Resampling techniques for statistical modeling

Gianluca Bontempi

Département d'Informatique

Boulevard de Triomphe - CP 212

http://www.ulb.ac.be/di

Motivations

- Classical statistics is largely based on idealized assumptions (e.g. linear relations, stationary and/or normal distribution)
- Time and cost considerations directed statistical work towards models and methods which kept down the amount of computation.
- The application of statistical methods to highly nonlinear, non stationary and non gaussian problems required the adoption of alternative statistical tools.
- The advance of computation makes possible a new perspective in statistics.
- Hard theoretical analysis is replaced by computationally intensive but simple methods.
- Resampling techniques has met recently an enormous success.

Computer intensive statistical methods

- **Computer intensive statistical methods** are methods which require a large number of repeated (and almost similar) computations on the data.
- They are automatic and generally easy to implement.
- They require very few assumption on the form of the model and/or uncertainty.
- The advance of processing power make them feasible and easy-to-use.
- **NB**: computational intensive methods **are not more precise** than classical methods when the **assumptions of classical methods are true.**
- However, they are **more powerful** when these assumptions are **not fulfilled.** In other terms, the new methods solve many problems which could not be properly attacked before.

Computer intensive statistical methods (II)

- Potential risk: a lot of computation does not guarantee that information has been well used.
- The larger availability of statistical software makes the potential misuse of statistical measures larger than ever.
- It is therefore important to be critical and have a theoretical view in order to understand when computer intensive methods have anything essential to offer compared to classical techniques.

Required statistical notions

Estimator.

Bias and variance of an estimator.

Hypothesis testing.

Confidence intervals.

Model.

Model assessment.

Model selection.