Performance Evaluation

A not so Short Introduction
Analyzis of experimental results and inference

Jean-Marc Vincent[1,2]

[1] Laboratoire LIG, projet Inria-Mescal
UniversitéJoseph Fourier
Jean-Marc.Vincent@imag.fr
[2] LICIA
Laboratoire International de Calcul Intensif
et d'Informatique Ambiante

CAPES/COFECUB Project
Gestion de Ressources pour le Calcul Parallèle sur Grilles

2011 May 27

Introduction

Aim of this lecture

Discuss about experiments in computer science

- Why experiencing ?
- Advantages and drawbacks of experiments
- Experiments = Modelling
- Scientific method

Interactive course : discussion about your own experiments

Outline

Outline

Why experiments ?

Design of architectures, softwares

- System debugging (!!)
- Validation of a proposition
- Qualification of a system
- Dimensioning and tuning
- Comparison of systems

Many purposes $\Rightarrow$ different methodologies

Experiments fundamentals

Scientific Method

Falsifiability is the logical possibility that an assertion can be shown false by an observation or a physical experiment. [Popper 1930]

Modelling comes before experimenting

Modelling principles [J-Y LB]

- (Occam:) if two models explain some observations equally well, the simplest one is preferable
- (Dijkstra:) It is when you cannot remove a single piece that your design is complete.
- (Common Sense:) Use the adequate level of sophistication.

Experiments fundamentals

Scientific Method

Falsifiability is the logical possibility that an assertion can be shown false by an observation or a physical experiment. [Popper 1930]

Modelling comes before experimenting

Modelling principles [J-Y LB]

- (Occam:) if two models explain some observations equally well, the simplest one is preferable
- (Dijkstra:) It is when you cannot remove a single piece that your design is complete.
- (Common Sense:) Use the adequate level of sophistication.

Experiments fundamentals

Scientific Method

Falsifiability is the logical possibility that an assertion can be shown false by an observation or a physical experiment. [Popper 1930]

Modelling comes before experimenting

Modelling principles [J-Y LB]

- (Occam:) if two models explain some observations equally well, the simplest one is preferable
- (Dijkstra:) It is when you cannot remove a single piece that your design is complete.
- (Common Sense:) Use the adequate level of sophistication.

Design of experiments (introduction)

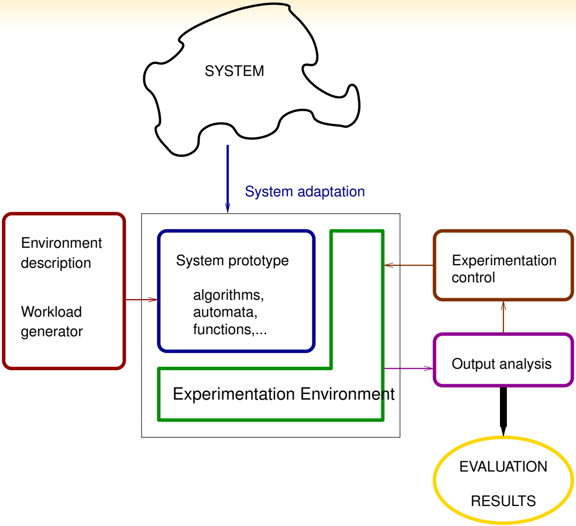
Formulation of the question

Give explicitly the question (specify the context of experimentation)

- Identify parameters (controlled and uncontrolled)
- Identify factors (set levels)
- Specify the response of the experiment

Minimize the number of experiments for a maximum of accuracy

Experimental Framework



Observation technique

Integrated environment : Benchmarks

- Qualification
- Comparison
- Standardization

No interpretation

Level of observation

- Instruction level (Papi)
- System level (OS probes)
- Middleware level (JVMTI)
- Application level (traced libraries, MPItrace)
- User level (own instrumentation point)

Build a semantic on events

Observation technique

Integrated environment : Benchmarks

- Qualification
- Comparison
- Standardization

No interpretation

Level of observation

- Instruction level (Papi)
- System level (OS probes)
- Middleware level (JVMTI)
- Application level (traced libraries, MPItrace)
- User level (own instrumentation point)

Build a semantic on events

Qualification of experiments

Qualification of measurement tools

- Correctness
- Accuracy
- Fidelity
- Coherence (set of tools)

Qualification on the sequence of experiments

- Reproducibility
- Independence from the environment
- Independence one with each others

