Learning Methods For Spectrum Estimation

Joshua Pohlkamp-Hartt

Queen's University

October 20, 2015

Chapter 2: Background - The Multitaper Method(MTM)

- Our primary tool for spectral estimation, the multitaper method (MTM) balances the variance and bias of the estimated spectrum.

$$\overline{S}(f) = \frac{1}{K}\sum_{k=0}^{K-1} |Y_k(f)|^2, \quad (1)$$

$$Y_k(f) = \sum_{t=0}^{N-1} \upsilon_t^{(k)} e^{-2i\pi ft} x_t, \quad (2)$$

where Y_k are the eigenspectra and $\upsilon_t^{(k)}$ are the Slepian sequences in the time domain.

- **Issue 1**: The Slepian sequences (and spectral estimate) are defined for a choice of bandwidth, NW. The parameters of NW and number of windows, K, are user selected. K can range from NW to $2NW$.

Background - F-test for line components

We model the eigenspectra from the MTM at each frequency by windowed line components. We then perform an F-test to determine if each model is significant. Significant models indicate the presence of signals.

- We assume a model of

$$Y_k(f) = \mu(f)V_k(0) + \epsilon(f) \tag{3}$$

- Then we test $H_0 : \mu(f) = 0$ with the statistic,

$$F(f) = (K-1)\frac{|\hat{\mu}(f)|^2 \sum_{k=0}^{K-1} |V_k(0)|^2}{\sum_{k=0}^{K-1} |(Y_k(f) - \hat{\mu}(f)V_k(0))^2}, \tag{4}$$

$$\hat{\mu}(f) = \frac{\sum_{k=0}^{K-1} V_k(0)Y_k(f)}{\sum_{k=0}^{K-1} |V_k(0)|^2}. \tag{5}$$

- The F-statistic will follow an $F(2, 2K-2, \alpha)$ distribution if H_0 is true.
- **Issue 2**: The F-test suffers from false and missed detection problems in some situations.

Chapter 3: Sphericity Tests

- The objective of these tests is to identify if the parameter choices in the MTM are appropriate. (**Issue 1**)
- The residuals from the F-test, $\hat{r}_k(f)$ should follow $CN(0, \sigma^2_{noise})$ if the eigenspectra are fully described by the Slepians and the assumption of normality is true.
- By testing if the real and imaginary parts of the residuals have uniform variance and no covariance, we can see if eigenspectra are well described.
- To ensure that the residuals are independent, we test a subset of the residuals with that are $2W$ away from each other in the frequency domain. We define this number of frequencies as M.

Sphericity Tests - Naïve Sphericity Test

- We apply John's Sphericity Test (1972) to the concatenated vector of the K sets of residuals and tested the real and complex parts for sphericity.

$$H_0 : Cov([A, B]) = \sigma^2 \mathbb{I}. \tag{6}$$

  With unknown common variance, σ^2 and the real and imaginary parts in vector form are A (real) and B (imaginary).

- We use the test statistic

$$\Xi_N = (MK - 2)[\frac{1}{2(1 - \hat{W})^{1/2}} - 1], \hat{W} = \frac{tr((M\hat{R})^2)}{(tr(M\hat{R}))^2}, \quad \hat{R} = Cov([A, B]). \tag{7}$$

- Under the null hypothesis we have

$$\Xi_N \sim \mathbb{F}_{2,2MK-4} \tag{8}$$

Sphericity Tests - Bagged Sphericity Test

- If we wish to include information on the noise variance of the time series we use the hypothesis

$$H_0 : Cov([A, B]) = \sigma^2 \mathbb{I}, \tag{9}$$

where σ^2 is a specified variance for the noise process.

