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

LEI HUANG and MALAY GHOSH

UNIVERSITY OF FLORIDA and UNIVERSITY OF FLORIDA

Supplementary Material

This supplementary material contains the proof of Theorem 1. The proof of Theorem 2 is very similar to that of Theorem 1 and thus omitted.

S1 Proof of Theorem 1

We have $X_1, \cdots, X_{n_1} \mid P \sim F(x)$ and $Y_1, \cdots, Y_{n_2} \mid P \sim H(x) = 1 - \{1 - F(x)\}^{\alpha}$ and want to test

$$H_0 : \alpha = 1 \quad vs \quad H_1 : \alpha > 1 \; (or \; \alpha < 1).$$

where $P \sim$ PT(G) is the probability measure induced by $F(x)$, and $\alpha > 0$. For simplicity, let P' be the probability measure induced by $H(x)$.

Proof. First we put these two samples together and denote them by $V_1, \cdots, V_{n_1+n_2}$ and define $Z_1, \cdots, Z_{n_1+n_2}$ as described in Section 3. Let $n = n_1 + n_2$.

In case that there is no censoring, take $m > n$, such that at level m, $V_1, ..., V_n$ are separated in different intervals. For $v \in [0, +\infty)$, let $\vec{\epsilon_m}(v) = \epsilon_1, ..., \epsilon_m$ such that $v \in B_{\epsilon_1,...,\epsilon_m}$. In addition, with appropriate parameters, P is continuous with probability 1. Thus without loss of generality, assume $V_1 < ... < V_n$. Write $\vec{\epsilon_m}^{\,i} = \vec{\epsilon_m}(V_i) = \epsilon_1^i, ..., \epsilon_m^i$.

Under the null hypothesis, $V_1, \cdots, V_n$ are independent and identially distributed conditional on P. Hence, at level m of the tree, given P, the joint pdf of $V_1, ..., V_n$ is

$$\begin{aligned} f_m(v_1, ..., v_n \mid P) &= \frac{\prod_{i=1}^n Pr(B_{\vec{\epsilon_m}^{\,i}} \mid P)}{\prod_{i=1}^n \lambda(B_{\vec{\epsilon_m}^{\,i}})} \\ &= \frac{\prod_{i=1}^n \prod_{\epsilon_j^i=0} Y_{\epsilon_1^i,...,\epsilon_j^i} \prod_{\epsilon_j=1} (1 - Y_{\epsilon_1^i,...,\epsilon_j^i})}{\prod_{i=1}^n \lambda(B_{\vec{\epsilon_m}^{\,i}})} \end{aligned} \tag{S1.1}$$

The exact marginal joint pdf is given by letting m goes to $+\infty$ and then taking the expectation. By repeated use of the Theorem 2 in Lavine (1992), the existence and finiteness of the limit is guaranteed. We denote this limit by f. By dominated convergence theorem, the order of expectation and limit can be exchanged.

The joint pdf under H_1 is much more complicated. Take any $v \in B_{\epsilon_1,\ldots,\epsilon_m}$. We need to find $P'(B_{\vec{\epsilon_m}})$. To do this, we have to sum all the probabilities of the intervals to the right of $B_{\vec{\epsilon_m}}$, plus probability of $B_{\vec{\epsilon_m}}$, and then raise to power α. This quantity is denoted by $P'(B_{\vec{\epsilon_m}}+ \mid P)$. At mth level of the tree, given P, a simple expression is provided by,

$$P'(B_{\vec{\epsilon_m}}+ \mid P) = \{\sum_{n=1}^{m} Pr(B_{\epsilon_1,\ldots,\epsilon_{n-1},1} \mid P)\delta_0(\epsilon_n) + Pr(B_{\vec{\epsilon_m}} \mid P)\}^{\alpha}.$$

