Supplement to "Calculating Confidence Intervals for Prediction Error in Microarray Classification Using Resampling"

Wenyu Jiang[1§]*, Sudhir Varma[1]* and Richard Simon[1]*

[1]Biometric Research Branch, National Cancer Institute, Bethesda MD, USA

*These authors contributed equally to this work
[§]Corresponding author

Supplement 1. Description of "Bias Corrected Accelerated Interval Using Bootstrap Case Cross-Validation (BCCV-BCa)"

With this approach, the $100(1-\alpha)\%$ level upper confidence interval becomes $(0, \hat{\theta}^*_{(1-\alpha_1)}]$ using the $1-\alpha_1$ empirical quantile of cross-validation estimate $\hat{\theta}^{*,b}$, b=1, …, B, calculated through the bootstrap case cross-validation (BCCV). Specifically, we calculate the bias correction term as $\hat{z}_0 = \Phi^{-1}$(Proportion of $\{\hat{\theta}^{*,b} \le \hat{\theta}^{LOOCV}\}$ among $b = 1,...B$) and the acceleration term as $\hat{a} = \sum_{i=1}^{n}(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(-i)}^{LOOCV})^3 \Big/ 6\{\sum_{i=1}^{n}(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(-i)}^{LOOCV})^2\}^{3/2}$ where $\hat{\theta}^{LOOCV}$ is the leave-one-out cross-validation estimate on the observed sample, $\hat{\theta}_{(-i)}^{LOOCV}$ is the leave-one-out cross-validation estimate on the sample with observation *i* removed and $\hat{\theta}_{(\cdot)} = \sum_{i=1}^{n}\hat{\theta}_{(-i)}^{LOOCV} / n$. The adjustment $1-\alpha_1 = \Phi[\hat{z}_0 + (\hat{z}_0 + z_{(1-\alpha)}) \big/ \{1 - \hat{a}(\hat{z}_0 + z_{(1-\alpha)})\}]$ follows directly from the BCa algorithm of Efron and Tibshirani (1998) where Φ, Φ^{-1} and $z_{(1-\alpha)}$ are the cumulative distribution function (cdf), inverse cdf and the $1-\alpha$ quantile of the standard normal distribution.

In traditional framework, the BCa algorithm makes use of an acceleration term exploiting the third moment of the estimate. But it is unclear in the literature how to approach the third moment of a prediction error estimate, although the conventional formulae of Efron and Tibshirani (1998) can be applied verbatim.

Supplement 2. Results on Simulations 5 to 8

Table S1. Description of Simulations 5 to 8. Number of simulation replications is 1000. The quantity $\tilde{\theta}_n$ is the "true" prediction error for each sample evaluated on 1000 independent test data. Average of $\tilde{\theta}_n$ is calculated across all simulation replications.

Simulation	n	p	Classifier	% differential genes	Average of $\tilde{\theta}_n$
5	40	1000	DLDA	10%	.143
6	100	1000	DLDA	2%	.183
7	40	3000	DLDA	2%	.162
8	40	1000	SVM	2%	.309

DLDA: Diagonal Linear Discriminant Analysis
SVM: Supported Vector Machines

Table S2. Upper Confidence Intervals for Simulations 5-8. Methods for confidence procedures include the bootstrap case cross-validation percentile interval (BCCVP), BCCVP with bias reduction (BCCVP-BR), bias corrected accelerated interval using BCCV (BCCV-BCa) , binomial interval based on leave-one-out cross-validation (LOOCV-Bin), binomial interval based on split-sample (Split-Bin) with 1/3 sample in the test set. Number of bootstrap repetitions is 100 in BCCVP, BCCVP-BR, BCCV-BCa.

