ON THE COMPUTER INTENSIVE METHODS IN MODEL SELECTION

D. Robert Iskander

Contact Lens & Visual Optics Lab
School of Optometry
Queensland University of Technology
Kelvin Grove Q4059, Brisbane, Australia
d.iskander@qut.edu.au

Weaam Alkhaldi & Abdelhak M. Zoubir

Signal Processing Group
Institute for Communications
Technische Universität Darmstadt
Merckstr. 25, Darmstadt 64283, Germany
{alkhaldi,zoubir}@spg.tu-darmstadt.de

ABSTRACT

Bootstrap-based model selection has been shown in many practical instances to be superior to classical methods such the AIC and MDL. This is particularly noticeable when the distribution of modelling noise is unknown and/or when the available data samples are small. One of the main problems of using bootstrap model selection with real data is the necessity of tuning the residual scaling parameter or estimating the length of a sub-sample. Recently, we have developed a new *hook and loop* (HL) resampling plane, in which the scaling of the residuals is avoided. Here, we compare the performance of the range of resampling planes that can be used in the context of model selection and show that the HL-based model selection is superior to its predecessors. Moreover, in the context of fitting parametric models to corneal data measured by videokeratoscopes, the HL provides results that are consistent with clinical expectations.

Index Terms— model/order selection, resampling planes, corneal surface modelling

1. INTRODUCTION

Traditional model selection techniques that are widely used in both linear and non-linear models include the Akaike Information Criterion (AIC) [1], Mallows' C_p criterion [6], and Rissanen's Minimum Description Length (MDL) [9]. In recent years, bootstrap-based model order selection techniques have been advocated as they showed superior performance to those achieved with classical methods in a range of simulated applications [12, 13, 14] and real life problems [3, 4, 2].

One of the main problems associated with the use of the bootstrap in model order selection is the necessity of selecting a suitable scaling parameter for the detrended residuals in the resampling procedure [10]. An alternative to scaling is to use subsampling [8]. Here again the difficulty remains in the estimation of subsample length. To avoid this problem, we have recently proposed the "hook and loop" (HL) resampling procedure that avoids the scaling [5]. In this work, we assess the performance of the HL resampling plane in comparison to non-parametric and parametric bootstrap techniques, the jackknife [7], the AIC, and the MDL in the context of selecting the optimal model.

2. THE RESAMPLING PLANES

For clarity of presentation, let us consider the case of finding the optimal model that can be described as linear in parameters (the generalisation to a non-linear model is essentially straightforward). Let the considered linear model be described by

$$X_t = \boldsymbol{h}_t'\boldsymbol{\theta} + W_t, \qquad t = 1, 2, \ldots, n,$$

where $\boldsymbol{\theta}$ is the unknown parameter vector of length p while W_t describes a noise sequence, assumed to be a collection of i.i.d. random variables of unknown distribution $F_W(w)$ with mean zero and variance σ_W^2. Alternatively, we can write our linear model as

$$\boldsymbol{X} = \boldsymbol{h}\boldsymbol{\theta} + \boldsymbol{W}\,.$$

Furthermore, let us denote the *model* β as a subset of $\{1, 2, \ldots, p\}$ that results in the following linear equation

$$\boldsymbol{X} = \boldsymbol{h}_\beta\boldsymbol{\theta}_\beta + \boldsymbol{W}.$$

The goal of model selection is to choose the optimal model β_o such that $\boldsymbol{\theta}_{\beta_o}$ contains all non-zero components of $\boldsymbol{\theta}$ only. A bootstrap-based procedure to achieve this goal has been described earlier [12, 14]. Since most of the steps involved in all resampling-based model selection procedures are common we will describe them only again in general terms and focus on the subtle differences that clearly differentiate them from each other.

1. Given observations $x_1, x_2, \ldots, x_n$, we first calculate the least-squares estimate $\hat{\boldsymbol{\theta}}_\alpha$ and derive the residual

   $$\hat{w}_t = x_t - \boldsymbol{h}_{\alpha t}'\hat{\boldsymbol{\theta}}_\alpha, \quad t = 1, 2, \ldots, n,$$

   where $\alpha = \{1, 2, \ldots, p\}$ is the full model and $\boldsymbol{h}_{\alpha t}'$ is the tth row of $\boldsymbol{h}_\alpha$.

