Statistica Sinica **24** (2014), 1717-1733
doi:http://dx.doi.org/10.5705/ss.2013.040

TWO-SAMPLE HYPOTHESIS TESTING UNDER LEHMANN ALTERNATIVES AND POLYA TREE PRIORS

Lei Huang and Malay Ghosh

University of Florida

Abstract: The paper revisits two-sample hypothesis testing problems under Lehmann alternatives. We consider the problem in a fully Bayesian nonparametric framework with Polya tree priors. Our findings are expected to be useful in life testing and survival analysis where Dirichlet process priors are not quite suitable. We derive Bayes factors for some fixed power in the Lehmann alternative and also for the case where the power is treated as a parameter. Our Bayesian solution has a closed form even for censored data. It can be calculated easily and has a ready interpretation.

Key words and phrases: Bayes factor, order statistics, ranks, spacings.

1. Introduction

Lehmann alternatives were introduced by Lehmann (1953) in the two-sample hypothesis testing context. Specifically, one considers independent random samples, $X_1, \ldots, X_{n_1}$ each distributed as $F(x)$ and $Y_1, \ldots, Y_{n_2}$ each distributed as $H(x)$. One tests $H_0 : F = H$ against the alternatives $H_1 : H(x) = 1 - \{1 - F(x)\}^\alpha$, where $\alpha > 0$ and $\alpha \neq 1$. The power α is usually treated as a prespecified value, $\alpha = \alpha_0$, say. Lehmann considered these tests not just for their mathematical simplicity, but also for straightforward interpretation of alternatives. For instance, when α_0 is an integer, $H(x)$ is the distribution function of the minimum of α_0 independent random variables each having distribution function F. Moreover, the alternatives, in general, introduce a very natural stochastic ordering of F and H. In particular, H is stochastically larger (smaller) than F when $0 < \alpha_0 < 1$ ($\alpha_0 > 1$). Clearly, Lehmann alternatives are also useful in survival analysis since the survival function is $1 - F(x)$ under the distribution function F. A well-known example is the Cox (1972) proportional hazards model where the null hypothesis corresponds to a zero regression vector. Another important application was pointed out by Davies (1971). If X_i and Y_i are lifetimes of similar articles, the hypothesis asserts that the failure rate of one is a constant multiple of the failure rate of the other.

Much of the literature on Lehmann alternatives is restricted to rank order tests. In particular, these alternatives are used to compare the performance of

locally most powerful rank tests against uniformly most powerful rank tests for specific alternatives, as in Lehmann (1953) and Savage (1956). Subsequently, Davies (1971) showed asymptotic equivalence of the approaches of Lehmman and Savage.

Brooks (1974) addressed the problem from a Bayesian perspective. The power α was regarded as an unknown parameter in his model. He began with the joint distribution of rank order statistics as derived by Savage (1956), and then assigned an F prior to α to complete the analysis.

A key feature of these articles is that the tests are based only on the ranks. Here not just the ranks, but also the spacings of order statistics are taken into account in the testing problem. We treat F as fully nonparametric and assign priors to F.

Bayesian nonparametric methods have received extensive attention because of their flexibility. Dirichlet process (DP) priors, introduced by Ferguson (1973), are the most commonly used nonparametric priors, and a large body of theory has been developed for them. Their scope is somewhat limited by the fact that they select, with probability 1, only discrete probability measures. Accordingly, it may not be reasonable to employ Dirichlet process priors, for example, in the survival context.

There have been other approaches to accommodate continuous problems. For instance, Kalbfleisch (1978) defines a family of random probability distributions called the Gamma process. Hjort (1990) discussed Beta processes in the context of survival analysis. Beta process priors were also used by Damien and Walker (2002) for testing the effects of two treatments. The Dirichlet process mixture model introduced by Escobar and M. West (1995) received great success in Bayesian nonparametrics. Antoniak (1974) proposed another approach, which he referred to as mixtures of Dirichlet processes to "smooth" a Dirichlet process that it gives positive probability to continuous distributions. However, the posterior probability of a tie continues to be positive. We are interested in Polya tree priors, originally introduced by Fabius (1964), while Ferguson (1974) termed them Polya tree processes. These priors can select continuous distributions with positive probability and, if necessary, even with probability 1. Lavine (1992, 1994) investigated the basic properties of Polya tree priors. Sufficient conditions for these priors to assign probability 1 to the set of continuous distributions are discussed in Mauldin, Sudderth, and Williams (1992) (MSW) and Lavine (1992). Muliere and Walker (1997) discussed how Polya tree priors could be used in survival analysis without covariates. Hanson (2006) and Hanson and Jara (2012) utilized mixtures of Polya tree (MPT) priors in a variety of survival models. The reason we use Polya tree priors instead of MPT is that we end up with an explicit expression of the Bayes factor by letting the Polya tree be data

dependent. Chen and Hanson (2014) proposed a test for $H_0 : F = H$ vs. $H_1 : F \neq H$ that uses a Polya tree prior centered at a distribution that is estimated from a parametric fit. However, they did not obtain the Bayes factor in a closed form, and required MCMC methods to carry out the computation.

The outline of the remaining sections is as follows. The definition and basic properties of Polya tree priors are reviewed in Section 2. Section 3 deals with the case when the power is set to be some pre-specified value. In Section 4, the power α is treated as an unknown parameter. We assign a spike and slab prior to α and compute the posterior probability of a null hypothesis. Section 5 provides a summary of our work and some pointers for future research.

2. Polya Tree Process

Polya tree processes form a large class of priors that includes the Dirichlet process as a special case. The tree is constructed by successive partitioning of the sample space. The partition plays a deterministic role in Polya trees, and a large collection of parameters makes it possible to incorporate a wide range of beliefs.

Let $E = \{0, 1\}$, E^m be the m-fold product $E \times E \times E \times \cdots \times E$, $E^0 = \emptyset$, and $E^* = \bigcup_0^\infty E^m$. Define a separating binary tree of partitions of Ω, $\Pi = \{\pi_m, m = 0, 1, 2, \ldots\}$, such that $\pi_0 = \Omega$. Here $\pi_0, \pi_1, \ldots$ forms a sequence of partitions such that $\bigcup_0^\infty \pi_m$ generates the measurable sets and every $B \in \pi_{m+1}$ is obtained by splitting some $B' \in \pi_m$ into two sets. Degenerate splits are permitted, i.e. some $B \in \pi_m$ can be split into $B \bigcup \emptyset$.