Qualification of experiments

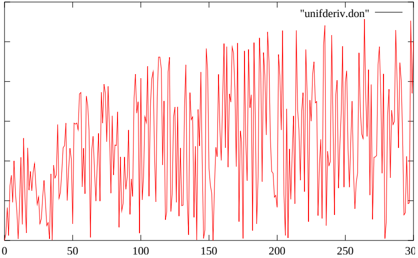
Qualification of measurement tools

- Correctness
- Accuracy
- Fidelity
- Coherence (set of tools)

Qualification on the sequence of experiments

- Reproducibility
- Independence from the environment
- Independence one with each others

Control of experiments (1)



Tendency analysis

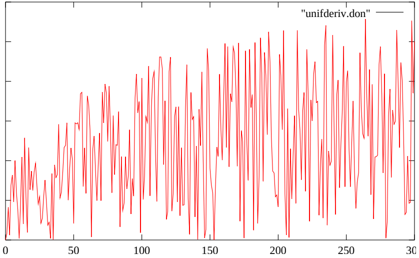
non homogeneous experiment

⇒ model the evolution of experiment

estimate and compensate tendency

explain why

Control of experiments (1)



Tendency analysis

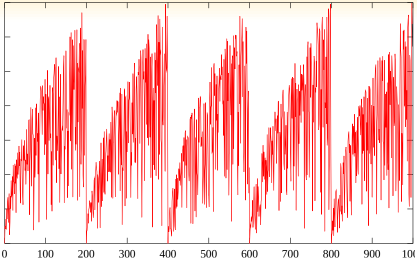
non homogeneous experiment

$\Rightarrow$ model the evolution of experiment

estimate and compensate tendency

explain why

Control of experiments (2)



Periodicity analysis

periodic evolution of the experimental environment ?

⇒ model the evolution of experiment

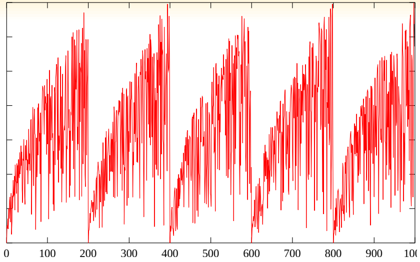
Fourier analysis of the sample

Integration on time (sliding window analysis) Danger : size of the window

Wavelet analysis

explain why

Control of experiments (2)



Periodicity analysis

periodic evolution of the experimental environment ?

⇒ model the evolution of experiment

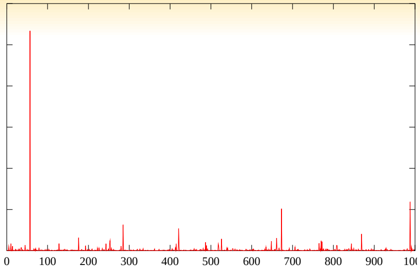
Fourier analysis of the sample

Integration on time (sliding window analysis) Danger : size of the window

Wavelet analysis

explain why

Control of experiments (3)



Non significant values

extraordinary behaviour of experimental environment

rare events with different orders of magnitude

⇒ threshold by value

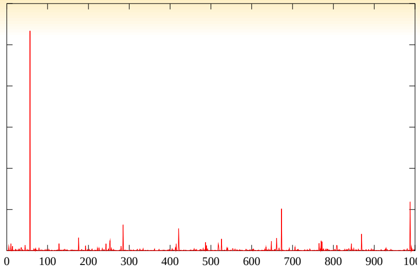
Danger : choice of the threshold : indicate the rejection rate

⇒ threshold by quantile

Danger : choice of the percentage : indicate the rejection value

explain why

Control of experiments (3)



Non significant values

extraordinary behaviour of experimental environment

rare events with different orders of magnitude

$\Rightarrow$ threshold by value

Danger : choice of the threshold : indicate the rejection rate

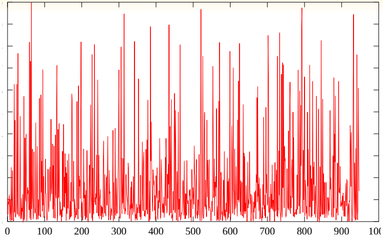
$\Rightarrow$ threshold by quantile

Danger : choice of the percentage : indicate the rejection value

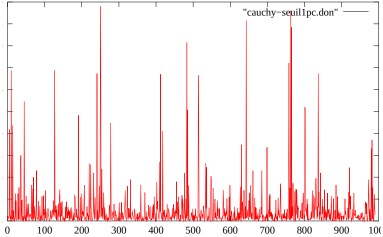
explain why

Control of experiments (4)

