Testing Hypotheses under Lehmann Alternatives with Polya Tree Priors Doctoral Dissertation Proposal

Lei Huang
Under Supervision of Dr Malay Ghosh
University of Florida

Dec. 5th, 2012

Outline of Presentation

1. INTRODUCTION
2. LITERATURE REVIEW
   - Review of Testing Lehmann Alternatives
   - Review of Estimation in Cox Proportion Hazards Model
   - Introduction to Polya Tree Processes
3. CURRENT RESULTS
   - Testing Lehmann Alternatives with Polya Tree Processes Priors
     - Two-sample Case
     - One-sample Case
4. FUTURE WORK

Lehmann Alternatives

Definition:
Suppose $X_1, \cdots, X_{n_1} \sim F(x)$ and $Y_1, \cdots, Y_{n_2} \sim H(x)$.

$$H_0 : F = H \text{ vs } H_1 : H(x) = 1 - \{1 - F(x)\}^\alpha = 1 - S^\alpha(x)$$

where $S(x) = 1 - F(x)$ is the survival function, $\alpha > 0$ and $\alpha \neq 1$.

Motivations:

- Mathematical simplicity.
- When α is an integer, $H(x)$ is the distribution function of the minimum of α independent random variables each having distribution function F.
- H_1 introduces a very natural stochastic ordering of F and H.
- Related to Cox proportional hazards model

Cox Proportional Hazards Model

Cox proportional hazards model introduced by Cox (1972):

$$h(t \mid x) = h_0(t) exp(x^T \beta) \tag{1}$$

where $h_0(t)$ is the baseline hazard function and β is a vector of coefficients. It follows that the survival function is

$$S(t \mid x) = [S_0(t)]^{exp(x^T \beta)}$$

β is generally estimated by maximizing partial likelihood function

$$PL(\beta) = \prod_{k=1; d_k=1}^{n} \frac{exp(x_k^T \beta)}{\sum_{j=k}^{n} exp(x_j^T \beta)} \tag{2}$$

where d_k is censoring indicator.

Lehmann (1953)

Suppose the ranks of Y's in the combined sample are denoted by $e_1, \cdots, e_{n_2}$. The complete set of rank is determined by the ranks of Y's alone. Lehmann (1953) derived

$$P(E_1 = e_1, \cdots, E_{n_2} = e_{n_2}) = \frac{\alpha^{n_2}}{\binom{n_1+n_2}{n_1}} \prod_{j=1}^{n_2} \frac{\Gamma(e_j + j\alpha - j)}{\Gamma(e_j)} \frac{\Gamma(e_{j+1})}{\Gamma(e_{j+1} + j\alpha - j)} \tag{3}$$

Using Equation (3), one can compute the power of various rank tests against the alternatives.

Savage (1956)

Savage (1956) considered a different but more general question:

$$H_L : F(x) = F_0(x)^{\alpha_1} \text{ and } H(x) = F_0(x)^{\alpha_2}$$

where $\alpha_2 > \alpha_1 > 0$ and $F_0(x)$ is a continuous cumulative distribution function.
Put these two samples together and denote it by $V_1, \cdots, V_{n_1+n_2}$.
Assume:

- WLOG, assume that V_k's are ordered.
- There are no ties.

Then, $V_1 < V_2 < \cdots < V_{n_1+n_2}$.

Savage (1956)

Define indicator functions Z_k, $k = 1, \cdots, n_1 + n_2$, as

$$Z_k = \begin{cases} 0, & \textit{if } V_k \in \mathfrak{X} = \{X_1, \cdots, X_{n_1}\} \\ 1, & \textit{if } V_k \in \mathfrak{Y} = \{Y_1, \cdots, Y_{n_2}\} \end{cases} \tag{4}$$

Under H_L, the probability of a rank order $z_1, \cdots, z_{n_1+n_2}$ is given by

$$Prob. = \frac{n_1!n_2!\alpha_1^{n_1}\alpha_2^{n_2}}{\prod_{i=1}^{n_1+n_2}\left(\sum_{j=1}^{i}\left[(1-z_j)\alpha_1 + z_j\alpha_2\right]\right)}.$$

Other Literature

- Davies (1971) showed asymptotic equivalence of the approaches of Lehmann (1953) and Savage (1956).
- Brooks (1974) assigned an F distribution prior to α and conducted a Bayesian test.
- Miura and Tsukahara (1993) discussed the estimation problem in one-sample generalized Lehmann alternative model

Summary:
The above-mentioned tests are based only on the ranks which remain invariant under one-to-one transformation.This considers only the positions, but not the magnitudes of differences of order statistics.