	Simulation 5		Simulation 6		Simulation 7		Simulation 8	
Nominal levels	80%	90%	80%	90%	80%	90%	80%	90%
BCCVP								
Coverage Probability	.993	1	.974	.993	1	1	0.998	1
Average Confidence Limit	.306	.380	.289	.334	0.469	0.559	0.545	0.626
SD of Confidence Limit	.070	.080	.060	.069	0.075	0.077	0.082	0.084
BCCVP-BR								
Coverage Probability	.877	.981	.857	.964	0.942	0.998	0.889	0.976
Average Confidence Limit	.247	.321	.243	.288	0.315	0.405	0.474	0.554
SD of Confidence Limit	.094	.100	.056	.062	0.111	0.111	0.147	0.148
BCCV-BCa								
Coverage Probability	.712	.827	.662	.799	0.515	0.653	0.685	0.802
Average Confidence Limit	.252	.309	.223	.259	0.210	0.272	0.434	0.504
SD of Confidence Limit	.164	.172	.080	.088	0.178	0.197	0.227	0.225
LOOCV-Bin								
Coverage Probability	.788	.883	.784	.888	0.744	0.830	0.746	0.813
Average Confidence Limit	.219	.249	.224	.242	0.237	0.266	0.400	0.433
SD of Confidence Limit	.094	.097	.049	.051	0.107	0.111	0.143	0.143
Split-Bin (1/3 in test set)								
Coverage Probability	.928	.981	.913	.959	0.969	0.997	0.942	0.969
Average Confidence Limit	.310	.365	.284	.319	0.415	0.478	0.531	0.581
SD of Confidence Limit	.114	.117	.080	.082	0.139	0.147	0.143	0.145

SD: Standard Deviation

Figures S1 and S2 compare the BCCVP-BR method to the binomial intervals based on split-sample (Split-Bin) and the multiple random validation percentile intervals (MRVP) under different learning-and-test-set allocations; only results for Simulations 5 and 6 are displayed.

Figure Legends

Figure S1. Comparison of coverage properties for the binomial intervals based on split-sample (Split-Bin) and the multiple random validation percentile intervals (MRVP) with 2/3, 1/3 and 1/10 samples in the test sets. Empirical coverage probabilities are plotted against nominal confidence levels of 70%, 80%, 90% and 95%. Also displayed are results using the bootstrap case cross-validation percentile interval with bias reduction (BCCVP-BR) and the 45° yellow line for reference.

Figure S2. Comparison of chances to reach conclusive confidence intervals using the binomial intervals based on split-sample (Split-Bin) and the multiple random validation percentile intervals (MRVP) with 2/3, 1/3 and 1/10 samples in the test sets. Proportions of simulated upper confidence intervals falling below 0.5 are plotted against nominal confidence levels of 70%, 80%, 90% and 95%. Also displayed are results using the bootstrap case cross-validation percentile interval with bias reduction (BCCVP-BR).

