Appendix

A.1. Implementing DP Mixtures in R

We introduce some R macros to implement inference in a DP mixture in (3.9). Using Gaussian kernels the DP mixture model becomes

$$y_i \mid G \sim F(y_i) = \int N(y_i;\ \theta_i, \sigma)\ dp(G). \tag{A.10}$$

We will use $f(x)$ to denote the p.d.f. The moel can equivalently be written as a hierarchical model

$$y_i \mid \theta_i \sim N(\theta_i, \sigma), \qquad \theta_i \mid G \sim G. \tag{A.11}$$

where $G \sim \mathsf{DP}(M, G_0)$. For the centering measure, we use $G_0 = N(m_0, B_0)$, and we take $m_0 = 0$ and $B_0 = 4$, while the precision parameter is set to $M = 1$. We complete the model with a gamma prior for the kernel width, $1/\sigma^2 \sim \mathsf{Ga}(a, b)$; in the example below we take $a = b = 1$. We implement posterior MCMC simulation using the methods described in §3.3. The complete R code is available at

`http://www.math.utexas.edu/users/pmueller/prog/BNPnotes/`

We briefly explain the main steps in the macros. We use the data from the Old Faithful geyser data in R and fix the hyperparameters for the model. The hyperparameters and the data are saved as global variables:

```
## DATA: Old Faithful geyser data
y <- round(faithful$eruptions, digits=2)
n <- length(y)
## hyperparameters
a  <- 1;   b <- 1    # 1/sig ~ Ga(a,b)
m0 <- 0;  B0 <- 4    # G0 = N(m0,B0)
M  <- 1
```

Collapsed Gibbs Sampler – Conjugate Models

We first implement the Gibbs sampler from §3.3.1. The algorithm consists of the following steps:

0. Initialization: We initialize s_i using (deterministic) hierarchical clustering. Initialize a plot by plotting a kernel density estimate of $f(y)$.
1. Update θ_i: sample from $p(\theta_i \mid \boldsymbol{\theta}_{-i}, \sigma, \mathbf{y})$, $i = 1, \ldots, n$. See (3.13).
2. Update $\theta^\star_j$: sample from $p(\theta^\star_j \mid s_i, \sigma, \mathbf{y})$, $j = 1, \ldots, k$, as in (3.14).
3. Update σ^2: sample from $p(\sigma^2 \mid \mathbf{s}, \boldsymbol{\theta}^\star, \mathbf{y})$.
4. Generate f: generate $f \sim p(f \mid \boldsymbol{\theta}, \phi, \sigma)$; add f to the plot.

In Step 4 we sample f conditional on the currently imputed values of $\boldsymbol{\theta}, \phi, \sigma$. For plotting it suffices to evaluate f on a grid that is fine enough to get a smooth plot. To sample f (on the grid) we use a minor approximation. Essentially we (i) sample $G \sim p(G \mid \boldsymbol{\theta})$, and (ii) evaluate $F = \int \mathsf{N}(\theta, \sigma)\, dG(\theta)$. The approximation