How to specify ann approximate numerical result

Nicolas Bouleau

How to specify an approximate numerical result

Nicolas Bouleau

—

classification : 60Hxx, 31C25, 94B70, 65Gxx.

Abstract

The Dirichlet forms methods, in order to represent errors and their propagation, are particularly powerful in infinite dimensional problems such as models involving stochastic analysis encountered in finance or physics, cf. [5]. Now, coming back to the finite dimensional case, these methods give a new light on the very classical concept of 'numerical approximation' and suggest changes in the habits. We show that for some kinds of approximations only an Ito-like second order differential calculus is relevant to describe and propagate numerical errors through a mathematical model. We call these situations *strongly stochastic*. The main point of this work is an argument based on the *arbitrary functions principle* of Poincaré-Hopf showing that the errors due to measurements with graduated instruments are strongly stochastic. Eventually we discuss the consequences of this phenomenon on the specification of an approximate numerical result.

1 The dichotomy of small errors.

Let us begin by showing that there are two kinds of small errors which do not propagate according to the same differential calculus.

Suppose two applied mathematicians A and B attempt to perform stochastic simulation rigourously. By means of the well known inversion and rejection methods, they are able to simulate any probability law provided that they can pick up a real number in the unit interval $[0, 1]$ randomly.

For this, the researcher A chooses the method of drawing the binary digits by heads or tails. The researcher B, instead, prefers using a Polya's urn.

Let us compare the biases and the variances of the errors in the two procedures.

In the case A, the real number

$$x = 0, a_1 a_2 a_3 \cdots = \sum_{k=1}^{\infty} \frac{a_k}{2^k} \qquad a_k \in \{0, 1\}$$

is approximated by $x_n = \sum_{k=1}^{n} \frac{a_k}{2^k}$.

Denoting $\mathcal{F}_n$ the σ-field generated by $a_1, \cdots, a_n$, the bias of the error is

$$b_n = \mathbb{E}[(x - x_n)|\mathcal{F}_n] = \sum_{k=n+1}^{\infty} \frac{1/2}{2^k} = \frac{1}{2^{n+1}}$$

and the variance of the error is

$$v_n = \mathbb{E}[(x - x_n)^2|\mathcal{F}_n] - (\mathbb{E}[(x - x_n)|\mathcal{F}_n])^2 = \frac{1}{3}\frac{1}{4^n} - \frac{1}{4}\frac{1}{4^n} = \frac{1}{12}\frac{1}{4^n}.$$

For the case B, let us recall the principle of Polya's urn : there is at the beginning a white ball and a black ball in the urn and each time a ball is drawn from the urn, it is put back into the urn together with an other ball of the same colour.

After n drawings, the proportion X_n of white balls in the urn is given by

$$(n + 2)X_n = (n + 1)X_{n-1} + 1_{\{U_n \leq X_{n-1}\}}$$

where U_n is uniformly distributed on $[0, 1]$ independent of $\mathcal{F}_{n-1} = \sigma(X_0, \ldots, X_{n-1})$. In other words

$$X_n = X_{n-1} + \frac{1}{n + 2}(1_{\{U_n \leq X_{n-1}\}} - X_{n-1}).$$

X_n is a bounded martingale which converges a.s. and in L^p, $p \in [1, \infty[$, to a random variable X_∞ uniformly distributed on $[0, 1]$ as easily seen when the initial configuration of the urn is one white ball and one black ball.

For the bias we have

$$b_n = \mathbb{E}[X_\infty - X_n|\mathcal{F}_n] = 0$$

and for the variance

$$v_n = \mathbb{E}[(X_\infty - X_n)^2 | \mathcal{F}_n]$$

we have $\mathbb{E}[v_n] = \frac{1}{6n} + o(1/n)$.

We see that in case A the variances are smaller than the biases, while in case B the biases are smaller than the variances.

How will these errors propagate through the simulations of our two Monte Carlo practioners ?

A Taylor expansion on a $\mathcal{C}^3$-function with bounded derivatives gives

$$\begin{aligned} f(X) - f(X_n) = (X - X_n) f'(X_n) \quad & + \tfrac{1}{2}(X - X_n)^2 f''(X_n) \\ & + \tfrac{1}{6}(X - X_n)^3 f'''(X_n + \theta(X - X_n)) \end{aligned}$$

$$\text{new bias } = \mathbb{E}[f(X_\infty) - f(X_n)|\mathcal{F}_n] = b_n f'(X_n) + \frac{1}{2}(v_n + b_n^2) f''(X_n) + o(v_n)$$

$$\begin{aligned} \text{new variance } & = \mathbb{E}[(f(X_\infty) - f(X_n))^2 | \mathcal{F}_n] - (\mathbb{E}[f(X_\infty) - f(X_n)|\mathcal{F}_n])^2 \\ & = (v_n - b_n^2) f'^2(X_n) + o(v_n). \end{aligned}$$

We can distinguish three cases

1) If the variance is negligible with respect to the bias, $v_n \ll b_n$, (case of researcher A), the dominant term for the bias is asymptotically the first one. $\mathbb{E}[(f(X_\infty) - f(X_n))^2 | \mathcal{F}_n]$ is negligible with respect to $\mathbb{E}[f(X_\infty) - f(X_n)|\mathcal{F}_n]$ and the situation will be carried on. It is enough to use the formula

$$\mathbb{E}[f(X_\infty) - f(X_n)|\mathcal{F}_n] = b_n f'(X_n) + o(b_n) \tag{1}$$

2) If the variance is of the same order of magnitude as the bias, the situation will persist. The propagation formulae are

$$\left.\begin{aligned} \mathbb{E}[f(X_\infty) - f(X_n)|\mathcal{F}_n] &= b_n f'(X_n) + \tfrac{1}{2} v_n f''(X_n) + o(v_n) \\ \mathbb{E}[(f(X_\infty) - f(X_n))^2 | \mathcal{F}_n] &= v_n f'^2(X_n) + o(v_n) \end{aligned}\right\} \tag{2}$$

3) If the bias is negligible in comparison to the variance, $b_n \ll v_n$, (case of researcher B), the main term in the bias becomes $\frac{1}{2} v_n f''(X_n)$ and we fall in the case 2) where biases and variances remain of the same order of magnitude.

