Supplementary material to "Bayesian Hierarchically Weighted Finite Mixture Models for Samples of Distributions"

Abel Rodriguez,[1,*] **David B. Dunson**[2] **and Jack Taylor**[3]

[1]Department of Applied Mathematics and Statistics, University of California, Santa Cruz

[2]Biostatistics Branch, National Institute of Environmental Health Sciences, U.S. National Institutes of Health

[3]Epidemiology Branch, National Institute of Environmental Health Sciences, U.S. National Institutes of Health

1. Simulation study

In this section we compare pairs of samples composed of 100 observations each, using the model discussed in the paper. For each simulated data set, we wish to determine whether both samples arise from the same underlying distributions, providing some insight into the frequentist properties of the testing procedure. As summarized in Table 1 and Figure 1, we simulated data under three different cases, with each simulated data set analyzed using three different models corresponding to $K \in \{3, 10, 20\}$. The first case corresponds to only a subtle change in shape, the second mimics the DNA damage data in exhibiting a shift in location and shape, and the third case case corresponds to the null hypothesis of equality between both true distribution.

The boxplots in Figure 2 show the performance of the model over 50 simulations carried out

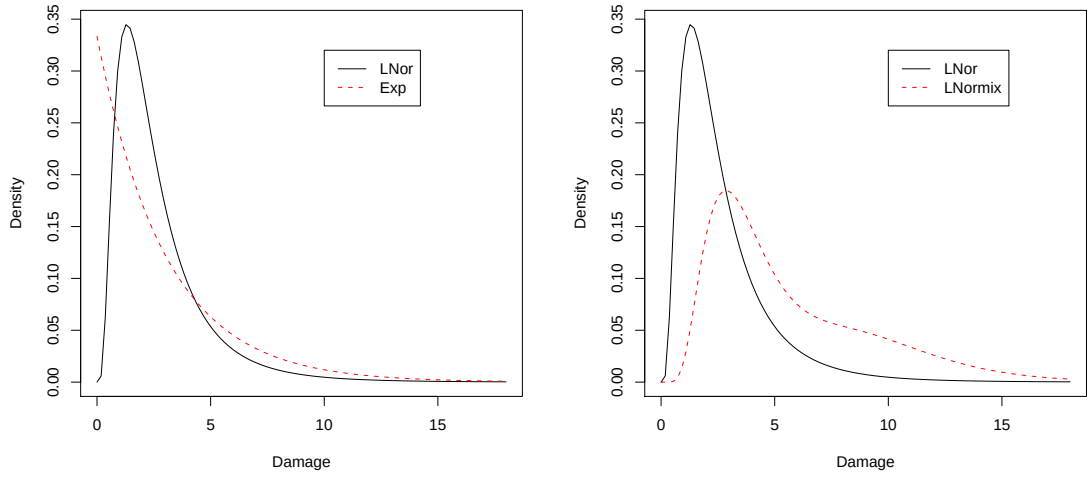
[*] *Corresponding author's email address:* abel@ams.ucsc.edu

Key words: Hierarchical functional data, Finite mixture model, Comet assay, Samples of distributions, Stochastic search, Latent factor, Genotoxicity

Table 1

True distributions in each case of the simulation study.

Case	Distribution 1	Distribution 2
1	$\mathsf{LN}(0.752, 0.693)$	$\mathsf{E}(3.000)$
2	$\mathsf{LN}(0.752, 0.693)$	$0.75\mathsf{LN}(1.300, 0.25) + 0.25\mathsf{LN}(2.300, 0.090)$
3	$\mathsf{LN}(0.752, 0.693)$	$\mathsf{LN}(0.752, 0.693)$

