Probabilistic Algorithms

Paul Cotofrei

information management institute

PA 2016

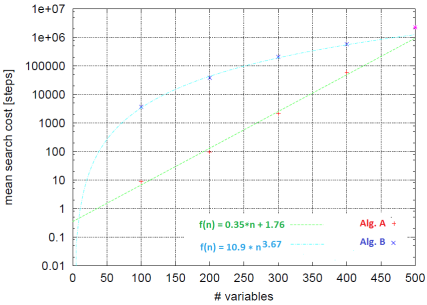
Outline

Empirical Analysis of Algorithms

- Goals:
  - Show that algorithm A improves state-of-the-art
  - Show that algorithm A is better than algorithm B
  - Show that algorithm A has property P
- Issues:
  - algorithm implementation (fairness)
  - selection of problem instances (benchmarks)
  - performance criteria (what is measured?)
  - experimental protocol
  - data analysis and interpretation
- Comparative empirical performance analysis of optimisation algorithms

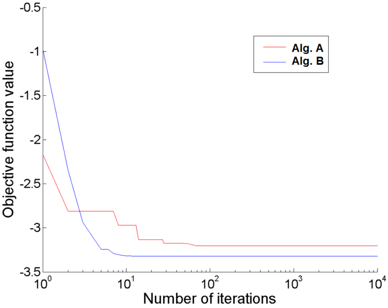
Type of Analysis

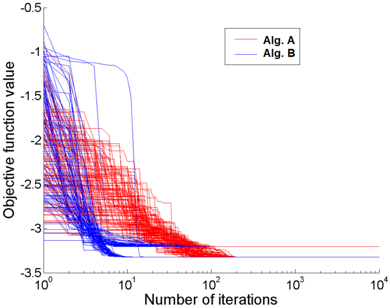
- Scaling Analysis : analyse scaling of performance with instance size
  - measure performance for various instance sizes
  - fit parametric model (e.g., $\alpha e^{\beta x}$) to data points
  - test interpolation / extrapolation
- Robustness Analysis: Measure robustness of performance w.r.t. ...
  - algorithm parameter settings
  - problem type



Solution Quality Distribution

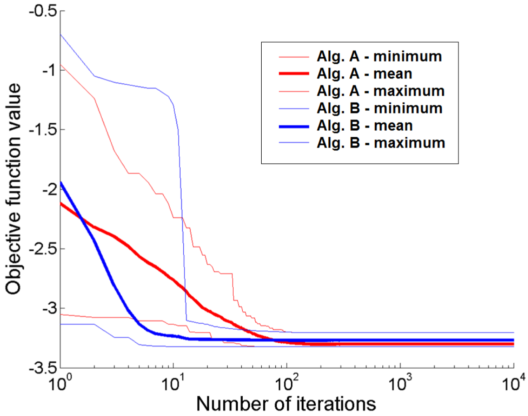
- Speed - measured as CPU time
  - depend on a number of variables (language, compiler, machine)
- Solution quality: absolute measures (if optimal solution is known) or relative measures
- Visual comparisons
  - The "Progress Plot" (a single run/multiple runs)





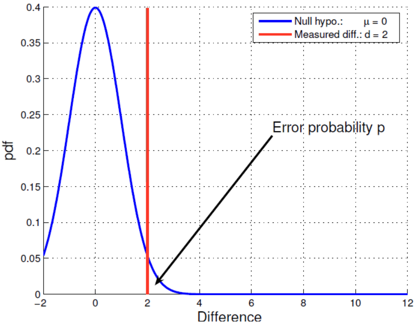
SQD-based methodology

- Run algorithm multiple times on given problem instance(s)
- Estimate empirical solution quality distributions (SQDs) for different run-times
- Get simple descriptive statistics (mean, median, stddev, percentiles, ...) from SQD data
- Check statistical significance using appropiate statistical tests



Statistical hypothesis

- Run algorithms A and B n times.
- Two result vectors: y_A and y_B that contain the best function values.
- Difference $d_{AB} = y_A - y_B$
- $Var(d_{AB}) \neq 0$.
- Null hypothesis: "There is no difference in means, i.e $E(d_{AB}) = \mu = 0$."
- Error probability