We see that a first order differential calculus is relevant for the researcher A. But instead, the researcher B (with Polya's urn) must perform an error calculus involving both biases and variances, and

- the error calculus on the variances is a first order differential calculus,

- the error calculus on the biases is a second order differential calculus and uses the calculus on variances.

The first case will be called the *weakly stochastic* case. Then the usual differential calculus is enough to propagate errors and to assess the sensitivity of the model outputs to data.

The second case will be called *strongly stochastic*. Then the propagation of biases (which is important in non-linear models) needs an Ito-like differential calculus given by formulae (1.2).

Comment. In practice, generally, we do not control the nature of the errors. In modelling, errors on the data are *exogenous*, we know few from where they come. It is wise to think according to the second case, especially to take in account the randomness of the errors through the non-linearities of the model.

Let us go deeper into the mathematical arguments by displaying the bias operators and the variance operator (the Dirichlet form) associated with an approximation.

2 The bias operators and the Dirichlet form associated with an approximation.

When two random variables Y and Y_n are close together, the asymptotic behaviour of

$$\mathbb{E}[(\phi(Y_n) - \phi(Y))\chi(Y)]$$

and of

$$\mathbb{E}[(\phi(Y_n) - \phi(Y))\chi(Y_n)]$$

where ϕ and χ are test functions, are generally different. As a consequence several bias operators have to be distinguished (cf. [6]) :

Let Y be a random variable defined on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a measurable space $(E, \mathcal{F})$ and let Y_n be approximations also defined on $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $(E, \mathcal{F})$. We consider an algebra $\mathcal{D}$ of bounded functions from E into $\mathbb{R}$ or $\mathbb{C}$ containing the constants and dense in $L^2(E, \mathcal{F}, \mathbb{P}_Y)$ and a sequence α_n of positive numbers. With $\mathcal{D}$ and (α_n) we consider the four

following assumptions defining the four bias operators

$$\text{(H1)} \qquad \begin{cases} \forall \varphi \in \mathcal{D}, \text{ there exists } \overline{A}[\varphi] \in L^2(E, \mathcal{F}, \mathbb{P}_Y) \quad s.t. \quad \forall \chi \in \mathcal{D} \\ \lim_{n\to\infty} \alpha_n \mathbb{E}[(\varphi(Y_n) - \varphi(Y))\chi(Y)] = \mathbb{E}_Y[\overline{A}[\varphi]\chi]. \end{cases}$$

$$\text{(H2)} \qquad \begin{cases} \forall \varphi \in \mathcal{D}, \text{ there exists } \underline{A}[\varphi] \in L^2(E, \mathcal{F}, \mathbb{P}_Y) \quad s.t. \quad \forall \chi \in \mathcal{D} \\ \lim_{n\to\infty} \alpha_n \mathbb{E}[(\varphi(Y) - \varphi(Y_n))\chi(Y_n)] = \mathbb{E}_Y[\underline{A}[\varphi]\chi]. \end{cases}$$

$$\text{(H3)} \qquad \begin{cases} \forall \varphi \in \mathcal{D}, \text{ there exists } \widetilde{A}[\varphi] \in L^2(E, \mathcal{F}, \mathbb{P}_Y) \quad s.t. \quad \forall \chi \in \mathcal{D} \\ \lim_{n\to\infty} \alpha_n \mathbb{E}[(\varphi(Y_n) - \varphi(Y))(\chi(Y_n) - \chi(Y))] = -2\mathbb{E}_Y[\widetilde{A}[\varphi]\chi]. \end{cases}$$

$$\text{(H4)} \qquad \begin{cases} \forall \varphi \in \mathcal{D}, \text{ there exists } \not{A}[\varphi] \in L^2(E, \mathcal{F}, \mathbb{P}_Y) \quad s.t. \quad \forall \chi \in \mathcal{D} \\ \lim_{n\to\infty} \alpha_n \mathbb{E}[(\varphi(Y_n) - \varphi(Y))(\chi(Y_n) + \chi(Y))] = 2\mathbb{E}_Y[\not{A}[\varphi]\chi]. \end{cases}$$

We first note that as soon as two of hypotheses (H1) (H2) (H3) (H4) are fulfilled (with the same algebra $\mathcal{D}$ and the same sequence α_n), the other two follow thanks to the relations

$$\widetilde{A} = \frac{\overline{A} + \underline{A}}{2} \qquad \not{A} = \frac{\overline{A} - \underline{A}}{2}.$$

When defined, the operator $\overline{A}$ which considers the asymptotic error from the point of view of the limit model, will be called *the theoretical bias operator*.

The operator $\underline{A}$ which considers the asymptotic error from the point of view of the approximating model will be called *the practical bias operator*.

Because of the property

$$< \widetilde{A}[\varphi], \chi >_{L^2(\mathbb{P}_Y)} = < \varphi, \widetilde{A}[\chi] >_{L^2(\mathbb{P}_Y)}$$

the operator $\widetilde{A}$ will be called *the symmetric bias operator*.

The operator $\not{A}$ which is often (see theorem 2.2 below) a first order operator will be called *the singular bias operator*.