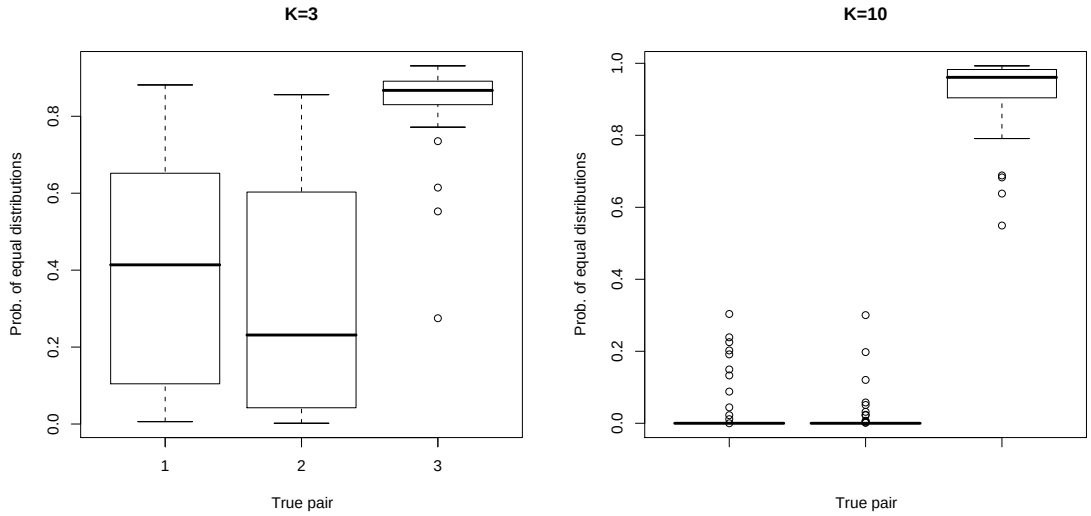


(a) (b)

Figure 1. True distributions in the simulation example. Panel (a) corresponds to case 1, while panel (b) corresponds to case 2. Case 3 uses the same lognormal distribution in panels (a) and (b) to generate both datasets.

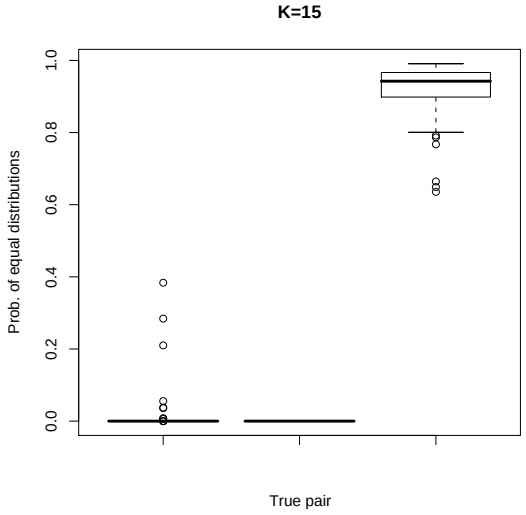
under each of 9 combinations discussed above. Panels (a), (b) and (c) correspond to $K = 3$, $K = 10$ and $K = 20$ respectively. Note that when both samples arise from a common distribution (case 3 in all three panels), the model correctly reports high posterior probabilities for the null model. Indeed, at least 75% of the simulations report a posterior probability over 0.8, and around 25% report values over 0.9. On the other hand, when the distributions are indeed different, this probability tends to drop dramatically, especially when a moderately large number of mixture components are used. In the case $K = 3$, when very few components are used, the model has a harder time differentiating across groups, but it still capable of selecting the correct model more than 50% of the time. When the number of component grows, the model is capable of finding differences most of time. Indeed, over 75% of the simulations for $K = 10$ and $K = 15$ have an estimated posterior probability for the null hypothesis of no difference in both of the first two simulation cases. This is in contrast with regular nonparametric tests like the Wilcoxon rank-sum test, which have power under 20% for the examples displayed here and do not allow for hierarchical structures. The plots also show that little additional advantages are obtained by increasing the number of components from 10 to 15.

The explanation is relatively straightforward; if too few components are used, the models need to compromise in order to fit all distributions as closely as possible, possibly masking their differences. On the other hand, too many components add little to the ability of the model to fit the empirical distributions. Similar results (not shown) arise when comparing more than two distributions. In particular, we note that the probability of false negatives does not fall significantly.



(a)

(b)



(c)

Figure 2. Simulation results, showing the probability that both samples arise from the same probability distributon. Each boxplot corresponds to 50 simulations, under the nine conditions in our simulation study .