RENAULT

The Gamma Test

Octobre 2004

The Gamma Test:

- What is it ?
- What can we do with it ?
- Some examples

The Gamma Test is an estimator of the modelisation noise

The data set: $(x^i_1, \ldots, x^i_M, y^i)$

The model: $y^i = f(x^i) + r$

The Gamma Test estimates the value of Var(r)

With some hypothesis on f:
$\exists$ A, B
$|\nabla f(x)|<A$
$|Hf(x)|<B$

Neighborhood defined using a KD-Tree

For p in [1, …, n] compute:

$$\Delta(p)=\frac{1}{M}\sum_{i=1}^{M}|x(N(i,p))-x(i)|^2$$

$$\Gamma(p)=\frac{1}{2M}\sum_{i=1}^{M}(y(N(i,p))-y(i))^2$$

Where:

- M is the number of points in the data set;
- N(i,p) is the pth nearest neighbor of point I (localised using kd-tree;
- n is a parameter defined by the user

Next, compute the regression line of these points:

$$\Gamma = A\,\Delta + B$$

$$B \simeq Var(r)$$

$$A \simeq \frac{1}{4}\langle|\nabla f|^2\rangle$$

Model learning:

- Using the value of Var(r) to stop the learning before over training

Model sizing:

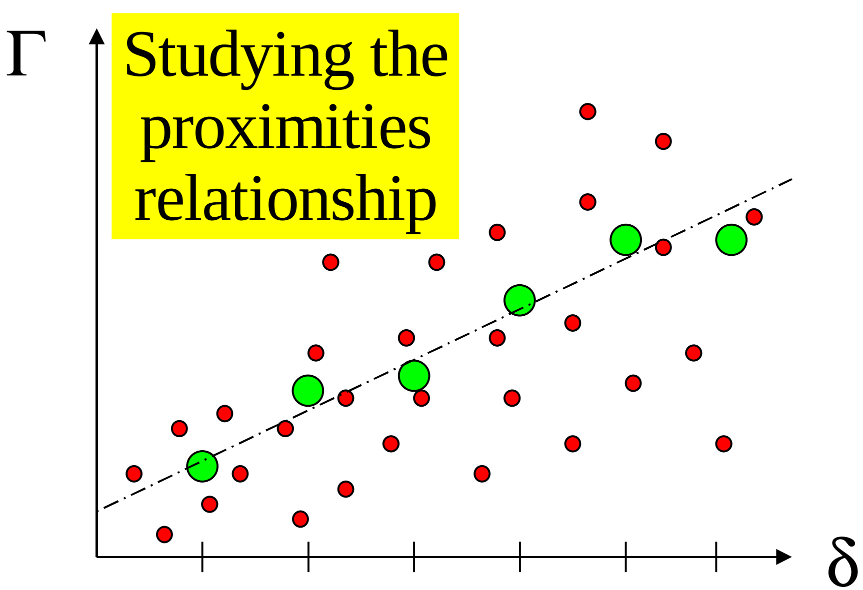
- Using the value of A to pre-determine the size of the model (see A. Pui Man Tsui and Koncar)

Inputs selection:

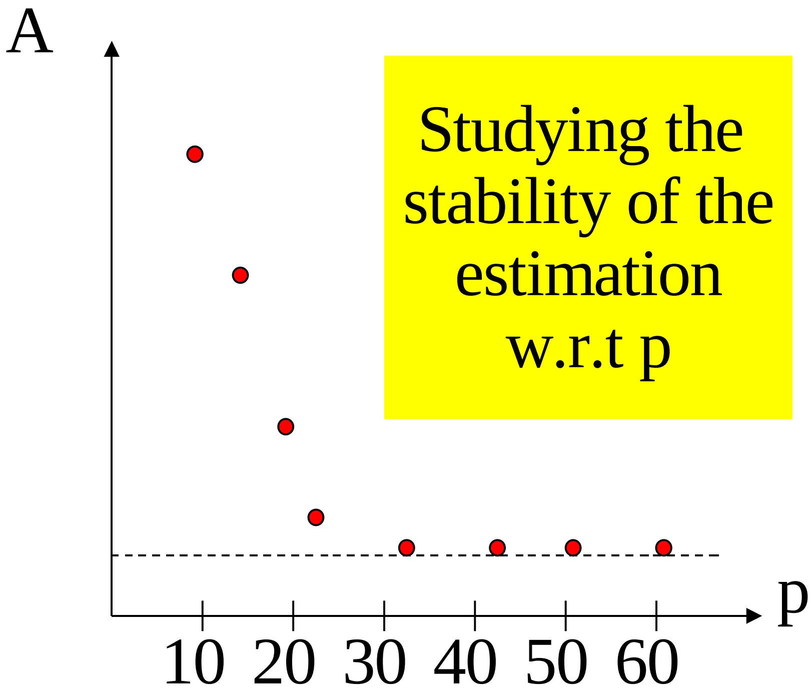
- What are the inputs that minimize the modelisation error

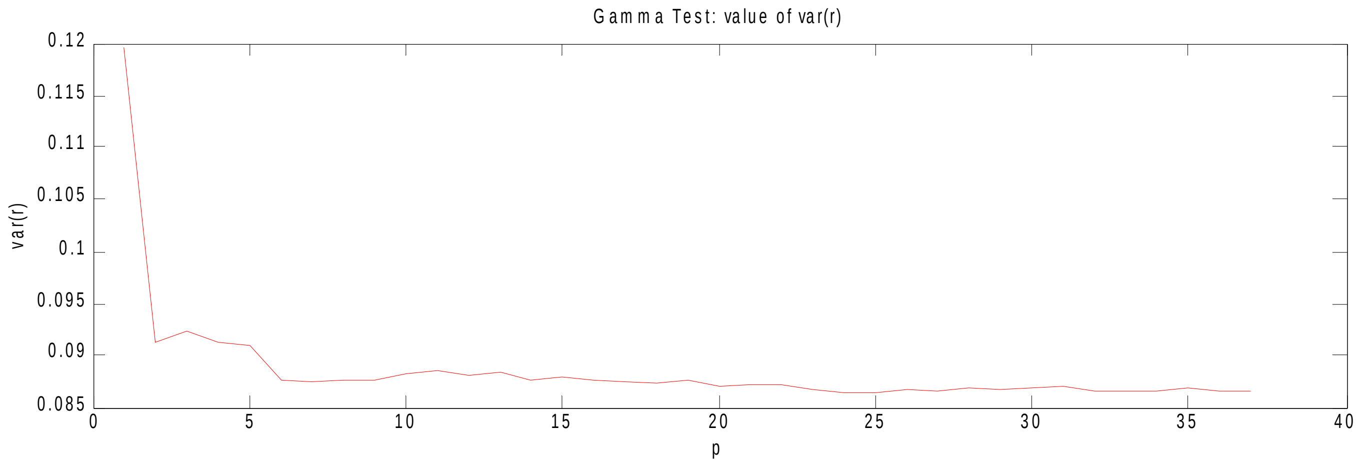
The Scatter Plot:

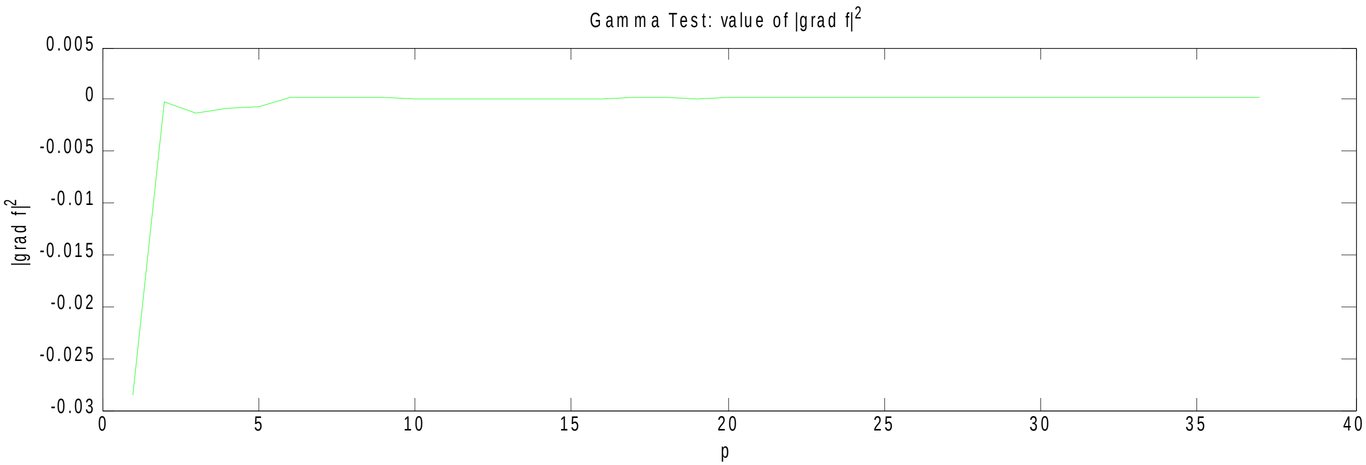
Plot $\left(|x(N(i,p))-x(i)|^2, 0.5(y(N(i,p))-y(i))^2\right)$ in the (Γ,Δ) plane for each i and p



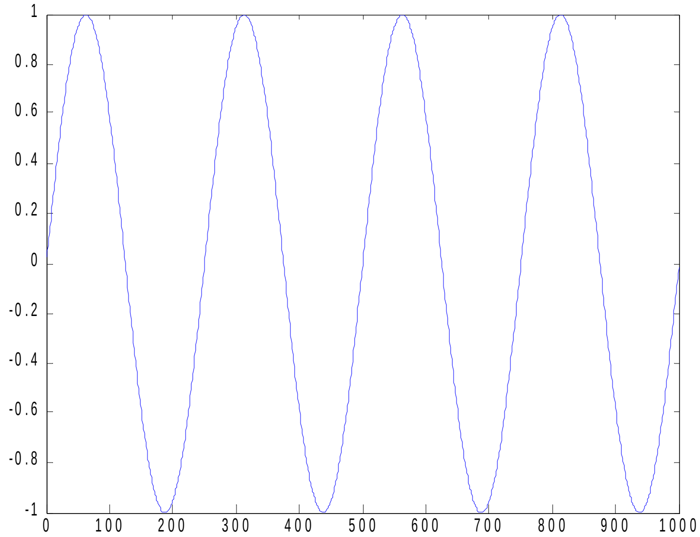
The M-test:
Plot the value of A and B
For various values of p



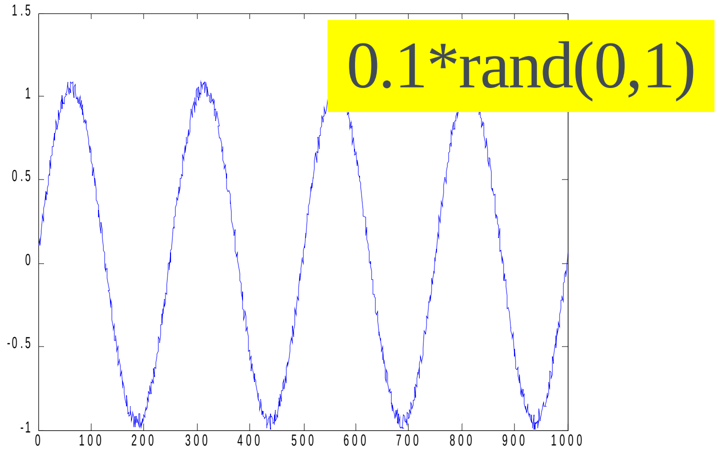

Gamma Test: value of var(r)
var(r)
0.12
0.115
0.11
0.105
0.1
0.095
0.09
0.085
0
5
10
15
20
25
30
35
40
p

