DM811

Heuristics for Combinatorial Optimization

Lecture 14
Race: A Configuration Tool

Marco Chiarandini

Department of Mathematics & Computer Science
University of Southern Denmark

Outline

Outline

Outline

A Motivating Example

- There is a competition and two stochastic algorithms $\mathcal{A}_1$ and $\mathcal{A}_2$ are submitted.
- We run both algorithms once on n instances.
  On each instance either $\mathcal{A}_1$ wins (+) or $\mathcal{A}_2$ wins (-) or they make a tie (=).

Questions:

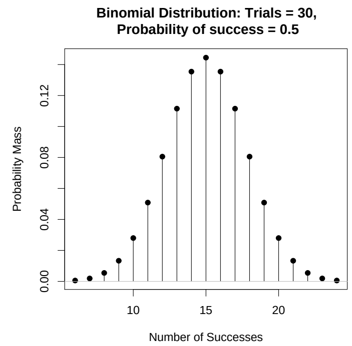
1. If we have only 10 instances and algorithm $\mathcal{A}_1$ wins 7 times how confident are we in claiming that algorithm $\mathcal{A}_1$ is the best?
2. How many instances and how many wins should we observe to gain a confidence of 95% that the algorithm $\mathcal{A}_1$ is the best?

A Motivating Example

- p: probability that $\mathcal{A}_1$ wins on each instance (+)
- n: number of runs without ties
- Y: number of wins of algorithm $\mathcal{A}_1$

If each run is indepenedent and consitent:

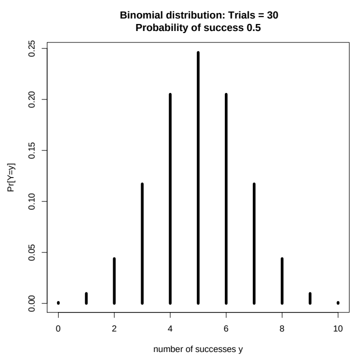
$$Y \sim B(n,p): \qquad \Pr[Y=y] = \binom{n}{y} p^y (1-p)^{n-y}$$



1 If we have only 10 instances and algorithm $\mathcal{A}_1$ wins 7 times how confident are we in claiming that algorithm $\mathcal{A}_1$ is the best?

Under these conditions, we can check how unlikely the situation is if it were $p(+) \leq p(-)$.

If $p = 0.5$ then the chance that algorithm $\mathcal{A}_1$ wins 7 or more times out of 10 is 17.2%: quite high!



2 How many instances and how many wins should we observe to gain a confidence of 95% that the algorithm $\mathcal{A}_1$ is the best?

To answer this question, we compute the 95% quantile, *i.e.*, $y : \Pr[Y \geq y] < 0.05$ with $p = 0.5$ at different values of n:

n	10	11	12	13	14	15	16	17	18	19	20
y	9	9	10	10	11	12	12	13	13	14	15

This is an application example of sign test, a special case of binomial test in which $p = 0.5$

Inferential Statistics

General procedure:

- Assume that data are consistent with a null hypothesis H_0 (e.g., sample data are drawn from distributions with the same mean value).
- Use a statistical test to compute how likely this is to be true, given the data collected. This "likely" is quantified as the p-value.
- Accept H_0 as true if the p-value is larger than an user defined threshold called level of significance α.
- Alternatively (p-value $< \alpha$), H_0 is rejected in favor of an alternative hypothesis, H_1, at a level of significance of α.

Preparation of the Experiments

Variance reduction techniques

- Same pseudo random seed

Sample Sizes

- 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
  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 [Birattari, 2004]

Experimental Design

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	$X_{b11}, \ldots, X_{b1r}$	$X_{b21}, \ldots, X_{b2r}$		$X_{bk1}, \ldots, X_{bkr}$

Outline

Unreplicated Designs

Procedure Race [Birattari 2002]:
repeat
Randomly select an unseen instance and run all candidates on it
Perform *all-pairwise comparison* statistical tests
Drop all candidates that are significantly inferior to the best algorithm
until only one candidate left or no more unseen instances ;

- F-Race use Friedman test
- Holm adjustment method is typically the most powerful

```
stat.test=c(‘‘friedman’’,’’t.bonferroni’’,’’t.holm’’,’’t.none’’)
first.test=3
```


Sequential Testing