BAYESIAN ESTIMATION OF UNIMODAL DISTRIBUTIONS

P. Damien
Duke University
Imperial College
Kimberly-Clark Corp.
P.W. Laud
Medical College of Wisconsin
A.F.M. Smyth
Imperial College

DP #94-34
Bayesian Estimation of Unimodal Distributions
P. Damien\textsuperscript{1}, P.W.Laud\textsuperscript{2} and A.F.M. Smith\textsuperscript{3}

Abstract

In this paper, we estimate unimodal distribution functions (DFs) from a Bayesian perspective. A DF is assumed to be ogive, that is to say one that switches from convex to concave with an unknown change-point. Monte Carlo methods are described which provide a full Bayesian solution for the modelling of unimodal DFs.

Keywords: Unimodal Distributions, Switch-Point, Convex, Concave, Gibbs Sampling.

1. Duke University, USA; Imperial College, UK; Kimberly-Clark Corp. USA
2. Medical College of Wisconsin USA
3. Imperial College, UK.
I. Introduction

Nonparametric methods of estimating concave DFs are addressed by Denby and Vardi (1986) and the references therein. These papers, however, do not take a Bayesian approach. A related context where the problem of estimating DFs reappears is the estimation of dose response curves in a bioassay. Also, one can estimate DFs based on attribute data. For example, suppose that the stress is the age of the device, say t, and that the response is the failure of the device to survive until t. Then, the response curve is the DF of the device.

We consider the problem of estimating any unimodal DF by exploiting the geometric properties of DFs. Previous Bayesian work in this area can be found in Shaked and Singpurwalla (1990) who develop distribution theory for convex and concave DFs. They are unable to provide a full Bayesian solution to the problem given that the posterior distribution is analytically intractable. Also, Ramgopal, Laud and Smith (1993; henceforward referred to as RL&S) consider the problem of nonparametric modelling of potency curves in a bioassay. The contribution below applies the model developed in RL&S to the class of unimodal distributions. In fact, this translates to addressing the question of estimating unimodal densities.

In Section II, we develop the Mathematical model. In Section III, we apply the model to estimate well-known classes of unimodal DFs, and a few concluding remarks are given in Section IV.

II. The Mathematical Model

Definition 2.1: Lukacs, 1970, Page 91 A distribution function, \( F(x) \) is said to be unimodal if there exists \( t \) in the domain of the function such that \( F(x) \) is convex for \( x < t \) and \( F(x) \) is concave for \( x > t \).

Correspondingly, if \( F = \int f(s)ds \), with \( t \) unknown, \( F \) is unimodal if \( f \) is nondecreasing on \( [S_0, S_t] \) and \( f \) is nonincreasing on \( (S_t, S_{M+1}] \) where \( S_0 \) and \( S_{M+1} \) are the endpoints of the support of \( F \). Lukacs notes that all unimodal densities correspond to DFs of the above type.

To motivate the nonparametric estimation of an arbitrary unimodal DF \( F(x) \), consider the following. Let \( S_1, \ldots, S_M \) denote a partition of the domain of \( F \). Suppose a sample of size \( n \) from a unimodal distribution is observed such that for \( j = 1, \ldots, n \), \( S_1 \leq x_j \leq S_M \). Let \( e_i \) denote the number of \( x_j \)'s in \( (S_{i-1}, S_i] \). Finally, let \( p_i = P(X \leq S_i) \), \( p_0 = 0 \), \( p_{M+1} = 1 \). The likelihood function for \( p | e \), where \( p = (p_1, \ldots, p_{M+1}) \), is

\[
L(p; e) = \prod_{i=1}^{M+1} (p_i - p_{i-1})^{e_i}. \tag{1}
\]

