Nonparametric Bayes methods using predictive updating

Michael A. Newton
Fernando A. Quintana
Yunlei Zhang

ABSTRACT Approximate nonparametric Bayes estimates calculated under a Dirichlet process prior are readily obtained in a wide range of models using a simple recursive algorithm. This chapter develops the recursion using elementary facts about nonparametric predictive distributions, and applies it to an interval censoring problem and to a Markov chain mixture model. S-Plus code is provided.

1 Introduction

Sampling models that enforce relatively weak assumptions are naturally favored in many applications, but it is well known that the corresponding posterior computations can become very intensive when a Dirichlet process encodes prior uncertainty in the weakly specified part of the model. In all but the most simple models, posterior calculations involve a mixture of Dirichlet processes. As evidenced by companion chapters, advances in Markov chain Monte Carlo (MCMC) provide critical methodology for enabling these calculations, and have opened up a wide range of interesting applications to Dirichlet-process-based nonparametric Bayesian analysis.

Although MCMC provides the most effective computational solution, other algorithms can be advantageous in certain situations, and these deserve further consideration. In this chapter we discuss a simple recursive algorithm derived by approximating posterior predictive distributions. Specifically, we review the recent proposal of Newton and Zhang (1996). It yields approximate Bayes estimates through a simple and efficient algorithm, and may be particularly helpful when the standard MCMC algorithm runs on a very high dimensional space, or if a fast approximate solution is helpful prior to full MCMC implementation.

In the class of models under consideration, an unobserved random sample \( \theta_1, \ldots, \theta_n \) from an unknown distribution \( G \) determines the conditional probability structure of observables \( Y_1, \ldots, Y_n \). We think of \( \theta_i \) as real or vector-valued; and observations can come in various types. The unobserved random sample may be missing data or it may represent parameters that
2 On $n = 1$

The most simple case motivates the general recursive approximation. The object of interest is the posterior predictive distribution

$$G_1(B) = P(\theta_2 \in B | Y_1)$$

$$= E \{ P(\theta_2 \in B | \theta_1) | Y_1 \}$$

$$= E \left\{ \frac{\alpha}{1 + \alpha} G_0(B) + \frac{1}{1 + \alpha} 1_B(\theta_1) \right\} | Y_1$$

$$= \frac{\alpha}{1 + \alpha} G_0(B) + \frac{1}{1 + \alpha} P(\theta_1 \in B | Y_1).$$

Critical in the above development is the representation of $P(\theta_2 \in B | \theta_1)$ as a mixture of prior opinion $G_0$ with information from $\theta_1$. A general argument supporting this claim comes from Polya sequence theory as developed in Blackwell and MacQueen (1973) and as discussed in Chapter **. A direct argument for the special case of $n = 1$ is to recall that the Dirichlet-process distributed $G$ may be expressed $G(B) = \sum_k u_k 1_B(v_k)$ where $v_1, v_2, \ldots$ are independent and identically distributed from $G_0$ and where the $u_k$'s arise from a simple stick-breaking exercise (Sethuraman, 1994). Specifically, $w_1 = b_1$, and for any $k \geq 2$, $w_k = b_k \prod_{i=1}^{k-1} (1 - b_i)$ where $b_1, b_2, \ldots$ are independent and identically distributed Beta$(1, \alpha)$ random variates. If $G_0$ is non-atomic, then conditional on $G$, the probability of a tie, that is that $\theta_1 = \theta_2$, is $\sum_k w_k^2$. Averaging over $G$, the tie probability is readily calculated to be $1/(1 + \alpha)$. On the other hand, if there is no tie, then $\theta_2$ must be distributed as $G_0$, hence (1.1). This argument requires slight elaboration if $G_0$ has atoms, but (1.1) continues to hold.
1. Nonparametric Bayes methods using predictive updating

Whereas the measure $m_0 = \alpha G_0$ represents uncertainty prior to observing data, the measure $m_1 = (\alpha + 1)G_1$ encodes updated uncertainties. We observe that Dirichlet-process based learning occurs by accumulating measure:

$$m_1(B) = m_0(B) + P_0(\theta_i \in B|Y_i).$$

(1.2)