Figure S1

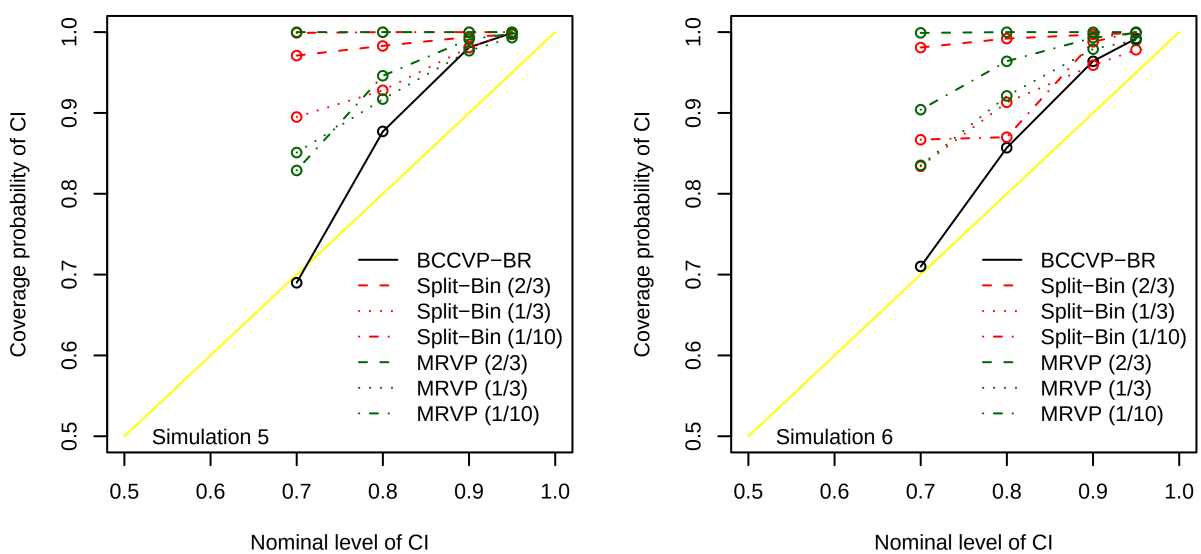
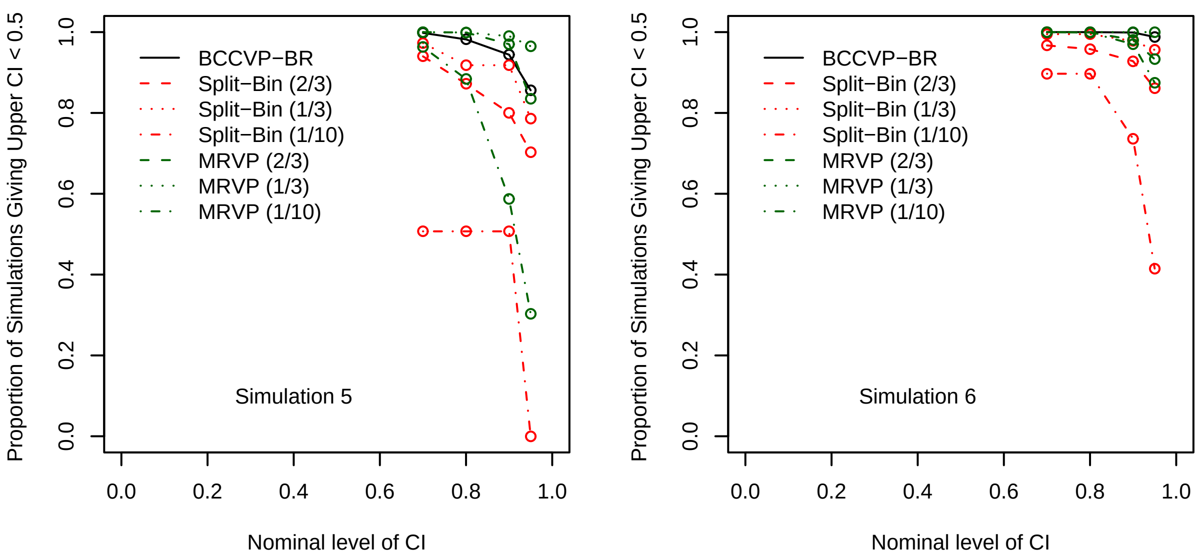

Coverage probability of CI
Nominal level of CI
Simulation 5
Simulation 6
BCCVP−BR
Split−Bin (2/3)
Split−Bin (1/3)
Split−Bin (1/10)
MRVP (2/3)
MRVP (1/3)
MRVP (1/10)
0.5
0.6
0.7
0.8
0.9
1.0

Figure S2


Proportion of Simulations Giving Upper CI < 0.5
BCCVP−BR
Split−Bin (2/3)
Split−Bin (1/3)
Split−Bin (1/10)
MRVP (2/3)
MRVP (1/3)
MRVP (1/10)
Simulation 5
Nominal level of CI
Simulation 6
0.0
0.2
0.4
0.6
0.8
1.0

Supplement 3. Confidence intervals based on Bootstrap Cross-Validation (BCV)

The bootstrap cross-validation (BCV) resampling of Fu, Carroll and Wang (2005) is described as follows.