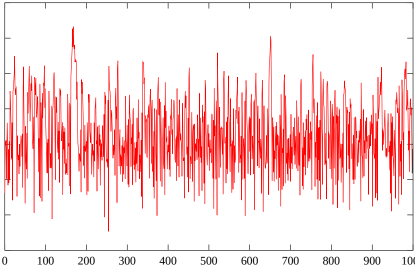
Threshold value : 10



Threshold percentage : 1%

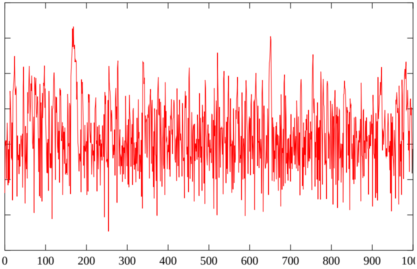


Control of experiments (5)



looks like correct experiments
Statistically independent
Statistically homogeneous

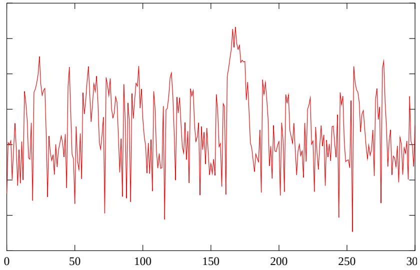
Control of experiments (5)



looks like correct experiments
Statistically independent
Statistically homogeneous

Control of experiments (5bis)

Zooming

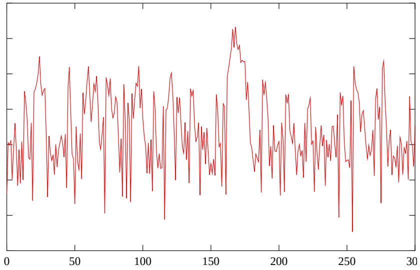


Autocorrelation

Danger time correlation among samples
experiments impact on experiments
$\Rightarrow$ stationarity analysis
autocorrelation estimation (ARMA)

Control of experiments (5bis)

Zooming



Autocorrelation

Danger time correlation among samples

experiments impact on experiments

⇒ stationarity analysis

autocorrelation estimation (ARMA)

Outline

Experimental results

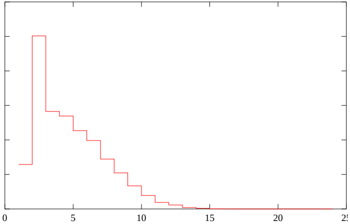
- Deterministic (controlled error non significant (white noise))
- Statistic (the system is non deterministic)

Sample analysis

- Identification of the response set
- Structure of the response set (measure)

Distribution analysis

Summarize data in a **histogram**

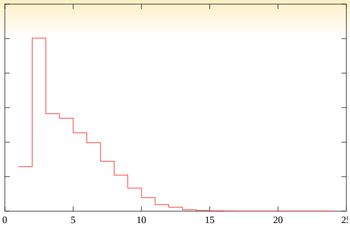


Shape analysis

- unimodal / multimodal
- variability
- symmetric / dissymmetric (skewness)
- flatness (kurtosis)

$\Longrightarrow$ **Central tendency analysis**
$\Longrightarrow$ **Variability analysis around the central tendency**

Mode value



Mode

- **Categorical data**
- Most frequent value
- highly unstable value
- for continuous value distribution depends on the histogram step
- interpretation depends on the flatness of the histogram

$\Longrightarrow$ **Use it carefully**
$\Longrightarrow$ **Predictor function**

Median value

Median

- **Ordered data**
- Split the sample in two equal parts

$$\sum_{i \leqslant Median} f_i \leqslant \frac{1}{2} \leqslant \sum_{i \leqslant Median+1} f_i.$$

- more stable value
- does not depends on the histogram step
- difficult to combine (two samples)

$\Longrightarrow$ **Randomized algorithms**

Mean value

Mean

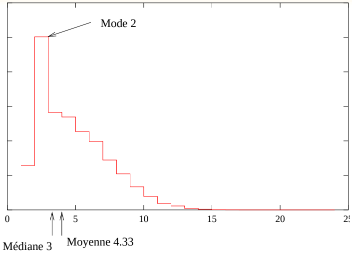
- **Vector space**
- Average of values

$$Mean = \frac{1}{Sample\_Size}\sum x_i = \sum_x x.f_x.$$

- stable value
- does not depends on the histogram step
- easy to combine (two samples $\Rightarrow$ weighted mean)