Theorem 2.1 *Under the hypothesis* (H3)*,*

a) the limit

$$\widetilde{\mathcal{E}}[\varphi, \chi] = \lim_n \frac{\alpha_n}{2} \mathbb{E}[(\varphi(Y_n) - \varphi(Y))(\chi(Y_n) - \chi(Y)] \qquad \varphi, \chi \in \mathcal{D} \qquad (3)$$

defines a closable positive bilinear form whose smallest closed extension is denoted $(\mathcal{E}, \mathbb{D})$*.*

b) $(\mathcal{E},\mathbb{D})$ *is a Dirichlet form* (cf. [4])
c) $(\mathcal{E},\mathbb{D})$ *admits a square field operator* Γ *satisfying* $\forall\varphi,\chi\in\mathcal{D}$

$$\Gamma[\varphi]=\widetilde{A}[\varphi^2]-2\varphi\widetilde{A}[\varphi] \tag{4}$$

$$\mathbb{E}_Y[\Gamma[\varphi]\chi]=\lim_n \alpha_n\mathbb{E}[(\varphi(Y_n)-\varphi(Y))^2(\chi(Y_n)+\chi(Y))/2] \tag{5}$$

d) $(\mathcal{E},\mathbb{D})$ *is local if and only if* $\forall\varphi\in\mathcal{D}$

$$\lim_n \alpha_n\mathbb{E}[(\varphi(Y_n)-\varphi(Y))^4]=0 \tag{6}$$

this condition is equivalent to $\quad\exists\lambda>2\quad \lim_n \alpha_n\mathbb{E}[|\varphi(Y_n)-\varphi(Y)|^\lambda]=0.$

e) If the form $(\mathcal{E},\mathbb{D})$ *is local, then the principle of asymptotic error calculus is valid on* $\widetilde{\mathcal{D}}=\{F(f_1,\ldots,f_p)\ :\ f_i\in\mathcal{D},\ \ F\in\mathcal{C}^1(\mathbb{R}^p,\mathbb{R})\}$ *i.e.*

$$\begin{aligned}&\lim_n \alpha_n\mathbb{E}[(F(f_1(Y_n),\ldots,f_p(Y_n))-F(f_1(Y),\ldots,f_p(Y)))^2]\\ &\qquad=\mathbb{E}_Y[\textstyle\sum_{i,j=1}^p F_i'(f_1,\ldots,f_p)F_j'(f_1,\ldots,f_p)\Gamma[f_i,f_j]].\end{aligned}$$

The proof of this theorem is given in [6] Theorem 1, Remark 3 and Theorem 2. The point e) of the theorem is a commutativity of limits, it means that the error on a function of Y may be directly obtained starting from the error on Y by functional calculus.

An operator B from $\mathcal{D}$ into $L^2(\mathbb{P}_Y)$ will be said to be a *first order operator* if it satisfies

$$B[\varphi\chi]=B[\varphi]\chi+\varphi B[\chi]\qquad\forall\varphi,\chi\in\mathcal{D}$$

Theorem 2.2 *Under* (H1) *to* (H4)

a) the theoretical variance $\lim_n \alpha_n\mathbb{E}[(\varphi(Y_n)-\varphi(Y))^2\psi(Y)]$ *and the practical variance* $\lim_n \alpha_n\mathbb{E}[(\varphi(Y_n)-\varphi(Y))^2\psi(Y_n)]$ *exist and we have* $\forall\varphi,\chi,\psi\in\mathcal{D}$

$$\begin{aligned}&\lim_n \alpha_n\mathbb{E}[(\varphi(Y_n)-\varphi(Y))(\chi(Y_n)-\chi(Y))\psi(Y)]\\ &\qquad=\mathbb{E}_Y[-\underline{A}[\varphi\psi]\chi+\underline{A}[\psi]\varphi\chi-\overline{A}[\varphi]\chi\psi]\\ &\lim_n \alpha_n\mathbb{E}[(\varphi(Y_n)-\varphi(Y))(\chi(Y_n)-\chi(Y))\psi(Y_n)]\\ &\qquad=\mathbb{E}_Y[-\overline{A}[\varphi\psi]\chi+\overline{A}[\psi]\varphi\chi-\underline{A}[\varphi]\chi\psi]\end{aligned}$$

b) These two variances coincide if and only if $\not{A}$ *is a first order operator, and then are equal to* $\mathbb{E}_Y[\Gamma[\varphi]\psi]$.

The proof of this result is given in [6] Proposition 2.

Example: Typical formulae of finite dimensional error calculus.

Let us consider a triplet of real random variables (Y, Z, T) and a real random variable G independent of (Y, Z, T) centered with variance one. We are interested in the approximation Y_ε of Y given by

$$Y_\varepsilon = Y + \varepsilon Z + \sqrt{\varepsilon} TG. \tag{7}$$

In the multidimensional case, Y is with values in $\mathbb{R}^p$ as Z, T is a $p \times q$-matrix and G is independent of (Y, Z, T) with values in $\mathbb{R}^q$, centered, square integrable, such that $\mathbb{E}[G_i G_j] = \delta_{ij}$.

Operator $\overline{A}$.

Proposition 2.1 *If Z and T are square integrable, if φ is $\mathcal{C}^2$ bounded with bounded derivatives of first and second orders ($\varphi \in \mathcal{C}^2_b$) and if χ is bounded,*

$$\frac{1}{\varepsilon}\mathbb{E}[(\varphi(Y_\varepsilon) - \varphi(Y))\chi(Y)] \to \mathbb{E}_Y[\overline{A}[\varphi]\chi]$$

where $\overline{A}[\varphi](y) = \mathbb{E}[Z|Y=y]\varphi'(y) + \frac{1}{2}\mathbb{E}[T^2|Y=y]\varphi''(y)$.