2. Next, we use either the non-parametric bootstrap, the parametric bootstrap, the jackknife, or the HL to calculate a set of pseudo-new residuals:

   NP-BOOT: In the non-parametric bootstrap procedure, we resample with replacement from

   $$\sqrt{n/m}(\hat{w}_t - \hat{w}.)/\sqrt{1-p/n}, \qquad (1)$$

   $n = 1, 2, \ldots, n$ to obtain $\hat{w}_t^*$. Here a scaling parameter m is introduced such that $m/n \to 0$ and

   $$\frac{n}{m} \max_{t \le n} \boldsymbol{h}'_{\beta t}(\boldsymbol{h}'_{\beta}\boldsymbol{h}_{\beta})^{-1}\boldsymbol{h}_{\beta t} \to 0$$

   for all β [10].

   P-BOOT: In the parametric bootstrap procedure, we assume a certain distribution of the residuals and use a pseudo-random number generator with parameters estimated from the residuals defined in Equation (1).

   JACK: In the jackknife, we use the detrended residuals

   $$\sqrt{n}(\hat{w}_t - \hat{w}.)/\sqrt{1-p/n}$$

   and form subsamples $\hat{w}^{(i)}$, in which the the i-th sample is omitted. This pseudo-new residuals are then additionally ordered according to the strength of the signal x_t as in the HL resampling plane. Note that we do not include here the scaling parameter m.

   HL: In the hook and loop procedure, we sort the detrended residuals

   $$\hat{w}_1 - \hat{w}., \hat{w}_2 - \hat{w}., \ldots, \hat{w}_n - \hat{w}.$$

   in an increasing order to obtain a set of residuals $\hat{w}_{(1)}, \hat{w}_{(2)}, \ldots, \hat{w}_{(n)}$ and generate a new HL sample using, for example,

   $$\hat{w}^*_{(i)} = \frac{1}{2}\left(w_{(i)} + w_{(i+1)}\right) + \varepsilon_i$$

   where

   $$\varepsilon_i \sim \mathcal{N}\left(0, \left[\frac{1}{6}\left(w_{(i+1)} - w_{(i)}\right)\right]^2\right).$$

   The HL residuals are then further ordered according to the strength of the signal x_t [5].

3. In the next step, we compute

   $$x_t^* = \boldsymbol{h}'_{\beta t}\hat{\boldsymbol{\theta}}_\beta + \hat{v}_t^*, \quad t = 1, 2, \ldots, n$$

   where $\hat{v}_t$, $t = 1, 2, \ldots, n$ denote either the bootstrap, the HL, or the weighted bootstrap residuals, and the least-squares estimate $\hat{\boldsymbol{\theta}}^*_{\beta,m}$ from $(x_t^*, \boldsymbol{h}_{\beta t})$.

4. Steps 2 and 3 are then repeated B times (or $B = n$ for the jackknife) to obtain $\hat{\boldsymbol{\theta}}^{*(i)}_{\beta,m}$ and the bootstrap estimate of the residual squred error

   $$\hat{\Gamma}^{*(i)}_{n,m}(\beta) = \frac{\|\boldsymbol{x} - \boldsymbol{h}_\beta \hat{\boldsymbol{\theta}}^{*(i)}_{\beta,m}\|^2}{n}, \quad i = 1, \ldots, B.$$

5. Finally, we average $\hat{\Gamma}^{*(i)}_{n,m}(\beta)$ over $i = 1, \ldots, B$ to obtain $\bar{\Gamma}^*_{n,m}$ and minimise over β to obtain $\hat{\beta}_0$.

The differences between the four resampling planes described above consist mainly of two aspects:

- Scaling the residuals for the resampling [NP-Boot] or for estimating the distributional parameters [P-BOOT].
- Ordering (sorting) the residuals according to the strength of the signal [JACK, HL].

The idea of sorting the pseudo-new residuals has been introduced in the HL resampling plane [5] to associate small amplitudes of the residuals with small values of the signal and large values of the residuals with larger values of the signal. In this work, we also discovered that such sorting is fundamental if a jakknife procedure is to be used for model selection.

3. SIMULATION RESULTS

The performances of the procedures described in section 2 were tested against the classical methods of model selection, namely the AIC, and the MDL. For bench-marking purposes, we considered a simple example of trend estimation from [14], in which the parameter vector $\boldsymbol{\theta} = (0, 0, 0.035, -0.0005)'$ with $n = 64$, $n = 32$, and $n = 16$. In our earlier works [14, 5] the number of bootstrap repetitions was set to $B = 100$. However, to make the comparative analysis possible with the inclusion of the jakknife procedure, we chose in this particular simulation $B = n$. In Table 1, we show the results of model selection for this particular trend estimation problem using the considered resampling planes as well as those achieved by the AIC, and the MDL for $n = 64$, $n = 32$, and $n = 16$ in the case the noise model is standard Gaussian and in the case where it is t_3-distributed. The results are based on 1000 independent Monte Carlo runs.