$\Longrightarrow$ **Additive problems (cost, durations, length,...)**

Central tendency



Complementarity

- Valid if the sample is "Well-formed"
- Semantic of the observation
- Goal of analysis

$\Longrightarrow$ **Additive problems (cost, durations, length,...)**

Central tendency (2)

Summary of Means

- Avoid means if possible
  Loses information
- Arithmetic mean
  When sum of raw values has physical meaning
  Use for summarizing times (not rates)
- Harmonic mean
  Use for summarizing rates (not times)
- Geometric mean
  Not useful when time is best measure of perf
  Useful when multiplicative effects are in play

Computational aspects

- Mode : computation of the histogram steps, then computation of max $O(n)$ "off-line"
- Median : sort the sample $O(nlog(n))$ or $O(n)$ (subtile algorithm) "off-line"
- Mean : sum values $O(n)$ "on-line" computation

Is the central tendency significant ?
$\Rightarrow$ Explain variability.

Computational aspects

- Mode : computation of the histogram steps, then computation of max $O(n)$ “off-line”
- Median : sort the sample $O(nlog(n))$ or $O(n)$ (subtile algorithm) “off-line”
- Mean : sum values $O(n)$ “on-line” computation

Is the central tendency significant ?
$\Rightarrow$ Explain variability.

Variability

Categorical data (finite set)

f_i : empirical frequency of element i
Empirical entropy

$$H(f) = \sum_i f_i \log f_i.$$

Measure the empirical distance with the uniform distribution

- $H(f) \geqslant 0$
- $H(f) = 0$ iff the observations are reduced to a unique value
- $H(f)$ is maximal for the uniform distribution

Variability (2)

Ordered data

Quantiles : quartiles, deciles, etc
Sort the sample :

$$(x_1, x_2, \cdots, x_n) \longrightarrow (x_{(1)}, x_{(2)}, \cdots, x_{(n)});$$

$$Q_1 = x_{(n/4)};\ Q_2 = x_{(n/2)} = Median;\ Q_3 = x_{(3n/4)}.$$

For deciles

$$d_i = argmax_i\{\sum_{j \leqslant i} f_j \leqslant \frac{i}{10}\}.$$

Utilization as quantile/quantile plots to compare distributions

Variability (3)

Vectorial data

Quadratic error for the mean

$$Var(X) = \frac{1}{n}\sum_{1}^{n}(x_i - \bar{x}_n)^2.$$

Properties:

$$\begin{aligned}
Var(X) &\geqslant 0; \\
Var(X) &= \overline{x^2} - (\bar{x})^2, \text{ où } \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2. \\
Var(X + cste) &= Var(X); \\
Var(\lambda X) &= \lambda^2 Var(X).
\end{aligned}$$

A simple example

Maximum value

int maximum (int * T, int n)
{*T array of distinct integers,*}
{n Size of T}
{
int max,i;
max= int_minimal_value;
for (i=0; i $<$ n; i++) **do**
  if (T[i] $>$ max)
  {
  max = T[i];
  Process(max); {Cost of the algorithm}
  }
end for
return(max)
}

Cost of the algorithm

Number of calls to **Process**

- minimum : 1
  example : T=[n,1,2,...,n-1]
  min cases : $(n-1)!$
- maximum : n
  example : T=[1,2,...,n]
  max case : 1

Bounded by a linear function $\mathcal{O}(n)$

But on average ?

A simple example

Maximum value

int maximum (int * T, int n)
{*T array of distinct integers,*}
{n Size of T}
{
int max,i;
max= int_minimal_value;
for (i=0; i $<$ n; i++) **do**
  if (T[i] $>$ max)
  {
  max = T[i];
  Process(max); {Cost of the algorithm}
  }
end for
return(max)
}

Cost of the algorithm

Number of calls to **Process**

- minimum : 1
  example : T=[n,1,2,...,n-1]
  min cases : $(n-1)!$
- maximum : n
  example : T=[1,2,...,n]
  max case : 1

Bounded by a linear function $\mathcal{O}(n)$

But on average ?

A simple example

Maximum value

int maximum (int * T, int n)
{*T array of distinct integers,*}
{n Size of T}
{
int max,i;
max= int_minimal_value;
for (i=0; i $<$ n; i++) **do**
  if (T[i] $>$ max)
  {
  max = T[i];
  Process(max); {Cost of the algorithm}
  }
end for
return(max)
}

Cost of the algorithm

Number of calls to **Process**

- minimum : 1
  example : T=[n,1,2,...,n-1]
  min cases : $(n-1)!$
- maximum : n
  example : T=[1,2,...,n]
  max case : 1

Bounded by a linear function $\mathcal{O}(n)$

But on average ?