Goal: Establish a testing method such that not only the ranks, but also the spacings of order statistics are taken into account.

Cox proportional hazards model

Bayesian nonparametric methods:

- Kalbfleisch (1978) — Gamma process
- Hjort (1990) — Beta process
- Muliere and Walker (1997) — Polya tree process (without covariates)
- Hanson (2006) — Mixture of finite Polya tree model
- Hanson and Jara (2012) — Mixture of Polya tree process

What is Polya tree process?

Notations:

- Let $E = \{0, 1\}$, $E^0 = \emptyset$. Let $E^m = E \times E \times E \cdots \times E$ be the m-fold product and $E^* = \bigcup_0^\infty E^m$.
- let Ω be a separable measurable space. Define a separating binary tree of partition of Ω, $\Pi = \{\pi_m, m = 0, 1, 2, ...\}$, such that :
    - $\pi_0 = \Omega$.
    - $\pi_0, \pi_1, ...$ form a sequence of partitions such that $\bigcup_0^\infty \pi_m$ generates the measurable sets.
    - 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 $B \bigcup \emptyset$.

What is Polya tree process?

Definition: For each m, $\pi_m = \{B_{\vec{\epsilon_m}} : \vec{\epsilon_m} = \epsilon_1, ..., \epsilon_m \in E^m\}$ 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 $\mathfrak{y} = \{W_{\vec{\epsilon_m}} : \vec{\epsilon_m} \in E^*\}$ be a collection of random variables. Then we say a random probability measure P on Ω have a Polya tree distribution with parameter (Π, A), written $P \sim PT(\Pi, A)$, if the following hold:

What is Polya tree process?

1. all the random variables in $\mathfrak{y}$ with subscripts ending with 0 are independent, i.e. $W_{\vec{\epsilon_m},0}$, for all $\vec{\epsilon_m} \in E^*$, are independent; $W_{\vec{\epsilon_m},1} = 1 - W_{\vec{\epsilon_m},0}$
2. for every $\vec{\epsilon_m} \in E^*$, $W_{\vec{\epsilon_m},0} \sim Beta(a_{\vec{\epsilon_m},0}, a_{\vec{\epsilon_m},1})$;
3. for every m=1,2,... and every $\vec{\epsilon_m} \in E^*$,

$$P(B_{\epsilon_1,\ldots,\epsilon_m}) = \left(\prod_{j=1;\epsilon_j=0}^{m} W_{\epsilon_1,\ldots,\epsilon_j}\right) \prod_{j=1;\epsilon_j=1}^{m} (1 - W_{\epsilon_1,\ldots,\epsilon_{j-1},0})$$

$$= \prod_{j=1}^{m} W_{\epsilon_1,\ldots,\epsilon_j} \tag{5}$$

Properties of Polya tree processes

Some established properties of Polya tree process that we need to use by Lavein (1992 & 1994), Mauldin et al. (1992) and Ghosh and Ramamoorthi (2003):

- Conjugacy: update when observe $X = x$
- Continuity: Polya trees can assign probability 1 to the set of continuous distributions. A sufficient condition is, for example, $a_{\vec{\epsilon}_m} = m^2$.

Properties of Polya tree processes

- Density: a Polya tree with partitions $\{B_{\vec{\epsilon_m}} : \vec{\epsilon_m} \in E^*\}$ and parameters A has predictive density at $x \in B_{\vec{\epsilon_m}}$ is

$$f(x) = \lim_{m \to +\infty} \frac{Pr(B_{\vec{\epsilon_m}})}{\lambda(B_{\vec{\epsilon_m}})} = \lim_{m \to +\infty} \frac{\prod_{i=1}^{m} \frac{a_{\epsilon_1,\ldots,\epsilon_j}}{a_{\epsilon_1,\ldots,\epsilon_{j-1},0} + a_{\epsilon_1,\ldots,\epsilon_{j-1},1}}}{\lambda(B_{\vec{\epsilon_m}})} \tag{6}$$

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

Properties of Polya tree processes

- Centering: A Polya tree can be constructed centering at an arbitrary distribution. Suppose $\Omega = \mathbb{R}$, and any pre-specified distribution function G. Two ways to do this:
  - M1 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 $k = 0, 1, ..., 2^m - 1$, with the obvious interpretation for $G^{-1}(0)$ and $G^{-1}(1)$, and $a_{\vec{\epsilon_m}} = m^2$.
  - M2 let the partition be data-dependent, i.e. $B_1 = [x_1, +\infty)$, $B_{11} = [x_2, +\infty)$,...,$B_{\underbrace{1,...,1}_{n}} = [x_n, +\infty)$. Parameters $a_{\vec{\epsilon_m}}$ need to satisfy

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

    and $a_{\vec{\epsilon_m}}$ should grow quickly enough to ensure the continuity property.