Therefore,

$$\begin{aligned} P'(B_{\vec{\epsilon_m}} \mid P) =& P'(B_{\vec{\epsilon_m}}+ \mid P) - P'((B_{\vec{\epsilon_m}}+) - B_{\vec{\epsilon_m}} \mid P) \\ =& \{\sum_{n=1}^{m} Pr(B_{\epsilon_1,\ldots,\epsilon_{n-1},1} \mid P)\delta_0(\epsilon_n) + Pr(B_{\vec{\epsilon_m}} \mid P)\}^{\alpha} \\ &- \{\sum_{n=1}^{m} Pr(B_{\epsilon_1,\ldots,\epsilon_{n-1},1} \mid P)\delta_0(\epsilon_n)\}^{\alpha} \end{aligned} \tag{S1.2}$$

where the "-" sign in the probability means exclusion. Now using the second order Taylor Expansion for function $h(t) = t^{\alpha}$,

$$h(t+\Delta) - h(t) = \alpha t^{\alpha-1}\Delta + \alpha(\alpha-1)(t+\theta)^{\alpha-2}\Delta^2$$

where $\theta \in (0, \Delta)$. It follows that

$$\begin{aligned} P'(B_{\vec{\epsilon_m}} \mid P) = \quad & \alpha \quad Pr(B_{\vec{\epsilon_m}} \mid P)\{\sum_{n=1}^{m} Pr(B_{\epsilon_1,\ldots,\epsilon_{n-1},1} \mid P)\delta_0(\epsilon_n)\}^{\alpha-1} \\ & + \quad \alpha(\alpha-1)Pr(B_{\vec{\epsilon_m}} \mid P)^2\{\sum_{n=1}^{m} Pr(B_{\epsilon_1,\ldots,\epsilon_{n-1},1} \mid P)\delta_0(\epsilon_n) + \theta\}^{\alpha-2} \end{aligned}$$

where $\theta \in (0, Pr(B_{\vec{\epsilon_m}} \mid P))$. For simplicity, write

$$W_m(x) = \sum_{n=1}^{m} Pr(B_{\epsilon_1,\ldots,\epsilon_{n-1},1} \mid P)\delta_0(\epsilon_n)$$

W_m depends on v because $\epsilon_1, \ldots, \epsilon_m$ depend on v.

Now we are in place to calculate the conditional joint pdf of $Y_1, \cdots, Y_{n_2}$ under H_1.

$$\begin{aligned} \bar{f}_m(y_1, \ldots, y_{n_2} \mid P) &= \frac{\prod_{i=1}^{n_2} P'(B_{\vec{\epsilon_m}\, i} \mid P)}{\prod_{i=1}^{n_2} \lambda(B_{\vec{\epsilon_m}\, i})} \\ &= \frac{NUM}{\prod_{i=1}^{n_2} \lambda(B_{\vec{\epsilon_m}\, i})} \end{aligned}$$

where NUM is the numerator of the fraction, namely

$$NUM = \prod_{i=1}^{n_2} P'(B_{\vec{\epsilon_m}^i} \mid P)$$
$$= \frac{\prod_{i=1}^{n_2} Pr(B_{\vec{\epsilon_m}^i} \mid P) \prod_{i=1}^{n_2} \{\alpha W_m(Y_i)^{\alpha-1} + \alpha(\alpha-1)[W_m(Y_i)+\theta_i]^{\alpha-2} Pr(B_{\vec{\epsilon_m}^i} \mid P)\}}{\prod_{i=1}^{n_2} \lambda(B_{\vec{\epsilon_m}^i})}$$

Again, the exact marginal joint pdf is found by letting $m \to +\infty$ and then taking expected value.

Note that

$$\lim_{m\to+\infty} E[\frac{NUM}{\prod_{i=1}^{n_2} \lambda(B_{\vec{\epsilon_m}^i})}] = \lim_{m\to+\infty} E[\frac{\prod_{i=1}^{n_2} Pr(B_{\vec{\epsilon_m}^i} \mid P) \prod_{i=1}^{n_2} \{\alpha W_m(Y_i)^{\alpha-1}\}}{\prod_{i=1}^{n_2} \lambda(B_{\vec{\epsilon_m}^i})}] \quad \text{(S1.3)}$$

Indeed, if we write the products in the numerator as summation, we have

$$\prod_{i=1}^{n_2} [\alpha W_m(Y_i)^{\alpha-1} + \alpha(\alpha-1)\{W_m(Y_i)+\theta_i\}^{\alpha-2} Pr(B_{\vec{\epsilon_m}^i} \mid P)]$$
$$= \sum_{S\subseteq\Omega} [\prod_{j\in S} \alpha W_m(X_j)^{\alpha-1} \prod_{k\in S^c} \alpha(\alpha-1)\{W_m(Y_k)+\theta_k\}^{\alpha-2} Pr(B_{\vec{\epsilon_m}^k} \mid P)] \quad \text{(S1.4)}$$

where $\Omega = \{1, ..., n\}$, and the summation is taken for all (proper and improper) subsets $S \subseteq \Omega$.