- Following Korin's test (1968) we get that under H_0,

$$\begin{aligned} \Xi_B &= \rho(MK-1)(\log \mid \sigma^2 \mathbb{I} \mid - \log \mid \hat{W} \mid + tr(\hat{W}(\sigma^2 \mathbb{I})^{-1}) - 2) \quad (10) \\ &\sim \chi^2_3 + \omega(\chi^2_7 - \chi^2_3) \quad (11) \end{aligned}$$

with

$$\omega = \frac{47}{432(MK-1)^2}, \; \rho = 1 - \frac{15}{18(MK-1)}. \tag{12}$$

Sphericity Tests - Bagged Sphericity Test cont.

- Direct implementation of this test is sensitive to the varying sample size of the concatenated residual vector from changes to K.
- To avoid the issue with sample size we implement a bagging algorithm, creating multiple sets of residuals with resampling and recombination with Fisher's combined probability test to determine an overall p-value.
- The procedure now is as follows:
  1) Sample the residuals with replacement O times to be used as separate data sets.
  2) Test the null hypothesis, $H_0 : Cov([A, B]) = \sigma^2\mathbb{I}$, for all O sets of samples using the statistic from equation 10 and then calculate the p-value for each set.
  3) Use Fisher's test to determine the overall p-value of the combination of the O sets of samples. Fisher's method states that $-2\sum_{i=1}^{O} ln(p_i) \sim \chi^2_{2O}$.

Sphericity Tests - Making parameter choices

- With these two tests we can identify how spherical the residuals are. The larger the p-value, the less evidence we have to reject the null hypothesis, which is analogous to having less evidence that our spectrum is poorly resolved.

  1. For a selected set of parameters we can use these tests to determine if they will reject the null at a set significance level to ensure these choices are reasonable.
  2. We can also test a variety of parameter choices and determine which has the largest p-value, selecting that as our ideal set of parameters.

Sphericity Tests - Simulations

- We wanted to check if the Sphericity tests would give reasonable choices of parameters under ideal conditions.
- We used a data set of 1000 points that was made up of 38 5-pronged sinusoidal signals of width .08Hz that are evenly spaced across the frequency band at .13Hz apart.

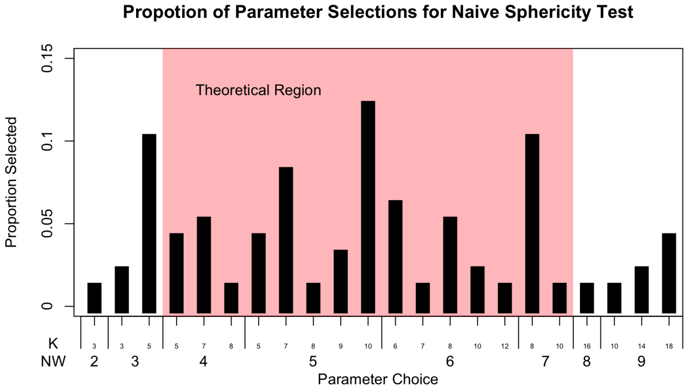
$$X(t) = \sum_{l=1}^{38} \alpha_l \sum_{j=-2}^{2} (.3 - .1 \mid j \mid) \sin(2\pi(.013l + .002j)t) + z_t, \quad (13)$$

  where α_l is a random amplitude for each signal that is taken from $U(.5, 1)$ and $z_t \sim N(0, 2)$.

- For the best MTM estimate, we wanted the frequency band to cover one signal, so ideally $NW \in [4, 7]$.
- We tested 1000 realization of 1000 data points from equation 13 and determined how often each test chose reasonable parameter values.