In the multidimensional case

$$\overline{A}[\varphi](y) = \mathbb{E}[Z^t|Y=y]\nabla\varphi(y) + \frac{1}{2}\sum_{ij}\mathbb{E}[(TT^t)_{ij}|Y=y]\varphi''_{ij}(y).$$

Proof. Let us give the argument with the notation of the case $q = p = 1$. The Taylor-Lagrange formula applied up to second order gives

$$\begin{aligned}\tfrac{1}{\varepsilon}\mathbb{E}[(\varphi(Y_\varepsilon) - \varphi(Y))\chi(Y)] = \mathbb{E}[Z\varphi'(Y)\chi(Y)] \\ +\tfrac{1}{2}\mathbb{E}[(\varepsilon Z^2 + 2\sqrt{\varepsilon}ZTG + T^2G^2) \\ \textstyle\int_0^1\int_0^1 \varphi''(Y + ab(\varepsilon Z + \sqrt{\varepsilon}TG))2a\,da\,db\,\chi(Y)]\end{aligned}$$

(note that ZTG and $T^2G^2 \in L^1$ because of the independence) and this tends by dominated Lebesgue theorem to $\mathbb{E}[Z\varphi'(Y)\chi(Y)] + \frac{1}{2}\mathbb{E}[T^2\varphi''(Y)\chi(Y)]$. □

Quadratic form and operator $\widetilde{A}$.

Proposition 2.2 *If Z and T are square integrable, if φ and χ are $\mathcal{C}_b^1$*

$$\frac{1}{\varepsilon}\mathbb{E}[(\varphi(Y_\varepsilon)-\varphi(Y)(\chi(Y_\varepsilon)-\chi(Y)] \to \mathbb{E}[T^2\varphi'(Y)\chi'(Y)]$$

and in the multidimensional case

$$\frac{1}{\varepsilon}\mathbb{E}[(\varphi(Y_\varepsilon)-\varphi(Y)(\chi(Y_\varepsilon)-\chi(Y)] \to \mathbb{E}[(\nabla\varphi)^t(Y)TT^t\nabla\chi(Y)].$$

Proof. The demonstration is similar with a first order expansion. □
In order to exhibit the operator $\widetilde{A}$, we must examine the conditions of an integration by parts in the preceding limit. Let us put $\theta_{ij}(y)=\mathbb{E}[(TT^t)_{ij}|Y=y]$ so that $\mathbb{E}[(\nabla\varphi)^t(Y)TT^t\nabla\chi(Y)]=\sum_{ij}\mathbb{E}_Y[\varphi_i'\theta_{ij}\chi_i']$.

Proposition 2.3 *If Z and T are square integrable, if for $i,j=1,\dots,p$ the measure $\theta_{ij}\mathbb{P}_Y$ on $\mathbb{R}^p$ possesses a partial derivative in the sense of distributions $\partial_j(\theta_{ij}\mathbb{P}_Y)$ which is a bounded measure absolutely continuous with respect to $\mathbb{P}_Y$, say $\rho_{ij}\mathbb{P}_Y$, then as soon as θ_{ij} and $\rho_{ij}\in L^2(\mathbb{P}_Y)$ the form $\widetilde{\mathcal{E}}[\varphi,\chi]=\frac{1}{2}\sum_{ij}\mathbb{E}_Y[\varphi_i'\theta_{ij}\chi_j']$ is closable on the algebra $\mathcal{D}=\mathcal{C}_b^2$, hypotheses* (H1) *to* (H4) *are fulfilled and*

$$\widetilde{A}[\varphi]=\frac{1}{2}\sum_{ij}\theta_{ij}\varphi_{ij}''+\frac{1}{2}\sum_{ij}\rho_{ij}\varphi_j'.$$

Proof. We have

$$\sum_{ij}\int\theta_{ij}\varphi_i'\chi_j'\,d\mathbb{P}_Y=\sum_{ij}\int\theta_{ij}(\partial_j(\varphi_i'\chi)-\varphi_{ij}''\chi)d\mathbb{P}_Y$$

and the equality

$$\int\theta_{ij}\partial_j(\varphi_i'\chi)d\mathbb{P}_Y=-\int\varphi_i'\chi\rho_{ij}d\mathbb{P}_Y$$

valid for $\varphi,\chi\in\mathcal{C}_K^\infty$ extends, under the assumptions of the statement, to $\varphi,\chi\in\mathcal{C}_b^2$. This yields $\frac{1}{2}\sum_{ij}\mathbb{E}[\varphi_i'\theta_{ij}\chi_j']=-\frac{1}{2}\int(\sum_{ij}\theta_{ij}\varphi_{ij}''+\sum_{ij}\rho_{ij}\varphi_j')\chi\,d\mathbb{P}_Y$. □

The operator $\widetilde{A}$ depends only on T, not on Z. We obtain $\underline{A}$ by difference :

$$\underline{A}[\varphi] = \frac{1}{2}\sum_{ij}\theta_{ij}\varphi''_{ij} + \sum_j(\sum_i \rho_{ij} - z_j)\varphi'_j$$

where $z_j(y) = \mathbb{E}[Z_j|Y = y]$. At last, $\not{A}$ is first order : $\not{A}[\varphi] = \sum_j(z_j - \frac{1}{2}\sum_i \rho_{ij})\varphi'_j$.

For infinite dimensional examples see [6].

We are now able to make more precise the dichotomy of §1: we shall say that the approximation is *weakly stochastic* if hypotheses H1 to H4 are fulfilled and $\widetilde{A} = 0$ and hence $\not{A} = \overline{A} = -\underline{A}$.