Bayes Factor

Setup:

- $X_1, \cdots, X_{n_1} \sim F(x)$ and $Y_1, \cdots, Y_{n_2} \sim H(x)$.
- Suppose $V_1 < V_2 < \cdots < V_{n_1+n_2}$ is the order statistic of combined sample when no ties occur. Indicator functions Z_k, $k = 1, \cdots, n_1 + n_2$ are defined as usual.
- Let $d_1, \cdots, d_{n_1+n_2}$ be the censoring indicators corresponding to $V_1, \cdots, V_{n_1+n_2}$.

By definition, the Bayes factor is

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

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.

Bayes Factor— Two-sample case

Then by simple calculation,

$$BF_{01} = \frac{\int f_0(v_1, ..., v_{n_1+n_2} \mid P)dPT(P)}{\int f_1(v_1, ..., v_{n_1+n_2} \mid P)dPT(P)} \tag{7}$$

Thus the Bayes factor is given by the ratio of the marginal distributions of $V_1, ..., V_{n_1+n_2}$ under H_0 to that under H_1.

Theorem 1:

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

Properties of Bayes Factor— Two-sample case

- For fixed t_i, BF_{01} is a decreasing function of r_i ($\alpha > 1$) .

$$\textit{small } BF_{01} \leftarrow \textit{large } r_i \rightarrow \textit{data clustered} \rightarrow H_1$$

- Monotonicity on t_i:
  - BF_{01} increase when $\alpha > 1$, then $F(x) < H(x)$. More likely to observe large t_i's under H_0.
  - BF_{01} decrease when $\alpha < 1$, then $F(x) > H(x)$. More likely to observe large t_i's under H_1.

Summary:
Both the rank order statistics and the spacings of order statistics are considered in the test.

Effects of α

- When $\alpha \to \infty$ ($\alpha > 1$), $BF_{01} \to \infty$ if there exists at least one uncencored observations in sample $\mathfrak{X}$ and such that t_{i+1} is nonzero.
- Similar results when $\alpha \to 0^+$ and $\alpha < 1$.

An extreme case: when all uncensored observations in sample $\mathfrak{Y}$ are clustered closely and they are uniformly smaller than any observations in sample $\mathfrak{X}$.

$$\begin{cases} t_1 = n_2,\ t_2 = n_2 - 1, \cdots, t_{n_2} = 1 \\ t_i = 0,\ for\ i = n_2 + 1, \cdots, n_1 + n_2 \end{cases} \tag{8}$$

$BF_{01} \to 0$
As $\alpha \to \infty$, H degenerates at the infimum of the support of F.

Effects of spacings

Suppose $V_{k_0-1} \approx V_{k_0}$ for some $k_0 \in \{2, \cdots, n_1 + n_2\}$ and they come from different samples. Switching the positions of V_{k_0-1} and V_{k_0} results in

$$BF_{01}^{new} = \frac{\sigma_{k_0} + n_1 + n_2 + 1 - k_0 + (\alpha - 1)t_{k_0}^{new}}{\sigma_{k_0} + n_1 + n_2 + 1 - k_0 + (\alpha - 1)t_{k_0}^{old}} BF_{01}^{old}$$
$$= (1 \pm \frac{\alpha - 1}{\sigma_{k_0} + n_1 + n_2 + 1 - k_0 + (\alpha - 1)t_{k_0}^{old}}) BF_{01}^{old}$$

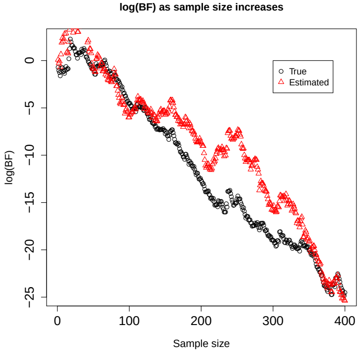
$\left| \frac{\alpha-1}{\sigma_{k_0}+n_1+n_2+1-k_0+(\alpha-1)t_{k_0}^{old}} \right|$ is called the *expansion rate*.