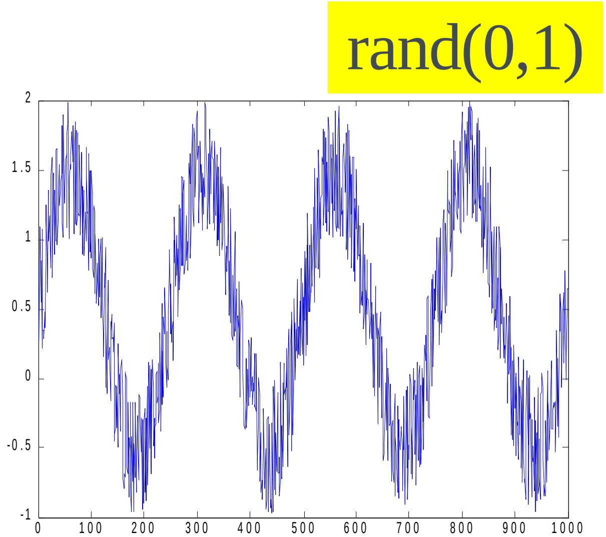

Gamma Test: value of |grad f|^2
|grad f|^2
0.005
0
-0.005
-0.01
-0.015
-0.02
-0.025
-0.03
0
5
10
15
20
25
30
35
40
p

