Using BDC and HSIC in Semi-Nonparametric Inference

S Teran Hidalgo*, S Hoberman*, M Wu† and MR Kosorok*

UNC-Chapel Hill Department of Biostatistics (*) and Fred Hutchinson Cancer Research Center (†)

May 1, 2014
Outline

- Using Brownian distance covariance (BDC) for testing independence in survival data
- Using Hilbert-Schmidt independence criterion (HSIC) for testing additivity
- Conclusions
Testing for Independence

- The goal: testing for independence of failure time $T$ and covariate vector $Z$ under right-censoring:
  - Observe only $X = T \wedge C$ and $\delta = 1\{T \leq C\}$.
  - $T$ and $C$ are independent given $Z$.
  - Under $H_0$: $T$ and $Z$ are independent, only $C$ and $Z$ may be dependent.
- The challenge: cannot directly compute BDC of $T$ and $Z$. 
BDC Definition

Let $a_{ij} = |T_i - T_j|$ and $b_{ij} = \|Z_i - Z_j\|$, and define

$$A_{ij} = a_{ij} - a_i - a_j - a.. \quad \text{and} \quad B_{ij} = b_{ij} - b_i - b_j - b...$$

Then

$$V_n^2(A, B) = \frac{1}{n^2} \sum_{i,j=1}^n A_{ij}B_{ij} \quad \text{and} \quad \text{BDC} = \frac{V_n^2(A, B)}{\sqrt{V_n^2(A, A)V_n^2(B, B)}}.$$

The main idea is to “estimate” $T_i$ with some $\hat{T}_i$ to create $\hat{a}_{ij}$ and $\hat{A}_{ij}$ and then compute BDC with these substitutions.
Two Proposals

- Let \( \hat{S}_n \) be the usual Kaplan-Meier estimator based on censored failure time data, ignoring \( Z \), ranging over \([0, \tau]\) where \( \tau \) is the upper limit of observation times.
- Let \( \hat{\Lambda}_n = -\log \hat{S}_n \).
- Method 1:
  - Let \( \tilde{T}_i(c) \) be a random draw \( T \) from \( \hat{F}_n = 1 - \hat{S}_n \) given \( T > c \).
  - Use \( \hat{T}_i = X_i + (1 - \delta_i) \tilde{T}_i(X_i) \) (truncated at \( \tau \)).
- Method 2:
  - Let \( E_i \) be a random draw from a standard exponential distribution.
  - Use \( \hat{T}_i = \hat{\Lambda}_n(X_i) + (1 - \delta_i)(X_i + E_i) \) (truncated at \( \tau \)).
Not hard to show that under $H_0$,

$$\max_{1 \leq i \leq n} |\hat{T}_i - T_i^*| \to 0,$$

in probability, where the $T_i^*$s are independent of the $Z_i$s, and

- Under Method 1: $T_i^*$ has the same distribution as $T_i$.
- Under Method 2: $T_i^*$ is a standard exponential random deviate truncated at $\tau$.

We permute $\hat{T}_i$ and $Z_i$ and recompute BDC to obtain p-values.
Simulations

- Four dependency relationships: none (null), linear, exponential, sinusoidal.
- Two levels of censoring: heavy (50%) and light (20%).
- Three statistics: BDC Method 1 (BDC1), BDC Method 2 (BDC2), and Cox Wald test (Cox).
## Simulation Results

<table>
<thead>
<tr>
<th>setting</th>
<th>null</th>
<th>linear</th>
<th>exponential</th>
<th>sinusoidal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>light censoring</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cox:</td>
<td>0.05</td>
<td>1.00</td>
<td>0.30</td>
<td>0.06</td>
</tr>
<tr>
<td>BDC1:</td>
<td>0.05</td>
<td>0.76</td>
<td>0.77</td>
<td>0.99</td>
</tr>
<tr>
<td>BDC2:</td>
<td>0.05</td>
<td>0.19</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td><strong>heavy censoring</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cox:</td>
<td>0.04</td>
<td>1.00</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td>BDC1:</td>
<td>0.05</td>
<td>0.58</td>
<td>0.61</td>
<td>0.97</td>
</tr>
<tr>
<td>BDC2:</td>
<td>0.06</td>
<td>0.11</td>
<td>0.10</td>
<td>0.15</td>
</tr>
</tbody>
</table>
Testing Additivity