A Bayesian approach would require specifying a prior distribution for the vector, \( p \). Shaked and Singpurwalla (1990) present ideas and distribution theory relating to the
specification of prior shape assumptions for potency curves that RL&$S$ exploit to define a prior for $p$. This prior has the interesting property that it allows prior shape assumptions that encapsulates convex, concave, or as in this context, ogive forms, that is to say, one that changes from convex to concave at some point in the domain of the DF. Given prior information about the shape of the DF, the idea is to find a suitable transformation, say $u = (u_1, \ldots, u_{M+1})$, of the vector $p$ which satisfies two conditions:

$$u_i \geq 0, \quad i = 1, \ldots, M + 1$$

$$\sum_{i=1}^{M+1} u_i = 1.$$ 

The motivation for such a transformation is to note that any vector $u$ which satisfies the above conditions can be regarded as a vector of probabilities, and so a meaningful prior distribution on $u$ would be a Dirichlet distribution. Thus, while it may be difficult to assign directly a prior distribution for the vector $p$, a suitable transformation of $p$ can overcome the problem. RL&$S$, following Shaked and Singpurwalla, identify such transformations. Moreover, as we will see later, such transformations will enable random variate generation from the posterior distribution of $p$ using Markov Chain Monte Carlo methods. With this in mind, we describe the ogive model.

Following RL&$S$, a prior for $p$ is developed by considering the nonnegative "slope parameters",

$$z_i = \frac{F(S_i) - F(S_{i-1})}{\Delta_i} \quad i = 1, \ldots, M + 1,$$

where $\Delta_i = S_i - S_{i-1}$. Define $y_i = z_i - z_{i-1}$ ($i = 1, \ldots, t$), and $y_i = z_i - z_{i+1}$ ($i = t + 1, \ldots, M + 1$), for the convex and concave part of the DF, respectively. We set $z_0 = 0$ and $z_{M+1} = 0$ by convention. To identify a prior distribution for $p$ under the ogive assumption, RL&$S$ develop priors for the convex and concave portions after finding a suitable transformation $u = (u_1, \ldots, u_{M+1})$ of $y = (y_1, \ldots, y_{M+1})$. They let $u$ be distributed Dirichlet with parameter $\alpha = (\alpha_1, \ldots, \alpha_{M+1})$; i.e., the density of $u$ is proportional to

$$\pi(u) \propto u_1^{\alpha_1-1} \cdots u_{M+1}^{\alpha_{M+1}-1}. \quad (2)$$

From (2), RL&$S$ arrive at a prior for $p = (p_1, \ldots, p_M)$ given by

$$p_i | p_1, \ldots, p_t, t \sim Beta(\alpha_i, \bar{\gamma}_{i+1}; \bar{c}_i, \bar{d}_i) \quad (3)$$

for $i = 1, \ldots, t$ and

$$p_i | p_1, \ldots, p_t, p_{M+1}, \ldots, p_{i+1}, t \sim Beta(\gamma_i - \gamma_t, \alpha_{i+1}; c_i, d_i) \quad (4)$$

3
for $i = t+1, \ldots, M$, where

$$
\tilde{c}_i = p_{i-1} + \frac{\Delta_i}{\Delta_{i-1}} (p_{i-1} - p_{i-2})
$$

$$
\tilde{d}_i = p_{i-1} + \frac{\Delta_i}{S_t - S_{i-1}} (1 - p_{i-1})
$$

$$
c_i = p_{i+1} - \frac{\Delta_{i+1}}{S_{i+1} - S_t} (p_{i+1} - p_t)
$$

$$
d_i = p_{i+1} - \frac{\Delta_{i+1}}{\Delta_{i+2}} (p_{i+2} - p_{i+1}).
$$

$\tilde{\gamma}_{i+1} = \sum_{j=i+1}^{M+1} \alpha_j$ and $\gamma_i = \sum_{j=1}^{i} \alpha_j$. Omitting the normalising constant, a $w \sim Beta(a, b; c, d)$ random variable has density

$$
g(w|a, b, c, d) \propto \left( \frac{w - c}{d - c} \right)^{a-1} \left( \frac{d - w}{d - c} \right)^{b-1} I_{(c,d)}(w). \tag{5}
$$

$I_{(c,d)}(w)$ denotes the indicator function defined to be equal to 1 if $w \in (c, d)$, equal to zero otherwise.

With $t$ unknown, RL&S complete the prior specification by assigning

$$
pr(t = i) = \pi_i; \quad \pi_i \geq 0 \quad (i = 0, \ldots, M + 1); \quad \sum_{i=0}^{M+1} \pi_i = 1. \tag{6}
$$

In what's to follow, the notation $[A|B]$ denotes the conditional probability density of $A$ given $B$, and likewise $[A]$ denotes the probability density of $A$.

