Iskander, D. Robert and Alkhaldi, Weaam (2007) The "Hook and Loop" Resampling Plane . In *Proceedings 2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSAP 2007.*, pages pp. 65-68, St. Thomas, Virgin Islands.

# THE "HOOK AND LOOP" RESAMPLING PLANE

*D. Robert Iskander*

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

*Weaam Alkhaldi*

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

**ABSTRACT**

We propose a new resampling scheme that takes literally the concept of the non-parametric bootstrap in which new samples are generated from the empirical distribution function. The introduced resampling concept is totally heuristic, but already shows promising results when applied to model selection. We show that for a range of linear models, the proposed resampling scheme outperforms the classical model selection techniques as well as its predecessor, the non-parametric bootstrap. It also simplifies the practical problem of choosing residual scaling or the length of the subsample that exists in the traditional bootstrap based model selection approach.

## 1. INTRODUCTION

The main motivation behind this work has been the problem in finding an appropriate scaling of the residuals in a practical setting of a bootstrap based model selection. To make this point clearer, let us briefly review the problem of linear model selection while for the general case and for more detail the reader is referred to [10, 11].

Consider the following linear model

$$Y_t = \boldsymbol{x}_t'\boldsymbol{\theta} + Z_t, \qquad t = 1, \ldots, n,$$

where $\boldsymbol{\theta}$ is an unknown $p$ vector-valued parameter and $Z_t$ is a noise sequence that is assumed to be a collection of i.i.d. random variables of unknown distribution $F_Z(z)$ with mean zero and variance $\sigma_Z^2$. In a vector form, the above equation can be written as

$$\boldsymbol{Y} = \boldsymbol{x}\boldsymbol{\theta} + \boldsymbol{Z}\,.$$

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

$$\boldsymbol{Y} = \boldsymbol{x}_\beta\boldsymbol{\theta}_\beta + \boldsymbol{Z}.$$

The problem of model selection is to choose the optimal model $\beta_o$ such that $\boldsymbol{\theta}_{\beta_o}$ contains all non-zero components of $\boldsymbol{\theta}$ only. A bootstrap based procedure to achieve this task usually consists of the following steps:

1. Given observations $y_1, \ldots, y_n$, we calculate the least-squares estimate $\hat{\boldsymbol{\theta}}_\alpha$ and derive the residual
$$\hat{z}_t = y_t - \boldsymbol{x}_{\alpha t}'\hat{\boldsymbol{\theta}}_\alpha, \quad t = 1, \ldots, n,$$
where $\alpha = \{1, \ldots, p\}$ is the full model and $\boldsymbol{x}_{\alpha t}'$ is the $t$th row of $\boldsymbol{x}_\alpha$.
2. Next, we resample with replacement from $\sqrt{n/m}(\hat{z}_t - \hat{z}_\cdot)/\sqrt{1-p/n}$ to obtain $\hat{z}_t^*$. Here a scaling parameter $m$ is introduced such that $m/n \to 0$ and
$$\frac{n}{m} \max_{t \le n} \boldsymbol{x}_{\beta t}'(\boldsymbol{x}_\beta'\boldsymbol{x}_\beta)^{-1}\boldsymbol{x}_{\beta t} \to 0$$
for all $\beta$ [9].
3. In the next step we compute
$$y_t^* = \boldsymbol{x}_{\beta t}'\hat{\boldsymbol{\theta}}_\beta + \hat{z}_t^*, \quad t = 1, \ldots, n$$
and the least-squares estimate $\hat{\boldsymbol{\theta}}_{\beta,m}^*$ from $(y_t^*, \boldsymbol{x}_{\beta t})$.
4. The steps 2 and 3 are then repeated $B$ times to obtain $\hat{\boldsymbol{\theta}}_{\beta,m}^{*(i)}$ and the bootstrap estimate of the residual squred error [2]
$$\hat{\Gamma}_{n,m}^{*(i)}(\beta) = \frac{\|\boldsymbol{y} - \boldsymbol{x}_\beta\hat{\boldsymbol{\theta}}_{\beta,m}^{*(i)}\|^2}{n}, \quad i = 1, \ldots, B\,.$$
5. Finally, we average $\hat{\Gamma}_{n,m}^{*(i)}(\beta)$ over $i = 1, \ldots, B$ to obtain $\bar{\Gamma}_{n,m}^*$ and minimise over $\beta$ to obtain $\hat{\beta}_0$.