The results clearly indicate that, in the context of model selection, the HL procedure is superior to all the other considered resampling planes. The case of the small sample length (i.e., $n = 16$) is of particular significance as all the other methods simply fail to find the correct model order while the performance of the HL resampling plane is still reasonable. Further simulation analyses excluding the jackknife routine showed that the increase in the number of resampling iterations, B, in the non-parametric and parametric bootstrap procedures does significantly improve their performances.

Table 1. The percentages of selecting the correct model evaluated over 1000 independent Monte Carlo runs for the example of trend estimation for the noise sequence modelled as $\mathcal{N}(0,1)$ (upper part of table) and as t_3 (lower part). In the bootstrap, the scaling parameter m was set to 2.

n	NP-BOOT	P-BOOT	JACK	HL	AIC	MDL
64	98.5	98.2	97.1	**99.9**	89.4	97.6
32	63.2	61	56.2	**94.7**	78.3	85.1
16	2.5	3.3	1.0	**28.8**	1.9	1.1
64	98.2	56.3	93.9	**99.2**	90.8	97.9
32	31.1	38.9	24.4	**83.8**	63.4	64.4
16	3.6	6.7	0.8	**20.9**	1.9	0.7

Table 2. The optimal order of the Zernike polynomial expansion selected by the bootstrap and the HL methods for several types of corneal surfaces and a range of corneal diameters. (1) normal healthy cornea, (2) astigmatic health cornea, (3) early keratoconic cornea, (4) post radial keratotomy cornea.

Cornea Type	Corneal Diameter [mm] 4		6		8	
	B	HL	B	HL	B	HL
(1)	6	20	11	14	11	11
(2)	6	12	11	21	11	21
(3)	12	13	12	28	14	21
(4)	14	19	14	21	12	14

4. MODELLING OF THE CORNEAL TOPOGRAPHY

The topography of the corneal surface is normally measured with videokeratoscopes. The 3D point-cloud in cylindrical coordinates (ρ_d, θ_d, S_d), $d = 1, 2, \ldots, D$ can be modelled by a finite series of Zernike polynomials [3]

$$S(\rho,\theta) = \sum_{p=1}^{P} a_p Z_p(\rho,\theta) + \varepsilon(\rho,\theta) \tag{2}$$

with $Z_p(\rho,\theta)$ being the single indexed p-th Zernike polynomial defined as

$$Z_p(\rho,\theta) = \begin{cases} \sqrt{2(n+1)} R_n^m(\rho)\cos(m\theta), & \text{even } p, m \neq 0 \\ \sqrt{2(n+1)} R_n^m(\rho)\sin(m\theta), & \text{odd } p, m \neq 0 \\ \sqrt{n+1} R_n^0(\rho), & m = 0 \end{cases}$$

where n is the radial degree, m is the azimuthal frequency, and

$$R_n^m(\rho) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s (n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!}\rho^{n-2s}.$$

In Equation (2), $\varepsilon(\rho,\theta)$ represents the measurement and modelling error. When D discrete samples of the surface $S(\rho_d, \theta_d)$ are available, the equation above can be easily written in a linear form as

$$\boldsymbol{S} = \boldsymbol{Z}\boldsymbol{a} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{S}$ is a D-element column vector, $\boldsymbol{Z}$ is a $D \times P$ matrix of discrete, orthogonalized Zernike polynomials, $\boldsymbol{a}$ is a P-element column vector of Zernike coefficients, and $\boldsymbol{\varepsilon}$ is a D-element column vector of measurement and modeling error. In such modeling, a fundamental problem arises on the number of Zernike terms to be used.

In essence, corneas from healthy normal eyes should be well represented with Zernike polynomial expansions up to the 4th radial order (up to 15 coefficients). For deformed corneas, however, such as encountered in keratoconus or for the corneas that have been surgically altered, a different number of terms is expected. One may suspect that a deformed cornea necessarily needs to be modelled by a larger number of Zernike terms. However, the larger the deformity, the larger the measurement error is encountered [11]. The number of Zernike terms fitted to the corneal topography data is also corneal diameter-depended [3].