Given the likelihood function in (1), the posterior joint distribution of $(p, t)$ is given by

$$
[p, t, |c] \propto [t|p] \prod_{i=1}^{M+1} (p_i - p_{i-1})^{e_i} \tag{7}
$$

where the appropriate forms for each of the densities appearing in (7) are given in (3), (4), and (6).

The complicated constraint region for the posterior in (7) makes it clear that one would have to resort to some type of numerical solution to calculate quantities of interest such as the posterior marginal distributions of $t$ and $p$. To this end, following the introduction of sample based methods discussed in Geman and Geman (1984), Gelfand and Smith (1990) and Smith and Roberts (1993), we describe the use of a Markov Chain Monte Carlo method, the Gibbs Sampler (Geman and Geman, 1984) in this context. The
Gibbs Sampler requires knowledge of the conditional distributions of each of the random quantities appearing in the posterior joint distribution in a Bayesian model.

In our context, the following conditionals are needed:

\[ p_i | p_j, t, \text{data}, i,j = 1, \ldots, M, i \neq j, \]

\[ [t | p, \text{data}] \]

A sampling approach to simulating from the joint posterior for \( p \) and \( t \) would proceed as follows. Given arbitrary starting values, a long iteration of successive random variate generations from each of the conditional distributions appearing in (8) results in eventual \((p, t)\) realizations which are close to being drawings from the joint posterior. Replication of this process then provides a sample from the joint posterior, which can be used as a basis for inference summaries of interest. Alternatively, ergodic averages from a single run of the process can be used to form consistent estimates of quantities of interest. Details of the Gibbs Sampler within the context of Bayesian inference can be found in Gelfand and Smith (1990) and Smith and Roberts (1993).

From RL&S, the conditional distributions \([p.|] \) and \([t.|] \) are obtained as follows.

\[ [t | p, \text{data}] \propto [t] f_i(p) I(t_{0-1}, t_{0})(i), \]

where

\[ t_0 = \min \{ i : (p_i - p_{i-1})/\Delta_i > (p_{i+1} - p_i)/\Delta_{i+1} \} \]

and

\[ f_i(p) = \prod_{j=1}^{t} g(p_j | \alpha_j, \bar{\gamma}_{j+1}, \bar{c}_j, \bar{d}_j) \prod_{j=t+1}^{m} g(p_j | \gamma_i - \gamma_i, \alpha_j, c_j, d_j), \]

with the function \( g(\cdot | \cdot) \) denoting the appropriate four parameter Beta distributions defined in (3) and (4), and \([t] \) appears in (6) of Section II.

Examination of the posterior in (7) shows that the conditional distribution for \( p \) when \( i = 1, \ldots, t - 2 \) is given by

\[ [p_i | t, e, p_j, i \neq j] \propto (p_i - p_{i-1})^{\alpha_i} (p_{i+1} - p_i)^{\bar{\alpha}_{i+1}} I(\bar{a}_i, \bar{b}_i)(p_i) \]

\[ \times \left( \prod_{j=1}^{i+2} g(p_j | \alpha_j, \bar{\gamma}_{j+1}, c_j, d_j) \right), \]

where

\[ \bar{a}_i = \max \left\{ p_{i-1} + \frac{\Delta_i}{\Delta_{i-1}} (p_{i-1} - p_{i-2}), p_{i+1} - \frac{\Delta_{i+1}}{\Delta_{i+2}} (p_{i+2} - p_{i+1}) \right\} \]

\[ \bar{b}_i = \Delta_{i+1} p_i - 1 + \Delta_i p_{i+1} / (\Delta_i + \Delta_{i+1}), \]
when \( i = t - 1, \bar{a}_i = \bar{c}_i \). For \( i = t + 2, \ldots, M \)
\[
[p_i|t, e, p_j \neq j] \propto (p_i - p_{i-1})^{\bar{e}_i}(p_{i+1} - p_i)^{\bar{e}_{i+1}} I(a_i, b_i)(p_i) \\
\times \left( \prod_{j=i+2}^{i} g(p_j|\gamma_j - \gamma_t, c_j, d_j) \right),
\]
where
\[
a_i = \Delta_i P_{i+1} + \Delta_{i+1} P_{i-1}/(\Delta_i + \Delta_{i+1})
\]
\[
b_i = \min \left\{ p_{i+1} - \frac{\Delta_{i+1}}{\Delta_{i+2}} p_{i+2} - p_{i+1}, p_{i-1} + \frac{\Delta_i}{\Delta_{i-1}} (p_{i-1} - p_{i-2}) \right\},
\]
when \( i = t + 1, a_i = c_i \), and \( b_i = d_i \).