In practical cases of model selection, the choice of the scaling parameter $m$ is not always obvious. In some cases this choice may become critical if one aims to show the superiority of the bootstrap methods to other classical model selection techniques [4]. A popular alternative to scaling

978-1-4244-1714-8/07/$25.00 

is based on subsampling [7] in which rather than using full data record of $n$ samples, subsamples of $m$ are used to achieve the consistency in the estimation of the residual squared error $\bar{\Gamma}^*_{n,m}$. Nevertheless, the choice of the subsample length $m$ may also be a problem in particular when $n$ is small.

In this work we introduce a new heuristic resampling scheme that shows promising results in the application of bootstrap based model selection while at the same time alleviates the problems of scaling residuals.

## 2. THE "HOOK AND LOOP" RESAMPLING SCHEME

In a classical non-parametric bootstrap scheme the data are resampled from the empirical distribution $\hat{F}_Z(z)$ with replacement to form a pseudo sample. In a parametric bootstrap, on the other hand, a particular form of the cumulative distribution function is chosen on the basis of its parameter estimates that are derived from the original sample. Subsequently, new sets of data are generated from this distribution using a pseudo-random number generator. In the proposed "hook and loop" (HL) resampling scheme, we suggest the concept of the non-parametric bootstrap of sampling from the empirical distribution function is taken rather literally. Specifically, given a collection of $n$ observations $z_1, z_2, \ldots, z_n$ of a random sample with unspecified distribution $F_Z(z)$ we first estimate the empirical cumulative distribution functions (ECDF)

$$\hat{F}_n(z) = \frac{1}{n}\sum_{i=1}^{n} I(z_i \leq z)$$

where $I(\cdot)$ is the indicator function. The ECDF is then used to generate a new set of samples of size $n-1$ in which each of the new samples lies between each of the two consecutive original samples that form the ECDF. Figuratively speaking, the original and the new sample form a kind of a zipper or a "hook and loop" fastener. Note also that similar to the classical case of jackknife [6], the new resample is of smaller size than that of the original sample.

The HL resamples can be generated in a following two steps:

1. Given the original sample $z_1, z_2, \ldots, z_n$, sort the data in an increasing order to obtain $z_{(1)}, z_{(2)}, \ldots, z_{(n)}$.

2. Generate a new HL sample using, for example,

$$z^*_{(i)} = \frac{1}{2}\left(z_{(i)} + z_{(i+1)}\right) + \varepsilon_i$$

where

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

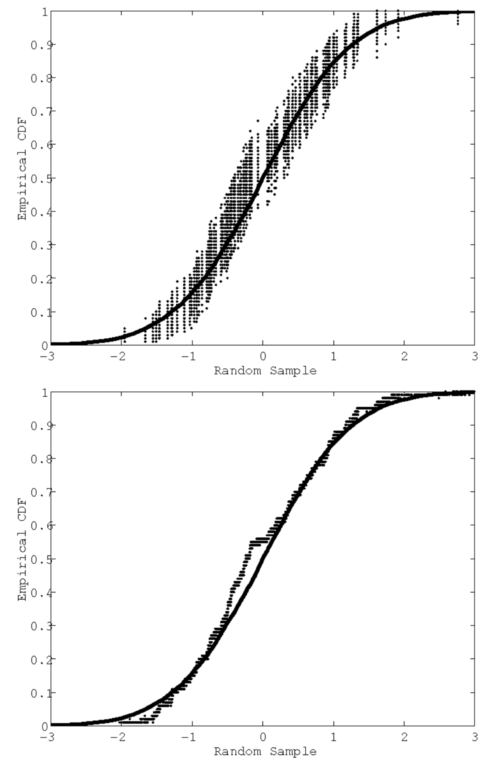


**Fig. 1**. The estimated ECDFs for a sample of length $n = 100$ from a standard Gaussian random variable using 100 bootstrap resamples (top) and 100 "hook and loop" resamples (bottom). The thick solid lines show the theoretical CDF.