From the original sample *x*, we draw a bootstrap sample of size n using simple random sampling with replacement. This is repeated *B* times and we denote the bootstrap samples by $x^{*,b} = (x_1^{*,b}, \ldots, x_n^{*,b})$ where $b = 1, \ldots B$. We then apply leave-one-out cross-validation procedure on a bootstrap sample, and the resulting LOOCV estimate is denoted by $(\hat{\theta}_{x^{*,b}})^{LOOCV}$. This resampling approach gives rise to a BCV estimate $\hat{\theta}^{BCV} = \frac{1}{B}(\hat{\theta}_{x^{*,b}})^{LOOCV}$ for the true prediction error. Confidence intervals can also be formed using BCV.

Bootstrap Cross-Validation Percentile Interval (BCVP)

We can approximate $100(1-\alpha)\%$ level upper confidence interval for the true prediction error by $(0, \hat{\theta}_{(1-\alpha)}^{BCV}]$ where $\hat{\theta}_{(1-\alpha)}^{BCV}$ is the $100(1-\alpha)$ empirical percentile of $(\hat{\theta}_{x^{*,b}})^{LOOCV}$, $b = 1, \ldots B$.

Bootstrap Cross-Validation Percentile Interval with Bias Reduction (BCVP-BR)

By reducing the bias in the BCVP method, we can obtain a $100(1-\alpha)\%$ level upper confidence interval, $(0, \hat{\theta}_{(1-\alpha)}^{BCV} - (\hat{\theta}^{BCV} - \hat{\theta}^{LOOCV})]$, where the term $\hat{\theta}^{BCV} - \hat{\theta}^{LOOCV}$ approximates the bias of the BCV estimate since the LOOCV procedure is known to give an almost unbiased estimate for the true prediction error.