For each m, $\pi_m = \{B_{\vec{\epsilon}_m} : \vec{\epsilon}_m = \epsilon_1, \ldots, \epsilon_m \in E^m\}$ is a partition of Ω such that for all $\vec{\epsilon}_m \in E^*$, $B_{\vec{\epsilon}_m,0}$, $B_{\vec{\epsilon}_m,1}$ is a partition of $B_{\vec{\epsilon}_m}$. Let $A = \{a_{\vec{\epsilon}_m} : \vec{\epsilon}_m \in E^*\}$ be a set of nonnegative real numbers and $\mathfrak{y} = \{Y_{\vec{\epsilon}_m} : \vec{\epsilon}_m \in E^*\}$ be a collection of random variables. Following Lavine (1992), we say a random probability measure P on Ω has a Polya tree distribution with parameter (Π, A), written as $P \sim PT(\Pi, A)$, if

I. $Y_{\vec{\epsilon}_m,0}$, for all $\vec{\epsilon}_m \in E^*$, are independent; $Y_{\vec{\epsilon}_m,1} = 1 - Y_{\vec{\epsilon}_m,0}$;
II. for every $\vec{\epsilon}_m \in E^*$, $Y_{\vec{\epsilon}_m,0}$ has a Beta distribution with parameters $a_{\vec{\epsilon}_m,0}$ and $a_{\vec{\epsilon}_m,1}$;
III. for every $m = 1, 2, \ldots$ and every $\vec{\epsilon}_m \in E^*$,

$$P(B_{\epsilon_1,\ldots,\epsilon_m}) = \Big(\prod_{j=1;\epsilon_j=0}^{m} Y_{\epsilon_1,\ldots,\epsilon_j}\Big) \prod_{j=1;\epsilon_j=1}^{m} (1 - Y_{\epsilon_1,\ldots,\epsilon_{j-1},0}) = \prod_{j=1}^{m} Y_{\epsilon_1,\ldots,\epsilon_j}. \quad (2.1)$$

Here the form of $P(B_{\epsilon_1,\ldots,\epsilon_m})$ differs from what is given in Lavine (1992) by re-arranging $Y_{\vec{\epsilon}_m}$ and defining $Y_{\vec{\epsilon}_m,1} = 1 - Y_{\vec{\epsilon}_m,0}$. With it, we get a compact expression for $P(B_{\epsilon_1,\ldots,\epsilon_m})$, noting that $Y_{\vec{\epsilon}_m,0}$ and $Y_{\vec{\epsilon}_m,1}$ are not independent.

Properties of Polya tree processes are listed below. For more properties, see Lavine (1992) and Ghosh and Ramamoorthi (2003).

1. Polya trees are conjugate. If P has a Polya tree distribution, and $X \mid P \sim P$, then $P \mid X$ has a Polya tree distribution. The posterior distribution is updated as follows: for every $\vec{\epsilon}_m$ such that $X \in B_{\vec{\epsilon}_m}$, add 1 to $a_{\vec{\epsilon}_m}$. If we only observe that an X is in some set I, then for every $\vec{\epsilon}_m$ such that $B_{\vec{\epsilon}_m} \supset I$, add 1 to $a_{\vec{\epsilon}_m}$.

2. Some Polya trees assign probability 1 to the set of continuous distributions. A sufficient condition for this can be found in Theorem 3.3.7 in Ghosh and Ramamoorthi (2003).

3. If we have a Polya tree with partitions $\{B_{\vec{\epsilon}_m} : \vec{\epsilon}_m \in E^*\}$ and parameters A, the predictive density at $x \in B_{\vec{\epsilon}_m}$ is

$$\begin{aligned} f(x) &= \lim_{m\to+\infty} \frac{Pr(B_{\vec{\epsilon}_m})}{\lambda(B_{\vec{\epsilon}_m})} \\ &= \lim_{m\to+\infty} \frac{\prod_{i=1}^{m} a_{\epsilon_1,\ldots,\epsilon_j}/(a_{\epsilon_1,\ldots,\epsilon_{j-1},0} + a_{\epsilon_1,\ldots,\epsilon_{j-1},1})}{\lambda(B_{\vec{\epsilon}_m})}, \end{aligned} \tag{2.2}$$

   where $\lambda(\cdot)$ is Lebesgue measure.

4. A Polya tree can be constructed with centering at an arbitrary distribution. There are two ways to do this. Suppose $\Omega = \mathbb{R}$ and we want a Polya tree to center at a pre-specified distribution function G. Let the partition be such that the elements of π_m are taken as the intervals $[G^{-1}(k/2^m), G^{-1}((k+1)/2^m))$ for $k = 0, 1, \ldots, 2^m - 1$, with the obvious interpretation for $G^{-1}(0)$ and $G^{-1}(1)$. We refer to this as Method 1. The other approach is to make the partition data-dependent, as mentioned in Muliere and Walker (1997). Suppose we specify a number of points $x_1 < \cdots < x_n$ as end points, and let $B_1 = [x_1, +\infty)$, $B_{11} = [x_2, +\infty), \ldots, B_{\underbrace{1,\ldots,1}_{n}} = [x_n, +\infty)$. We need the parameters $a_{\vec{\epsilon}_m}$ to satisfy

$$\frac{a_{\epsilon_1,\ldots,\epsilon_{j-1},0}}{a_{\epsilon_1,\ldots,\epsilon_{j-1},1}} = \frac{G(B_{\epsilon_1,\ldots,\epsilon_{j-1},0})}{G(B_{\epsilon_1,\ldots,\epsilon_{j-1},1})} \tag{2.3}$$

   with $a_{\vec{\epsilon}_m}$ growing quickly enough to ensure the continuity property. (Here and later, any unspecified subintervals are generated by splitting their parent intervals into two equal parts with respect to the G measure.) For example, $B_0 = (0, x_1)$, and $B_{00} = (0, xx)$ and $B_{01} = [xx, x_1)$ are obtained by specifying a $xx \in (0, x_1)$ such that $G(B_{00}) = G(B_{01}) = G(B_0)/2$. This we refer to as Method 2. We take $a_{\vec{\epsilon}_m} \propto cm^2$ to ensure continuous priors.

Method 2 has some benefits. First, the expectation of this Polya tree is G. Second, as seen by (2.2), it assigns probability 1 to the set of continuous probability measures. We carried out calculations based on the partitions described in Method 2. Generally, results only depend on finitely many parameters, $a_{\underbrace{1,\ldots,1}_{k}}$ and $a_{\underbrace{1,\ldots,1}_{k-1},0}$ for $k = 1,\ldots,n$. Calculations go through as long as Polya trees select continuous distributions with probability 1, and $a_{\epsilon_1,\ldots,\epsilon_m}$ grows to infinity as $m \to +\infty$.