Sine wave without noise

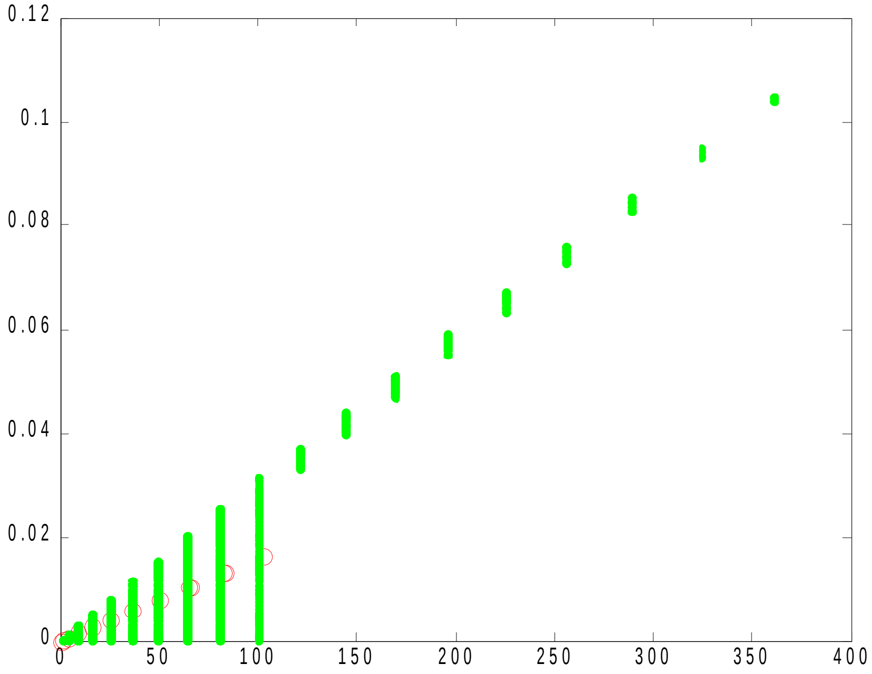


Sine wave with noise



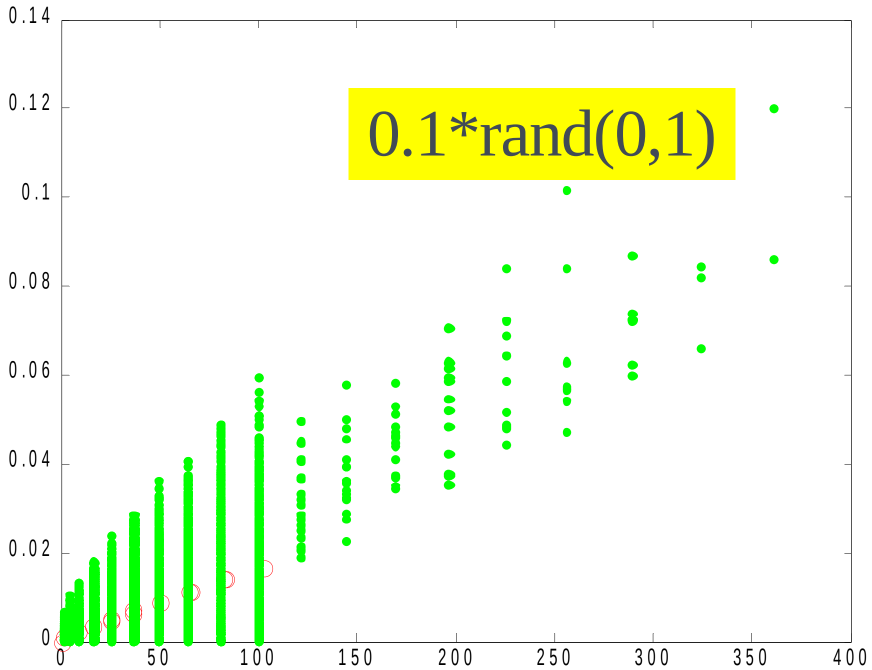


Sine wave without noise



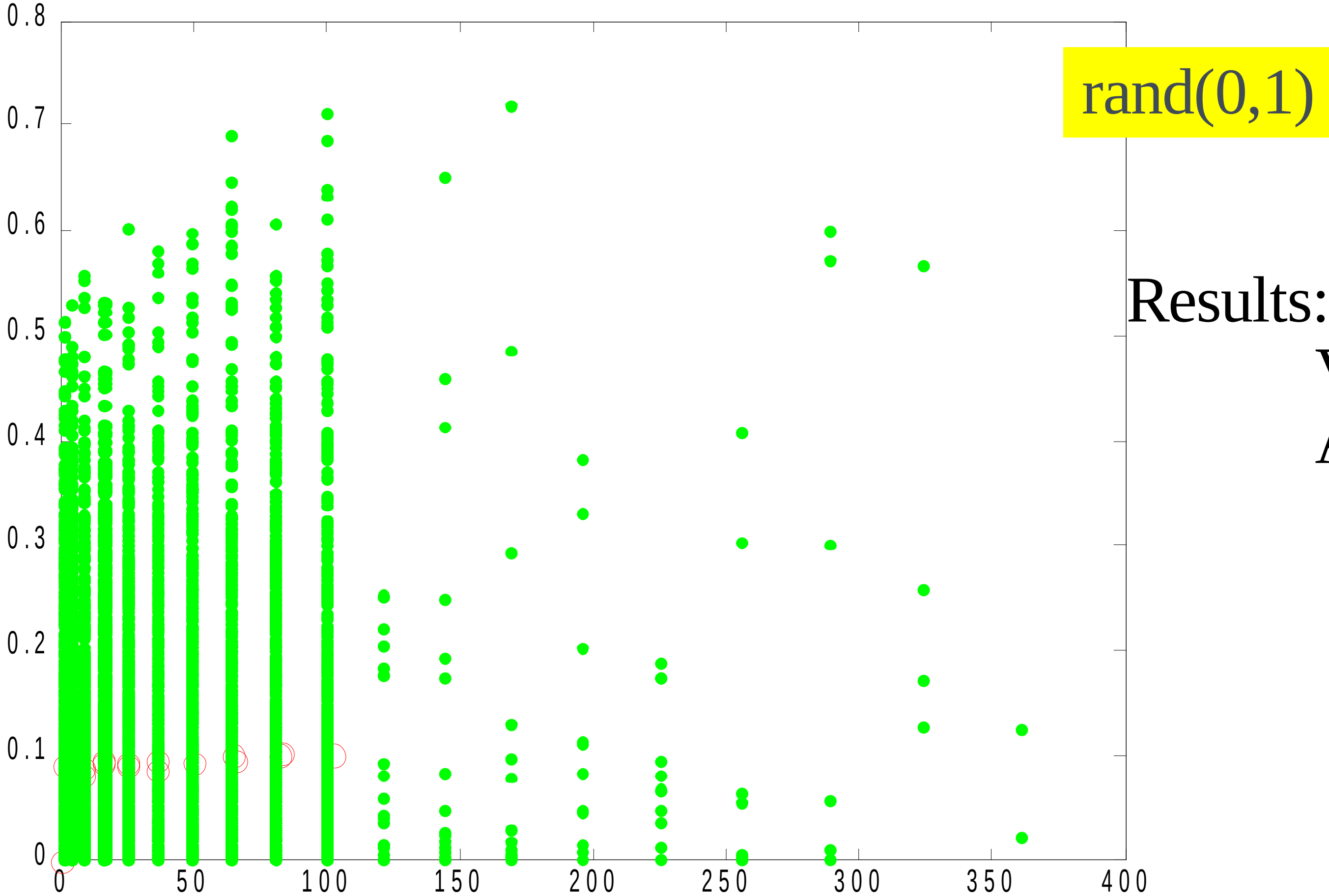
Results: Estim 3.95E-6
Value 0.0
A = 0.00015