It is interesting to note that Bayes rule enters the second term in (1.2). The corresponding posterior calculations are driven by the prior $G_0$ on $\theta$ (hence the subscripting in $P_0$). If no information is lost going from $\theta_1$ to $Y_1$, then the added measure is simply a point mass at $\theta_1$, and we have the familiar Polya sequence rule: after sampling with replacement from an urn, add another bit of mass at the observed value. More generally, the rule then is to return an entire probability distribution to the urn. It is also interesting that the so-often criticized discreteness property of $G$ coincides with the additive accumulation of measure.

3 A recursive algorithm

The idea that predictive uncertainty can be encoded in a single measure, and that learning occurs by adding measure lead us to the following recursion generalizing (1.2)

$$m_i(B) = m_{i-1}(B) + P_{i-1}(\theta_i \in B|Y_i)$$

(1.3)

for $i \geq 1$. In terms of probability distributions

$$G_i(B) = (1 - w_i)G_{i-1}(B) + w_iP_{i-1}(\theta_i \in B|Y_i)$$

(1.4)

where again Bayes rule enters the second term taking $G_{i-1}$ as the updated prior distribution for $\theta_i$. The nominal weights are $w_i = 1/(\alpha + i)$, though we consider some alternatives later.

As constructed, $G_i$ is not the exact posterior predictive distribution for $\theta_{i+1}$ in general, even though it is so when $i = 1$. It will be only in the case of no information loss, that is when $Y_i = \theta_i$. Thus the value of $G_i$ depends on the order in which $Y_1, \ldots, Y_i$ are processed. What gives some credence to the recursion is that the dependence on order can be relatively weak. The suggested algorithm then is to arrange in some order $Y_1, \ldots, Y_n$ and to process them through (1.4) to produce an approximate Bayes estimate $G_n$. Calculations being $O(n)$, we can easily re-evaluate $G_n$ over a random sample of orderings and average the results.

That (1.4) is order-dependent and approximate for $n > 1$ may not be obvious though the calculations in the next sections bear this out. In Section 7, we dissect a particular example with $n = 2$ to study this phenomenon.
The proposed recursion has the form of a stochastic approximation algo-

rithm (Kushner and Yin, 1997), one motivated by Dirichlet-process based
learning and one formally residing in a function space.

An important special case of the proposed recursion is the quasi-Bayes
sequential procedure discussed by Smith and Makov (1978) and elsewhere.
The canonical quasi-Bayes problem concerns finite mixture models. In our
notation, \( \theta_i \) indicates a component population from which feature data \( Y_i \)
are generated. The relevant sets \( B \) indicate the different component pop-
ulations, and \( G(B) \) is the mixing probability; i.e., the probability that an
observation is from population \( B \). Then, with a Dirichlet distribution prior
for the mixing probabilities, a quasi-Bayes procedure arises by approximating
the posterior distribution of these mixing probabilities in a certain way.
The recursive approximation proposed by Newton and Zhang and reviewed
here differs in several respects from the quasi-Bayes recursions, although
they do coincide in the finite mixture case. Aside from a difference in scope,
a general feature of the present recursive approximation is its emphasis on
posterior predictive distributions rather than on posterior distributions over
a parameter space. There are also some issues of implementation special to
the recursion discussed here, as we see in Sections 5 and 6.