3. Two-Sample Tests with Fixed $\boldsymbol{\alpha = \alpha_0}$

Consider the two-sample case with fixed $\alpha = \alpha_0$. There are some Polya tree based nonparametric hypothesis tests in the two sample case, such as in Holmes et al. (2009), Ma and Wong (2012), and Chen and Hanson (2014), focusing primarily on testing the difference of the distributions of the two samples. We limit our attention to testing Lehmann alternatives. Suppose we have samples of sizes n_1 and n_2 drawn from distributions F and H, respectively. We discuss the case $H(x) = 1 - \{1 - F(x)\}^{\alpha}$. Let $X_1,\ldots,X_{n_1}$ and $Y_1,\ldots,Y_{n_2}$ be the two samples, and denote the combined sample by $V_1,\ldots,V_{n_1+n_2}$. Thus V_k, $k = 1,\ldots,n_1+n_2$, is a sample from $F(x)$ under the null hypothesis $H_0:\ \alpha = 1$ and a mixture of two samples from different distributions under the alternative $H_1:\ \alpha = \alpha_0$, where α_0 is some fixed known real number not equal to 1. Without loss of generality we assume that the V_k's are ordered, in the absence of ties, $V_1 < V_2 < \cdots < V_{n_1+n_2}$. Let, for $k = 1,\ldots,n_1+n_2$,

$$Z_k = \begin{cases} 0, \text{ if } V_k \in \mathfrak{X} = \{X_1,\ldots,X_{n_1}\} \\ 1, \text{ if } V_k \in \mathfrak{Y} = \{Y_1,\ldots,Y_{n_2}\} \end{cases}. \tag{3.1}$$

We consider censoring and take $d_1,\ldots,d_{n_1+n_2}$ to be the censoring indicators corresponding to $V_1,\ldots,V_{n_1+n_2}$.

3.1. Derivation of the Bayes factor

Suppose that a Polya tree prior is applied to $F(x)$. For Bayesian hypothesis testing, the Bayes factor is a widely used tool. Kass and Raftery (1995) reviewed and summarized its uses. Here the Bayes factor is

$$BF_{01} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{P(H_0 \mid V_1,\ldots,V_{n_1+n_2})/P(H_1 \mid V_1,\ldots,V_{n_1+n_2})}{\pi(H_0)/\pi(H_1)}.$$

Let $\pi(H_0) = p_0 = 1 - \pi(H_1)$, where $0 < p_0 < 1$. Suppose the joint pdf of

$V_1, \ldots, V_{n_1+n_2}$ is f_0 under H_0 and f_1 under H_1. Then

$$P(H_0 \mid V_1, \ldots, V_{n_1+n_2}) = \frac{\int p_0 f_0(v_1, \ldots, v_{n_1+n_2}) dPT(P)}{\int p_0 f_0(v_1, \ldots, v_{n_1+n_2}) dPT(P) + \int p_1 f_1(v_1, \ldots, v_{n_1+n_2}) dPT(P)},$$
$$P(H_1 \mid V_1, \ldots, V_{n_1+n_2}) = \frac{\int p_1 f_1(v_1, \ldots, v_{n_1+n_2}) dPT(P)}{\int p_0 f_0(v_1, \ldots, v_{n_1+n_2}) dPT(P) + \int p_1 f_1(v_1, \ldots, v_{n_1+n_2}) dPT(P)}.$$

Therefore

$$BF_{01} = \frac{\int f_0(v_1, \ldots, v_{n_1+n_2}) dPT(P)}{\int f_1(v_1, \ldots, v_{n_1+n_2}) dPT(P)}. \tag{3.2}$$

And the Bayes factor is the ratio of the marginal distributions of $V_1, \ldots, V_{n_1+n_2}$ under H_0 to that under H_1. For more details, see Yuan and Johnson (2008).

We have a closed form expression of the Bayes factor for testing H_0 against H_1 under Polya tree priors.

Theorem 1. *Suppose that a Polya tree prior is applied to $F(x)$ with partitions $B_1 = [v_1, +\infty)$, $B_{11} = [v_2, +\infty), \ldots, B_{\underbrace{1,\ldots,1}_{n_1+n_2}} = [v_{n_1+n_2}, +\infty)$, and G is a strictly increasing baseline measure (with respect to the Polya tree). Then the Bayes factor of the test is*

$$BF_{01} = \frac{1}{\alpha_0^{\sum_{i=1}^{n_1+n_2} d_i Z_i}} \frac{\Gamma(a_1 + n_1 + n_2)\Gamma(a_0 + a_1 + n_1 + n_2\alpha_0)}{\Gamma(a_0 + a_1 + n_1 + n_2)\Gamma(a_1 + n_2\alpha_0)} \prod_{i=1}^{n_1+n_2-1} \left[\frac{\Gamma(a_{\underbrace{1,\ldots,1}_{i},1} + n_1 + n_2 - i)}{\Gamma(a_{\underbrace{1,\ldots,1}_{i},1} + n_1 + n_2 - i + (\alpha_0 - 1)t_{i+1})} \right.$$
$$\left. \frac{\Gamma(a_{\underbrace{1,\ldots,1}_{i},0} + a_{\underbrace{1,\ldots,1}_{i},1} + n_1 + n_2 - i + (\alpha_0 - 1)t_{i+1} + d_i)}{\Gamma(a_{\underbrace{1,\ldots,1}_{i},0} + a_{\underbrace{1,\ldots,1}_{i},1} + n_1 + n_2 - i + d_i)} \right] \tag{3.3}$$

where $t_i = \sum_{j=i}^{n_1+n_2} Z_j$.

The proof of the theorem is provided in the supplementary material. As BF_{01} depends on α_0, we write it as $BF_{01}(\alpha_0)$. Intuitively, t_i shows how many observations on the tail of the mixed sample are coming from sample $\mathfrak{Y}$ at time V_i. Thus if $F(x) > H(x)$, we would expect larger t_i's because sample $\mathfrak{Y}$ is stochastically larger than sample $\mathfrak{X}$.

The assumption that the baseline measure G is strictly increasing is not used explicitly in the proof of the theorem. It is made to ensure that when $m \to +\infty$, $\lambda(B_{\vec{\epsilon}_m}) \to 0$ in a way such that the limit in (2.2) makes sense.

The Bayes factor does not depend on m, and involves only a finite number of parameters, $a_{\underbrace{1,\ldots,1}_{k}}$ and $a_{\underbrace{1,\ldots,1}_{k-1},0}$, for $k = 1, \ldots, n_1 + n_2$. We re-parameterize as

$$\begin{cases} \sigma_k = a_{\underbrace{1,\ldots,1}_{k}} + a_{\underbrace{1,\ldots,1}_{k-1},0}, \\ r_k = \dfrac{G([v_k,+\infty))}{G([v_{k-1},+\infty))} = \dfrac{a_{\underbrace{1,\ldots,1}_{k}}}{a_{\underbrace{1,\ldots,1}_{k-1},0} + a_{\underbrace{1,\ldots,1}_{k}}}, \end{cases} \tag{3.4}$$

where we have used the assumption that $a_{\underbrace{1,\ldots,1}_{k}} \propto G([v_k, +\infty))$ and $a_{\underbrace{1,\ldots,1}_{k-1},0}$ $\propto G([v_{k-1}, v_k))$.