And the approximation will be said *strongly stochastic* if hypotheses H1 to H4 are fulfilled and $\widetilde{A} \neq 0$ hence the Dirichlet form (cf. Thm 2.1) is not nought.

3 The usual norm-based numerical analysis revisited.

For boundary value problems or optimization problems etc. the resolution by approximation is often displayed in numerical analysis in the following manner :

The data are represented by a function f in some functional space F, the parameters of the problem are represented by a point λ in a suitable space Λ and the mathematical solution of the problem writes

$$g = \Phi(f, \lambda).$$

The solution belongs to the space G when $f \in F$ and $\lambda \in \Lambda$. Then, the analysis of the functional Φ yields norm estimates of the form :

$$\left.\begin{array}{l}\|f_n - f\|_F \leq \alpha \\ \|\lambda_n - \lambda\|_\Lambda \leq \beta\end{array}\right\} \Longrightarrow \|g_n - g\|_G \leq \xi(\alpha, \beta, n) \tag{8}$$

for some function ξ, which assures the convergence of the resolution procedure.

It has to be emphasized that such a reasoning supposes that the premises of (3.1) be fullfilled. The error $f - f_n$ is thought deterministically. The

possible randomness of the error and the behaviour of its bias through the functional Φ are not taken in account in this approach.

Remark. a) When the problem is *purely mathematical*, the above difficulty may, most often, be considered of secondary importance. Indeed, if the function f and parameters λ are random, we may consider that the problem is solved as soon as we are able to compute *the law* of the output g or to have an approximation of it (if there were also randomness in the functional Φ, our aim would be to get the joint law of (f, λ, g), we move away this case which is similar for simplicity). Now for this, it is enough that an approximation f_n and λ_n of f and λ yields an approximation g_n of g in probability. In other words, estimates of the form (3.1) in probability are sufficient to solve the problem:

$$\left.\begin{array}{r} \forall \delta > 0,\ \exists \varepsilon > 0, \quad \text{s.t.} \\ \mathbb{P}\{\|f_n - f\|_F \leq \alpha; \|\lambda_n - \lambda\|_\Lambda \leq \beta\} \geq 1 - \varepsilon \\ \Rightarrow \mathbb{P}\{\|g_n - g\|_G \leq \xi(\alpha, \beta, n)\} \geq 1 - \delta \end{array}\right\} \quad (9)$$

then the law of g may be approximated by Monte Carlo methods, because we are allowed to choose the sample f and parameter λ as we want provided they follow the right probability law.

This can be said otherwise : from a mathematical point of view, most often, the sensitivity of g to the input f may be thought *globally*. Estimate like (3.2) will be usually obtained by inequalities similar to (3.1) but in the sense of spaces like $L^p(\Omega, \mathcal{A}, \mathbb{P}; F)$, $L^p(\Omega, \mathcal{A}, \mathbb{P}; \Lambda)$ and $L^p(\Omega, \mathcal{A}, \mathbb{P}; G)$.

b) But different is the situation where f comes from an experiment. For example the temperatures, the wind velocities, etc. in a meteorological model. In such cases, the data f is imposed, known with some precision, and the question whether the errors are weakly or strongly stochastic is relevant. In the first case the sensitivity analysis reduces to a derivation (in a suitable sense between suitable spaces), in the second case a second order Ito-like calculus is compulsory.

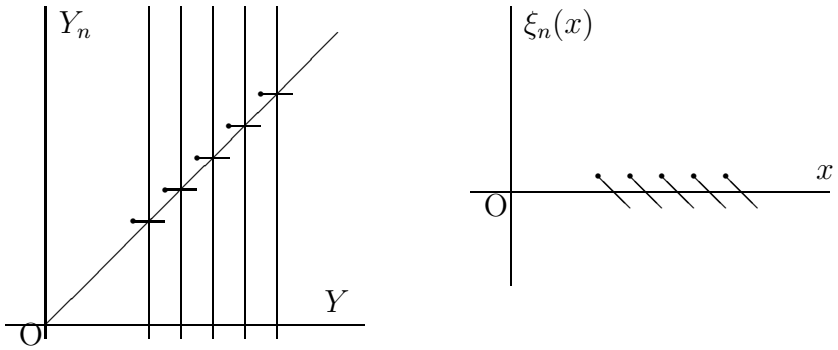
The importance of this discussion is reinforced by the results of the next section.

4 The errors due to the graduation of measuring instruments are strongly stochastic.

Suppose Y is a real quantity measured with a graduated instrument. Let Y_n be the approximation of Y to the nearest graduation, i.e.

$$Y_n = \frac{[nY]}{n} + \frac{1}{2n}$$

($[x]$ denotes the integral part of x, and $\{x\} = x - [x]$ the fractional part).



Let us put $Y_n = Y + \xi_n(Y)$ where the function $\xi_n(x) = \frac{[nx]}{n} - \frac{1}{2n} - x$ is periodic with period $\frac{1}{n}$ and may be written $\xi_n(x) = \frac{1}{n}\theta(nx)$ with $\theta(x) = \frac{1}{2} - \{x\}$. Let $\mathbb{P}_Y$ be the law of Y, we have