4 Interval Censoring

The form of information loss that we have studied most extensively is inter-
val censoring. Rather than observing \( \theta_i \in \mathcal{X} \) we observe an interval \( Y_i \subset \mathcal{X} \)
that is known to contain \( \theta_i \). Interval censoring arises frequently in statis-
tical practice, and methods exist for obtaining nonparametric maximum
likelihood estimates of \( G \) (e.g., Groeneboom and Wellner, 1992; Gentleman
and Geyer, 1994). Less seems to have been done concerning nonparamet-
ric Bayesian approaches to this problem, although the MCMC methods in
Doss (1994) certainly apply. As an illustrative example, we consider in the
next section a much studied data set on cosmetic deterioration after radio-
therapy of the breast. Each patient is monitored periodically, and a time
to deterioration \( \theta_i \) is known to occur either in an interval between hospital
visits, or is right-censored and known to occur only beyond some maximum
observed time. A second example concerns the weight at one year of age
of calves on a large ranch in Brazil. At a time of round-up, calves ages are
known and weights are obtained, but the ages at round-up vary around one
year. Assuming that weight is nondecreasing during this period of growth,
the weight at one year \( \theta_i \) is known to exceed the measured weight if calf \( i \)
is younger than one year, and is known to be smaller than the measured
weight if calf \( i \) is older than one year. Data on this example are currently
being compiled.

The recursive formula (1.4) is particularly simple for interval censored
1. Nonparametric Bayes methods using predictive updating

data:

\[ G_i(B) = (1 - w_i) G_{i-1}(B) + w_i \frac{G_{i-1}(B \cap Y_i)}{G_{i-1}(Y_i)} \quad (1.5) \]

Newton and Zhang (1996) reported an interesting theoretical property of (1.5) in the restricted case that \( \mathcal{X} \) is a finite set. (In his thesis, Zhang extended this result to the countable support case.) Suppose that \( Y_1, Y_2, \ldots \) are independent and identically distributed random subsets of \( \mathcal{X} \) (for example, a sequence of observed intervals), and that the prior guess \( G_0 \) has support \( \mathcal{X} \). Then, with probability one, the sequence of approximate predictive distributions \( G_1, G_2, \ldots \) converges weakly to a distribution \( G^* \), say, which satisfies, for any \( B \subset \mathcal{X} \),

\[ G^*(B) = E \left\{ \frac{G^*(B \cap Y)}{G^*(Y)} \right\}. \quad (1.6) \]

Randomness on the right comes through the generic random subset \( Y \) which is distributed as the other \( Y_i \).

Note that (1.6) holds for any distribution of the random subsets and so further consideration of the problem is needed to connect this to samples from \( G \). In what some authors call case I and case II interval censoring, random censoring times partition \( \mathcal{X} \) into some number of subsets, and the observed \( Y_i \) is that subset which contains a random \( \theta_i \) from \( G \). To simplify the discussion, suppose that this censoring process partitions \( \mathcal{X} \) into two random subsets \( \mathcal{X} = A_i \cup A_i^c \), independently from \( \theta_i \). Then

\[ Y_i = \begin{cases} A_i & \text{if } \theta_i \in A_i \\ A_i^c & \text{if } \theta_i \in A_i^c \end{cases} \quad (1.7) \]

and (1.6) becomes

\[ G^*(B) = E \left\{ \frac{G^*(B \cap Y)}{G^*(Y)} \right| A \right\} \]

\[ = E \left\{ \frac{G^*(B \cap A)}{G^*(A)} 1_{A}(\theta) + \frac{G^*(B \cap A^c)}{G^*(A^c)} 1_{A^c}(\theta) \right| A \right\} \]

\[ = E \left\{ \frac{G^*(B \cap A)}{G^*(A)} G(A) + \frac{G^*(B \cap A^c)}{G^*(A^c)} G(A^c) \right\}. \]

Certainly one solution to this system is \( G^* = G \), owing to the independent censoring. Multiple solutions may exist depending on the censoring process, but if this is sufficiently rich, then \( G \) is the only solution, and hence the recursive approximations are consistent.

We present a simple numerical example to illustrate these points. Figure 1 shows the results of applying recursion (1.5) to random subsets \( Y_i \) of \( \mathcal{X} = \{0, 1, \ldots, 49\} \) formed as follows. A random partition of \( \mathcal{X} \) is formed
by independent coin tossing on elements of \( \mathcal{X} \), with heads going into a set \( A_i \) and tails into \( A_i^c \). Then a random \( \theta_i \) is sampled from a distribution \( G \) having masses \( g(\theta) \propto \tan(c\theta) \), where \( c = \pi/100 \); the cumulative distribution function is indicated in dotted lines in Figure 1. Data \( Y_1, \ldots, Y_n \) arise as in (1.7). We applied the recursive approximation (1.5) to these simulated data, obtaining approximate predictive distributions \( G_n \) shown in solid lines in Figure 1. The recursion was evaluated for sets \( B = [0, \theta] \) with \( \theta \in \mathcal{X} \). The reader may be surprised that so many samples are required for \( G_n \) to approach \( G \), but remember that there is extreme information loss in this example. Also, cube-root rather than square-root asymptotics govern standard estimators with such information loss (Groeneboom and Wellner, 1992).