The Bayes factor now simplifies to

$$BF_{01}(\alpha_0) = \frac{1}{\alpha_0^{\sum_{i=1}^{n_1+n_2} d_i Z_i}} \frac{\Gamma(\sigma_1 r_1 + n_1 + n_2)\Gamma(\sigma_1 + n_1 + n_2\alpha_0)}{\Gamma(\sigma_1 + n_1 + n_2)\Gamma(\sigma_1 r_1 + n_1 + n_2\alpha_0)}$$
$$\times \prod_{i=1}^{n-1} \frac{\Gamma(\sigma_{i+1} r_{i+1} + n_1 + n_2 - i)\Gamma(\sigma_{i+1} + n_1 + n_2 - i + (\alpha_0 - 1)t_{i+1} + d_i)}{\Gamma(\sigma_{i+1} + n_1 + n_2 - i + d_i)\Gamma(\sigma_{i+1} r_{i+1} + n_1 + n_2 - i + (\alpha_0 - 1)t_{i+1})}.$$

Note that the data appear in the expression of the Bayes factor through r_i's and t_i's. Basically, the t_i's here correspond to the effects of rank order statistics and the r_i's explain the effect of the spacings of order statistics.

3.2. Properties of the Bayes factor

3.2.1. Monotonicity of the Bayes factor

Fix t_j, then $BF_{01}(\alpha_0)$ is a function of r_i. For $i = 1, \ldots, n-1$, if

$$q_{i+1}(r_{i+1}) = \log(\Gamma(\sigma_{i+1} r_{i+1} + n_1 + n_2 - i)) - \log(\Gamma(\sigma_{i+1} r_{i+1} + n_1 + n_2 - i + (\alpha_0 - 1)t_{i+1})),$$

its derivative is

$$q'_{i+1}(r_{i+1}) = \psi(\sigma_{i+1} r_{i+1} + n_1 + n_2 - i) - \psi(\sigma_{i+1} r_{i+1} + n_1 + n_2 - i + (\alpha_0 - 1)t_{i+1})),$$

where $\psi(\cdot)$ is Digamma function. Hence, $q_{i+1}(r_{i+1})$ is decreasing if $\alpha_0 > 1$ and increasing if $\alpha_0 < 0$, because $\psi(\cdot)$ is strictly increasing in $(0, +\infty)$. The same result holds for the term $\log(\Gamma(\sigma_1 r_1 + n_1 + n_2)) - \log(\Gamma(\sigma_1 r_1 + n_1 + n_2\alpha_0))$. This

makes sense because large r_i's imply that the data are highly clustered, and more likely under the alternative hypothesis. An analogous statement is true when $\alpha_0 < 1$.

Similar calculations show the monotonicity of the Bayes factor on the t_i's when the r_i's are fixed; $BF_{01}(\alpha_0)$ is increasing in the t_j's if $\alpha_0 > 1$ and decreasing if $\alpha_0 < 1$. This make sense since, when $\alpha_0 > 1$, it is less likely to observe large t_i's under the alternative than under the null hypothesis, for one.

3.2.2. Effects of spacings

The proposed test considers order statistics and their spacings; the effect of the ordering is reduced if two observations are close. This clearly enhances the robustness of the test.

Consider, for example, that for some $k_0 \in \{2, \ldots, n_1 + n_2\}$, V_{k_0-1} and V_{k_0} are close, then

$$r_{k_0} = \frac{G([v_{k_0}, +\infty))}{G([v_{k_0-1}, +\infty))} \approx 1,$$

and hence

$$\frac{\Gamma(\sigma_{k_0} r_{k_0} + n_1 + n_2 + 1 - k_0)\Gamma(\sigma_{k_0} + n_1 + n_2 + 1 - k_0 + (\alpha_0 - 1)t_{k_0} + d_{k_0-1})}{\Gamma(\sigma_{k_0} + n_1 + n_2 + 1 - k_0 + d_{k_0-1})\Gamma(\sigma_{k_0} r_{k_0} + n_1 + n_2 + 1 - k_0 + (\alpha_0 - 1)t_{k_0})} \tag{3.5}$$

$$\approx \frac{\sigma_{k_0} + n_1 + n_2 + 1 - k_0 + (\alpha_0 - 1)t_{k_0}}{\sigma_{k_0} + n_1 + n_2 + 1 - k_0}. \tag{3.6}$$

Here, reversing the order of V_{k_0-1} and V_{k_0} leaves all t_i's intact except for t_{k_0}. Hence, if the order is changed, the new Bayes factor differs from the original one only by one term. If V_{k_0-1} and V_{k_0} are from the same sample, exchanging their positions does not affect the Bayes factor. If not, $t_{k_0}^{new} = t_{k_0}^{old} \pm 1$. The plus (minus) sign corresponds to the case where $V_{k_0-1} \in \mathfrak{Y}$ ($\mathfrak{X}$) and $V_{k_0} \in \mathfrak{X}$ ($\mathfrak{Y}$). To compute the new Bayes factor, (3.6) is replaced by

$$\frac{\sigma_{k_0} + n_1 + n_2 + 1 - k_0 + (\alpha_0 - 1)t_{k_0}^{new}}{\sigma_{k_0} + n_1 + n_2 + 1 - k_0}.$$

Therefore,

$$\begin{aligned} BF_{01}(\alpha_0)^{new} &= \frac{\sigma_{k_0} + n_1 + n_2 + 1 - k_0 + (\alpha_0 - 1)t_{k_0}^{new}}{\sigma_{k_0} + n_1 + n_2 + 1 - k_0 + (\alpha_0 - 1)t_{k_0}^{old}} BF_{01}(\alpha_0)^{old} \\ &= \Big(1 \pm \frac{\alpha_0 - 1}{\sigma_{k_0} + n_1 + n_2 + 1 - k_0 + (\alpha_0 - 1)t_{k_0}^{old}}\Big) BF_{01}(\alpha_0)^{old}. \end{aligned}$$

We call $\mid (\alpha_0 - 1)/(\sigma_{k_0} + n_1 + n_2 + 1 - k_0 + (\alpha_0 - 1)t_{k_0}^{old}) \mid$ the *expansion rate*. The flexibility of choosing parameters enables one to test H_0 vs H_1 according to one's needs. For instance, if $\sigma_1 = \cdots = \sigma_{n_1+n_2}$, the expansion rate

goes up as k_0 increases, which implies that the order statistics play a more important role in the tail than at the beginning. The expansion rate grows as α_0 increases for $\alpha_0 > 1$ or $\alpha_0 \to 0^+$ for $\alpha_0 < 1$. Thus the effect of ordering is enhanced when the distance from the alternative to the null is increased. General selection of $\sigma_{k_0} = ck_0^2$ ensures that the expansion rate is small when k_0 is large; When k_0 is small, with reasonable sample sizes, $n_1 + n_2 - k_0$ is reasonably large, and the expansion rate does not expand or shrink the original Bayes factor overwhelmingly.