Theorem 4.1 *a) If $\mathbb{P}_Y$ has a density,*

$$(n(Y_n - Y), Y) \quad \stackrel{d}{\Longrightarrow} \quad (V, Y) \tag{10}$$

where V is uniform on $(-\frac{1}{2}, \frac{1}{2})$ independent of Y, and for $\varphi \in \mathcal{C}^1 \cap Lip$

$$n^2 \mathbb{E}[(\varphi(Y_n) - \varphi(Y))^2] \quad \to \quad \frac{1}{12}\mathbb{E}_Y[\varphi'^2]. \tag{11}$$

b) If $\mathbb{P}_Y$ has a density satisfying one of the following conditions :

i) the derivative in distribution sense $\partial\mathbb{P}_Y$ is a measure $\ll \mathbb{P}_Y$ of the form $\rho\mathbb{P}_Y$ with $\rho \in L^2(\mathbb{P}_Y)$,

ii) $\mathbb{P}_Y = h.1_G \frac{dy}{|G|}$ with G open set, $h \in H^1 \cap L^\infty(G)$, $h > 0$,

then hypotheses H1 to H4 are fulfilled on the algebra $\mathcal{D} = \mathcal{C}_b^2$ of bounded

functions with bounded derivatives up to order 2 with $\alpha_n = n^2$ *and*

$$\begin{aligned}\overline{A}[\varphi] &= \tfrac{1}{24}\varphi'' \\ \widetilde{A}[\varphi] &= \tfrac{1}{24}\varphi'' + \tfrac{1}{24}\rho\varphi' \qquad \textit{case i)} \\ \widetilde{A}[\varphi] &= \tfrac{1}{24}\varphi'' + \tfrac{1}{24}hh'\varphi' \qquad \textit{case ii).}\end{aligned}$$

Here $\stackrel{d}{\Longrightarrow}$ denotes the weak convergence, i.e. the convergence of probability measures on continuous bounded functions, $\mathbb{E}_Y$ is the expectation under $\mathbb{P}_Y$.

Proof. a) It is equivalent to study the weak convergence of $(\frac{1}{2} + n(Y_n - Y), Y) = (\frac{1}{2} + \theta(nY), Y)$. Since $\frac{1}{2} + \theta$ takes its values in the unit interval, it is enough to study the convergence on the characters of the group $\mathbb{T}^1 \times \mathbb{R}$, i.e.

$$\mathbb{E}[e^{2i\pi k(\frac{1}{2}+\theta(nY))}e^{iuY}] = \mathbb{E}[e^{-2i\pi knY}e^{iuY}] = \Psi_Y(u - 2\pi kn)$$

where Ψ_Y is the characteristic function of Y. This tends to $\Psi(u)1_{\{k\neq 0\}}$ by the Riemann-Lebesgue lemma since $\mathbb{P}_Y$ has a density.

If $\varphi \in \mathcal{C}^1 \cap Lip$, the relation $\varphi(Y_n) - \varphi(Y) = (Y_n - Y)\int_0^1 \varphi'(Y + \alpha(Y_n - Y))d\alpha$ gives

$$n^2\mathbb{E}[(\varphi(Y_n) - \varphi(Y))^2] = \mathbb{E}[\theta^2(nY)\varphi'^2(Y)] + o(1)$$

and $\mathbb{E}[\theta^2(nY)\varphi'^2(Y)] \to \int_{-\frac{1}{2}}^{\frac{1}{2}} \theta^2(t)dt\mathbb{E}[\varphi'^2(Y)]$ what proves the second assertion.

b) We postpone the proof of b) after theorem 4.2. □

Comments. 1) The result a) is the classical *arbitrary functions principle* (cf. [1] [2]), it would be still valid if $\mathbb{P}_Y$ were a *Rajchman measure* (see [8]). For extensions of the arbitrary functions principle to infinite dimensional cases see [7] and [8]. A summary of the history of this principle is given in [8] section I.3.

2) The b) of the theorem shows that when the law $\mathbb{P}_Y$ is smooth, the approximation Y_n of Y to the nearest graduation is *strongly stochastic*.

The results of theorem 4.1 extend to the finite dimensional case: Let us suppose Y is $\mathbb{R}^d$-valued, measured with an equidistant graduation corresponding to an orthonormal rectilinear coordinate system, and estimated to the nearest graduation component by component. Thus we put

$$Y_n = Y + \frac{1}{n}\theta(nY)$$

with $\theta(y) = (\frac{1}{2} - \{y_1\}, \cdots, \frac{1}{2} - \{y_d\})$.

Theorem 4.2 *a) If $\mathbb{P}_Y$ has a density and if X is $\mathbb{R}^m$-valued*

$$(X, n(Y_n - Y)) \quad \stackrel{d}{\Longrightarrow} \quad (X, (V_1, \ldots, V_d)) \tag{12}$$

where the V_i's are i.i.d. uniformly distributed on $(-\frac{1}{2}, \frac{1}{2})$ independent of X. For all $\varphi \in \mathcal{C}^1 \cap Lip(\mathbb{R}^d)$

$$(X, n(\varphi(Y_n) - \varphi(Y))) \quad \stackrel{d}{\Longrightarrow} \quad (X, \sum_{i=1}^{d} V_i \varphi_i'(Y)) \tag{13}$$

$$n^2 \mathbb{E}[(\varphi(Y_n) - \varphi(Y))^2 | Y = y] \to \frac{1}{12} \sum_{i=1}^{d} \varphi_i'^2(y) \qquad \text{in } L^1(\mathbb{P}_Y) \tag{14}$$

in particular

$$n^2 \mathbb{E}[(\varphi(Y_n) - \varphi(Y))^2] \to \mathbb{E}_Y[\frac{1}{12} \sum_{i=1}^{d} \varphi_i'^2(y)]. \tag{15}$$

b) If φ is of class $\mathcal{C}^2$, the conditional expectation $n^2 \mathbb{E}[\varphi(Y_n) - \varphi(Y) | Y = y]$ possesses a version $n^2(\varphi(y + \frac{1}{n}\theta(ny)) - \varphi(y))$ independent of the probability measure $\mathbb{P}$ which converges in the sense of distributions to the function $\frac{1}{24} \triangle \varphi$.