At \( t \),
\[
[p_t|t, e, p_j \neq t] \propto (p_t - p_{t-1})^{\bar{e}_t}(p_{t+1} - p_t)^{\bar{e}_{t+1}} I(a^*_t, b^*_t)(p_t) \\
\times g(p_t|\alpha_t, \bar{\gamma}_{t+1}, \bar{c}_t, \bar{d}_t) \left( \prod_{j=t+1}^{M} g(p_j|\gamma_j - \gamma_t, c_j, d_j) \right),
\]
where
\[
a^*_t = p_{t-1} + \frac{\Delta_t}{\Delta_{t-1}} (p_{t-1} - p_{t-2})
\]
\[
b^*_t = p_{t+1} + \frac{\Delta_{t+1}}{\Delta_{t+2}} (p_{t+2} - p_{t+1}).
\]

**Specification of Prior Parameter Values**

Shaked and Singpurwala provide the following ways to define the \( \alpha_i \)'s in the Dirichlet distribution. Firstly, from equation 3.16 page 11 of Shaked and Singpurwala,
\[
\alpha_i = S_i[\Delta_i^{-1}(p_i^* - p_{i-1}^*) - \Delta_{i+1}^{-1}(p_{i+1}^* - p_i^*)], \quad i = 1, \ldots, M + 1,
\]
where \( p_i^* \) is the prior guess at each component of the \( p \) vector, \( p_0^* = 0 \) and \( p_{M+1}^* = p_{M+2}^* = 1. \) Secondly, if the user is unable to guess the \( p \) vector, then they suggest setting \( p_i^* = 1 - \exp(-S_i) \), and then solving for \( \alpha_i \) using the above equation. Thirdly, suppose the grid is partitioned into \( m + 1 \) grid points. Let \( r = \lfloor \frac{m+1}{2} \rfloor \) where \( \lfloor . \rfloor \) denotes the greatest integer less than or equal to \( \frac{m+1}{2} \). Now, set
\[
\alpha_i^* = \begin{cases} 
    r - (i-1) & \text{if } i = 1, \ldots, r; \\
    i - r & \text{if } i = r + 1, \ldots, m + 1.
\end{cases}
\]

Define
\[
\alpha_i = \frac{K \alpha_i^*}{\sum_{j=1}^{m+1} \alpha_j},
\]
\[
6
\]
where $K$ is a positive constant. For $K$ large, Shaked and Singpurwalla note that the prior distribution tends to become diffuse. Finally, the probability of the switch-point is specified by the discrete distribution, $\Pr(t = i) = \frac{1}{m+2} i = 0, \ldots, m + 1$.

Markov Chain Monte Carlo Methods (MCMC) for the Estimation of Distribution Functions

Given $p$, the posterior form for the switch-point in (9) specifies a simple discrete distribution over the two points $t_0 - 1$ and $t_0$, as the contenders for the possible 'switch points' (from convex to concave) if, given $p$, $t_0$ is the smallest stimulus value at which the 'slope' is subsequently decreasing. Simulation from this two point distribution is straightforward.

It remains only to sample efficiently from each of the one-dimensional (truncated) forms that comprise the vector $p$. In fact, a number of methods are available and easily implemented, including classical rejection and ratio-of-uniforms techniques (Devroye, 1987). Yet another possibility, which is particularly appealing if $n_1, \ldots, n_m$ are not too small, is the sampling-importance-resampling method (Rubin, 1988; Smith and Gelfand, 1992). This proceeds as follows, for any component of $p$.

