Bootstrap analysis methods for linear model and its applications in GNSS data processing

Jianqing Cai
Geodätisches Institut, Universität Stuttgart

Main Topics

1. Motivation
2. Bootstrap methods for the confidence domains
3. Case Study: Testing with GPS observation sets
4. Conclusion
The statistical property of the fractional phase measurements of the GPS double difference carrier phase is validated as von Mises distribution.

The classical testing theory (such as, t-test, Chi-squared test, F-test and related ratio-test) cannot be simply applied to the GPS data analysis since the GPS carrier phase observables are not Gauss normally distributed anymore.

The proper hypothesis tests for the estimators based on the von Mises normal distribution should be further studied and applied in GPS data analysis. One possibility: to develop and investigate new efficient bootstrap algorithms for the confidence domain/hypothesis tests on the parameters of the GPS mixed integer linear models.
Bootstrap methods for the confidence domains/hypothesis tests

- Bootstrap methods: A data-based simulation method derived from the phase to pull

5

Schematic of the bootstrap process for estimating the standard error of a statistic \( s(x) \).

Bootstrap samples are generated from the original data set. (After Efron and Tibshirani, 1993)

useful in the directional context.

Since the distributions of the statistics commonly used for inference on directional distributions are more complex than those arising in standard Gauss normal theory, bootstrap methods are particularly useful in the directional context.

One of the principal goals is to produce good confidence intervals.

Yielding more accurate results than Gaussian approximation:

(a) The sampling distribution of a statistic is not known?
(b) The sampling procedures are not valid and/or not available

Two of the modern statistical techniques since 1980s:

- Computer-intensive (Resampling) statistical procedures, which is
- In statistics the phase 'Bootstrap method' refers to a class of

2. Bootstrap methods for the confidence domains/
Just the case for the validation and hypothesis tests of the linear and fixed procedures (e.g., confidence sets) that are more accurate than those produced by the other methods. In the linear model context, these bootstrap methods provide

The bootstrap algorithm for estimating the standard error of a statistic \( \hat{y} \) from \( \hat{y} \) is:

\[
\theta = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})
\]

where

\[
\hat{y} = (X'X)^{-1}X'y
\]

The bootstrap algorithm for estimating the standard error of a

![Diagram of bootstrap algorithm]
Normally we can perform the resampling iterations with 1000 times.

But bootstrapping pairs can also be used for deterministic $g$. Whereas bootstrapping pairs is more appropriate for the case of random $g$. Bootstrap residuals is more suitable for the case of deterministic $g$.

Iteractions:

Corresponding values can be examined after any desired number of iterations.

$\hat{y} = \hat{G}x + e$

To obtain $\hat{y}$ and $\hat{G}$, The linear model selected with the probability $1/n$, and sample with replacement, to each $y_i$, is the ith observation and $x_i$ is the ith row of the $G$ matrix, each $\gamma$ corresponds is the ith row of design matrix o and is the sample mean and sample standard deviation of corresponding row of design matrix.

Choose a sample of size $n$ from the residuals generated with the Bootstrap Residuals - Fit the linear model and obtain the $n$ residuals.

Normally we can perform the resampling iterations with 1000 times.

The estimated parameters $\hat{y}$, the linear model $\hat{G}$ and $\hat{e}$ can be found, which allows constructing confidence domains of these sampled values to the predicted $\hat{y}$ to give a resampled set of $\hat{y}$.'
Bootstrap-t confidence interval:

\[ (\bar{\theta} - q_{\alpha/2} \cdot \hat{\sigma}, \bar{\theta} + q_{\alpha/2} \cdot \hat{\sigma}) \]

To calculate the bootstrap-t confidence interval:

\[ \bar{\theta} \mp q_{\alpha} \frac{\hat{\sigma}}{\sqrt{n}} \]

Quantile, denoted by \( q_{\alpha} \), is estimated by \( \hat{\theta} \).

To estimate the quantiles needed for the endpoints of interval, the \( q \)-th quantile is computed:

To generate bootstrap samples and for each bootstrap sample the following quantity is computed:

Bootstrap-t confidence interval:
Residuals methods (5 epochs).

The comparison of the float estimates and their confidence intervals with the LS and bootstrapping methods.

For the testing observation period 5~20 epochs there are 11 unknown parameters, including 3 coordinate differences and 8 ambiguities.

There are total 320 L1 double difference phase observables.

Phase baseline lengths were calculated using observations above 10°.

Sampling rate at one baselines (~3.6 km).

Short baselines test data: about 2 hour observations with 20 second.

GPS observation sets:

Case Study: Testing with GPS observation sets.
The comparison of the float estimates and their confidence intervals with the LS and bootstrapping residuals methods (20 epochs).
The comparison of the float estimates and their confidence intervals with the LS and bootstrapping pairs methods (10 epochs).

The comparison of the float estimates and their confidence intervals with the LS and bootstrapping pairs methods (5 epochs).
The comparison of the float estimates and their confidence intervals with the LS and bootstrapping pairs methods (20 epochs).

Analysis of the Bootstrapping confidence intervals for the float solutions:

- Bootstrapping confidence intervals are consistent with the LS confidence intervals based on the t-test.
- The bootstrapping confidence intervals are consistent with the LS confidence intervals derived without any assumption about the probability distribution of the observations.
- Both kinds of the confidence intervals all cover the potential correct fixed ambiguity integers, which are important for searching processes.
- The bootstrapping pairs method depends on the condition number of the design matrix.
- The bootstrapping pairs method provides us an efficient and accurate algorithm to construct the confidence domains of the GPS float solutions.
- Note: The Bootstrapped confidence sets are slightly varied among the every resampling (simulation) process.

But the bootstrapping confidence intervals are derived without any and fixed solution.
We have studied the bootstrap algorithms and successfully applied the efficient bootstrapping residuals and pairs methods to construct the confidence domains of the GPS float solutions. This answers the open question mentioned above and provides a complete solution for the estimation and hypothesis testing of the parameters of the GPS mixed integer linear models in the directional context.

Some selected References:


GPS mixed integer linear models in the directional context. This answers the open question mentioned above and provides a complete solution for the estimation and hypothesis testing of the parameters of the observations and its effects on the hypothesis testing of the related estimators, ION GNSS 2007 Meeting Proceedings "ION GNSS 20th International Technical Meeting of the Satellite Division, 25-28, Sep. 2007, Fort Worth, TX, 331-338.


Thank you!