3.2.3. Asymptotic results

We can track the asymptotic behavior of the Bayes factor when α_0 approaches a limit, for example when $\alpha_0 \to \infty$. The power of α_0 in $BF_{01}(\alpha_0)$ is the negative of $\sum_{i=1}^{n_1+n_2} d_i Z_i$, which is the number of uncensored observations in sample $\mathfrak{Y}$. Given one uncensored observation in sample $\mathfrak{Y}$, there is a corresponding $d_i \neq 0$ for some $i \in 1, \ldots, n_1 + n_2$. Then, for any i such that $d_i Z_i \neq 0$, we have $d_i \neq 0$, $d_i = 1$, and one term in the expression of Bayes factor associated with this particular d_i is

$$\begin{aligned}
&\frac{\Gamma(\sigma_{i+1}r_{i+1} + n_1 + n_2 - i)\Gamma(\sigma_{i+1} + n_1 + n_2 - i + (\alpha_0 - 1)t_{i+1} + d_i)}{\Gamma(\sigma_{i+1} + n_1 + n_2 - i + d_i)\Gamma(\sigma_{i+1}r_{i+1} + n_1 + n_2 - i + (\alpha_0 - 1)t_{i+1})} \\
&\quad = \text{const.} \times \frac{\Gamma(\sigma_{i+1} + n_1 + n_2 - i + (\alpha_0 - 1)t_{i+1} + d_i)}{\Gamma(\sigma_{i+1}r_{i+1} + n_1 + n_2 - i + (\alpha_0 - 1)t_{i+1})} \\
&\quad \geqslant \text{const.} \times \frac{\Gamma(\sigma_{i+1} + n_1 + n_2 - i + (\alpha_0 - 1)t_{i+1} + d_i)}{\Gamma(\sigma_{i+1} + n_1 + n_2 - i + (\alpha_0 - 1)t_{i+1})} \\
&\quad = \text{const.} \times \{\sigma_{i+1} + n_1 + n_2 - i + (\alpha_0 - 1)t_{i+1}\}, \qquad (3.7)
\end{aligned}$$

which is $\mathcal{O}(\alpha_0)$ when $t_{i+1} \neq 0$.

These terms offset the power of negative $\sum_{i=1}^{n_1+n_2} d_i Z_i$. Thus, if there exists at least one uncensored observation in the sample $\mathfrak{X}$ and the indicator function in (3.7) is nonzero, $BF_{01} \to \infty$. When all uncensored observations in sample $\mathfrak{Y}$ are clustered closely and they are uniformly smaller than any observations in sample $\mathfrak{X}$, the t_i's attain their minimums, denoted by t_i^{min}'s, where

$$\begin{cases} t_1 = n_2,\ t_2 = n_2 - 1, \ldots, t_{n_2} = 1, \\ t_i = 0, \quad \text{for } i = n_2 + 1, \ldots, n_1 + n_2. \end{cases} \qquad (3.8)$$

Also, $r_i \approx 1$ for $i = 2, \ldots, n_2$. One can show that in this case $BF_{01} \to 0$, as expected.

Table 1. Ovarian Cancer.

Treatment	Survival Time (Days)
Treatment 1	59, 115, 156, 268, 329, 431, 448+, 477+, 638, 803+, 855+, 1040+, 1106+
Treatment 2	353, 365, 377+, 421+, 464, 475, 563, 744+, 769+, 770+, 1129+, 1206+, 1227+

Table 2. Cox proportional hazards model.

	coef	**exp(coef)**	**se(coef)**	**P-value**	**lower .95**	**upper .95**
Treatment 2	-0.5964	0.5508	0.5870	0.31	0.1743	1.74

3.2.4. Impact of prior parameters

Technically, specifying a Polya tree prior requires selection of infinitely many parameters, $a_{\vec{\epsilon}_m}$, or σ_k and r_k. Our partition is essentially based on order statistics, putting larger observations at a higher level of the Polya tree. This leads to larger σ_k and causes smaller impact to the Polya tree. Therefore, these tests are insensitive in the tails. One can alter the situation by assigning $a_{\vec{\epsilon}_m}$ accordingly. For instance, one can use a common value for all corresponding σ_k's, which leads to a test equally sensitive on $\mathbb{R}^+$. As long as one restricts the assignment of σ_k's to a finite number, it is not going to affect the continuity property of the Polya tree priors because one then modifies only a finite number of parameters.

3.3. Data analysis

We take the ovarian cancer dataset in the Survival package of R software as a data example. The data set was originally reported by Edmunson et al. (1979), and was analyzed in a number of papers (e.g. Collett (2003)). The study included $n = 26$ patients with advanced ovarian carcinoma (stages IIIB and IV). Treatment of patients using either cyclophosphamide alone (1 g/m2) or cyclophosphamide (500 mg/m2) plus adriamycin (40 mg/m2) by i.v. injection every 3 weeks produced partial improvement in approximately one third of the patients. The objective of the trial was to see if the two treatments differentiate in prolonging the time of survival.

As an illustration, we used the Treatment 1 group as the baseline group. A simple Cox proportional hazards model regression gives an estimated $\hat{\alpha} = 0.55$. We used this as the power parameter in the alternative.

3.3.1. Bayes factor using the proposed method

The maximum likelihood estimates of a and b are $(0.947, 980.4)$, based on the combined sample, where the Weibull (a, b) has density

$$f(x) = \frac{a}{b}(\frac{x}{b})^{a-1}e^{-(x/b)^a}. \tag{3.9}$$

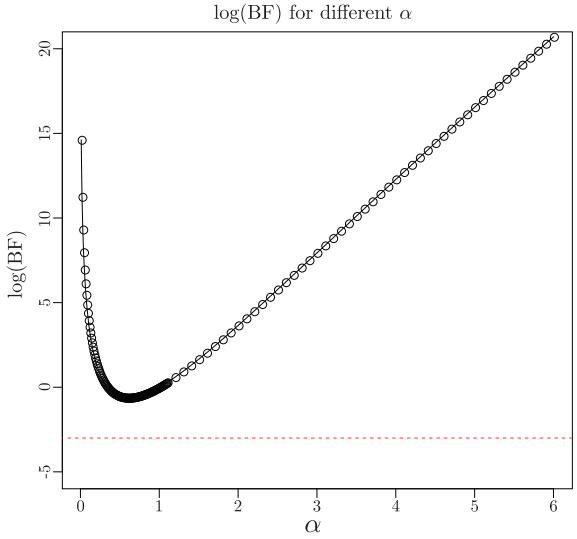


Figure 1. Ovarian Cancer: log(BF) for different α_0.