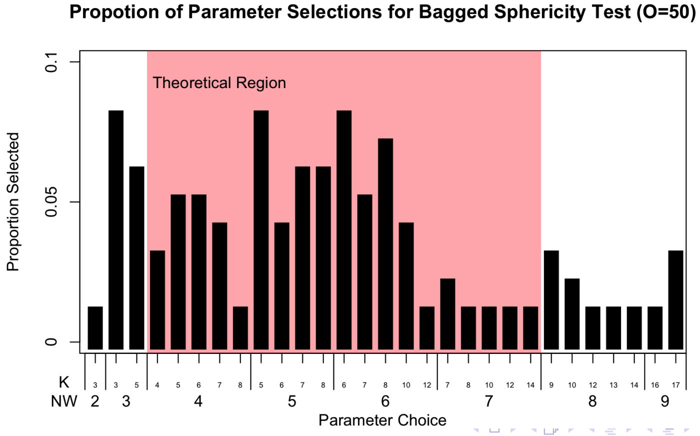
Sphericity Tests - Naïve Sphericity Test simulation results

- We tested the naïve sphericity of the residuals for parameters $NW \in [2, 10]$ and $K \in [2, 20]$. The test found $NW = 5$ and $K = 10$ was selected most often and that 64% of the choices were within our theoretical range.



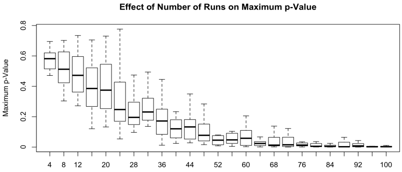
Sphericity Tests - Bagged Sphericity Test simulation results

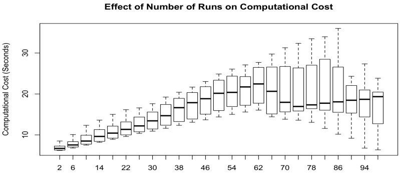
- For the bagged sphericity test with $O = 50$ on the parameters $NW \in [2, 10]$ and $K \in [2, 20]$ the results were that the choices of $(NW, K) = (3, 3)$ or $(5, 5)$ or $(6, 6)$ were equally most often selected and overall 73% of choices were in the theoretical range.

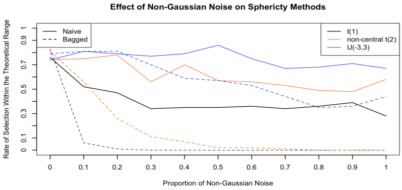


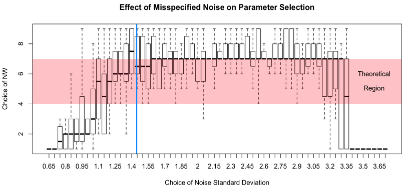
Sphericity Tests - Bagging Test Drawbacks

- The bagging test does run into several drawbacks; p-values are dependent on O, computationally expensive for consistent results, less robust to non-Gaussian noise, and sensitivity to noise variance choices.









Sphericity Tests - Comparison of Tests

- Both tests can provide a reasonable choice of parameters when no theoretical background is possible.
- The bagged sphericity test provides significant improvement in situations where the noise level is known.
- The bagged sphericity test does suffer from several drawbacks.
- For quick checks for parameter selection we recommend the naïve approach but if the noise level is well estimated and computer power is a non-issue then the bagged test is ideal.

Chapter 4: Bootstrapped F-Test

- The F-test suffers lowered detection rates and higher rates of false detection (**Issue 2**) when there is a low signal to noise ratio or the choice of parameters NW and K are made incorrectly.
- Ideally, if all the structure in the residuals were removed (we have chosen NW and K correctly), they should be random variables that are independent of frequency.
- Treating them as independent realizations, we can perform a bootstrap to get a better estimate of the distribution of the F-statistic at each frequency.
- We do so by re-sampling the residuals with replacement and computing a new F-value. We then take the mean of a number of re-sampled F-values as our test statistic.