One popular, but not rigorous, way of selecting the number of Zernike terms is to minimise the residual variance and determine a suitable cut-off threshold value. In the past, however, this approach led to over-parameterization, where sometimes unrealistically large numbers of Zernike terms (hundreds) were used. Alternatively, a suitable penalty function could be used that increases with the number of parameters to form a model order selection criterion. In videokeratoscopy, where the measurement errors are several orders smaller than the effective signals (i.e. few microns vs. few millimetres) [11], the use of classical model selection criteria such as the AIC and the MDL results in an unrealistic situation where hundreds or even thousands of Zernike polynomial terms need to be fitted.

This problem has been resolved in [3] with a bootstrap based model order selection procedure, although it was later found that it sometimes underestimates the clinically expected model order. This bootstrap procedure has been recently improved by incorporating the knowledge of the spatially nonuniformity of the measurement noise in the resampling procedure [4]. This was achieved by performing resampling in semi-rings of data, where the noise distribution can be assumed constant. However, this procedure is numerically very complex and hence, has a very limited practical applicability.

In Table 2, we show the results of estimating the optimal order of the Zernike polynomial expansion fitted to the range of corneal topography data. Four topographically different corneas were chosen: (1) a normal healthy cornea of an emmetrope, (2) a normal healthy cornea with astigmatism (approximately -2.0 Diopters), (3) the cornea of a subject with early keratoconus, and (4) the cornea of a subject that have undergone a refractive surgery (radial keratotomy). Videokeratocopic discrete data consisting of 256 semi-meridians and 26 rings resulting in $D = 6656$ were used.

5. CONCLUSIONS

We have performed a comparative analysis of four resampling planes in the context of model selection. We showed that a model selection technique based on the recently proposed hook and loop resampling scheme clearly outperforms the traditional model selection techniques such as the AIC and the MDL as well as the other computer intensive methods based on the parametric and non-parametric bootstrap and the jackknife. The superior performance of the HL-based method was evident in both cases where the noise is Gaussian or non-Gaussian distributed. It is remarkable that the HL procedure is able to provide a reasonable result for samples as little as 16 data points and only 16 repetitions.

The HL-based selection of the optimal order of Zernike polynomial expansion to corneal topography data measured with videokeratoscopes has been compared to the traditional bootstrap technique [3]. Unlike the traditional method, which showed deficiencies in terms of underestimating the Zernike model order, the HL-based routine provides results that are consistent with clinical expectations. At the same time, the HL routine is computationally less complex than the block-based resampling technique reported in [4].

6. REFERENCES

[1] Akaike, H. Statistical predictor identification. *Ann. Inst. Statist. Math.*, 22: 203–217, 1970.

[2] Chong-Yaw, W., Paramesran, R., and Takeda, F. Bootstrap model order selection of Zernike polynomial expansion for classification of rice. In *Proceedings of TENCON'04*, vol. 1, pages 203–206, 2004.

[3] Iskander, D. R., Collins, M. J., and Davis, B. Optimal modeling of corneal surfaces with Zernike polynomials. *IEEE Trans. Biomed. Eng.*, 48(1): 87–95, 2001.

[4] Iskander, D. R., Morelande, M. R., Collins, M. J., and Buehren, T. A refined bootstrap method for estimating the Zernike polynomial model order for corneal surfaces. *IEEE Trans. Biomed. Eng.*, 51(12): 2203–2206, 2004.

[5] Iskander, D. R. and Alkhaldi, W. The ”hook and loop” resampling plane. In Proceedings of CAMSAP'07, St. Thomas, U.S. Virgin Islands, December 2007 (in press).

[6] Mallows, C. Some comments on C_p. *Technometrics*, 15: 661–675, 1973.

[7] Miller, R. G. The jackknife - a review. *Biometrika*, 61:1–15, 1974

[8] Politis, D. N., Romano, J. P., and Wolf, M. *Subsampling*, Springer, 1999.

[9] Rissanen, J. A universal prior for integers and estimating by minimum description length. *The Ann. Statist.*, 11: 416–431, 1983.

[10] Shao, J. and Tu, D. *The jackknife and bootstrap*. Springer Verlag, New York, 1995.

[11] Tang, W., Collins, M. J., Carney, L., and Davis, B. The accuracy and precision performance of four videokeratoscopes in measuring test surfaces. *Opt. Vis. Sci.*, 77: 483–491, 2000.

[12] Zoubir, A. M. Model selection: a bootstrap approach. In *Proceedings of ICASSP'99*, pages 1377–1380, 1999.

[13] Zoubir, A. M. and Iskander, D. R. Bootstrap modeling of a class of nonstationary signals. *IEEE Trans. Sig. Process*, 48: 399–408, 2000.

[14] Zoubir, A. M. and Iskander, D. R. *Bootstrap techniques for signal processing*. Cambridge University Press, Cambridge, 2004.