Effects of spacings

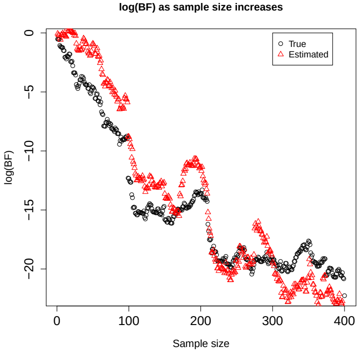
One can choose suitable parameters to meet his needs:

- When $\sigma_1 = \cdots = \sigma_{n_1+n_2}$, the expansion rate goes up as k_0 increases
  $\Rightarrow$ the order statistics play a more important role in the tail than in the beginning.
- When σ_{k_0} grow fast with k_0, the expansion rate could be a decreasing function of k_0
  $\Rightarrow$ a test more sensitive in the beginning.
- The expansion rate grows when α increases if $\alpha > 1$ or $\alpha \to 0^+$ if $\alpha < 1$
  $\Rightarrow$ the effect of order is enlarged when the distance of the alternative from the null is increased.

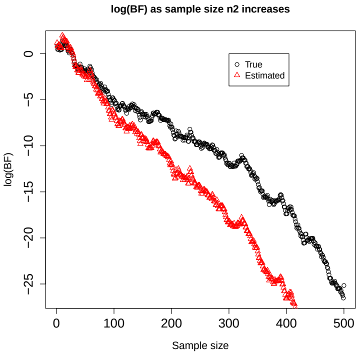
Simulations



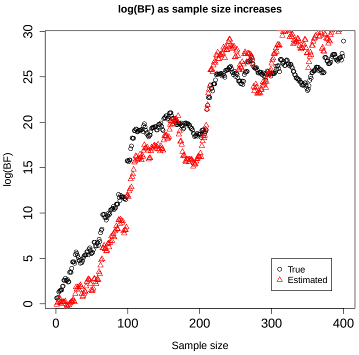
(a) $\alpha = 1.5$



(b) $\alpha = 0.7$



(c) $\alpha = 1.5$



(d) $\alpha = 0.7$

Real data application

Acute myelogenous leukemia (AML)

Table : **Acute Myelogenous Leukemia**

Treatment	Survival Time
Maintained	9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+
Nonmaintained	5, 8, 12, 16+, 23, 27, 30, 33, 43, 45

Table : Cox proportional hazards model

	coef	**exp(coef)**	**se(coef)**	**P-value**	**lower .95**	**upper .95**
Nonmaintained	-0.9155	0.4003	0.5119	0.0737	0.1468	1.092

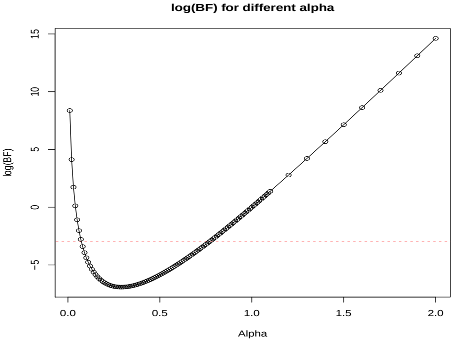
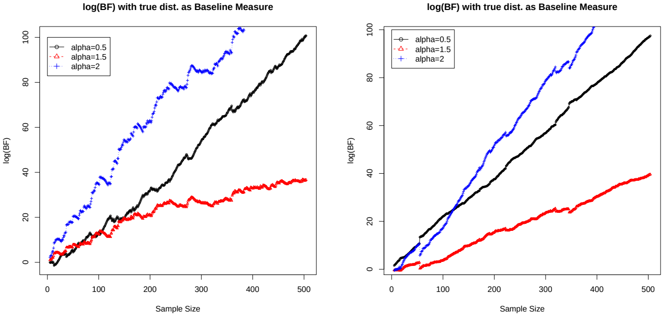


Figure : Leukemia: log(BF) for different α

Strong evidence interval based on Kass and Raftery (1995) criteria: $(0.10, 0.75)$.

Simulations- One-sample Case



(a) Baseline Measure: True Weibull Distribution

(b) Baseline Measure: Estimated Weibull Distribution

Figure : Log(BF) grows drastically as sample size increases under the null hypothesis

Simulations

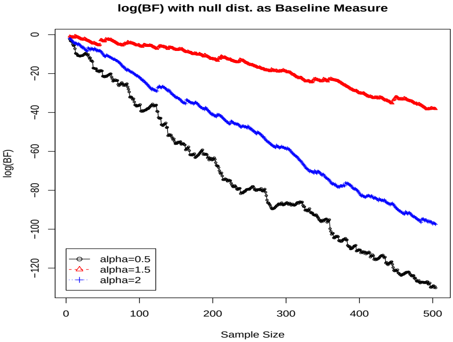


Figure : Log(BF) decreases drastically as sample size increases under the alternative hypothesis

Future Work

- Modification for ties
- Comparison to other Bayesian nonparametric estimators
- Fully Bayesian analysis for Cox model
- Reliability analysis and life testing

THANK YOU!