However, if $\mid S^c \mid \geq 1$, i.e. there exists $k_0 \in S^c$, then

$$\mid [\prod_{j\in S} \alpha W_m(Y_j)^{\alpha-1} \prod_{k\in S^c} \alpha(\alpha-1)\{W_m(Y_k)+\theta_k\}^{\alpha-2} Pr(B_{\vec{\epsilon_m}^k} \mid P)] \mid$$
$$\leq \mid \alpha^n (\alpha-1)^{|S^c|} Pr(B_{\vec{\epsilon_m}^{k_0}} \mid P) \mid$$

The above inequality uses the fact that

$$0 \leq W_m(Y) \leq W_m(Y) + \theta \leq W_m(Y) + Pr(B_{\vec{\epsilon_m}^k} \mid P) \leq \sum_{n=1}^{m} Pr(B_{\epsilon_1,\ldots,\epsilon_n} \mid P)$$
$$= Pr([0, +\infty) \mid P) = 1.$$

Hence, the expectation of the corresponding term in summation satisfies

$$E[| \frac{\prod_{i=1}^{n_2} Pr(B_{\vec{\epsilon_m}^i} \mid P)[\prod_{j\in S} \alpha W_m(Y_j)^{\alpha-1} \prod_{k\in S^c} \alpha(\alpha-1)\{W_m(Y_k)+\theta_k\}^{\alpha-2} Pr(B_{\vec{\epsilon_m}^k} \mid P)]}{\prod_{i=1}^{n_2} \lambda(B_{\vec{\epsilon_m}^i})} |]$$
$$\leq E[| \frac{\prod_{i=1}^{n_2} Pr(B_{\vec{\epsilon_m}^i} \mid P) \alpha^n (\alpha-1)^{|S^c|} Pr(B_{\vec{\epsilon_m}^{k_0}} \mid P)}{\prod_{i=1}^{n_2} \lambda(B_{\vec{\epsilon_m}^i})} |]$$
$$= const.E[\frac{\{\prod_{i\neq k_0} Pr(B_{\vec{\epsilon_m}^i} \mid P)\} Pr(B_{\vec{\epsilon_m}^{k_0}} \mid P)^2}{\prod_{i=1}^{n_2} \lambda(B_{\vec{\epsilon_m}^i})}] \quad \text{(S1.5)}$$

Comparing (S1.5) to (S1.1), it follows that

$$\begin{aligned} E & \quad [\frac{\{\prod_{i\neq k_0} Pr(B_{\vec{\epsilon_m}^i} \mid P)\} Pr(B_{\vec{\epsilon_m}^{k_0}} \mid P)^2}{\prod_{i=1}^{n_2} \lambda(B_{\vec{\epsilon_m}^i})}] \\ = & \quad E[f_m(y_1,...,y_{n_2} \mid P)] \prod_{j=1}^{m} \frac{a_{\epsilon_1^{k_0},...,\epsilon_j^{k_0}} + n_{\epsilon_1^{k_0},...,\epsilon_{j-1}^{k_0}} + 1}{a_{\epsilon_1^{k_0},...,\epsilon_{j-1}^{k_0},0} + a_{\epsilon_1^{k_0},...,\epsilon_{j-1}^{k_0},1} + n_{\epsilon_1^{k_0},...,\epsilon_{j-1}^{k_0}} + 1} \end{aligned}$$

where $n_{\epsilon_1^{k_0},...,\epsilon_{j-1}^{k_0}} = \sharp\{j : Y_j \in B_{\epsilon_1^{k_0},...,\epsilon_{j-1}^{k_0}}\}$.