Given a response and a set of variables, one may want to determine whether any interactions are present in the model beyond the main effects, without having to specify:

- which features are involved in the interaction
- whether the interaction happens to be a two-way or three-way interaction or higher
- what functional form the interaction has
Specifically, we have $p$ covariates and we wish to evaluate the semi-nonparametric additive model

$$Y = f_1(X_1) + \ldots + f_p(X_p) + \varepsilon,$$

where the $\varepsilon$ are independent and mean zero with finite variance, and we are interested in knowing whether this model is sufficient or there exists some arbitrary interactions.
This SS-ANOVA model can be estimated by solving the following penalized least squares problem:

\[
\frac{1}{n} \sum_{i=1}^{n} (Y_i - f(X_i))^2 + \lambda J(f),
\]

with \( f : [0, 1]^p \rightarrow \mathbb{R}, f = f_1 + \ldots + f_p, f_j : [0, 1] \rightarrow \mathbb{R} \) and

\[
J(f) = \sum_{j=1}^{p} \theta_j^{-1} \int |f_j^{(m)}|^2. \quad \lambda \text{ and } \theta_j \text{ are tuning parameters selected through GCV.}
\]
The specific hypothesis we want to test is

\[ H_0 : Y = f_1(X_1) + \ldots + f_p(X_p) + \varepsilon \]

versus

\[ H_A : Y = f_1(X_1) + \ldots + f_p(X_p) + f_{\{1,\ldots,p\}}(X_1, \ldots, X_p) + \varepsilon \]

with \( f_{\{1,\ldots,p\}}(X_1, \ldots, X_p) \) being any possible interaction or combination of interactions (left undefined).
Method

We propose an interaction test in SS-ANOVA. This is done through the use of the Hilbert-Schmidt Independence Criterion (HSIC) test statistic (Gretton, et al., 2005). We are able to test if any interactions exist beyond the main effects. The procedure works as follow:

- Fit a SS-ANOVA model with p main effects.
- Calculate the HSIC statistic between the estimated residuals and the p variables.
- Use a semi-nonparametric bootstrap to estimate the distribution of the HSIC statistic under the null.
- Derive a p-value from this distribution.
This same method can be extended to a Goodness-of-fit test for SS-ANOVA.

Bodhi Sen did similar work but for the linear model in the paper *On Testing Independence and Goodness-of-fit in Linear Models* (Sen and Sen, 2013).

The basic bootstrap approach from that paper is extended to our setting.
Proposed Bootstrap Procedure

Step 1
Create an empirical distribution $P_{n,e^o}$ from the centered estimated residuals \( \hat{\varepsilon}_i = Y_i - \hat{f}(X_i) \) rescaled to correct for large $p$ (see below).

Step 2
Draw a bootstrap sample $\eta^*$ from the empirical distribution $P_{n,e^o}$ and draw a bootstrap sample $X^*$ from the empirical distribution $P_{n,X}$ of the X's independently of the $\eta^*$. Then set $Y_i^*$ as
\[
Y_i^* = \hat{f}(X_i^*) + \eta_i^*
\]

Step 3
We estimate $\hat{f}^*(X^*)$ from $Y_i^*$ and $X^*$, and create new bootstrap residuals as
\[
\varepsilon_i^* = Y_i^* - \hat{f}^*(X_i^*)
\]

Step 4
Calculate the test statistic as $nT_n(X^*, \varepsilon^*)$. 
The Hilbert-Schmidt independence criterion (HSIC) between two random vectors $X$ and $Y$ with joint distribution $P_{x,y}$ is defined as

$$HSIC(X, Y) := E[k(X, X')I(Y, Y')] + E[k(X, X')]E[I(Y, Y')] - 2E[k(X, X')I(Y, Y'')]$$

with $k(X, X') = I(X, X') = \exp(-||X - X'||^2)$. 
$HSIC(X, Y) = 0$ if and only if $P_{x,y} = P_x \times P_y$.