Bootstrapped F-Test - Algorithm

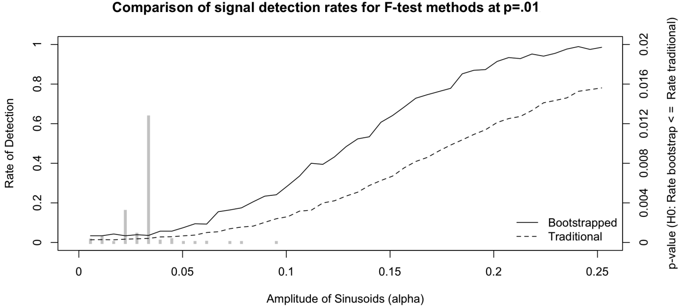
- The Bootstrapped residuals F-test algorithm is as follows,
  1. Compute the F-test.
  2. Re-sample the residuals.
  3. Compute new values for $\hat{Y}_k^{(1)}(f)$, $\hat{\mu}^{(1)}(f)$, $\hat{r}_k^{(1)\prime}(f)$, and $\hat{F}^{(1)}(f)$.
  4. Repeat steps 2 and 3 O times and take the mean of F-statistics found for each re-sampling.
  5. Lastly to test for signal detection we check if this value exceeds the empirical cut-off. The empirical cut-off for a given significance level is determined by the noise level of the time series and MTM parameter choices, NW and K.

Bootstrapped F-Test - Simulations

- We wanted to compare the ability of the Bootstrapped F-test ($O = 30$) to identify a single sinusoid with varying levels of amplitude within Gaussian noise of constant variance to that of the F-test.
- We simulated a data set, $Y_t = \alpha \sin(2\pi(.125t)) + z_t$, with $z_t \sim N(0,1)$, the amplitude of the signal, α varied across $(0, .5]$.
- We used $N = 1000$, $NW = 4$, and $K = 7$ for both tests.

Bootstrapped F-Test - Test comparison results

- The bootstrapped F-test had higher detection rates at both significance levels tested, p (.05 and .01).



- Overall, the bootstrapped F-test outperforms the traditional F-test for detecting signals and when the additional computational cost is available we recommend using it.

Chapter 5: Periodic Data Reconstruction

- From the MTM we have a method for estimating the periodic trends in a time series. We can apply this to the practical applications of interpolation and prediction.
- First demonstrated by Dr. David Thomson in 1990 we can reconstruct the periodic components found within a time series. The steps are:
  1. Multitaper Spectrum Estimation (mean of multiple windowed Fourier transforms)
  2. F-test & Complex Mean Values (regression in the frequency domain)
  3. Inverse Fourier Transform of Complex Mean Values (transform model back to time domain)

Periodic Data Reconstruction - Inverse Fourier Transform Reconstruction

- After identifying the significant frequencies for a set significance level, α, we set the complex mean values of all non-significant frequencies to zero.

$$\hat{\mu}_\alpha(f) = \begin{cases} \hat{\mu}(f), & \hat{F}(f) > F_{(2nw-1,2,\alpha)}. \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

- We now can take the inverse Fourier transform of the complex mean values to produce the periodic reconstruction of the time series.

$$\hat{x}_{t,\alpha} \approx \mathcal{F}^{-1}(\hat{\mu}_\alpha)(t). \tag{15}$$

- This reconstruction will suffer from bias at the end points of the series when we use $K < 2NW$. This bias is not a significant issue in our applications.

Periodic Data Reconstruction - Application to Interpolation and Prediction

Thomson's periodic reconstruction method can be directly applied to the problems of interpolation and prediction.

- For interpolating a gap in the data, we simply fill the gap with a simple interpolation (linear or mean value works well) and perform Thomson's method. This will produce a periodic reconstruction across the gap. Be warned that the choice of starting interpolation is important and can affect the estimates.

- As for prediction, we zero-pad (add zeros after last data point) the windowed sets of data and perform Thomson's method. This will result in predictions for the points we added as zeros and has the added bonus of improving the frequency resolution.

Periodic Data Reconstruction - Areas of Improvement

Within Thomson's method we have identified 3 areas that we feel can be addressed to improve the reconstruction of a time series.