The large scale parameter is due to the fact that one third of all patients in the study showed improvement. We assigned a Weibull (0.947, 980.4) distribution as the baseline of Polya tree for illustrationl. The centering of the Polya tree process at a Weibull distribution should not compromise the result because the Polya tree itself has large variability, and thus it can accommodate many distributions other than Weibull distributions so that the centering of the Polya tree is not that critical. For testing $H_0 : \alpha = 1$ vs $H_1 : \alpha = \alpha_0 = 0.55$, we find $\log(BF_{01}(0.55)) = -0.604$. Based on the criteria described in Kass and Raftery (1995), this shows little evidence against H_0. Repeating the test with α_0 ranging from 0 to 6 on the data set, we obtain the curve as shown in Figure 1. We took $c = 1$, and have a discussion of this below.

The lower dotted line is at level -3, which shows the region for strong evidence against H_0 with respect to Kass and Raftery (1995) criteria. For this particular data set, there appears to be no evidence against H_0 at any α.

3.3.2. Sensitivity to the Choice of c

The choice of c is generally subjective. It serves as a precision parameter in that it affects the prior variance of $P(B_{\vec{\epsilon}_m})$, as derived in Hanson (2006)

$$\mathrm{Var}\,(P(B_{\vec{\epsilon}_m})) = 4^{-m}[\prod_{j=1}^{m} \frac{2cj^2 + 2}{2cj^2 + 1} - 1].$$

The variance converges to 0 as $c \to \infty$. In this case, the Polya tree places all mass on distributions which more closely resemble the centering distribution.

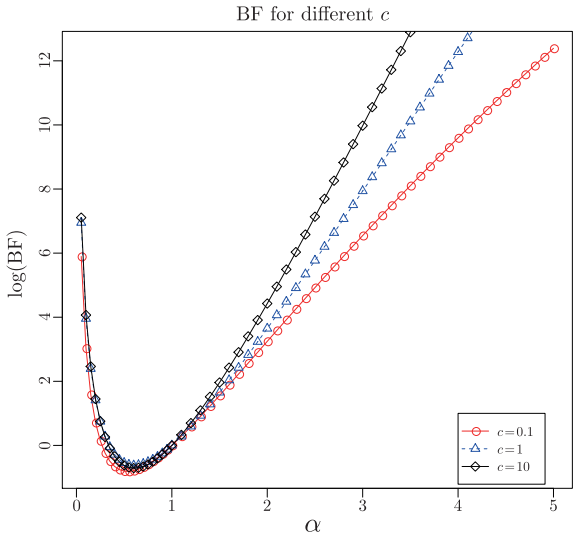


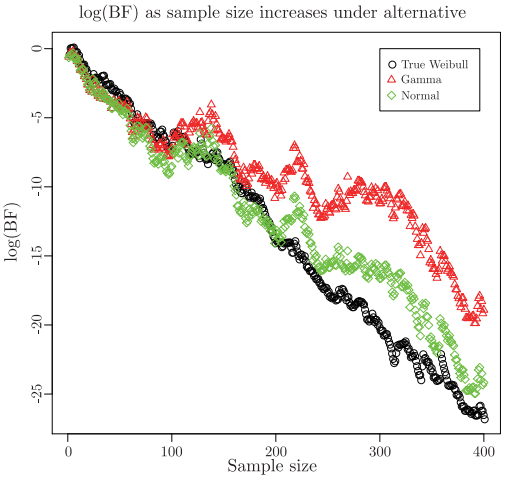
Figure 2. Ovarian Cancer: log(BF) for different α_0 and c.

Holmes et al. (2009), Berger and Guglielmi (2001) and a few others look at the Bayes Factors for different values of c to investigate sensitivity. Generally speaking, the parameter c does not impact the results too much. As an illustration, the Ovarian data were used to calculate the Bayes Factors for $c = 0.1$, $c = 1$ and $c = 10$. From Figure 2, the Bayes Factors (logarithm transformed) are close to each other.

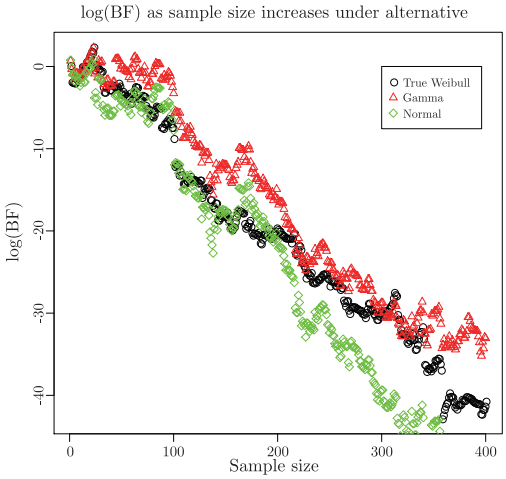
3.4. Robustness study through simulations

Simulation studies were carried out to investigate the robustness of misspecifiying the center of the Polya tree priors. The results of many parametric methods depend heavily on the distributional assumptions. We alleviate this dependence by assigning a prior on a large family of distributions. When the Polya tree process covers a large set of distributions, it can capture the true distribution even though the center of the process is incorrectly specified. To illustrate this, we generated the $\mathfrak{X}$ sample from a Weibull (3,12) distribution, with the distribution of sample $\mathfrak{Y}$ chosen according to H_1.