We can estimate HSIC with $T_n$:

$$T_n(X, Y) = \frac{1}{n^2} \sum_{i,j}^{n} k_{ij} l_{ij} + \frac{1}{n^4} \sum_{i,j,q,r}^{n} k_{ij} l_{qr} - 2 \frac{1}{n^3} \sum_{i,j,q}^{n} k_{ij} l_{iq}$$

where $k_{ij} = \exp(-\|X_i - X_j\|^2)$ and $l_{ij} = \exp(-\|Y_i - Y_j\|^2)$ and $\| \cdot \|$ is the euclidean distance.
Proposed Bootstrap Procedure

Step 5
Iterate Step 1 through 4 until enough bootstrap samples of $nT_n(X^*, \varepsilon^*)$ have been generated. This distribution approximates the distribution of $nT_n(X, \hat{\varepsilon})$ under the null.
In **Step 1**, \( P_{n,e} \) is the distribution with mass of \( \frac{1}{n} \) at each 
\[
\frac{\hat{\sigma}}{\hat{\sigma}'} (\hat{\varepsilon}_i - \bar{\varepsilon}),
\]
where

- \( \bar{\varepsilon} = \sum_{i=1}^{n} \frac{\hat{\varepsilon}_i}{n} \),
- \( \hat{\sigma}'^2 = \frac{\sum_{i=1}^{n} (\hat{\varepsilon}_i - \bar{\varepsilon})^2}{n} \)
- \( \hat{\sigma}^2 = \frac{||Y - AY||^2}{Tr(I - A)} \):
  - \( \hat{\sigma}'^2 \) is the variance of the distribution of the estimated residuals.
  - This tends to decrease with respect to the true \( \sigma^2 \) when \( p \) increases, and \( \hat{\sigma}^2 \) accounts for the increase in \( p \).
Theoretical Results

**Lemma 1**
Let \( f : [0, 1] \rightarrow \mathbb{R} \) and \( \int_0^1 (f^{(m)}(u))^2 du < \infty \). Then,

\[
\| f \|_\infty = O \left( \| f \|_2^{\frac{2m-2}{2m-1}} \right).
\]

**Proof:** Polynomial approximation in Sobolev Spaces.
Theoretical Results

**Theorem 1**
If $\hat{f}$ is the SS-ANOVA estimator of $f$, then

$$||\hat{f} - f||_2 = O_P \left( [n(\log n)^{1-r}]^{-2m/(2m+1)} \right),$$

where $r$ is the highest degree of interaction. For the current presentation we will only consider the additive model, hence $r = 1$.

**Proof:** Yi Lin (2000, *AOS*), *Tensor product Space ANOVA Models.*
Theoretical Results

**Lemma 2**
Let $m = 2$ and $r = 1$, then

$$||\hat{f} - f||_{\infty} = O_P(n^{-8/15}).$$

**Proof:** Apply Lemma 1 to Theorem 1.
Theoretical Results

Lemma 3
Let \( \mathcal{F} = \{ f : [0, 1]^p \to \mathbb{R}, f = f_1 + \ldots + f_p, \sum_{j=1}^{p} \int_{0}^{1} |f_j^{(m)}|^2 < M \} \). Then,
\[
\mathcal{H}_n(\delta, \mathcal{F}) \leq p A \left( \frac{pM}{\delta} \right)^{1/m} = A^* \left( \frac{M}{\delta} \right)^{1/m},
\]
where \( \mathcal{H}_n(\delta, \mathcal{F}) \) is the \( \delta \)-entropy with respect to the empirical norm \( ||f||_n^2 = \frac{1}{n} \sum_{i=1}^{n} |g(x_k)|^2 \).

Proof: \( \mathcal{F} = \mathcal{F}_1 + \ldots + \mathcal{F}_p \) and we know the entropy of each \( \mathcal{F}_j \).
Theoretical Results

**Theorem 2**
If \( \hat{f} \) is the SS-ANOVA estimator of \( f \) for only main effects, then

\[
\| \hat{f} - f \|_n^2 = O_p(\lambda^*)
\]

provided \( n^{2m/(2m+1)} \lambda^* \geq 1 \) and \( \lambda^* \) is the slowest converging tuning parameter \( \lambda \theta_j^{-1} \).