A simple example (2)

Theoretical complexity

On average the complexity of the algorithm is : ....

Build the program

Put probes on the program

Questions :

1. Given $n = 1000$ does the observed cost follows the theoretical value ?
2. Does the average cost follows the theoretical complexity for all n ?
3. Does the average execution time linearly depends on the average cost ?

Modelling

Basic assumptions :

- Data are considered as random variables
- Mutually independent
- Same probability distribution

Check Check Check

The distribution is given by

- Probability density function (pdf) (asymptotic histogram)

$$f_X(x) = \mathbb{P}(x \leqslant X \leqslant x + dx)/dx = F'_X(x).$$

- Cumulative distribution function

$$F_X(x) = \mathbb{P}(X \leqslant x);$$

- Moments : $M_n = \mathbb{E}X^n$, Variance

Modelling

Basic assumptions :

- Data are considered as random variables
- Mutually independent
- Same probability distribution

Check Check Check

The distribution is given by

- Probability density function (pdf) (asymptotic histogram)

$$f_X(x) = \mathbb{P}(x \leqslant X \leqslant x + dx)/dx = F_X'(x).$$

- Cumulative distribution function

$$F_X(x) = \mathbb{P}(X \leqslant x);$$

- Moments : $M_n = \mathbb{E}X^n$, Variance

Average convergence

Law of large numbers

Let $\{X_n\}_{n \in \mathbb{N}}$ be a iid random sequence with finite variance, then

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mathbb{E}X, \quad \text{almost surely an in } L^1.$$

$\rightarrow$ convergence of empirical frequencies
$\rightarrow$ for any experience we get the same result
$\rightarrow$ fundamental theorem of probability theory

$$\text{Notation : } \overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

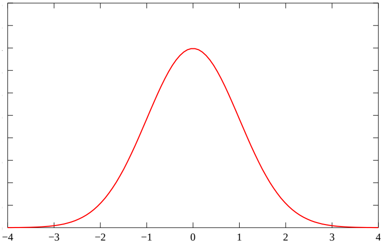
Law of errors

Central limit theorem (CLT

Let $\{X_n\}_{n\in\mathbb{N}}$ be a iid random sequence with finite variance σ^2, then

$$\lim_{n\to+\infty} \frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - \mathbb{E}X\right) \stackrel{\mathcal{L}}{=} \mathcal{N}(0,1).$$

$\to$ error law (Gaussian law, Normal distribution, Bell curve,...)
$\to$ Normalized mean = 0, variance = 1



Distribution

$$\mathbb{P}(X \in [-1,1]) = 0.68;$$

$$\mathbb{P}(X \in [-2,2]) = 0.95;$$

$$\mathbb{P}(X \in [-3,3]) \geqslant 0.99.$$

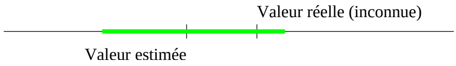
Confidence intervals

Confidence level α compute ϕ_α

$$\mathbb{P}(X \in [-\phi_\alpha, \phi_\alpha]) = \alpha$$

For n sufficiently large ($n > 50$)

$$\mathbb{P}\left([\overline{X}_n - \frac{\phi_\alpha \sigma}{\sqrt{n}}, \overline{X}_n + \frac{\phi_\alpha \sigma}{\sqrt{n}}] \ni \mathbb{E}X\right) = \alpha.$$



Confidence intervals (2)

Need an estimator of the variance

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2.$$

Danger n too small $\rightarrow$ with a normal hypothesis take Student statistic
Three step method

1. In a first set of experiments check that the hypothesis is valid
2. Estimate roughly the variance
3. Estimate the mean and control the number of experiment by a confidence interval

Outline

Architecture comparison

Performance characterization

Distributed protocol (consensus)

- List of benchmarks (with some parameters)
- Several types of architecture

Problem: decide which architecture is the best one

Comparison of results

Decision problem

Two hypothesis :
- $\mathcal{H}_0$: (null hypothesis) A is equivalent to B
- $\mathcal{H}_1$: (alternative hypothesis) A is better than B

Decision error:
type 1 error : reject $\mathcal{H}_0$ when $\mathcal{H}_0$ is true
type 2 error : accept $\mathcal{H}_0$ when $\mathcal{H}_1$ is true.