Figure 3 shows how $\log(BF_{01}(\alpha_0))$ decreases as n_1 and n_2 increase when α_0 is set to be 1.5, and 0.5. In each graph, three Bayes factors were calculated: one with the Weibull distribution as the center of Polya tree process prior (as shown by circles), one with the Gamma distribution with estimated parameters based on sample $\mathfrak{X}$ as the center (as shown by triangles), and one with the Normal distribution with estimated parameters (as shown by diamonds). The test is consistent in the sense that the Bayes factor decreases to $-\infty$ as the

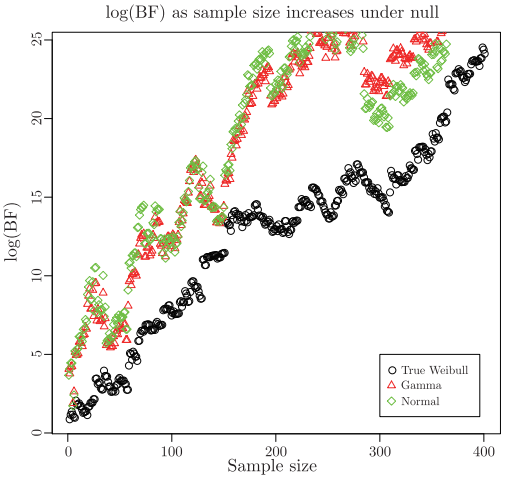


(a) Baseline Measure: True and Est. Weibull Distribution with $\alpha_0 = 1.5$.

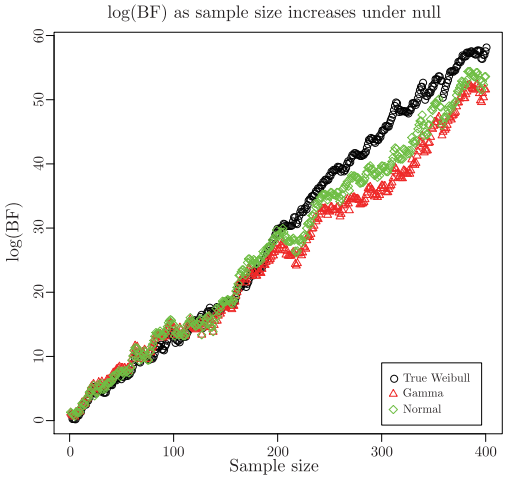


(b) Baseline Measure: True and Est. Weibull Distribution with $\alpha_0 = 0.5$.

Figure 3. log(BF) as sample size increases under the alternative hypothesis.



(a) Baseline Measure: True and Est. Weibull Distribution with $\alpha_0 = 1.5$.



(b) Baseline Measure: True and Est. Weibull Distribution with $\alpha_0 = 0.5$.

Figure 4. log(BF) as sample size increases under the null hypothesis.

sample size grows. Robustness is suggested by the fact that the curves with Gamma and Normal distributions do not depart much from the one with the true distribution. Figure 4 demonstrates the patterns of $\log(BF_{01}(\alpha_0))$ when testing H_1 with $\alpha_0 = 1.5$ and $\alpha = 0.5$ under the null hypothesis.

4. Bayesian Analysis with α Unknown

4.1. Posterior probability calculation

Brooks (1974) treated the power α as a unknown parameter and assigned a $F(2\nu_1, 2\nu_2, \delta)$ prior to it,

$$p(\alpha) \propto \frac{\alpha^{\nu_1 - 1}}{(\alpha + \delta)^{\nu_1 + \nu_2}}.$$

Then the posterior distribution of α is analyzed under the usual Bayesian paradigm.

Similar analysis is possible under our setup. Instead of a continuous prior over $(0, +\infty)$ for α, we propose a spike and slab prior so that both the prior and posterior probability of H_0 are non-zero and can be evaluated explicitly. To test $H_0 : \alpha = 1$ vs $H_1 : \alpha \neq 1$, suppose π_0 is fixed, with $0 < \pi_0 < 1$. Let $p(\alpha)$ be the density of a pre-specified continuous prior on $(0, +\infty)$. A spike and slab prior for α is $\pi(\alpha) = \pi_0 I_{[\alpha=1]} + (1 - \pi_0)p(\alpha)$, where $\pi(H_0) = \pi_0 > 0$. It is straightforward that

$$\begin{aligned} BF_{01} &= \frac{\int f_0(v_1, \ldots, v_{n_1+n_2} \mid \alpha = 1) dPT(P)}{\int\int f_1(v_1, \ldots, v_{n_1+n_2} \mid \alpha) dPT(P) p(\alpha) d\alpha} \\ &= \left(\int \frac{\int f_1(v_1, \ldots, v_{n_1+n_2} \mid \alpha) dPT(P)}{\int f_0(v_1, \ldots, v_{n_1+n_2} \mid \alpha = 1) dPT(P)} p(\alpha) d\alpha \right)^{-1} \\ &= \left(\int \frac{1}{BF_{01}(\alpha)} p(\alpha) d\alpha \right)^{-1}. \end{aligned} \tag{4.1}$$

We can also calculate the posterior probability as follows.

$$\begin{aligned} &\pi(H_0 \mid v_1, \ldots, v_{n_1+n_2}) \\ &\quad = \frac{\pi(H_0) f(v_1, \ldots, v_{n_1+n_2} \mid H_0)}{f(v_1, \ldots, v_{n_1+n_2})} \\ &\quad = \frac{\pi_0 f_0(v_1, \ldots, v_{n_1+n_2})}{\pi_0 f_0(v_1, \ldots, v_{n_1+n_2}) + (1 - \pi_0) \int f(v_1, \ldots, v_{n_1+n_2} \mid \alpha) p(\alpha) d\alpha} \\ &\quad = \left(1 + \frac{1 - \pi_0}{\pi_0} \int \frac{f(v_1, \ldots, v_{n_1+n_2} \mid \alpha) p(\alpha)}{f_0(v_1, \ldots, v_{n_1+n_2})} d\alpha \right)^{-1} \\ &\quad = \left(1 + \frac{1 - \pi_0}{\pi_0} \int \frac{p(\alpha)}{BF_{01}(\alpha)} d\alpha \right)^{-1}. \end{aligned} \tag{4.2}$$

Hypothesis tests can be made based on $\pi(H_0 \mid v_1, \ldots, v_{n_1+n_2})$. A naive rejection region can be taken as $\{v_1, \ldots, v_{n_1+n_2} : \pi(H_0 \mid v_1, \ldots, v_{n_1+n_2}) < 0.5\}$. The exact posterior distribution of α is analytically untractable. However, standard MCMC methods can be used to approximate its posterior distribution. We do not address this issue.

Table 3. Posterior probability with a spike and slab prior.

$\mathbf{p}(\boldsymbol{\alpha})$	$\boldsymbol{\pi}(H_0 \mid v_1, \ldots, v_{n_1+n_2})$ $\boldsymbol{\pi}_0 = 0.1$	$\boldsymbol{\pi}(H_0 \mid v_1, \ldots, v_{n_1+n_2})$ $\boldsymbol{\pi}_0 = 0.5$	$\boldsymbol{\pi}(H_0 \mid v_1, \ldots, v_{n_1+n_2})$ $\boldsymbol{\pi}_0 = 0.9$
Gamma(2,3)	0.0985	0.4958	0.8985
Gamma(5,2)	0.4038	0.8591	0.9821
F(3,5)	0.1453	0.6047	0.9323
F(6,3)	0.1426	0.5995	0.9309
Beta(3,3)	0.0746	0.4204	0.8672

4.2. Data application: Revisting ovarian cancer

We assigned three prior probabilities to the null hypothesis H_0, $\pi_0 = 0.1$, 0.5, 0.9. A spike and slab prior with Gamma distribution, central F distribution and Beta distribution were used. The resulting posterior probabilities are shown as follows.