c) If $\mathbb{P}_Y \ll dy$ on $\mathbb{R}^d$, $\forall \psi \in L^1([0,1])$

$$(X, \psi(n(Y_n - Y))) \quad \stackrel{d}{\Longrightarrow} (X, \psi(V)). \tag{16}$$

d) We consider the bias operators on the algebra $\mathcal{C}_b^2$ of bounded functions with bounded derivatives up to order 2 with the sequence $\alpha_n = n^2$. If $\mathbb{P}_Y$ has a density and if one of the following condition is fulfilled

i) $\forall i = 1, \ldots, d$ the partial derivative $\partial_i \mathbb{P}_Y$ in the sense of distributions is a measure $\ll \mathbb{P}_Y$ of the form $\rho_i \mathbb{P}_Y$ with $\rho_i \in L^2(\mathbb{P}_Y)$

ii) $\mathbb{P}_Y = h 1_G \frac{dy}{|G|}$ with G open set, $h \in H^1 \cap L^\infty(G)$, $h > 0$

then hypotheses (H1) *to* (H4) *are satisfied and*

$$\begin{array}{lll} \overline{A}[\varphi] & = \frac{1}{24} \triangle \varphi & \\ \widetilde{A}[\varphi] & = \frac{1}{24} \triangle \varphi + \frac{1}{24} \sum \varphi_i' \rho_i & \text{case i)} \\ \widetilde{A}[\varphi] & = \frac{1}{24} \triangle \varphi + \frac{1}{24} \frac{1}{h} \sum h_i' \varphi_i' & \text{case ii)} \\ \Gamma[\varphi] & = \frac{1}{12} \sum \varphi_i'^2. & \end{array}$$

Proof. The argument for relation (4.3) is similar to one dimensional case. The relation (4.4) comes from the Taylor expansion $\varphi(Y_n) - \varphi(Y) =$
$= \sum_{i=1}^d (Y_{n,i} - Y_i) \int_0^1 \varphi_i'(Y_{n,1}, \ldots, Y_{n,i-1}, Y_i + t(Y_{n,i} - Y_i), Y_{i+1}, \ldots, Y_d)\, dt$
and the convergence

$$(X, \sum_i \theta(nY_i)\varphi_i'(Y)) \quad \stackrel{d}{\Longrightarrow} \quad (X, \sum_i \varphi_i'(Y)V_i)$$

thanks to (4.3) and the following approximation in L^1

$$\mathbb{E}\left| \sum_i \theta(nY_i)\varphi_i'(Y) - \sum_i \theta(nY_i) \int_0^1 \varphi_i'(\ldots, Y_i + t(Y_{n,i} - Y_i), \ldots)dt \right| \to 0.$$

To prove the formulae (4.5) and (4.6) let us remark that

$$n^2\mathbb{E}[(\varphi(Y_n) - \varphi(Y)^2 | Y = y] =$$

$$= \mathbb{E}\left[\left| \sum_i \theta(nY_i) \int_0^1 \varphi_i'(\ldots, Y_i + t(Y_{n,i} - Y_i), \ldots)dt \right|^2 | Y = y \right]$$

$$= \left| \sum_i \theta(ny_i) \int_0^1 \varphi_i'(y_1 + \frac{1}{n}\theta(ny_1), \ldots, y_i + t\frac{1}{n}\theta(ny_i), \ldots)dt \right|^2 \quad \mathbb{P}_Y - a.s.$$

each term $(\theta(ny_i) \int_0^1 \varphi_i'(\ldots)dt)^2$ converges to $\int \theta^2 \varphi_i'^2(y) = \frac{1}{12}\varphi_i'^2$ in L^1 and each term $\theta(ny_i)\theta(ny_j) \int_0^1 \ldots \int_0^1 \ldots$ goes to zero in L^1 what proves the a) of the statement.

The point b) is obtained following the same lines with a Taylor expansion up to second order and an integration by part thanks to the fact that φ is now supposed to be $\mathcal{C}^2$.

In order to prove c) let us suppose first that $\mathbb{P}_Y = 1_{[0,1]^d}.dy$. Considering a sequence of functions $\psi_k \in \mathcal{C}_b$ tending to ψ in L^1 we have the bound

$$\begin{aligned}
&|\mathbb{E}[e^{i<u,X>}e^{iv\psi(\theta(nY))}] - \mathbb{E}[e^{i<u,X>}e^{iv\psi_k(\theta(nY))}]| \\
&\leq |v| \int |\psi(\theta(ny)) - \psi_k(\theta(ny))|dy \\
&= |v| \sum_{p_1=0}^{n-1} \cdots \int_{p_1}^{p_1+1} \cdots |\psi(\theta(ny_1)\ldots) - \psi_k(\theta(ny_1)\ldots)|dy_1 \ldots dy_d \\
&= |v| \sum \cdots \sum \int \cdots \int |\psi(\theta(x_1), \ldots) - \psi_k(\theta(x_1), \ldots)| \tfrac{dx_1}{n} \cdots \tfrac{dx_d}{n} \\
&= |v| \|\psi - \psi_k\|_{L^1}.
\end{aligned}$$

And this yields (4.7) in this case. Now if $\mathbb{P}_Y \ll dy$ then $\mathbb{P}_{\{Y\}} \ll dy$ on $[0,1]^d$ and the weak convergence under dy on $[0,1]^d$ implies the weak convergence under $\mathbb{P}_{\{Y\}}$ what yields the result.