(a) for suitable $N$, sample $p_i^1, \ldots, p_i^N$ from a $Beta(c_i + 1, c_{i+1} + 1)$ distribution truncated to $(\bar{c}_i, \bar{d}_i)$ or $(\bar{c}_i, \bar{d}_i)$ as appropriate (Gelfand and Kuo, 1991);

(b) evaluate $q_k = g(p_i^k | \ldots) / \sum_k g(p_i^k | \ldots), k + 1, \ldots, N$, using the appropriate $g(p_i | \ldots)$ form for the convex or concave part of the joint posterior density (see earlier discussion for explicit forms of these four parameter Beta densities);

(c) sample one of $p_i^1, \ldots, p_i^N$ using the $q_1, \ldots, q_N$ distribution.

As $N \to \infty$, it can be proved that the sampled $p_i$ tends to a drawing from the required conditional distribution (Smith and Gelfand, 1992).

Marginal inferences for the individual $p_i$'s are straightforwardly obtained from the corresponding marginal samples, which can be used as the basis for moment or quantile estimates, or graphical representation via box-plots, histograms or density estimates.

However, in order to determine the unimodality of the distribution function, interest focusses on the marginal distribution for $t$. In particular, a high concentration of posterior mass on $t = S_{M+1}$ or $t = S_0$ may indicate a concave or convex rather than ogive form. If the distribution function is indeed unimodal, then the posterior mass of $t$ will be concentrated about the point on the grid where the distribution function is most likely to change from convex to concave.

III. Illustrative Analyses

In order to illustrate the mathematical models described in the previous sections, two standard distributions that are known to be unimodal are considered. These are the standard Gaussian distribution that is unimodal at zero and a chi-squared distribution with 11 degrees of freedom. The latter is unimodal with mode at 9.0.
Example 1 (The Gaussian Distribution)

The density, up to proportionality, of the standard Gaussian distribution is given by

\[ f(x) \propto e^{-\frac{1}{2}x^2} \]

A random sample of size 100 was drawn from this likelihood. For illustrative purposes, we performed the simulation on two different grid lengths. We considered two partitions of the grid from -4.0 to +4.0. The first partition consisted of equally spaced intervals of length 0.5, resulting in a total of 17 grid points. The second partition consisted of equally spaced intervals of length 0.1 each, resulting in 80 grid points. In terms of the notation in this paper, \( M = 16 \) and \( M = 79 \), respectively, for this example. The prior parameters of the Dirichlet distribution were calculated along the lines described in (13) and (14). Thus, for example, for the simulation involving 17 grid points, \( r \) equals eight and \( \sum_{j=1}^{17} \alpha_j^* = 91 \). Likewise, for the simulation involving 80 grid points, \( \sum_{j=1}^{80} \alpha_j^* = 1640 \). Finally, we set \( K = 1 \), and equal weight was assigned to each of the grid points to reflect vague prior belief about the location of the switch-point along the grid.

The posterior distribution of the switch-point \( t \) for varying grid lengths appears in Table 1. The distribution of the switch point in both of the simulations clearly identifies the ogive nature of the distribution function at or about zero.

Example 2: The Chi-Square Distribution

A random sample of size 100 was drawn from a chi-square distribution with 11 degrees of freedom. Two grid lengths starting at zero, and up to 15 were considered. The increments were set at one and 0.1 respectively for the two simulations, resulting in simulating on a grid consisting of \( M + 1 = 15 \) points and \( M + 1 = 100 \) points. For this example, \( \sum_{j=1}^{15} \alpha_j^* = 72 \) and \( \sum_{j=1}^{100} \alpha_j^* = 1275 \). \( K \) was set equal to one, and equal weight was assigned to each of the grid points to reflect vague prior belief about the location of the switch-point along the grid. The posterior distribution of the switch-point for varying grid lengths (15 and 100) appears in Table 2. The unimodal nature of the distribution function at about 9 is clearly identified by noting that the posterior mass of the distribution of the switch-point appears to be highly concentrated around the grid point of nine.

For both examples considered above, the estimates reported in the Tables are based on 5000 iterations and 200 cycles of the Gibbs Sampler along with 500 sampling-resampling drawings. While for the sake of brevity we have reported the results of one set of iterations/cycles for two different grids, we experimented with numerous iterations/cycles combinations to ensure numerical stability in the results.
IV. Concluding Remarks

In this paper we model unimodal distribution functions using a Bayesian nonparametric approach. A Dirichlet prior distribution is defined on slope transformations of the unknown and underlying distribution function. The resulting posterior quantities are calculated using MCMC methods. The model is illustrated using standard unimodal distributions such as the Gaussian.