1. Choice of significance level, α. (Cross Validation)
2. Improve the estimate of the periodic reconstruction. (Gradient Boosting)
3. Find distribution for reconstruction. (Bootstrapping)

Periodic Data Reconstruction - Choosing α

We would like an unsupervised method for identifying the best α value for interpolation or prediction. Using cross-validation as a framework for our decision making we should be able to find an optimal choice.

1. Split the data into bins.
2. Remove one bin and reconstruct the data for a set α.
3. Find the mean squared reconstruction error for the removed bin.
4. Perform steps $2-3$ across all bins and calculate the mean of the mean squared errors.
5. Repeat $1-4$ for all α values in our potential set.
6. Select the α with the minimum mean-mean squared error.

Periodic Data Reconstruction - Improving Reconstruction

There may still be additional signals left over after the periodic reconstruction. We will use a gradient boosting approach on the residuals to determine if any more signals are present.

1. Treat the residuals, r_t, as a new time series and find the periodic reconstruction.
2. Now make a greedy model by finding $\underset{\gamma}{\operatorname{argmin}} \sum_{t \in T} (y_t - (\hat{y}_t + \gamma \hat{r}_t))^2$.
3. Define the new reconstruction as $\hat{y}_t' = \hat{y}_t + \gamma \hat{r}_t$ and test for significance with an F-test.
4. If the model is significant we repeat steps $1 - 3$ with the updated reconstruction's residuals, $y_t - \hat{y}_t'$, as our new series.
5. Continue repeating steps $1 - 3$ with the new residuals until the updated reconstruction is not considered significant under the F-test. Then consider the reconstruction from the previous iteration as the final boosted periodic reconstruction.

Periodic Data Reconstruction - Distribution Estimation

Using a bootstrapping methodology we are able to produce an estimate of the distribution for our periodic reconstructions.

1. After creating a periodic reconstruction, sample with replacement the residuals and create a new simple interpolation or zero-padding with the sampled residuals added to the initial reconstruction.
2. Find a periodic reconstruction with the new time series.
3. Repeat steps 1 and 2, n times ($n > 60$).
4. The resulting set of reconstructions allows us to estimate the distribution at each time. This estimation can be parametric or not, depending on your assumptions.
5. We can also obtain estimates of the distribution of the noise from the residuals from each periodic reconstructions.
6. For reporting purposes we can produce overall confidence intervals by adding the confidence intervals for the noise process and periodic reconstruction together.

Periodic Data Reconstruction - Simulation

- We simulated 1200 samples of sinusoidal data and examined the mean-squared error for predicting the final 100 points and interpolating the middle 100 points for each of our proposed methods.

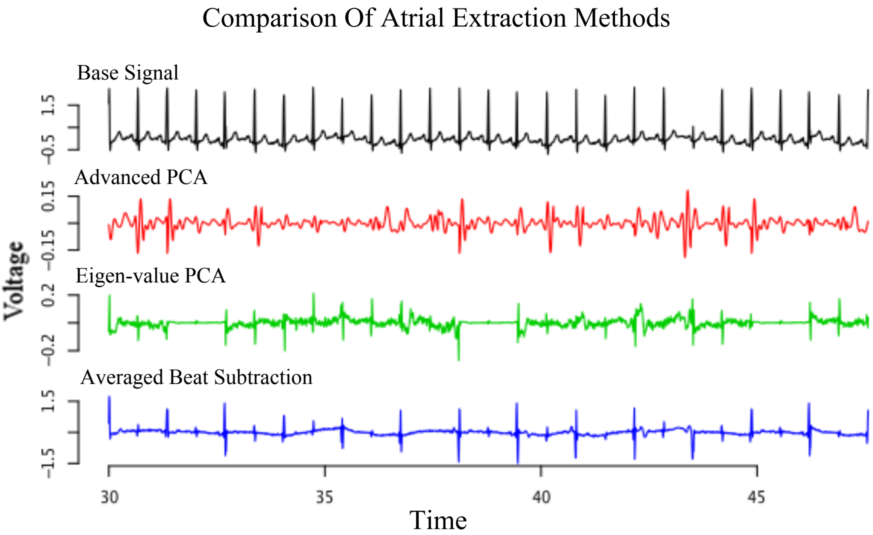
	Method			
Metric	Optimized α	Boosted	Bootstrap	Boosted bootstrap
Mean-squared interpolation error	1.186223	1.185132	1.132029	1.126720
Interpolation time (seconds)	22.12	35.54	546.77	$1,678.77$
Mean-squared prediction error	1.267972	1.251046	1.216298	1.200918
Prediction time (seconds)	6.65	15.48	288.77	586.6