For the calculations reported in Figure 1 we did not use nominal weights \( w_i = 1/(i + \alpha) \) suggested by (1.1). Inspection shows that these give relatively high weight to the first observations and quite low weight to later observations. It turns out that convergence holds as long as the positive weights satisfy \( \sum_i w_i = \infty \) and \( \lim_{k \to \infty} w_i = 0 \). We use \( w_i = 5_i^{-1/3} \) in Figure 1.

5 Censoring Example

The theory outlined in the last section does not say how well the recursive approximations match the actual posterior predictive distributions of interest. We study in this section a small example concerning the effects of radiotherapy on cosmetic deterioration of the breast. This example has been considered by a number of authors. Our goal is not to provide further insight into this application, but rather to illustrate how the recursive approximations work in a typical problem. The data are reported in Table 1 of Finkelstein and Wolfe (1985), in Table 1 of Gentleman and Geyer (1994), and are available at the first author’s web site through http://www.stat.wisc.edu/. There are \( n = 46 \) intervals in the data set. Associated with each woman in the study is a time \( \theta_i \) (days) indicating the time after radiotherapy treatment that a defined amount of change occurs in the treated tissue, and the \( i \)th interval is known to contain \( \theta_i \).

Figure 2 shows the results of our recursive approximation (1.5) using the nominal weighting scheme. We take \( G_0 \) to be an exponential prior with mean equal to one half a year, and we consider two values of \( \alpha \); 1 and 5 (corresponding to the left and right sides in Figure 2). Panels (a) and (b) show \( G_{40} \) for 100 random orderings of the data. We use sets \( B = [0, \theta] \) to drive the recursion, for \( \theta \) in the grid \( \{0, 1, 2, \ldots, 100, 1000\} \). The outlying grid point at 1000 days is included to account for mass beyond the range of data. (On a related point, the nonparametric MLE is typically a subdistribution function, with positive mass at an arbitrary point beyond the
FIGURE 1. Extreme censoring: The true distribution function $G$ is the dotted curve, and recursive approximations $G_n$ are marked as solid curves after various sample sizes. Vertical axis is cumulative probability and horizontal axis is $X$. 

\begin{align*}
n &= 0 \\
n &= 2800 \\
n &= 5600 \\
n &= 8000
\end{align*}
FIGURE 2. Interval censoring, nominal weights: Vertical axis is cumulative probability and horizontal axis is \( X \) in days in all panels. Panels (a) and (c) refer to calculations with prior mass \( \alpha = 1 \), and panels on the right correspond to \( \alpha = 5 \). Upper panels (a) and (b) show \( G_\tau \) calculated for 100 different random orderings of the \( n = 46 \) cases, with dashed line indicating the prior guess \( G_0 \). Solid lines in lower panels are pointwise averages of \( G_\tau \), and dotted lines are Gibbs sampler output.
FIGURE 3. Interval censoring, square root weights. Descriptions are as in Figure 2.
largest censoring time.) The lower panels (c) and (d) compare the pointwise average over orderings to a Gibbs sampler approximation. Similar results are shown in Figure 3, but here we use weights \( w_i = 1/[(1 + \alpha)/\sqrt{2}] \). The average over orderings \( G_n \) is a very accurate approximation to the correct posterior predictive distribution in this example.

The Gibbs sampler calculation used for comparison here is an adaptation of the algorithm given by Escobar (1994) (see Chapter **). It was run for 500\(^2\) complete scans, subsampled to produce 500 vectors \( (\theta_1, \ldots, \theta_{40}) \) drawn approximately from their posterior distribution. Time series output analysis indicated that this represents an informative posterior sample. The dotted curves in panels (c) and (d) of Figures 2 and 3 show the marginal empirical distribution of the entire collection.