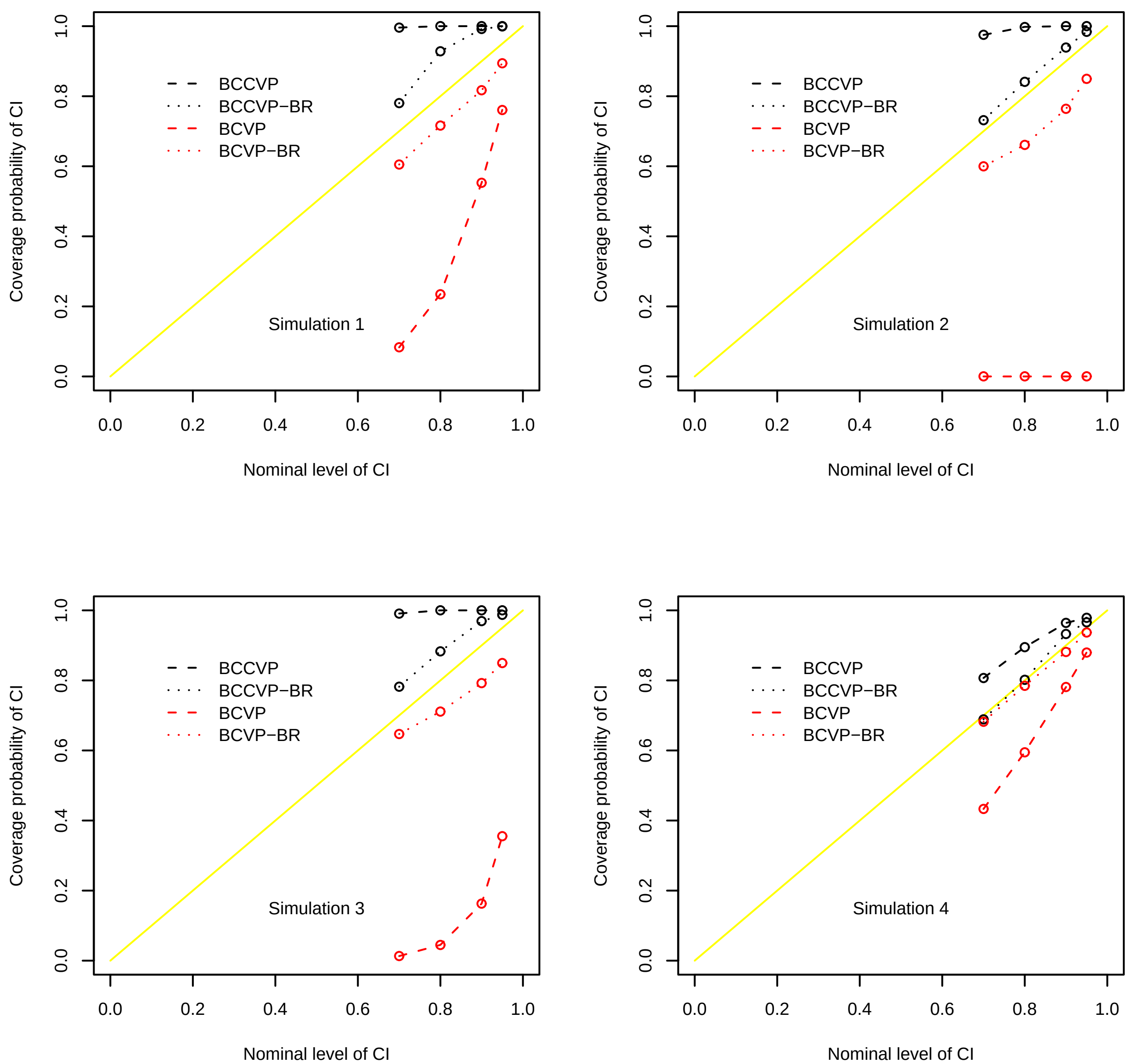
In Figure S3, we compare the BCV based methods versus the bootstrap case cross-validation (BCCV) based methods through empirical coverage probabilities in

Simulations 1-4. The BCVP and the BCCVP are percentile confidence intervals on the basis of B=100 bootstrap replications using BCV and BCCV techniques respectively. We plot the empirical coverage probabilities (CP) of the upper confidence intervals against the nominal coverage levels of 70%, 80%, 90% and 95%. The BCVP has serious under-coverage because of the overlaps between the resampled learning and test sets and the problem is quite striking in the $n \ll p$ situations in Simulations 1-3. The BCCVP gets over the under-coverage by avoiding such overlaps in the resampling but results in substantial over-coverage. Through bias reduction on the percentile intervals, the BCCVP-BR effectively brings down the over-coverage of the BCCVP intervals. A similar bias reduction approach is applied to correct the BCVP interval by subtracting the difference between the BCV estimate and the leave-one-out cross-validation estimate. The resulting BCVP-BR intervals have much better coverage than the BCVP intervals. But the coverage percentages are still lower than the nominal levels and the problem is more obvious in the $n \gg p$ situations; this is an undesirable property in microarray class prediction.

Figure Legend

Figure S3. Comparison of the bootstrap case cross-validation (BCCV) and the bootstrap cross-validation (BCV) schemes. Empirical coverage probabilities are plotted against nominal confidence levels of 70%, 80%, 90% and 95% for the BCCV percentile method (BCCVP), BCCVP with bias reduction (BCCVP-BR), BCV percentile method (BCVP) and BCVP with bias reduction (BCVP-BR). The yellow line is the 45° line for reference.

Figure S3


BCCVP
BCCVP−BR
BCVP
BCVP−BR
Simulation 1
Simulation 2
Simulation 3
Simulation 4
Coverage probability of CI
Nominal level of CI
0.0
0.2
0.4
0.6
0.8
1.0