When $m > n$, we specify the parameters as

$$a_{\epsilon_1,...,\epsilon'_{m-1}0} = a_{\epsilon_1,...,\epsilon_{j-1},1} = m^2$$

which implies that when m is large,

$$\frac{a_{\epsilon_1^{k_0},...,\epsilon_j^{k_0}} + n_{\epsilon_1^{k_0},...,\epsilon_{j-1}^{k_0}} + 1}{a_{\epsilon_1^{k_0},...,\epsilon_{j-1}^{k_0},0} + a_{\epsilon_1^{k_0},...,\epsilon_{j-1}^{k_0},1} + n_{\epsilon_1^{k_0},...,\epsilon_{j-1}^{k_0}} + 1} \to \frac{1}{2}.$$

Note that $E[f_m(y_1,...,y_{n_2} \mid P)]$ is finite. Therefore,

$$E[\frac{\{\prod_{i\neq k_0} Pr(B_{\vec{\epsilon_m}^i} \mid P)\} Pr(B_{\vec{\epsilon_m}^{k_0}} \mid P)^2}{\prod_{i=1}^{n_2} \lambda(B_{\vec{\epsilon_m}^i})}] \to 0 \quad as \quad m \to +\infty.$$

So all the terms with $\mid S^c \mid \geq 1$ in (S1.4) eventually goes to 0. The only term left is when $\mid S^c \mid = 0$, i.e. $S = \Omega$, which completes the proof to the claim.

Furthermore, the conditional joint pdf of $V_1, \cdots, V_n$ under H_1 is given by

$$\begin{aligned} \bar{f}_m(v_1,...,v_n \mid P) &= f_m(x_1,\cdots,x_{n1} \mid P)\bar{f}_m(y_1,\cdots,y_{n_2} \mid P) \\ &= \frac{\prod_{i=1}^{n_1} P(B_{\vec{\epsilon_m}^i} \mid P)}{\prod_{i=1}^{n_1} \lambda(B_{\vec{\epsilon_m}^i})} \frac{\prod_{i=1}^{n_2} Pr(B_{\vec{\epsilon_m}^i} \mid P) \prod_{i=1}^{n_2} \{\alpha W_m(Y_i)^{\alpha-1}\}}{\prod_{i=1}^{n_2} \lambda(B_{\vec{\epsilon_m}^i})} \\ &= \frac{\prod_{i=1}^{n} Pr(B_{\vec{\epsilon_m}^i} \mid P) \prod_{i=1}^{n_2} \{\alpha W_m(Y_i)^{\alpha-1}\}}{\prod_{i=1}^{n} \lambda(B_{\vec{\epsilon_m}^i})} \end{aligned}$$

Now we compute the Bayes factor. Before we do that, let us figure out what $\vec{\epsilon_m}(v_i)$ is. By the mechanism of partition, for $m > n$, and $i = 1,...,n$, v_i is an end point at level i of the tree. Before the ith level, v_i lies in the right subinterval every time the current interval splits into two; after ith level, v_i would be always in the left subinterval generated by splitting the current interval that contains v_i. Thus,

$$\vec{\epsilon_m}(v_i) = \underbrace{1,...,1}_{i}, \underbrace{0,...,0}_{m-i}$$

Clearly,

$$Pr(B_{\vec{\epsilon_m}^i} \mid P) = Y_1 Y_{11} ... Y_{\underbrace{1,...,1}_{i}} Y_{\underbrace{1,...,1}_{i},0} ... Y_{\underbrace{1,...,1}_{i},\underbrace{0,...,0}_{m-i}}.$$

By definition of $W_m(v)$,

$$W_m(v_1) = Y_1,\ W_m(v_2) = Y_1 Y_{11},\ \cdots,\ W_m(v_n) = Y_1 Y_{11} ... Y_{\underbrace{1,...,1}_{n}}.$$

Intuitively, W_m is the survival function at level m.