Appendix I provides some code to implement the recursion using built-in functions in S-Plus (e.g., Venables and Ripley, 1994).

6 Mixing Example

In a related class of models, the observed \( Y_i \) has some conditional density or mass function \( p(y|\theta) \) given that \( \theta_i = \theta \). The general recursion (1.4) may be expressed in terms of predictive densities instead of distribution functions:

\[
g_i(\theta) = (1 - w_i)g_{i-1}(\theta) + w_i g_{i-1}(\theta)p(\theta_i|\theta)/c_i
\]

where \( c_i \) ensures that the posterior distribution on the right integrates to 1, that is \( c_i = \int g_{i-1}(\theta)p(\theta_i|\theta)\,d\theta \).

To illustrate (1.8), we consider modeling a set of survey data concerning employment status of youth in the United States. We confine attention to a sample of \( n = 2390 \) individuals from the National Longitudinal Survey of Youth, as in Quintana and Newton (1998). Briefly, annual employment history of each individual is recorded over time for up to thirteen years. For analysis here, data are summarized into binary indicators of employment during each year, yielding a binary time series for each individual. We account for positive correlation by supposing that each \( Y_i = (Y_{i,1}, \ldots, Y_{i,13}) \) is a binary Markov chain, with some person-specific transition matrix

\[
\begin{pmatrix}
p_i & 1-p_i \\
1-q_i & q_i
\end{pmatrix},
\]

determined by the parameter vector \( \theta_i = (p_i, q_i) \). Now we take as prior assumptions that \( \theta_i \sim G \), and that \( G \) is a Dirichlet process, so the recursive equations can be invoked to approximate \( E(G(B)|D) = P(\theta_{n+1} \in B|D) \).

We work with densities, noting that the Markov assumption implies that in (1.8),

\[
p(Y_i|\theta_i) \propto p_i^{t_{0,0}}(1-p_i)^{t_{0,1}}q_i^{t_{1,1}}(1-q_i)^{t_{1,0}}
\]
FIGURE 4. Markov chain mixtures. Shown are contours of recursive approximations to the posterior predictive distribution of \( \theta_{n+1} = (p_{n+1}, q_{n+1}) \), the non-transition probabilities. The four panels correspond to results from different random orderings of the \( n = 2300 \) binary sequences. Contours define probability content at the levels \( .10, .25, .50, .75, .90, .95, .99 \).
where $t_{j,k}$ counts the transitions from $j$ to $k$ in the binary time series $Y_t$.

The move from one to two dimensions on $\theta$ creates no significant problems. A simple way to invoke the recursion (1.8) is to work on a grid in the unit square, which we do in the following calculations. (A speedier solution takes advantage of the smoothness, using Gauss-Legendre quadrature, as in Tao et al., 1997.) Starting with a uniform prior guess $G_0$, so $g_0(\theta)$ is constant, we run (1.8) using the weight sequence $w_t = \sqrt{t}$. We simply sum over the $100 \times 100$ grid after each step to calculate $\alpha_t$. Figure 4 shows contours of $G_n$ from four different random orderings of the data. Evidently, there is very little variation created by the processing order.

For comparison purposes, we also approximate the posterior predictive distribution of $\theta_{n+1}$ using MCMC. We adapted the Markov chain constructed in Bush and MacEachern (1996) (and discussed in Chapter **), running it for 10,000 complete scans after a short burn-in period. This chain moves through the $2n$ dimensional space of all $\theta$ values, nearly 5,000 dimensions in this example. Independent runs were performed for different values of the prior mass $\alpha$, and we were encouraged to see reasonably rapid mixing as measured by simple time-series diagnostics on a few one-dimensional summaries. The posterior predictive distribution of interest is obtained from the Monte Carlo sample by collapsing all dimensions and recording the marginal empirical distribution of the $\theta$'s. Little information is lost if we simply accumulate counts in bins defined by the same $100 \times 100$ grid used in the recursive approximation. We smoothed by a very small bit of local averaging this empirical distribution before plotting contours, as shown in Figure 5.