- We found that bootstrapping provided the greatest increase in performance but also had the largest increase in computational costs. Boosting was also a significant improvement although the effect is small.
- Overall, we recommend if only using one method, use bootstrapping if you have the computer power as it provides more information about your reconstruction and is significant improvement.

Chapter 6: Atrial ECG Extraction

- Atrial fibrillation (AF) is a serious cardiac problem. Early detection is a key factor in prolonging the life of people with this disease. The main method for detection is evaluation of Electrocardiogram (ECG) data recorded for a patient. To look for AF we need to extract the atrial signal from the ECG series.
- In some situations we are not able to get the full compliment of 12 leads of ECG data for a patient. Thus, The objective of this method is to extract atrial signals from a single series of ECG data.
- Following previously designed methods for single lead extraction we employ a five-stage principal components based extraction method.
  1. Data cleanup: pre-whitener and band-pass filter
  2. Beat extraction: transform data into a heartbeat matrix
  3. Principal components expansion: transform the heartbeat matrix
  4. Component Detection: (1) clustering for the kurtosis, (2) two-stage clustering for the eigenvalues, (3) bootstrapped F-test
  5. Recombination: remove non-atrial components, inverse the principal components transform, and concatenate

Atrial Extraction Tests - Example Extraction



Atrial Extraction Tests - Results

- We tested two common extraction methods against our APCA method for two metrics, QRS residual peak (false component detection) and F-wave power (missed component detection).

	Advanced PCA	Eigenvalue PCA	Average beat subtraction
Mean residual QRS peak power	.473	1.836	1.391
F-wave peak power	.0377	.0097	.0074
Relative run time	1672.928s	.166s	.047s

- We found our APCA method outperformed both other extraction methods, although it was considerably more computationally costly. If the extraction is being done in an off-line setting we believe this method is a useful tool for practitioners.

Chapter 7: Modeling Major Junior Hockey

Maple Leafs shake up front office, hire **stats guru** *Kyle Dubas, 28, as assistant GM* (Kevin McGran, Toronto Star)

There is no statistic to accurately quantify neutral-zone play. (Steve Simmons, Toronto Star)

Modeling Hockey - Estimated Neutral Zone Differential Formula

- To give an overall assessment of neutral zone play we developed a statistic to take into account both a player's offensive and defensive neutral zone skills. Known as the Estimated Neutral Zone Differential (END), we referenced a player's success rate to the average rate for similar players. For a player X and reference group G, we have

$$END(X, G) = (\hat{p}_O(X) - \hat{p}_O(G)) + (\hat{p}_D(X) - \hat{p}_D(G)). \tag{16}$$

- The group, G, of players referenced against could be other teammates that play the same position (forward or defense) or league wide. This choice is dependent on what you would like to investigate.

Modeling Hockey - Estimated Neutral Zone Differential Analysis

- To examine if the *END* statistic was a good metric of player quality, we examined the relationship between where in the lineup a player would play and their *END* statistic relative to the team.
- We modeled the probability of a player being on the top two lines given the p-value for the hypothesis that the player's *END* being non-positive. The fitted model for this past season's data was $\hat{P}(\mathit{Top}\,2\,\mathit{lines}) = L(.59 - 1.59 \times \text{p-value})$ with p-value on the significance for the coefficient being .0947.
- There appears to be evidence that a positive *END* statistic indicates a higher quality player.