These lead to

$$\prod_{i=1}^{n} Pr(B_{\vec{\epsilon_m}^i} \mid P) = \{Y_1 Y_{10} Y_{100} ... Y_{1,\underbrace{0,...,0}_{n-1}}\}\{Y_1 Y_{11} Y_{110} Y_{1100} ... Y_{11,\underbrace{0,...,0}_{n-2}}\} \cdots$$
$$\{Y_1 Y_{11} ... Y_{\underbrace{1,...,1}_{n}} Y_{\underbrace{1,...,1}_{n},0} Y_{\underbrace{1,...,1}_{n},00} \cdots Y_{\underbrace{1,...,1}_{n},\underbrace{0,...,0}_{m-n}}\} \tag{S1.6}$$

Also,

$$\prod_{i=1}^{n} Pr(B_{\vec{\epsilon_m}^i} \mid P) \prod_{i=1}^{n_2} \{\alpha W_m(Y_i)^{\alpha-1}\} = \prod_{i=1}^{n} \{Pr(B_{\vec{\epsilon_m}^i} \mid P)\}^{1-Z_i} \{Pr(B_{\vec{\epsilon_m}^i} \mid P) \alpha W_m(V_i)^{\alpha-1}\}^{Z_i} \tag{S1.7}$$

$$= \alpha^{n_2} Y_1^{n_1+n_2\alpha} Y_{11}^{n-1+(\alpha-1)t_2} ... Y_{\underbrace{1,...,1}_{k}}^{n+1-k+(\alpha-1)t_k} ... Y_{\underbrace{1,...,1}_{n}}^{1+(\alpha-1)t_n}$$
$$\{Y_{10} Y_{100} ... Y_{1,\underbrace{0,...,0}_{n-1}}\}\{Y_{110} Y_{1100} ... Y_{11,\underbrace{0,...,0}_{n-2}}\} \cdots$$
$$\{Y_{\underbrace{1,...,1}_{n},0} Y_{\underbrace{1,...,1}_{n},00} \cdots Y_{\underbrace{1,...,1}_{n},\underbrace{0,...,0}_{m-n}}\} \tag{S1.8}$$

Where $t_k = \sum_{i=k}^{n} Z_i$ evaluates the ordering of $\mathfrak{X}$ sample and $\mathfrak{Y}$ sample. It turns out that all these $Y's$ are not independent because

$$Y_{\epsilon_1,...,\epsilon_j,1} = 1 - Y_{\epsilon_1,...,\epsilon_j,0} \tag{S1.9}$$

for all $j = 1, ..., n$ and all $(\epsilon_1, ..., \epsilon_j)$. Taking into account that the denominators in (S1.1) and (S1.3) are the same, the Bayes factor reduces to

$$BF_{01} = \lim_{m \to +\infty} \frac{E[\prod_{i=1}^{n} Pr(B_{\vec{\epsilon_m}^i} \mid P)]}{E[\prod_{i=1}^{n} Pr(B_{\vec{\epsilon_m}^i} \mid P) \prod_{i=1}^{n_2} \{\alpha W_m(Y_i)^{\alpha-1}\}]} \tag{S1.10}$$

Combining (S1.6), (S1.8) and (S1.9), and canceling the independent common terms in (S1.10), we end up with

$$BF_{01} = \frac{1}{\alpha^{n_2}} \frac{E_1}{E_2} \tag{S1.11}$$

where

$$E_1 = E[(1-Y_0)^n Y_{10}(1-Y_{10})^{n-1} \ldots Y_{\underbrace{1,\ldots,1}_{i},0}(1-Y_{\underbrace{1,\ldots,1}_{i},0})^{n-i} \ldots Y_{\underbrace{1,\ldots,1}_{n-1},0}(1-Y_{\underbrace{1,\ldots,1}_{n-1},0})]$$

$$E_2 = E[(1-Y_0)^{n_1+n_2\alpha} Y_{10}(1-Y_{10})^{n-1+(\alpha-1)t_2} \ldots Y_{\underbrace{1,\ldots,1}_{i},0}(1-Y_{\underbrace{1,\ldots,1}_{i},0})^{n-i+(\alpha-1)t_i}$$

$$\cdots Y_{\underbrace{1,\ldots,1}_{n-1},0}(1-Y_{\underbrace{1,\ldots,1}_{n-1},0})^{1+(\alpha-1)t_n}]$$