Outline

Statistical tests

- Consider a hypothesis H_0 about a statistical fact (the value of a distribution parameter, of a set of parameters, the form of a distribution, etc..)
- To test this hypothesis we need a statistics $S(x_1, x_2, ..., x_n)$ such that:
  1. The distribution of S is known if H_0 is true
  2. Extreme values of S are arguments that H_0 is "maybe" false

Statistical tests - examples

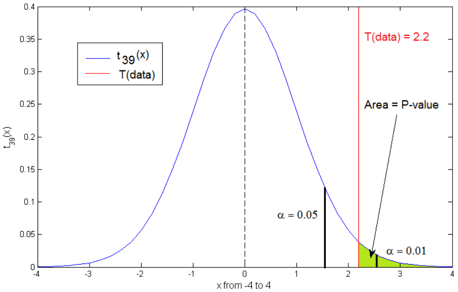
- Example1: let be $x_1, x_2, ..., x_n$ a sequence of n i.i.d. scalar measurements, from a variable X with distribution $N(\mu, \sigma^2)$
  1. Standard deviation σ known, mean μ unknown
  2. Hypothesis H_0: $\mu = 3$
  3. Consider $S(x_1, .., x_n) = \frac{\bar{x}-3}{\sigma/\sqrt{n}}$.
     - If H_0 is true, then S follows a distribution $N(0, 1)$
     - Very high/low values for S implies that H_0 is maybe false.
- Example2: $x_1, x_2, ..., x_n$ a sequence of n i.i.d. scalar measurements, from a variable X with distribution $N(\mu, \sigma^2)$
  1. Standard deviation σ unknown, mean μ unknown
  2. Hypothesis H_0: $\mu = 3$
  3. Consider $T(x_1, .., x_n) = \frac{\bar{x}-3}{s/\sqrt{n}}$, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.
     - If H_0 is true, then T follows a distribution Student-t with $n-1$ degree of freedom
     - Very high/low values for S implies that H_0 is maybe false.

P-value and significance level

- *P-value*: provides a measure of the extent to which the statistic appears inconsistent with the null hypothesis.
  - if extreme high values inconsistent with the null hypothesis - *P-value* = $P(S \geq S(\text{data}))$
  - if extreme low values inconsistent with the null hypothesis - *P-value* = $P(S \leq S(\text{data}))$
  - if extreme high and low values inconsistent with the null hypothesis - *P-value* = $P(S \geq |S(\text{data})|)$
- Only for having common references, standard significance error levels have been introduced ($\alpha = 0.1, \alpha = 0.05, \alpha = 0.01$)
  - The hypothesis is accepted if $P - value \leq \alpha$
- Accepting/rejecting an hypothesis depends on
  - Size of data (usually, more data, more confidence in final decision)
  - Selected significance level

Example

- Suppose $X = N(\mu, \sigma^2)$, μ, σ unknown
- Hypothesis $H_0 : \mu \leq 0$
- Data $= x_1, x_2, \ldots, x_{40}$; $T(data) = T(x_1, .., x_{40}) = \frac{\bar{x}}{s/\sqrt{39}} = 2.2$, where
  $\bar{x} = \frac{1}{40} \sum_{i=1}^{40} x_i$ and $s = \sqrt{\frac{1}{39} \sum_{i=1}^{40} (x_i - \bar{x})^2}$
- If H_0 true then T follows student t distribution with 39 degree of freedom; large values of T suggest rejecting the hypothesis
- *P-value* = $P(t_{39} \geq T(data)) = P(t_{39} \geq 2.2) = 0.017$
- If the significance level for this test is fixed at $\alpha = 0.05$ then, as *P-value* is less than 0.05, the hypothesis H_0 is rejected. If the significance level is fixed at $\alpha = 0.01$, the hypothesis H_0 is accepted.