According the observation find the decision function minimizing some risk criteria
Rejection region : if $(x_1, \cdots, x_n) \in C$ reject H_0

Danger : errors are not symmetric

Testing Normal Distributed Variables

Observations : $\mathcal{N}(m_0, \sigma_0^2))$ under hypothesis $\mathcal{H}_0$ and $\mathcal{N}(m_1, \sigma_1^2))$ under hypothesis $\mathcal{H}_1$ with $m_1 > m_0$

$$\text{Rejection region } C = \left\{ \frac{1}{n}(x_1 + \cdots + x_n) \geqslant K \right\}.$$

Computation of the rejection region type 1 error : choose α

$$\begin{aligned}
\alpha &= \mathbb{P}_{\mathcal{H}_0}(\frac{1}{n}(X_1 + \cdots + X_n) \geqslant K_\alpha) \\
&= \mathbb{P}_{\mathcal{H}_0}\left((\frac{\sqrt{n}}{\sigma}(\frac{1}{n}(X_1 + \cdots + X_n) - m_0) \geqslant \frac{\sqrt{n}}{\sigma}(K_\alpha - m_0) \right) \\
&= \mathbb{P}(Y \geqslant \frac{\sqrt{n}}{\sigma}(K_\alpha - m_0)) \text{ with } Y \sim \mathcal{N}(0, 1).
\end{aligned}$$

$$\Phi_\alpha = \frac{\sqrt{n}}{\sigma}(K_\alpha - m_0) \text{ then } K_\alpha = m_0 + \frac{\sigma}{\sqrt{n}}\Phi_\alpha.$$

Numerical example

- $\alpha = 0.05$ (a priori confidence)
  $\Phi_\alpha = 1.64$ (read on the table of the Normal distribution)
- Under $\mathcal{H}_0$, $m_0 = 6$ and $\sigma_0 = 2$
  Sample size $n = 100$

$$K_\alpha = 6 + \frac{2}{10}1.64 = 6.33.$$

If $\frac{1}{n}(x_1 + \cdots + x_n) \geqslant 6.33$ reject $\mathcal{H}_0$ (accept $\mathcal{H}_1$), else accept $\mathcal{H}_0$

Type 2 error: Depends on the alternative hypothesis

- $m_1 = m'$ (known) σ_1 known

$$\beta = \mathbb{P}_{\mathcal{H}_1}(\frac{1}{n}(X_1 + \cdots + X_n) \leqslant K_\alpha) = \mathbb{P}(Y \leqslant \frac{\sqrt{n}}{\sigma_1}(K_\alpha - m_1)).$$

- $m_1 > m_0$ or $m_1 \neq m_0$: cannot compute

Numerical example

- $\alpha = 0.05$ (a priori confidence)
  $\Phi_\alpha = 1.64$ (read on the table of the Normal distribution)
- Under $\mathcal{H}_0$, $m_0 = 6$ and $\sigma_0 = 2$
  Sample size $n = 100$

$$K_\alpha = 6 + \frac{2}{10} 1.64 = 6.33.$$

If $\frac{1}{n}(x_1 + \cdots + x_n) \geqslant 6.33$ reject $\mathcal{H}_0$ (accept $\mathcal{H}_1$), else accept $\mathcal{H}_0$

Type 2 error: Depends on the alternative hypothesis

- $m_1 = m'$ (known) σ_1 known

$$\beta = \mathbb{P}_{\mathcal{H}_1}(\frac{1}{n}(X_1 + \cdots + X_n) \leqslant K_\alpha) = \mathbb{P}(Y \leqslant \frac{\sqrt{n}}{\sigma_1}(K_\alpha - m_1)).$$

- $m_1 > m_0$ or $m_1 \neq m_0$: cannot compute

Application example (1)

Test if algorithm 1 is better than algorithm 0

- Generate n random inputs $i_1, \cdots, i_n$
- Compute $A_0(i_k)$ $A_1(i_k)$
- $x_k = A_1(i_k) - A_0(i_k)$
- Reject the hypothesis $m = 0$ if $\frac{1}{n}(x_1 + \cdots + x_n) \geqslant K_\alpha$

Application example (2)

Test if system 1 is better than system 0

- Generate n_0 random inputs $i_1, \cdots, i_{n_0}$
- Compute $S_0(i_k)$
- Generate n_1 random inputs $i_1, \cdots, i_{n_1}$
- Compute$S_1(i_k)$
- Compute the mean difference
- Compute the standard deviation of the difference
- Reject the hypothesis $m = 0$ if $\bar{x}_1 - \bar{x}_0 \geqslant K_\alpha$