The posterior probabilities in Table 3 are fairly sensitive to π_0: $\log(BF_{01})$ is relatively large for different α for this data set as shown in Figure 1. There is no clear evidence here that the survival curves for the two treatment groups are different.

5. Discussion

Our results on Bayes factors work only in cases without ties, and there are situations where ties occur. Here there are ties of censored data and ties of event times. For the first of these, our calculations still work. However, this is not so for the second type.

The problem is that the limit (2.2) may not exist or is infinity for some $x \in \mathbb{R}^+$ because the density does not necessarily exist when the underlying distribution is not absolutely continuous. In addition, when multiple events are observed at the same time, we know that the underlying distribution must be non-continuous. In this case, to keep the prior reasonable, we should not assign parameters such that the Polya tree gives probability 1 to the set of continuous distributions. We need to assign parameters to Polya tree accordingly, depending on whether it gives positive probability to the set of discrete distributions, or to the set of partly discrete, and partly continuous distributions. Taking into account the fact that the Dirichlet Process is a special case of a Polya tree, one would like to utilize it in cases where ties occur. We did calculate the Bayes factor under this assumption. No asymptotic properties could be developed, since the Bayes factor depends not only on the number of tied observations, but also on the location on $\mathbb{R}^+$ of the occurrences of ties.

We can borrow some ideas from how people deal with ties in the partial likelihood function. Suppose observed data set (t_k, δ_k), $k = 1, 2, \ldots, n$, sorted by

t_k, has tied event times $t_{k_0} = t_{k_0+1} = t$. If the underlying distribution of times is continuous, ties are attributable to measurement error. Take $t_{k_0} > t_{k_0+1}$ and $t_{k_0} < t_{k_0+1}$ as equally likely. We can calculate an approximate BF_{01} by letting both $t_{k_0} = t + \epsilon$, $t_{k_0+1} = t$ and $t_{k_0+1} = t + \epsilon$, $t_{k_0} = t$ for some ϵ . Then the overall BF_{01} can be taken as the average of the two resulting Bayes factors. The continuity of BF_{01} on r_{k_0} guarantees that the approximate BF_{01} is close to its true value for small ϵ.

Acknowledgement

This research was partially supported by NSF Grants: DMS-1007494 and SES-1026165. We are also thankful for the insightful comments provided by the anonymous referees.

References

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems, *Ann. Statist.* **2**, 1152-1174.

Berger, J. O. and Guglielmi, A. (2001). Bayesian testing of a parametric model versus nonparametric alternatives. *J. Amer. Statist. Assoc.* **96**, 174-184.

Brooks, R. J. (1974). Bayesian analysis of the two-sample problem under the Lehmann alternatives. *Biometrika* **61**, 501-507.

Chen, Y. and Hanson, T. E. (2014). Bayesian nonparametric k-sample tests for censored and uncensored data. *Comput. Statist. Data Anal.* **71**, 335-346.

Collett, D. (2003). *Modelling Survival Data in Medical Research*. Chapman and Hall.

Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

Damien, P. and Walker, S. (2002). A Bayesian non-parametric comparison of two treatments. *Scand. J. Statist.* **29**, 51-56.

Davies, R. B. (1971). Rank tests for Lehmann's alternative. *J. Amer. Statist. Assoc.* **66**, 879-883.

Edmunson, J. H., Fleming, T. R., Decker, G. D., Malkasam, G. D. and Kvols, L. K. (1979). Different chemotherapeutic sensitivities and host factors affecting prognosis in advanced ovarian carcinoma versus minimal disease residual. *Cancer Treatment Reports* **63**, 241-247.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90**, 577-588.

Fabius, J. (1964). Asymptotic behavior of Bayes' estimates. *Ann. Math. Statist.* **35**, 846-856.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615-629.

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*. Springer.

Hanson, T. E. (2006). Inference for mixtures of finite Polya tree models. *J. Amer. Statist. Assoc.* **101**, 1548-1565.

Hanson, T. E. and Jara, A. (2012). Surviving fully Bayesian nonparametric regression models. In *Bayesian Theory and Applications* (Edited by P. Damien, P. Dellaportas, N. Polson, D. Stephens), 592-615. Oxford University Press, Oxford.

Hjort, N. L. (1990). Nonparametric Bayes estimators based on Beta processes in models for life history data. *Ann. Statist.* **18**, 1259-1294.

Holmes, C. C., Caron, F., Griffin, J. E. and Stephens, D. A. (2009). Two-sample Bayesian nonparametric hypothesis testing. *ARXIV*.

Kalbfleisch, J. D. (1978). Non-parametric Bayesian analysis of survival time data. *J. Roy. Statist. Soc. Ser. B* **40**, 214-221.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *J. Amer. Statist. Assoc.*, **90**, 773-795.

Lavine, M. (1992). Some aspects of Polya tree distributions for statistical modelling. *Ann. Statist.* **20**, 1222-1235.

Lavine, M. (1994). More aspects of Polya tree distributions for statistical modelling. *Ann. Statist.* **22**, 1161-1176.

Lehmann, E. L. (1953). The power of rank tests. *Ann. Math. Statist.* **24**, 23-43.

Ma, L. and Wong, W. H. (2012). Coupling optional Polya trees and the two sample problem. *J. Amer. Statist. Assoc.* **106**, 1553-1565.

Mauldin, R. D., Sudderth, W. D. and Williams, S. C. (1992). Polya trees and random distributions. *Ann. Statist.* **20**, 1203-1221.

Muliere, P. and Walker, S. (1997). A Bayesian non-parametric approach to survival using Polya trees. *Scand. J. Statist.* **24**, 331-340.

Savage, I. R. (1956). Contribution to the theory of rank order statistics — the two-sample case. *Ann. Math. Statist.*, **27**, 590-616.

Yuan, Y. and Johnson, V. E. (2008). Bayesian hypothesis tests using nonparametric statistics, *Statist. Sinica* **18**, 1185-1200.

Department of Statistics, University of Florida, Gainesville, FL 32611-8545, USA.

E-mail: leihuang@stat.ufl.edu

Department of Statistics, University of Florida, P.O. Box 118545 Gainesville, FL 32611-8545, USA.

E-mail: ghoshm@stat.ufl.edu

(Received February 2013; accepted December 2013)