Outline

Confidence interval

- Let be p a parameter of interest. We can either
  1. Test the statistical hypothesis $H_0 : p = \beta$; or
  2. Construct a confidence interval for p
- Confidence interval for p with a confidence level α if exist statistics S_1, S_2 such that

$$P(S_1 \leq p \leq S2) = \alpha$$

- The upper and lower bounds of a confidence interval: random variables
- The interval "cover" the true value of parameter p with a probability α

Confidence interval for mean

- $x_1, x_2, ..., x_{40}$ a sequence of 40 independent, i.i.d. scalar measurements, from a variable $X = N(\mu, \sigma^2)$ with unknown σ.
- The parameter of interest: $\mu = E(X)$
- Consider $S = \frac{\bar{x}-\mu}{s/\sqrt{40}}$ which follows a Student distribution with 39 degree of freedom (t_{39}).
- Let α_1, α_2 two constants such that

$$P(\alpha_1 \leq S \leq \alpha_2) = \alpha$$

- $P(\alpha_1 \leq \frac{\bar{x}-\mu}{s/\sqrt{40}} \leq \alpha_2) = P(\frac{\alpha_1 s}{\sqrt{40}} \leq \bar{x} - \mu \leq \frac{\alpha_2 s}{\sqrt{40}}) =$
  $= P(\bar{x} - \frac{\alpha_2 s}{\sqrt{40}} \leq \mu \leq \bar{x} - \frac{\alpha_1 s}{\sqrt{40}}) = \alpha$
- $S_1 = \bar{x} - \alpha_2 s/\sqrt{40}$ and $S_2 = \bar{x} - \alpha_1 s/\sqrt{40}$
- In practice, for a confidence level α, we may take

$$\alpha_1 = F^{-1}\left(\frac{1-\alpha}{2}\right) \Rightarrow P(t_{39} \leq \alpha_1) = \frac{1-\alpha}{2}$$

$$\alpha_2 = F^{-1}\left(\frac{1+\alpha}{2}\right) \Rightarrow P(t_{39} \leq \alpha_2) = \frac{1+\alpha}{2}$$

Outline

Two-sample tests

- Let be an optimization problem related to a loss function $L(\mathbf{d})$ and **A**, **B** two optimization randomized heuristics
- Consider the sequence $x_1, x_2, ..., x_n$, where x_i is the value of the optimal solution $L(\hat{\theta})$ returned by the algorithm **A** during the i^{th} run
- Consider $y_1, y_2, ..., y_m$ a similar sequence, but for the algorithm **B**
- For a randomized algorithm, the returned optimal solution $L(\hat{\mathbf{d}})$ is a random variable. Consider
  - $X = L(\hat{\mathbf{d}})_{\text{algorithm A}}$, so x_i are samples of X
  - $Y = L(\hat{\mathbf{d}})_{\text{algorithm A}}$, so y_i are samples of Y
  - $\mu_X = E(X)$
  - $\mu_Y = E(Y)$
- Hypothesis $H_0 : \mu_X = \mu_Y$ (the two algorithms provide similar solutions)
- $\{x_i\}$ i.i.d., $\{y_j\}$ i.i.d., but σ_x and σ_y unknown

Case 1

- **Matched pair tests**
- The data are collected as n pairs $\{x_i, y_i\}$ and each pair shares the same randomness (the two algorithms, **A** and **B**, use the *same* sequence of random numbers; the final solution $\hat{\mathbf{d}}$ differ because the algorithms process this sequence in different ways)
- The two sequences $\{x_i\}$ and $\{y_i\}$ are highly dependent
- Calculate $d_i = x_i - y_i$, $i = 1..n$, representing n samples from $D = X - Y$
- The hypothesis $H_0 : \mu_X = \mu_Y \Leftrightarrow \mu_D = 0$
- Variance estimate for the variable $D = X - Y$ is

$$s^2_{X-Y} = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d}) = \frac{1}{n-1}\sum_{i=1}^{n}[x_i - y_i - (\bar{x} - \bar{y})]^2$$

where $\bar{x}$ and $\bar{y}$ are the sample means of $\{x_i\}$ and $\{y_i\}$

Case 2

- **Unmatched pair tests with common variance**
- The sequence $\{x_i\}, i = 1..n_X$ is independent of $\{y_j\}, j = 1..n_Y$, but both variables (X and Y) have the same variance $\sigma_X^2 = \sigma_Y^2$
- Generally arises when algorithms using the same approach are tested by different individuals
- The two sample sizes (n_X and n_Y) are not necessary equal
- The estimate s_P^2 of common variance is calculated using the individual samples variance estimates s_X^2 and s_Y^2

$$s_p^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}$$

where

$$s_X^2 = \frac{1}{n_X - 1} \sum_{i=1}^{n_X} (x_i - \bar{x})^2$$

$$s_Y^2 = \frac{1}{n_Y - 1} \sum_{i=1}^{n_Y} (y_i - \bar{y})^2$$

Case 3

- **Unmatched pair tests with different variances**
- Similar as case 2, but $\sigma_X^2 \neq \sigma_Y^2$
- Generally arise when two fundamental different algorithms are tested
- The two estimates of variances are:

$$s_X^2 = \frac{1}{n_X - 1} \sum_{i=1}^{n_X} (x_i - \bar{x})^2$$

and

$$s_Y^2 = \frac{1}{n_Y - 1} \sum_{i=1}^{n_Y} (y_i - \bar{y})^2$$

Generic test statistic

- Calculate $\bar{x} = \frac{1}{n_X} \sum_{i=1}^{n_X} x_i$ and $\bar{y} = \frac{1}{n_Y} \sum_{i=1}^{n_Y} y_i$
- Case 1: the statistics is

  $$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s^2_{X-Y}}{n}}}$$

  Under H_0, T follows Student (t_{n-1}) with $n - 1$ degree of freedom
- Case 2: the statistics is

  $$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_X} + \frac{s_p^2}{n_Y}}}$$

  Under H_0, T follows Student ($t_{n_X+n_Y-2}$) with $n_X + n_Y - 2$ degree of freedom

Generic test statistic

- Case 3: the statistics is

$$T = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}}$$

Under H_0, T follows Student (t_m) with m degree of freedom, where

$$m = \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{(s_X^2/n_X)^2}{n_X+1} + \frac{(s_Y^2/n_Y)^2}{n_Y+1}} - 2$$

- Extreme positive values for T indicate that algorithm **B** is better than algorithm **A**
- Extreme negative values for T indicate that algorithm **A** is better than algorithm **B**

Acceptance intervals

- An acceptance interval is an interval guaranteed to contain the difference $\bar{x} - \bar{y}$ with a probability $1 - \alpha$ under the hypothesis $\mu_X = \mu_Y$.

Step 1 Set the significance level α of the statistical test H_0

Step 2 Construct the acceptance interval $[l, u]$ for the statistic T, given α

Step 3 Using data $\{x_i\}, \{y_j\}$ calculate T

  - If $T < l$, accept hypothesis "A is better than B"
  - If $T \in [l, u]$, accept hypothesis "A and B are equivalent"
  - If $T > u$, accept hypothesis "B is better than A"

- If we denote $t_n^{(\alpha/2)} = F^{-1}(\alpha/2)$ where F is the cumulative distribution for random variable following a Student distribution with n degree of freedom, then the acceptance interval is $[-\beta, \beta]$, where

1. Case 1: $\beta = t_{n-1}^{(\alpha/2)} \sqrt{\frac{s^2_{X-Y}}{n}}$
2. Case 2: $\beta = t_{n_X+n_Y-2}^{(\alpha/2)} \sqrt{\frac{s_p^2}{n_X} + \frac{s_p^2}{n_Y}}$
3. Case 3: $\beta = t_m^{(\alpha/2)} \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}$

Example

- Two algorithms using random search, running 8 times
- L measured with noise; same function, so variance is the same for both algorithms ($\sigma_X = \sigma_Y$)

Run. no	Algorithm A	Algorithm B
1	4.0	2.5
2	2.9	3.6
3	2.8	0.2
4	2.0	2.3
5	3.3	0.3
6	4.0	0.7
7	3.6	3.1
8	1.7	2.1

- $\bar{x} = 3.04, \bar{y} = 1.85, s_X = 0.86, s_Y = 1.296, s_p = 1.1$
- Let $\alpha = 0.1$, so $t_{8+8-2}^{(0.05)} = 1.76$ and the acceptance interval is $[-0.97, 0.97]$
- $\bar{x} - \bar{y} = 1.19 \notin [-0.97, 0.97]$, so we may reject the hypothesis of "algorithms A and B have the same performance" in favor of "algorithm B is better than algorithm A"
- Remark: $T = 1.19/\sqrt{0.303}$ and *P-value* for T is 0.049, which is not a strong indication of rejection; more than, the size of samples is very low (8), so "B is better than A" must be consider with reserve!

Statistical tests in Matlab

Statistical formulae	MatLab function
$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	$\gg$ mean([x1, x2,...,xn])
$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$	$\gg$ std([x1,..,xn])
$t_n^\alpha = F^{-1}(\alpha)$	$\gg$ tinv(α, n)