Outline

Experiment with one factor

Evaluate complexity as a function of the size of data
Response time as function of the message sizes
Load of a web server function of the number of connexion
etc

Observations

Couple (x, y) paired observations

- x predictor variable (known without error or noise)
- y response variable

Methodology

1. Plot data and analyse separately x and y (histogram, central tendency,...)
2. Plot the cloud of points (x, y)
3. Analyse the shape of the cloud
4. Propose a dependence function (fix the parameters $y = ax + b$, $y = be^{ax}$,...)
5. Give the semantic of the function
6. Give an error criteria with its semantic
7. Compute the parameters minimizing a criteria
8. Compute the confidence intervals on parameters (precision of the prediction)
9. Explain the unpredicted variance (ANOVA)
10. Analyse the result

Linear regression

Theoretical model

(X, Y) follows a correlation model

$$Y = \alpha X + \beta + \epsilon;$$

with ϵ a white noise $\epsilon \sim \mathcal{N}(0, .)$

Objective function

Find estimator $(\hat{a}, \hat{b})$ minimizing the SSE (sum of square errors)

$$\sum_{i=1}^{n}(y_i - ax_i - b)^2 = \sum_{i=1}^{n} e_i^2.$$

$e_i = y_i - ax_i - b$ is the error prediction when the coefficients are a and b
$(\hat{a}, \hat{b})$ is the estimator of (α, β) minimizing SSE

Coefficients estimation

Statistics

- Empirical mean of x: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
- Empirical mean of y: $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.
- Empirical variance of x: $S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \overline{x^2} - \overline{x}^2$.
- Empirical variance of y: $S_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \overline{y^2} - \overline{y}^2$.
- Empirical Covariance of (x, y): $S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \overline{x \cdot y} - \overline{x} \cdot \overline{y}$.

Estimators

$$y_i = \frac{S_{XY}}{S_X^2}(x_i - \overline{x}) + \overline{y}$$

$$\hat{a} = \frac{S_{XY}}{S_X^2} \text{ and } \hat{b} = \overline{y} - \frac{\overline{x} \cdot S_{XY}}{S_X^2} = \overline{y} - \hat{a} \cdot \overline{x}$$

Coefficients estimation

Statistics

- Empirical mean of x: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
- Empirical mean of y: $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.
- Empirical variance of x: $S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \overline{x^2} - \overline{x}^2$.
- Empirical variance of y: $S_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \overline{y^2} - \overline{y}^2$.
- Empirical Covariance of (x, y): $S_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \overline{x \cdot y} - \overline{x} \cdot \overline{y}$.

Estimators

$$y_i = \frac{S_{XY}}{S_X^2}(x_i - \overline{x}) + \overline{y}$$

$$\hat{a} = \frac{S_{XY}}{S_X^2} \text{ and } \hat{b} = \overline{y} - \frac{\overline{x} \cdot S_{XY}}{S_X^2} = \overline{y} - \hat{a} \cdot \overline{x}$$

Error analysis

Total error :

$$SST = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2 = SSY - SS0.$$

Prediction error:

$$SSE = \sum_{i=1}^{n}(y_i - \hat{a}x_i - \hat{b})^2 = n(\overline{y^2} - \hat{b}\overline{y} - \hat{a}\overline{x \cdot y}$$

Residual error (that has not been predicted): $SSR = SST - SSE$
Determination coefficient:

$$R^2 = \frac{SSR}{SST}$$

Prediction quality

- $R^2 = 1$ perfect fit
- $R^2 = 0$ no fit

Usually we accept the model when $R^2 \geqslant 0.8$

Planning experiments

- One factor :
  $\Rightarrow$ estimate residuals
- Check homoskedasticity of data (homogeneous variance)
- Explain trends
- Replicate sample with x to reduce variance

Optimize the experiment such that for each estimation we get the same variance

Outline

Time dimensioning problems

Time out estimation

Distributed protocol (consensus)

- Crash of processes
- Variable communications (wireless network)
- Failure detection mechanism (parametrized)