References


Table 1. *Posterior distribution of the switch-point for the Gaussian density*

<table>
<thead>
<tr>
<th>Grid</th>
<th>Probability of switch-point</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.0</td>
<td>0</td>
</tr>
<tr>
<td>-3.5</td>
<td>0</td>
</tr>
<tr>
<td>-3.0</td>
<td>0</td>
</tr>
<tr>
<td>-2.5</td>
<td>0</td>
</tr>
<tr>
<td>-2.0</td>
<td>0</td>
</tr>
<tr>
<td>-1.5</td>
<td>0</td>
</tr>
<tr>
<td>-1.0</td>
<td>.07</td>
</tr>
<tr>
<td>-0.5</td>
<td>.19</td>
</tr>
<tr>
<td>0.0</td>
<td>.65</td>
</tr>
<tr>
<td>0.5</td>
<td>.04</td>
</tr>
<tr>
<td>1.0</td>
<td>.03</td>
</tr>
<tr>
<td>1.5</td>
<td>.02</td>
</tr>
<tr>
<td>2.0</td>
<td>0</td>
</tr>
<tr>
<td>2.5</td>
<td>0</td>
</tr>
<tr>
<td>3.0</td>
<td>0</td>
</tr>
<tr>
<td>3.5</td>
<td>0</td>
</tr>
<tr>
<td>4.0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: The probability of switch point was calculated based on a sample of size 5000 using the Gibbs Sampler. The Grid was partitioned into 17 parts starting at -4.0 with increments of 0.5.
Table 1 - continued

Posterior distribution of the switch-point for the Gaussian density

<table>
<thead>
<tr>
<th>Grid</th>
<th>Probability of switch-point</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-4.0 \leq S \leq .30$</td>
<td>0</td>
</tr>
<tr>
<td>-.20</td>
<td>.03</td>
</tr>
<tr>
<td>-.10</td>
<td>.03</td>
</tr>
<tr>
<td>0.0</td>
<td>.85</td>
</tr>
<tr>
<td>0.1</td>
<td>.09</td>
</tr>
<tr>
<td>$0.2 \leq S \leq 4.0$</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: The probability of the switch-point was calculated based on a sample of 5000 using the Gibbs Sampler. The Grid was partitioned into 80 parts starting at -4.0 with increments of 0.1.
Table 2

Posterior distribution of the switch-point for the Chi-square density with 11 degrees of freedom

<table>
<thead>
<tr>
<th>Grid</th>
<th>Probability of switch-point</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>.09</td>
</tr>
<tr>
<td>8</td>
<td>.22</td>
</tr>
<tr>
<td>9</td>
<td>.52</td>
</tr>
<tr>
<td>10</td>
<td>.08</td>
</tr>
<tr>
<td>11</td>
<td>.08</td>
</tr>
<tr>
<td>12</td>
<td>.01</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: The probability of switch point was calculated based on a sample of size 5000 using the Gibbs Sampler. The Grid was partitioned into 15 parts starting at 0.0 with increments of 1.0.
Table 2 - continued

Posterior distribution of the switch-point for the Chi-square density

with 11 degrees of freedom

<table>
<thead>
<tr>
<th>Grid</th>
<th>Probability of switch-point</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0 \leq S \leq 8.6$</td>
<td>0</td>
</tr>
<tr>
<td>8.7</td>
<td>0.04</td>
</tr>
<tr>
<td>8.8</td>
<td>0.01</td>
</tr>
<tr>
<td>8.9</td>
<td>0.05</td>
</tr>
<tr>
<td>9.0</td>
<td>0.77</td>
</tr>
<tr>
<td>9.1</td>
<td>0.08</td>
</tr>
<tr>
<td>9.2</td>
<td>0.05</td>
</tr>
<tr>
<td>$9.3 \leq S \leq 10$</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: The probability of the switch-point was calculated based on a sample of 5000 using the Gibbs Sampler. The Grid was partitioned into 100 parts starting at 0 with increments of 0.1.