All the Y's appearing in the above equation are independent with Beta distributions, that is,

$$Y_{\underbrace{1,\ldots,1}_{i},0} \quad independently \quad \sim Beta(a_{\underbrace{1,\ldots,1}_{i},0}, a_{\underbrace{1,\ldots,1}_{i},1})$$

for $i = 1, \ldots, n$. Hence simple calculations of moments of Beta distributions yield the final result

$$BF_{01} = \frac{1}{\alpha^{n_2}} \frac{\Gamma(a_1+n)\Gamma(a_0+a_1+n_1+n_2\alpha)}{\Gamma(a_0+a_1+n)\Gamma(a_1+n_1+n_2\alpha)} \prod_{i=1}^{n-1} \frac{\Gamma(a_{\underbrace{1,\ldots,1}_{i},1}+n-i)\Gamma(a_{\underbrace{1,\ldots,1}_{i},0}+a_{\underbrace{1,\ldots,1}_{i},1}+(\alpha-1)t_{i+1}+1)}{\Gamma(a_{\underbrace{1,\ldots,1}_{i},0}+a_{\underbrace{1,\ldots,1}_{i},1}+n-i+1)\Gamma(a_{\underbrace{1,\ldots,1}_{i},1}+(\alpha-1)t_{i+1})} \qquad \text{(S1.12)}$$

When data are censored, simply replace the term $\frac{Pr(B_{\vec{\epsilon_m}^i}|P)}{\lambda(B_{\vec{\epsilon_m}^i})}$ or $\frac{P'(B_{\vec{\epsilon_m}^i}|P)}{\lambda(B_{\vec{\epsilon_m}^i})}$ by $Pr(B_{\vec{\epsilon_i}^i} \mid P)$ or $P'(B_{\vec{\epsilon_i}^i} \mid P)$ respectively depending on whether the observation is coming form sample $\mathfrak{X}$ or $\mathfrak{Y}$.

Analogous calculation leads to

$$BF_{01} = \frac{1}{\alpha^{n_2'}} \frac{E_1^c}{E_2^c} \qquad \text{(S1.13)}$$

where

$$E_1^c = E[(1-Y_0)^n Y_{10}^{d_1}(1-Y_{10})^{n-1} \ldots Y^{d_i}_{\underbrace{1,\ldots,1}_{i},0}(1-Y_{\underbrace{1,\ldots,1}_{i},0})^{n-i} \ldots Y^{d_{n-1}}_{\underbrace{1,\ldots,1}_{n-1},0}(1-Y_{\underbrace{1,\ldots,1}_{n-1},0})]$$

$$E_2^c = E[(1-Y_0)^{n_1+n_2\alpha} Y_{10}^{d_1}(1-Y_{10})^{n-1+(\alpha-1)t_2} \ldots Y^{d_i}_{\underbrace{1,\ldots,1}_{i},0}(1-Y_{\underbrace{1,\ldots,1}_{i},0})^{n-i+(\alpha-1)t_{i+1}}$$

$$\cdots Y_{\underbrace{1,\ldots,1}_{n-1},0}(1-Y_{\underbrace{1,\ldots,1}_{n-1},0})^{1+(\alpha-1)t_n}],$$

and $n_2' = \sum_n^{i=1} d_i Z_i$ is the total number of uncensored observations in sample $\mathfrak{Y}$.

Integration yields the Bayes factor for data with censoring as follows

$$
\begin{aligned}
BF_{01} =& \frac{1}{\alpha^{n_2'}} \frac{\Gamma(a_1 + n)\Gamma(a_0 + a_1 + n_1 + n_2\alpha)}{\Gamma(a_0 + a_1 + n)\Gamma(a_1 + n_1 + n_2\alpha)} \\
& \prod_{i=1}^{n-1} \frac{\Gamma(a_{\underbrace{1,...,1}_{i},1} + n - i)\Gamma(a_{\underbrace{1,...,1}_{i},0} + a_{\underbrace{1,...,1}_{i},1} + (\alpha - 1)t_{i+1} + d_i)}{\Gamma(a_{\underbrace{1,...,1}_{i},0} + a_{\underbrace{1,...,1}_{i},1} + n - i + d_i)\Gamma(a_{\underbrace{1,...,1}_{i},1} + (\alpha - 1)t_{i+1})}
\end{aligned}
\tag{S1.14}
$$

□