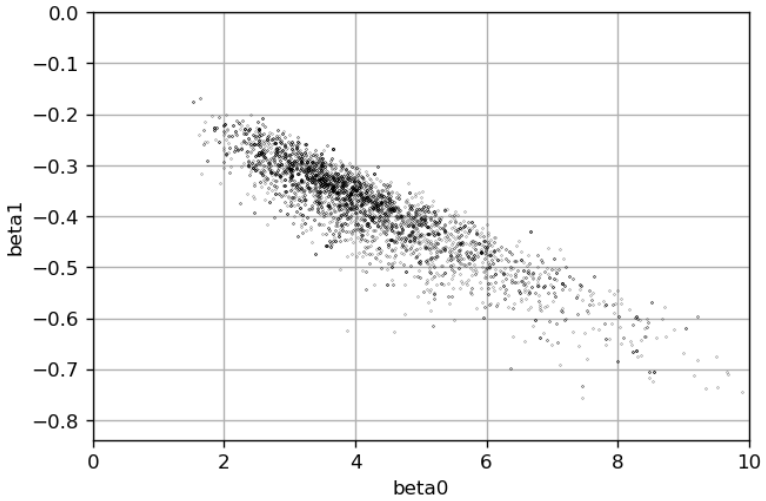



**Figure 4**
*Random Points Sampled From the Prior Distribution*

Each of the 5,000 points plotted in Figure 4 represents a single logistic function curve as given in the data model of Equation 1 and as illustrated in Figure 1. Figure 5 depicts 100 of these curves randomly drawn from the sample of 5,000 points. The solid blue bars represent the 95% interval of both 1-D priors from Figure 2. As expected, approximately 95% of the curves go through the blue bars. These curves are more easily interpreted by the analyst and the SMEs than the point cloud of Figure 4.

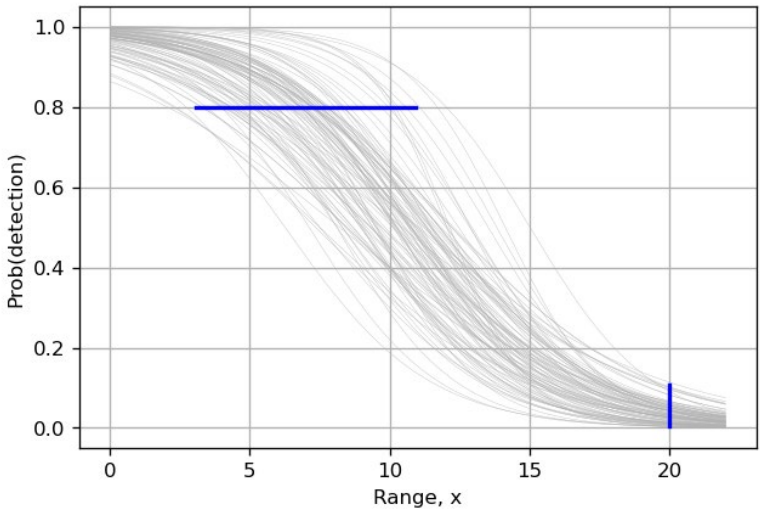



**Figure 5**

*Subsample of Curves Sampled from the Prior. Blue Lines Depict Where 95% of the Curves Should go, Based on the Selected Priors*

As a final verification that the MCMC procedure accurately sampled the prior, Figure 6 shows histograms of $x_q$ and $r$ from the 5,000 points overlaid with the selected prior probability density functions (PDFs). The plots verify that the constructed prior does not significantly negate the statements the SMEs provided.

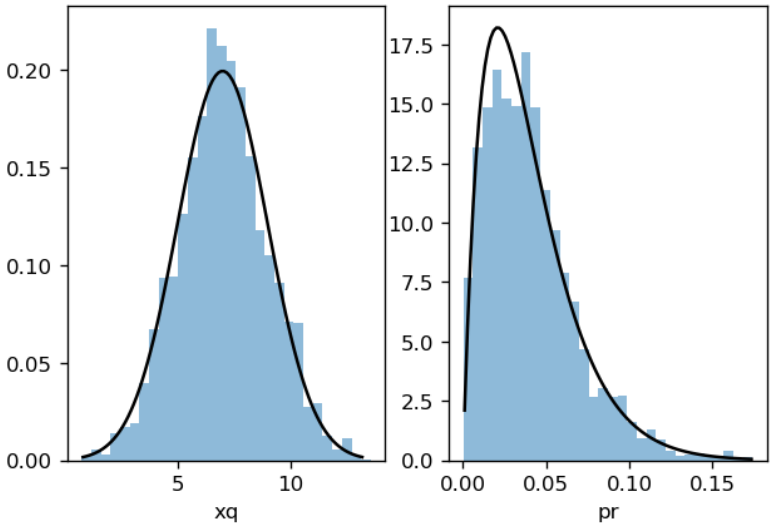



**Figure 6**

*Histograms of Parameter Values Produced Using the Joint Prior $f_\beta$*

**Technique #2: Repeated with a Different Prior**

Suppose SME analysis of Figure 5 reveals that the selected priors lead to a joint prior that is too informative; in other words, there is concern that the prior will dominate the data of the planned test and bias the ultimate test results. (There are tools for sizing a test to avoid this

situation which are outside the scope of this paper: see "Bayesian Model Checking" [Theimer, 2022]). After the analyst and SMEs review and refine their predictions, they decide to replace the normal and beta distributions used as priors in the previous section with priors that impart more uncertainty: a semicircle distribution and a uniform distribution with the following parameters

