DM63
HEURISTICS FOR
COMBINATORIAL OPTIMIZATION

Lecture 13

Empirical Analysis and Sequential Testing

Marco Chiarandini

Outline

1. Inferential Statistics (contd)
   - Multiple Comparisons
   - Sequential Testing: the Racing Algorithm
   - Other Methods
2. Design of Experiments

Outline

1. Inferential Statistics (contd)
   - Multiple Comparisons
   - Sequential Testing: the Racing Algorithm
   - Other Methods

2. Design of Experiments

Multiple Comparisons Analysis

Post-hoc analysis: Once the effect of factors has been recognized a finer grained analysis is performed to distinguish where important differences are.

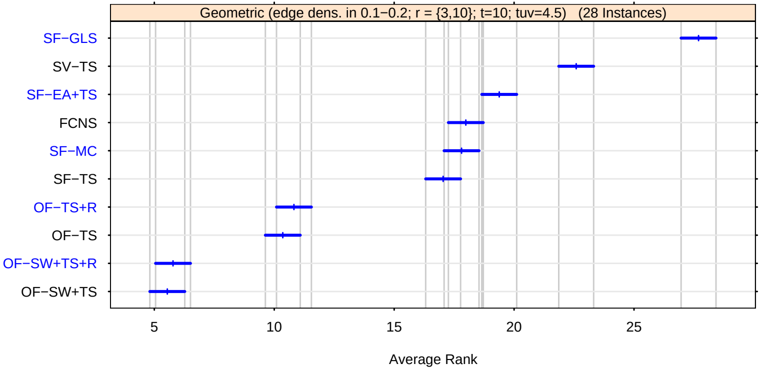
Statistical tests

- Parametric analysis: Pairwise t test
- Nonparametric analysis:
  - One Factor AOV: Pairwise Kruskall Wallis or Mann Whitney
  - Two Factors AOV: Pairwise Friedman Test or Wilcoxon Test
  - Two Factors Repeated Measures AOV: Pairwise Friedman Test

Adjustment method

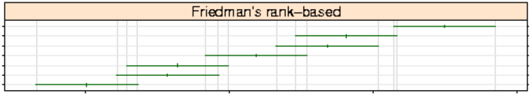
- Tukey Honest Significance Method (for parametric analysis)
- Bonferroni $\alpha = \alpha_{EX}/c$ (conservative)
- Holm (step-wise)

An Example with Confidence Intervals

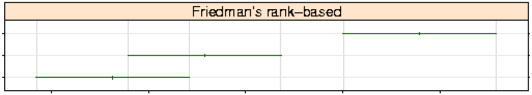


Sequential Testing

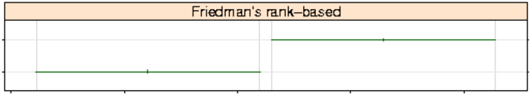
⇒ 7 algorithms, 5 instances, 3 runs



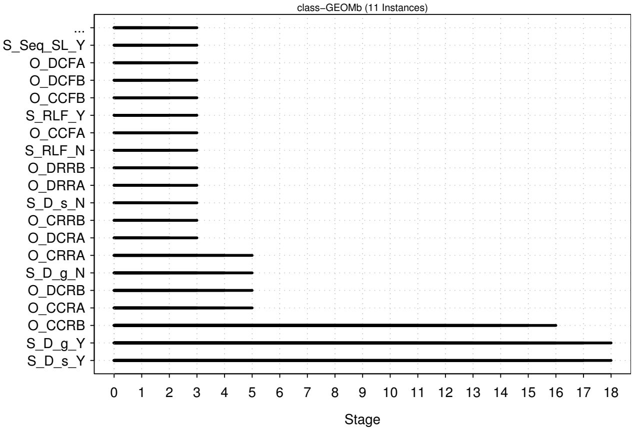
⇒ 3 algorithms, 5 instances, 9 runs



⇒ 2 algorithms, 5 instances, 30 runs



Sequential Testing



Unreplicated Designs

procedure *Race*
  repeat only one candidate left or no more unseen instances
    Randomly select an unseen instance and test all candidates on it
    Perform *all-pairwise comparison* statistical tests
    Drop all candidates that are significantly inferior to the best algorithm
  end
end *Race*

Replicated Designs

procedure *Race*
  repeat only one candidate left or maximal number of runs exceeded
    Run all candidates once on all instances
    Perform *all-pairwise comparison* statistical tests
    Drop all candidates that are significantly inferior to the best
      algorithm
  end
end *Race*

Distribution Fitting

Parameters of theoretical distributions to fit the data through the Maximum Likelihood Method can be found by in R

```
fitdistr(x, ``exp'')
```

Kolmogorov-Smirnov Tests and Goodness of Fit

The test compares empirical cumulative distribution functions.