In d) the point i) is proved by the approach already used in [6] consisting of proving that hypothesis (H3) is fulfilled by displaying the operator $\widetilde{A}$ thanks to an integration by parts. The point ii) is an application of a Girsanov theorem for Dirichlet forms (cf. [8]). □

5 Conclusion.

The question of the *specification* of an approximate numerical result may be made more precise : it is a description of the error on the inputs in such a way that it is possible (in smooth cases) to obtain the same kind of description on the output.

• To give the result with an interval for the error is a specification. But it is unsatisfactory for several reasons :

(i) the law of the error may have a non compact support, or a support not decreasing to a point (cf. Polya's urn),

(ii) with such a description we can manage neither the variances nor the biases.

• To give the result with an interval and a probability that the error be inside this interval is also a specification. It is a triplet $(x_n, \alpha, \varepsilon)$ where x_n is the proposed result, with the condition

$$\mathbb{P}\{|x - x_n| < \alpha\} \geq 1 - \varepsilon.$$

As already discussed, such a specification may be used when we are only concerned by the law of the output. If the probability $\mathbb{P}\{|x - x_n| < \alpha\}$ is known for every α, this gives the knowledge of $\|x - x_n\|^2_{L^2(\mathbb{P})}$, but the critique (ii) still holds.

• The Dirichlet theoretical specification used in our argumentation deals with the following mathematical objects :

. $\mathbb{P}_Y$ the law of the quantity to be approximated,

. the sequence α_n giving the order of magnitude,

. the algebra $\mathcal{D}$,

. $\overline{A}$, $\underline{A}$ the theoretical and practical bias operators,

. Γ the square field operator of the associated Dirichlet form.

This specification seems to be too sophisticated to be used by engineers in usual cases, and the question remains to simplify it, preserving the main ideas.

Here we will just give a comment on this question in the finite dimensional case $Y = \Phi(X)$ with Φ regular from $\mathbb{R}^p$ into $\mathbb{R}^q$. If the input is measured with a graduated instrument, the square field operator on the input Γ_{in} is yielded by the size of the graduation and do not depend on the probability law of the input provided that this law be regular, by the arbitrary functions principle. A natural hypothesis is to suppose that the law of the input is uniform in a neighbourhood of the numerical data. Then (theorem 4.2 d)) the approximation of the input satisfies

$$\overline{A}_{in} = \underline{A}_{in} = \widetilde{A}_{in} \qquad (= \frac{1}{12}\Delta)$$

and this equality will be transported to the output

$$\overline{A}_{out} = \underline{A}_{out} = \widetilde{A}_{out}$$

(see definitions H1 to H3). The generator $\widetilde{A}_{out}$ and the square field operator Γ_{out} will be given by *the image* of the input Dirichlet structure by the map Φ (cf. [3] Chap V, [5] Chap IV). The formulae are

$$\left.\begin{array}{ll} (\Gamma_{out}[u])(y) & = \mathbb{E}[\Gamma_{in}[u \circ \Phi](X)|Y = y] \\ (\widetilde{A}_{out}[u])(y) & = \mathbb{E}[\widetilde{A}_{in}[u \circ \Phi](X)|Y = y] \end{array}\right\} \tag{17}$$

where $\Gamma_{in}[u \circ \Phi]$ and $\widetilde{A}_{in}[u \circ \Phi]$ are obtained by the functional calculus in Dirichlet structures (cf. [3] Chap I section 6) with natural notation this writes

$$\left.\begin{array}{ll} \Gamma_{in}[u \circ \Phi] & = (\nabla u)^t \circ \Phi \ \ \Gamma_{in}[\Phi] \ (\nabla u) \circ \Phi \\ \widetilde{A}_{in}[u \circ \Phi] & = (\nabla u)^t \circ \Phi \ \ \widetilde{A}_{in}[\Phi] + \frac{1}{2} \sum_{ij} \partial^2_{ij} u \ \Gamma_{in}[\Phi_i, \Phi_j] \end{array}\right\} \tag{18}$$

We see that, in order to obtain the coefficients of the bias differential operator $\widetilde{A}_{out}$, by formulae (5.2) we have to compute $\Gamma_{in}[\Phi_i, \Phi_j]$ and $\widetilde{A}_{in}[\Phi]$ which involves the *Jacobian* and the *Hessian* matrices of the map Φ and then by formulae (5.1) to average in X on the level manifolds of Φ.

In conclusion, we have attempted to convince the reader that errors have to be thought in terms of second order differential operators. In order that this language be convenient for practical engineering use, more simplicity has to be looked for, taking in account the specific form of the different problems.

References

[1] Poincaré, H., *Calcul des Probabilités* Gauthier-Villars, 1912.

[2] Hopf E. Über die Bedeutung der willkürlichen Funktionen für die Wahrscheinlichkeitstheorie, *Jahresbericht der Deutschen Math. Vereinigung* XLVI, I, 9/12, 179-194, (1936).

[3] Bouleau N. and Hirsch F., *Dirichlet Forms and Analysis on Wiener Space,* De Gruyter, 1991.

[4] Fukushima, M.; Oshima, Y.; Takeda, M. *Dirichlet forms and symmetric Markov processes*, De Gruyter 1994.

[5] Bouleau, N., *Error Calculus for Finance and Physics, the Language of Dirichlet Forms*, De Gruyter, 2003.

[6] Bouleau, N., When and how an error yields a Dirichlet form, *Jour. Functional Analysis* 240, (2006), 445-494.

[7] Bouleau, N., An extension to the Wiener space of the arbitrary functions principle, *C. R. Acad. Sci. Paris*, ser I 343 (2006) 329-332.

[8] Bouleau, N., On the errors related to the graduation of measuring intruments, `http://hal.ccsd.cnrs.fr/ccrd-00105452`