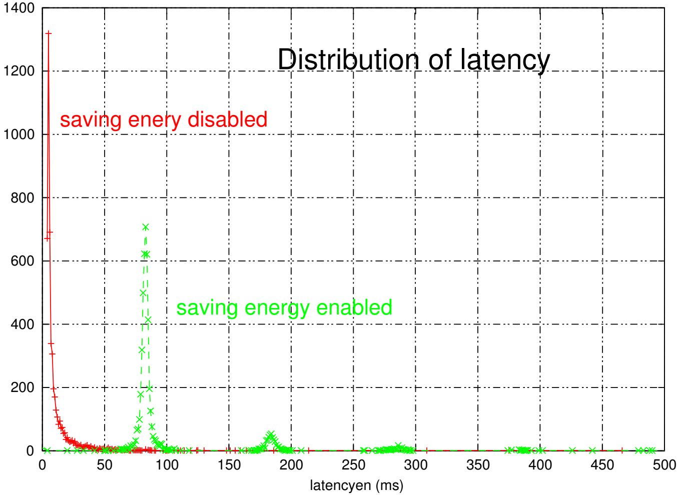
Factors

- Crash of processes
- Variable communications (wireless network)
- Failure detection mechanism (parametrized)

$\Rightarrow$ **Evaluation of the latency**

Latency estimation

PDA → PDA communication (ping)

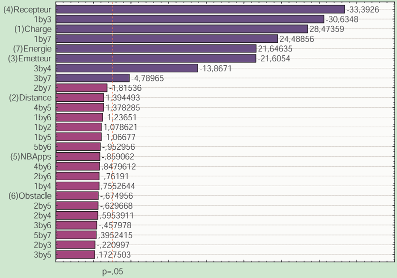


Factors Analysis

Factors (a priori)

- Distance
- Number of obstacles
- Number of nodes
- Network load
- Sender type
- Receiver type
- Saving energy

Tagushi analysis

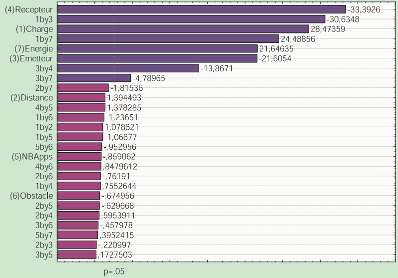


Factors Analysis

Significant factors

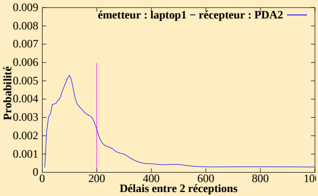
- Distance
- Number of obstacles
- Number of nodes
- **Network load (2)**
- **Sender type (4)**
- **Receiver type (1)**
- **Saving energy (3)**

Tagushi analysis

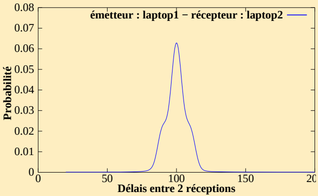


Time out estimation

Laptop → PDA



Laptop → laptop



Outline

Trace analysis example

Presentation of the paper available on `http://fta.inria.fr`

Mining for Statistical Models of Availability in Large-Scale Distributed Systems: An Empirical Study of SETI@home

Bahman Javadi[1], Derrick Kondo[1], Jean-Marc Vincent[1,2], David P. Anderson[3]

[1] Laboratoire d'Informatique de Grenoble, MESCAL team, INRIA, France
[2] University of Joseph Fourier, France
[3] UC Berkeley, USA

IEEE/ACM International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS 2009)

INRIA

Outline

Synthesis : principles

1. Formulate the **hypothesis**
2. Design the experiment to **validate** the hypothesis
3. Check the validity of the experience
4. Analyse the experiments to validate or invalidate the hypothesis
5. Report the arguments in a convincing form

Synthesis : Steps for a Performance Evaluation Study [Jain]

1. State the goals of the study and define system boundaries.
2. List system services and possible outcomes.
3. Select performance metrics.
4. List system and workload parameters
5. Select factors and their values.
6. Select evaluation techniques.
7. Select the workload.
8. Design the experiments.
9. Analyze and interpret the data.
10. Present the results. Start over, if necessary.

Common mistakes in experimentation [Jain]

1. The variation due to experimental error is ignored
2. Important parameters are not controlled
3. Simple one-factor-at-a-time designs are used
4. Interactions are ignored
5. Too many experiments are conducted

References

Bibliography

- **The Art of Computer Systems Performance Analysis : Techniques for Experimental Design, Measurment, Simulation and Modeling.** Raj Jain *Wiley 1991* http://www.rajjain.com/
- **Measuring Computer Performance: A Practitioner's Guide** David J. Lilja Cambridge University Press, 2000.
- **Performance Evaluation of Computer and Communication Systems** Jean-Yves Le Boudec EPFL http://perfeval.epfl.ch/lectureNotes.htm

Common tools

- Mathlab, Matematica
- Scilab http://www.scilab.org/
- gnuplot http://www.gnuplot.info/
- R http://www.r-project.org/