**Proof:** Use lemma 3 and modified Theorem 6.2 of Van De Geer (1990, AOS), *Estimating a Regression Function.*
Theoretical Results

Theorem 3
Under $H_0$ in a new probability space,

$$n T_n (X^*, \varepsilon^*) - n T_n (X, \hat{\varepsilon}) \xrightarrow{p} 0 \text{ as } n \to \infty.$$ 

Hence $n T_n (X^*, \varepsilon^*)$ is a good approximation to the asymptotic distribution of the test statistic. Remember that $X^*$ and $\varepsilon^*$ are the bootstrapped versions of $X$ and $\hat{\varepsilon}$. 
In our setting we have the original model,
\[ Y_i = f(X_i) + \varepsilon_i. \]
The bootstrap version,
\[ Y_i^* = \hat{f}(X_i^*) + \eta_i^*. \]
And the bootstrap residuals
\[ \varepsilon_i^* = Y_i^* - \hat{f}^*(X_i^*). \]
Which gives us
\[ \varepsilon_i^* - \eta_i^* = \hat{f}(X_i^*) - \hat{f}^*(X_i^*). \]
The statistic can be written as

\[ T_n(X^*, \varepsilon^*) = \frac{1}{n^2} \sum_{i,j} k_{ij} l_{ij}^* + \frac{1}{n^4} \sum_{i,j,q,r} k_{ij} l_{qr}^* - 2 \frac{1}{n^3} \sum_{i,j,q} k_{ij} l_{iq}^*, \]

with \( k_{ij} = \exp(-\|X_i^* - X_j^*\|^2) \) and \( l_{ij}^* = \exp(-\|\varepsilon_i^* - \varepsilon_j^*\|^2) \).
Sketch of the Proof

We do a taylor expansion of $T(X^*, \varepsilon^*)$ by expanding $l_{ij}^*$ at $\eta_i^*$ and $\eta_j^*$. We only use the expansion up to the second partial derivatives with the rest being the remainder.

By using Lemma 2 and Theorem 3 we can show the remainder goes to 0 in probability as $n \to \infty$. 
Now, if we look at the difference

\[ nT_n(X^*, \varepsilon^*) - nT_n(X, \hat{\varepsilon}) \]

We can use Lemma 2, the fact that the remainder goes to 0 and some of the techniques in Sen and Sen (2013), and we obtain that the difference goes to 0 in probability.
Example 1

We simulate the following hypotheses

\[ H_0 : Y = 5\sin(\pi X_1) + 2X_2^2 + \varepsilon \]

\[ H_A : Y = 5\sin(\pi X_1) + 2X_2^2 + 0.75\cos(\pi(X_1 - X_2)) + \varepsilon \]

with \( X_1 \) and \( X_2 \) distributed standard uniform and \( \varepsilon \) standard normal. We simulate this with sample size ranging from 100 to 500.
Example I

- .01 Error
- .05 Error
- 0.01 Power
- 0.05 Power

- n=100
- n=200
- n=300
- n=400
- n=500
Example II

We simulate the following hypotheses

\[ H_0 : Y = 5\sin(\pi X_1) + 2X_2^2 + 2\sin(\pi X_3) + X_4^2 + \varepsilon \]

\[ H_A : Y = 5\sin(\pi X_1) + 2X_2^2 + 2\sin(\pi X_3) + X_4^2 + 0.5\cos(\pi(X_1 - X_2)) + 0.5\cos(\pi(X_3 - X_4)) + \varepsilon \]

with \( X_1, \ldots, X_4 \) distributed standard uniform and \( \varepsilon \) standard normal. We simulate this with sample size ranging from 100 to 500.
Example II
Example II

Using BDC and HSIC in Semi-Nonparametric Inference

S Teran Hidalgo*, S Hoberman*, M Wu† and MR Kosorok*

* Corresponding authors

†Affiliated with the Department of Biostatistics, School of Public Health, University of North Carolina at Chapel Hill.
We are finishing an extension of this method for a function in a Tensor Product Space SS-ANOVA and in the setting where $p$ is allowed to grow with $n$.

This allows us to answer the question *Are they any interactions beyond the two-way interactions?* and also create a Goodness-of-Fit Test for general SS-ANOVA models with any kind of interactions.
THE END