Sine wave with noise



Results: Estim 0.00083
Value 0.00084
A = 0.00015

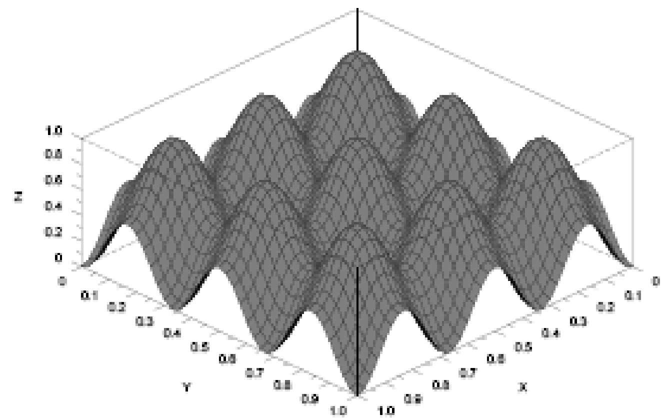
Sine wave with noise



Results: Estim 0.0086
Value 0.0084
A = 0.00012

Dimensionnement d'un Réseau de Neurones

$$f(x_1,\ldots,x_n)=\frac{1}{n}\sum_{i=1}^{n}\sin^2(a\pi x_i)$$



Calcul analytique d'un paramètre estimé par le Gamma Test

$$\langle|\nabla f|^2\rangle=\frac{a^2\pi^2}{2n}\left(1-\frac{\sin(4a\pi)}{4a\pi}\right)$$

Estimation donnée par le Gamma Test

$$A\simeq 1/4\langle|\nabla f|^2\rangle$$

Nombre estimé de maxima locaux

$$h=a^n=\left[\left(\frac{2\sqrt{2}}{\pi}\right)^n(nA)^{\frac{n}{2}}\right]$$

Nombre de neurones

$$n_{\text{neurones}}=(2n+1)h+1$$

Conclusions

Outil récent d'estimation non paramétrique de bruit
Nécessite un certain nombre de points dans la base de données
Permet de se passer d'une base de validation
Donne une indication sur le dimensionnement d'un réseau

Bruit effectif: bruit total sur la sortie (prend un compte un niveau de bruit sur les entrées en le reportant sur la sortie)
Conclusion d'une mesure du bruit effectif via le GammaTest: la présence de bruit sur les entrées déforme la relation entrée / sortie

La présence seul de bruit sur la sortie ne déforme pas cette relation entrée / sortie