It computes the maximal difference between the two curves and assess how likely is this value by permutation methods or approximation to the χ^2 distribution.

The test can be done in R with `ks.test`.

The test can be used as a two-samples or single-sample test.
In single-sample the second distribution is a specified continuous distribution, eg, ``pweibull''. The parameters of the distribution must be pre-specified and not estimated from the data.

Multiple Regression, Non Linear Regression and Smoothing

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_p x_{i2} + \ldots + \beta_p x_{ip} + \epsilon_i$$

$$(y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i)$$

In R: `lm`

R^2 tells the fraction of the variance explained by the model

---

$$y = f(x, \beta)\epsilon$$

In R: `nls`

Smooth curves in scatter plots (no idea on the functional form of the curve)

In R: `poly`, `smooth.spline`, `loess.smooth`, `supsmu`, `ksmooth`

Then add to a scatter plot with `lines`

Outline

Steps in the Design of Experiments for Algorithms

- Statement of the objectives of the experiment
  - Comparison of different algorithms
  - Impact of algorithm components
  - How instance features affect the algorithms
- Identification of the sources of variance
  - Treatment factors (qualitative and quantitative)
  - Controllable nuisance factors ⇐ blocking
  - Uncontrollable nuisance factors ⇐ measuring
- Definition of factor combinations to test

  Easiest design: Unreplicated or Replicated Full Factorial Design
- Running a pilot experiment and refine the design
  - Bugs and no external biases
  - Ceiling or floor effects
  - Rescaling levels of quantitative factors
  - Detect the number of experiments needed to obtained the desired power.

Experimental Design (2)

Algorithms $\Rightarrow$ Treatment Factor; Instances $\Rightarrow$ Blocking Factor

Design A: One run on various instances (Unreplicated Factorial)

	Algorithm 1	**Algorithm 2**	...	**Algorithm k**
Instance 1	X_{11}	X_{12}		X_{1k}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
Instance b	X_{b1}	X_{b2}		X_{bk}

Design B: Several runs on various instances (Replicated Factorial)

	Algorithm 1	**Algorithm 2**	...	**Algorithm k**
Instance 1	$X_{111}, \ldots, X_{11r}$	$X_{121}, \ldots, X_{12r}$		$X_{1k1}, \ldots, X_{1kr}$
Instance 2	$X_{211}, \ldots, X_{21r}$	$X_{221}, \ldots, X_{22r}$		$X_{2k1}, \ldots, X_{2kr}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$
Instance b	$X_{b11}, \ldots, X_{b1r}$	$X_{b21}, \ldots, X_{b2r}$		$X_{bk1}, \ldots, X_{bkr}$

Experimental Design (3)
How Many Runs?

- If the sample size is large enough (infinity) any difference in the means of the factors, no matter how small, will be significant
- Real vs. Statistical significance (Montgomery and Runger, 2002) Study factors until the improvement in the response variable is deemed small
- Desired statistical power + practical precision $\Rightarrow$ sample size

Note: If resources available for N runs then the optimal design is one run on N instances

Experimental Design (4)
Measures for Solution Quality

Different instances = different scales ⇒ need for an invariant measure

- Distance or error from a reference value:

$$z(q,i) = \frac{q(i) - \bar{q}(i)}{s(i)} \qquad \text{standard score} \tag{1}$$

$$e_1(q,i) = \frac{|q(i) - q_{opt}(i)|}{q_{opt}(i)} \qquad \text{relative error} \tag{2}$$

$$e_2(q,i) = \frac{q(i) - q_{opt}(i)}{q'(i) - q_{opt}(i)} \tag{3}$$

  - optimal value computed exactly or known by instance construction
  - surrogate value such bounds or best known values
- Rank (no need for normalization but loss of information)