Modeling Hockey - Line Optimization Formulation

- A useful tool for teams is a method to identify strong line combinations. To do this we modeled the probability of a goal being scored for and against in the next minute with players and their lines (2 or 3 term interactions). We performed this for goals for and against separately then evaluated the difference in cumulative pessimistic confidence intervals for the coefficients of our model. That is for a line combination we get the metric,

$$\begin{aligned}\Delta_\alpha = (&\sum_{i=1}^{18}(\beta_{i,gf} + \Phi(\alpha)S(\beta_{i,gf})) + \sum_{j=1}^{3}(\gamma_{j,gf} + \Phi(\alpha)S(\gamma_{j,gf})) \\ &+ \sum_{k=1}^{4}(\zeta_{k,gf} + \Phi(\alpha)S(\zeta_{k,gf}))) - (\sum_{i=1}^{18}(\beta_{i,ga} + \Phi(1-\alpha)S(\beta_{i,ga}) \\ &+ \sum_{j=1}^{3}(\gamma_{j,ga} + \Phi(1-\alpha)S(\gamma_{j,ga})) + \sum_{k=1}^{4}(\zeta_{k,ga} + \Phi(1-\alpha)S(\zeta_{k,ga}))).\end{aligned} \tag{17}$$

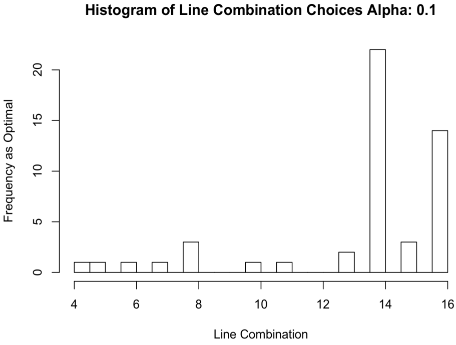
- Where α is how pessimistic we are.

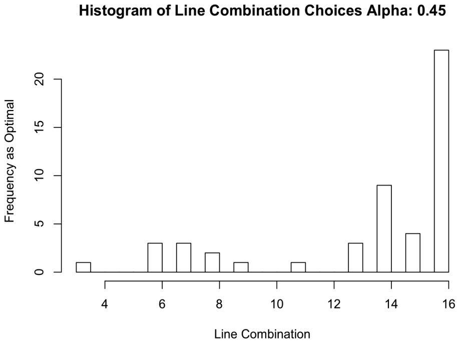
Modeling Hockey - Line Optimization Application

- To apply this method we start by identifying the line combinations we are interested in using. These line combinations can be given by the coaching staff or taken from historical data.
- The selection of the optimal line combination will depend on the confidence intervals we use. The choice of α should be made to meet the needs of the team. Larger values of α will give us lines that have a strong impact when we are less concerned about their variability or lack of evidence.
- Then to avoid issues with unequal use of line combinations in the data, we employed a bagging algorithm. For each sampling from our data we selected the line combination with the largest Δ_α. Then we reported the optimal line combination to the combination selected most often across the samples.

Modeling Hockey - Line Optimization Example

- We wanted to determine the optimal line combination for use in the second round of the playoffs from the data collected in the last 10 games of the season and first round of the playoffs.