Generally, there is a close agreement between the MCMC approximation and the recursive approximation. There certainly are differences. For small values of $\alpha$, the recursive approximation oversmooths slightly, missing what may be distinct modes in the true posterior predictive distribution. For larger values of $\alpha$ the two approximations agree quite well. We have reported in Figure 5 just an intermediate case, $\alpha = 5$. Of course an advantage of the recursive approximation is its computational simplicity.

A well-studied and somewhat simpler example has observations $Y_t$ binomially distributed with success probability $b_t$. Liu (1996) among others has studied the nonparametric Bayesian analysis of this problem, illustrating calculations on an interesting set of data on rolling tasks. In Appendix II we provide S-Plus code to implement the recursive approximation for this example.

7 On $n = 2$

It is difficult to make a direct comparison of the recursive approximation with the exact Bayes estimate unless we consider particular numerical ex-
amples or asymptotic properties. When \( n = 2 \), however, a comparison is quite feasible. To avoid technicalities, we work with densities as in Section 6. We can compute \( P(\theta_3 \in B|Y_1, Y_2) \) exactly by noting that it equals the conditional expectation of the prior probability \( P(\theta_3 \in B|\theta_1, \theta_2) \) given \( Y_1 \) and \( Y_2 \), and then by noting the mixture structure of the Polya sequence prior. Calculations reveal that this distribution is a mixture, with density

\[
a_0f_0(\theta) + a_1f_1(\theta) + a_2f_2(\theta) + a_{12}f_{12}(\theta),
\]

(1.9)

Here, \( f_0(\theta) = g_0(\theta) \), the prior guess, \( f_1(\theta) \) is the posterior of \( \theta \) if we were to observe \( Y_1 \) only, \( f_2(\theta) \) is the same given \( Y_2 \), and \( f_{12}(\theta) \) is the posterior of \( \theta \) if \( Y_1 \) and \( Y_2 \) are independent and identically distributed from the common \( p(y|\theta) \). Furthermore, the mixing proportions are

\[
a_0 = \frac{\alpha}{\alpha + 2} \quad \quad \quad a_1 = a_2 = \frac{\alpha}{(\alpha + 1)(\alpha + 2)} \quad \frac{p(Y_1)p(Y_2)}{p(Y_1, Y_2)}
\]

and \( a_{12} = 1 - a_0 - a_1 - a_2 \). These are prior predictive probabilities in the mixing weights. Consistent with our understanding of the role played by the prior mass \( \alpha \), we see that the \( f_0 \) component dominates for large \( \alpha \) and the \( f_{12} \) component dominates for \( \alpha \) near 0. Also, symmetry in \( Y_1 \) and \( Y_2 \) is evident in (1.9).

Now we turn to the recursive approximation (1.8). Obviously,

\[
\begin{align*}
g_1(\theta) &= (1 - w_1)g_0(\theta) + w_1g_0(\theta)p(Y_1|\theta)/c_1 \\
g_2(\theta) &= (1 - w_2)g_1(\theta) + w_2g_1(\theta)p(Y_2|\theta)/c_2
\end{align*}
\]

and so by direct substitution, solving for \( g_2 \), we get

\[
g_2(\theta) = b_0f_0(\theta) + b_1f_1(\theta) + b_2f_2(\theta) + b_{12}f_{12}(\theta),
\]

(1.10)

The recursive approximation produces a mixture of the same type as the correct answer (1.9), but with different mixing proportions. Using the nominal weights, \( b_0 \equiv a_0 \), but

\[
b_1 = \frac{\alpha + 1}{(\alpha + 2)^2} \quad \quad \quad b_2 = b_1c_2/c_2,
\]

where we recall that \( c_2 \) is the normalizing constant above. Interestingly, \( c_2^* \) is a slightly different normalizing constant, being \( \int p(Y_1|\theta)p(Y_2|\theta)g_0(\theta) \, d\theta \), i.e., the normalizer in \( f_{12} \). Also the asymmetry in (1.10) with respect to \( Y_1 \) and \( Y_2 \) is clear.