To see the differences between bootstrap samples and the HL samples consider the following simple simulation

```
>> x=randn(5,1);
>> x_b=bootrsp(x);
>> x_hp=hprsp(x)
>> [x x_b [x_hp;NaN]]
ans =
   -0.3851    0.9915   -0.0588
    0.2574    0.2574    0.2442
    0.7475    0.2310    0.4301
    0.2310    0.7475    0.8805
    0.9915    0.9915       NaN
```

and the estimated ECDFs for a sample of length $n = 100$ from a standard Gaussian random variable and 100 resamples as shown in Figure 1.

The estimated bootstrap ECDFs follow closely a result that would be achieved with a Monte Carlo simulation. The HP resampling scheme, on the other hand, follows closely the ECDF estimated from the original sample. This brings the requirement of the original sample being representative to a stricter level than that in the case of bootstrap resampling [12]. At the same time, however, the HP resampling scheme results in the ECDF variance that is less dependent on the length and the variance of the original sample, a characteristic that is useful in model selection.

## 3. EXAMPLE OF TREND ESTIMATION

Consider first a simple example of trend estimation from [11], in which the parameter vector $\boldsymbol{\theta} = (0, 0, 0.035, -0.0005)'$ and $n = 64$. For the scaling parameter $m = 2$ it was shown that the bootstrap based approach outperforms the classical model selection techniques such as the Akaike information criterion (AIC) [1] and the one based on Rissanen's minimum distance length (MDL) [8]. However, for a different sample length, one would need to tune the parameter $m$ to achieve equally impressive result. Often, for a smaller number of samples this task is not possible.

The HL resampling routine described in Section 2 is used to generate new residuals $\hat{z}_t^*$, $t = 1, 2, \ldots, n$ that are used in the model selection in a similar way to those used in the bootstrap model selection as described in Section 1. The only difference is that the HL residuals are additionally ordered according to the strength of the signal $y_t$ resembling the weighted bootstrap [3].

In Table 1, we show the results of model selection for this particular trend estimation problem using the bootstrap, the HL, AIC, and 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, and the percentages of selecting the correct model are rounded.

It is clear that the proposed HL resampling procedure achieves much better results than the classical model selection techniques as well as its bootstrap predecessor. Note also that in the case of the HL approach we did not need to scale the residuals. There are two additional steps that we had to employ in this new resampling scheme. First, we had to make the HL resample length equal to that of the original sample. This has been achieved by simply adding randomly one point from the original residual sample. Second, we had to re-order it according to the strength of the original noisy signal. This second step is performed to match the strength of the HL residual that is sorted to that of the signal.

## 4. MODELLING OPHTHALMIC SURFACES

An ophthalmic surface such as the cornea or a wavefront error can be modelled by a finite series of Zernike polynomials [4]

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

**Table 1**. The percentages of selecting the correct model 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.

| $n$ | Boot. | HL | AIC | MDL |
|---|---|---|---|---|
| 64 | 100 | 100 | 91 | 98 |
| 32 | 73 | 94 | 79 | 84 |
| 16 | 1 | 31 | 3 | 2 |
| 64 | 100 | 97 | 89 | 97 |
| 32 | 27 | 81 | 61 | 61 |
| 16 | 0 | 19 | 2 | 1 |

where $Z_p(\rho, \theta)$ is the single indexed $p$-th Zernike polynomial and $\varepsilon$ 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 of how many Zernike terms one should use. Traditionally, vision researchers have chosen to use the first 15 or 32 Zernike terms although in some recent reports hundreds of terms have been used.

A simple, although not rigorous, way of selecting the number of Zernike terms is to minimize the residual variance and determine a suitable cut-off threshold value. This, however, often leads to over-parameterization. An alternative approach is to use a suitable penalty function that increases with the number of parameters to form a model order selection criterion. In ophthalmic application of corneal modelling, the use of classical model selection criteria such as AIC and the MDL results in an unrealistic situation where hundreds or even thousands of Zernike polynomial terms can be fitted to the corneal elevation data. This problem has been resolved in [4] with a bootstrap based model order selection procedure, although it was later found that it often underestimated the clinically expected model order.