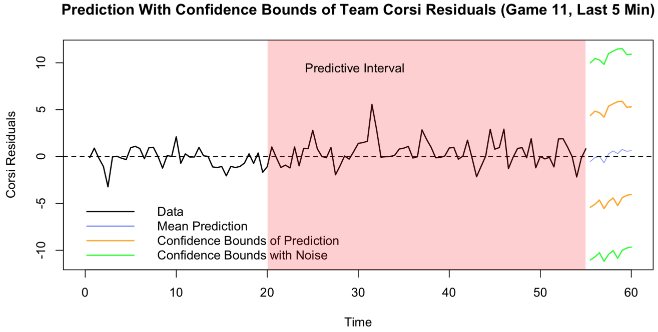


Modeling Hockey - Modeling and Prediction Formulation

- The last method we developed was to model player effects on different aspects of game play and attempt to predict the residual unexplained temporal correlations.
- For this we used LASSO regression to identify the optimal model and then used David Thomson's periodic reconstruction method to predict the upcoming residuals.
- This method can be used to identify upcoming trends in game play, which can be used with the player effects given from the LASSO model to make personnel decisions for upcoming shifts.

Modeling Hockey - Modeling and Prediction Example

- Here we wanted to model the final 5 minutes of puck possession using the data from the 2^{nd} and 3^{rd} periods.



- Care must be taken in selecting the prediction interval. Selecting an interval where stationarity does not hold will cause significant performance issues.

Modeling Hockey - Conclusions on Methods

- All three methods have been shown to have use for improving the decision making of teams in hockey.
- We would like more input variables to accurately model and predict game play.
- We have encouraged the Frontenacs to use these methods in the future and hope that this serves as an example of how statistical modeling can be used for hockey.

Chapter 8: Conclusions

- We showed several examples of ways we can improve on existing methods in spectrum analysis through the use of statistical learning methods.
- A major focus was removing supervised decisions when possible. We were able show in Chapters 3 and 5 several ways we can avoid introducing potential selection bias.
- Another key area was the introduction of methods that have increased robustness towards selection bias. Methods from Chapters 4 and 5 do this.
- We performed two data studies to highlight some of our new methods and improve upon the areas of cardiology and hockey analytics.

Future Works

Some areas where we believe future application of statistical learning and other methods can be used to improve spectrum estimation are:

1. More work on the using the distribution of Bootstrapped F-test values(confidence intervals, etc.).
2. Regularization on the basis expansions in quadratic inverse theory to produce noise-reduced time-frequency estimates.
3. Clustering within the F-test to produce a non-supervised signal detection method.
4. Cross-validation and bootstrapping to produce optimized confidence intervals on multitaper spectrum estimates where sub-sampling is available.
5. More methods for analyzing the residuals from time series and spectrum analysis methods. This could include methods for outlier identification, assumption checking, or model tuning.

Acknowledgments

This research is supported by the Kingston Frontenacs, Queen's University and my supervisors Dr. Takahara and Dr. Thomson.

References

D. Freedman, *Bootstrapping regression models*, The Annals of Statistics **9** (1981), no. 6, 1218–1228.

S. Lahiri, *Resampling methods for dependent data*, vol. 14, Springer, 2003.

D. Montgomery, *Introduction to statistical quality control*, John Wiley & Sons, 2007.

T Purdy, *Shots, fenwick and corsi*, February 2011.

M. Schuckers and J. Curro, *Total hockey rating (thor): A comprehensive statistical rating of national hockey league forwards and defensemen based upon all on-ice events*, Proceedings of the 2013 MIT Sloan Sports Analytics Conference, 2013.

S. Simmons, *Why hockey's trendy advanced stats are a numbers game*, May 2014.

D. Slepian and H.O. Pollack, *Prolate spheroidal wave functions, fourier analysis and uncertainty - I*, Bell System Technical Journal **40** (1961), no. 1, 43–64.

R. Tibshirani T. Hastie and J. Friedman, *The elements of statistical learning*, vol. 1, Springer, 2001.

D.J. Thomson, *Spectrum estimation and harmonic analysis*, Proceedings of the IEEE **70** (1982), no. 09, 1055–1096.

D.J. Thomson, *Quadratic-inverse spectrum estimates: applications to paleoclimatology*, Philosophical Transactions: Physical Sciences and Engineering **332** (1990), no. 1627, 536–597.

R. Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society. Series B (Methodological) **58** (1996), no. 1, 267–288.