8 Concluding Remarks

Recursive approximations are readily obtained for the posterior predictive distributions in Dirichlet-process-based nonparametric Bayesian analysis.
A major advantage of this approach is that computations are extremely simple and thus can be deployed rapidly in a wide range of applications. They do not require keeping track of cluster structure among unobserved \( \theta \) values, and thus have very low coding, storage, and CPU requirements. As a practical matter, the recursions might be useful as advance tools prior to full-blown implementation of MCMC. The accuracy of the recursive approximations is high in the examples studied here and in Newton and Zhang (1996), although in general this accuracy will depend on the particular data and prior, and further investigation is certainly warranted. The censoring examples we have studied exhibit quite high accuracy. In the missing examples considered so far accuracy is also high but there is some indication of oversmoothing by the recursive approximation when the prior mass is small.

9 References


1. Nonparametric Bayes methods using predictive updating


FIGURE 5. Comparing Methods: The left panel shows an average of distributions as in Figure 4 from 10 random orderings of the data. The right panel shows an MCMC approximation when $\alpha = 5$. Contours are defined as in Figure 4.
Appendix I: Splus code for censoring example

# Have left endpoints of intervals in a vector
# 'lefts' and right endpoints in 'right' and
# set 'N' equal to the common length.

# Partition support (0,infty) into sets B
grid <- c( seq(0,100,by=1), 1000 )
ngrid <- length(grid)
grt <- grid[2:ngrid]; glt <- grid[1:(ngrid-1)]

# Exponential prior guess and prior sample size
gg <- exp( -glt/(365/2) ) - exp( -grt/(365/2) )
alpha <- 1

# Identify partition elements that may contain
# each survival time
less <- function(x,y){ return( x<y ) }
ma <- t( outer(grt,rights,FUN="less") )
mb <- outer(lefts,glt,FUN="less")
dmat <- ma&mb

# Recursion yields approximate Bayes estimate gg
weight <- 1/(alpha+1:N) # Nominal weight sequence
ord <- sample( 1:N ) # Process in random order
for( i in 1:N ) # Process in random order
{
  ok <- dmat[ord[i],] # A
  numer <- rep(0,ngrid-1)
  numer[ok] <- gg[ok] # G(B and A)
  denom <- sum( numer ) # G(A)
  gg <- gg*( 1-weight[i] ) + weight[i]*numer/denom
}

# Repeat loop to see variation over orderings.

# Approximate Bayes estimate of distribution function.
plot(grid, cumsum( c(0,gg) ), type="s", xlim=c(0,100),
     xlab="time (days)", ylab="cumulative probability")
Appendix II: Splus code for binomial mixture example

# Beckett Diaconis Tack Data
success <- c(rep(1,3), rep(2,13), rep(3,18), rep(4,28),
    rep(5,47), rep(6,87), rep(7,54), rep(8,51), rep(9,19))
ntials <- 9; N <- length(success)

# Support of mixing distribution
grid <- seq(0.01,.99,length=100); delta <- grid[2]-grid[1]

# Beta prior guess and prior sample size
gg <- dbeta(grid, shape1=.5, shape2=.5); alpha <- 1/3

# Binomial likelihood
db2 <- function(y, prob, n){return(dbinom(y, n, prob))}
lik <- outer(success, grid, FUN="db2", n=ntials)

# Recursion yields approximate Bayes estimate gg
weight <- 1/sqrt((alpha+1)*(alpha+1:N)) # A weight sequence
ord <- sample(1:N) # Process tacks in random order
for( i in 1:N )
{
    post <- lik[ord[i],]*gg
    post <- (post/sum(post))/delta
    gg <- gg*(1-weight[i]) + weight[i]*post
}
# Repeat loop to see variation over orderings.

# Estimated predictive density for success probability
# of a future tack. Compare to Fig. 2, Liu (1996).
plot( grid, gg, type="l", xlab="tack success probability",
ylab="posterior predictive density" )