The bootstrap routine form [4] has been recently improved by incorporating the knowledge of the spatially non-uniformity of the measurement noise in the resampling procedure [5]. This was achieved by performing resampling in semi-rings of data where the noise distribution can be assumed constant. However, this procedure is even more numerically complex than its predecessor and has limited practical applicability. We were interested whether the HL resampling procedure used for selecting the optimal model

**Table 2**. The optimal order of the Zernike polynomial expansion selected by the bootstrap and the HL (number in brackets) methods for several types of corneal surfaces and a range of corneal diameters. For a keratoconic subject the maximum available corneal diameter was 7.1 mm, hence no data in the last column of the table.

| Cornea | Corneal Diameter [mm] | | | | |
|---|---|---|---|---|---|
| Type | 4 | 5 | 6 | 7 | 8 |
| Normal | 6(20) | 11(15) | 11(14) | 11(16) | 11(11) |
| Astigmat. | 11(15) | 11(11) | 11(21) | 11(18) | 11(12) |
| Keratoc. | 19(27) | 19(28) | 19(28) | 19(21) | – |
| RK | 14(19) | 14(21) | 14(21) | 12(14) | 11(14) |

order would result in a number of coefficients that could be clinically justified.

In Table 2, we show the results of optimal model order selection for four different corneal surfaces. Corneal elevation data was measured with a Keratron Optikon videokeratoscope. We chose two corneas of healthy subjects: one normal and another with significant amount of corneal astigmatism; and two pathological cases: a keratoconic cornea and one that undergone an unsuccessful refractive surgery known as radial keratotomy (RK). We note that the model order selection based on HL resampling scheme overcomes the underestimations problem of the traditional bootstrap based model order selection technique [4] and provides similar results to that reported in [5].

## 5. CONCLUSIONS

We proposed a new "hook and loop" (HL) resampling scheme based on non-parametric generation of samples from the estimated empirical distribution function. In a simulation study, we showed that this currently heuristic technique could outperform the traditional bootstrap based model selection procedures in cases where the noise data are Gaussian or non-Gaussian distributed. A theoretical study is currently underway to identify the potentials and limitations of this technique. Since it is based on resampling from ordered data it may find its potential in other application of order statistics such as, for example, in the estimation of median for which the bootstrap methods need refinements [2].

In a practical setting of selecting an appropriate Zernike expansion model for corneal surface, the estimated optimal model orders using the traditional bootstrap technique [4] are too conservative and have been disputed in the vision community. This study shows that the HL resampling based selection of the optimal model leads to results very similar to those obtained with the refined bootstrap method presented in [5] but with a computational cost similar to that of the standard bootstrap procedure.

In this view, we predict that the proposed new resampling plane will be useful in applications in which the noise data shows heteroskedasticity that is somehow correlated with the signal amplitude.

## 6. ACKNOWLEDGMENT

D. Robert Iskander thanks Jonathon Ralston for the humorous chat on bootstraps, shoe-strings and Velcro which inspired the developments of the "hook and loop" resampling procedure. Thanks also go to Abdelhak M. Zoubir for his continuing encouragement and support.

## 7. REFERENCES

[1] Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control*, 19: 716–723, 1974.

[2] Efron, B. and Tibshirani, R. *An Introduction to the Bootstrap*. Chapman and Hall, New York, 1993.

[3] Guillou, A. Weighted bootstraps for the variance. *J. Stat. Plan. Infer.*, 81:113–120, 1999.

[4] 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.

[5] 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.

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

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

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

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

[10] Zoubir, A. M. Model selection: a bootstrap approach. In *Proceedings of ICASSP*, pp. 1377-1380, 1999.

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

[12] Zoubir, A. M. and Iskander, D. R. Bootstrap methods and applications: a tutorial for the signal processing practitioner. *IEEE Sig. Proc. Mag.*, 24(4), July 2007 (in press).