$$f_{xq}(x_q) = \text{semicircle}(\text{center} = 10, \text{width} = 18)$$

$$f_{pr}(p_r) = \text{uniform}(min = 0, max = 0.2)$$

One advantage of these priors is that they have bounds that we can control. For example, when using a normal distribution, the parameter can be any number from $-\infty$ to $\infty$. In the real world, however, the parameter in question may have physical bounds–in this paper's working example, all ranges must be non-negative, so our parameter $x_q$ must be in the region $x_q \geq 0$. Another complication comes from our transformation formulas, where we must have $x_q \neq r$, and for our scenario with a decreasing probability with range ($\beta_1 < 0$), we must further have $x_q < r$ and $p_r < q$. These constraints can inadvertently be violated during an MCMC sampling procedure with our previous priors unless special rules are coded into it. However, the constraints are naturally satisfied by the newly selected semicircle and uniform priors.

This new prior $f_\beta(\beta_0, \beta_1)$ is sampled using MCMC as in the previous example to generate 10,000 points. Figure 7 shows the analytical priors and the histograms from the MCMC procedure. Figure 8 shows 100 of the resulting prediction curves. The curves are much more spread out, indicating more uncertainty is built into this joint prior. The SMEs can readily compare Figures 5 and 8 to decide which prior is more realistic.

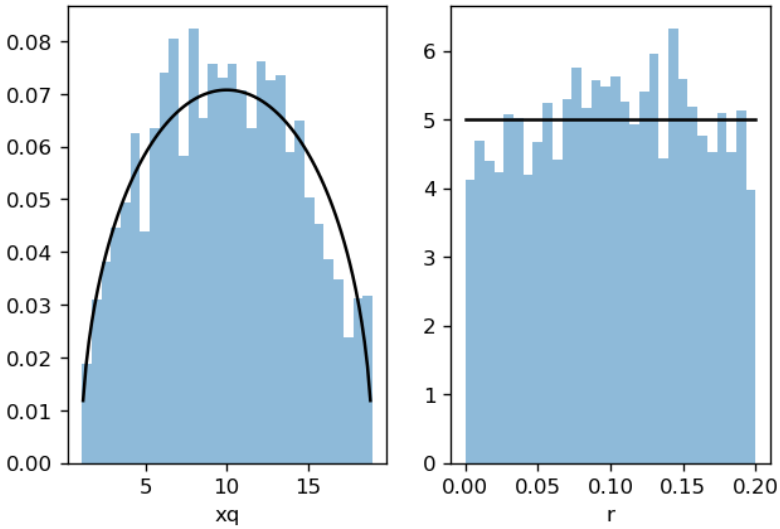



**Figure 7**
*Histograms of Parameter Values Produced Using the Less-Informative Prior*

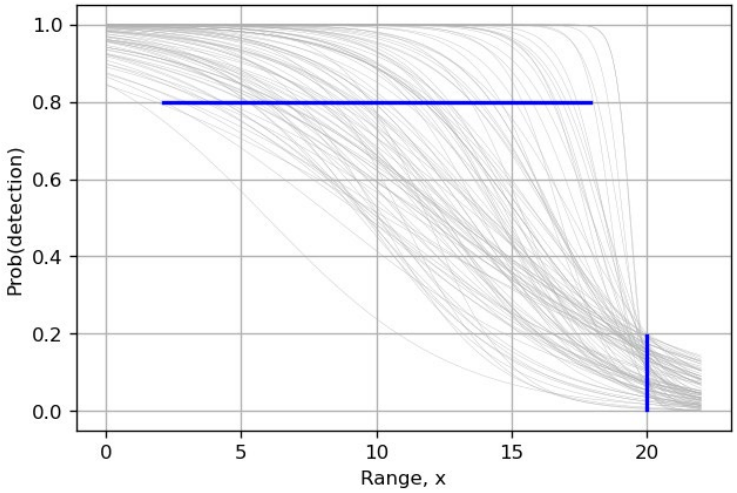



**Figure 8**
*A Sample of 100 Curves Using the Less-Informative Prior*

## Conclusion

The prior distribution is critical to successful application of Bayesian statistics. This Best Practice demonstrated the use of two techniques for translating insight elicited from SMEs into a prior: transformation of the original model formula parameters into new parameters that have meaning to the SMEs, and sampling the prior to create graphical products that are familiar and interpretable to the SMEs. Analysts may use either technique or both as desired.

## References

Christensen, R., & Johnson, W., & Brunscum, A., & Hanson, T. E. (2011). *Bayesian ideas and data analysis: An introduction for scientists and engineers.* CRC Press, 2011.

Garthwaite, P. H., & Kadane, J. B., & O'Hagan, A. (2005). Statistical Methods for Eliciting Probability Distributions. *Journal of the American Statistical Association* 100(470), 680–701.

Hogg, D., & Foreman-Mackey, D. (2018) Data Analysis Recipes: Using Markov Chain Monte Carlo. *The Astronomical Journal Supplemental Series*, 236(1), 1–18. https://doi.org/10.3847/1538-4365/aab76e

Sieck, V., & Kolsti, K. (2022). *Bayesian Methods in Test and Evaluation: A Decision-Maker's Perspective.* Best Practice. Scientific Test & Analysis Techniques Center of Excellence. https://www.afit.edu/STAT/statdocs.cfm?page=763

Theimer, James (2022). *Bayesian Model Checking*. Best Practice. Scientific Test & Analysis Techniques Center of Excellence. https://www.afit.edu/STAT/statdocs.cfm?page=763