STATISTIQUES/STATISTICS
PROBABILITES/PROBABILITIES

Error Structures and Parameter Estimation

Structures d'Erreur et Estimation Paramétrique

Nicolas Bouleau [a] Christophe Chorro [b]

[a] *Ecole des Ponts, Paris*

[b] *Cermics, Ecole des Ponts & Cermsem, Université Paris 1.*

Abstract

The error calculus based on the theory of Dirichlet forms is an extension of Gauss' approach to error propagation. The aim of this paper is to derive error structures from measurements. The links with Fisher's information lay the foundations of a strong connection with experiment. Here we show that this connection behaves well towards changes of variables and is related to the theory of asymptotic statistics. Finally the study of products permits to lay the foundation of an infinite dimensional empirical error calculus.

To cite this article: N. Bouleau, C. Chorro, C. R. Acad. Sci. Paris, Ser. I 337 (2003).

Résumé

Le calcul d'erreur fondé sur la théorie des formes de Dirichlet est une extension naturelle des calculs proposés par Gauss au $19^{ème}$ siècle. Il utilise jusqu'à présent des hypothèses *a priori* mais l'exploration de ses liens avec l'information de Fisher permet de le relier solidement à l'expérience. Nous testons ici la robustesse de cette relation vis-à-vis des changements de variables et montrons son lien avec la théorie des statistiques asymptotiques. Enfin, l'étude d'une grandeur produit permet de jeter les bases d'un calcul d'erreur empirique infini-dimensionnel.

Pour citer cet article : N. Bouleau, C. Chorro, C. R. Acad. Sci. Paris, Ser. I 337 (2003).

Version française abrégée

Une structure d'erreur est un terme $(W, \mathcal{W}, m, \mathbb{D}, \Gamma)$ où $(W, \mathcal{W}, m)$ est un espace de probabilité, $\mathbb{D}$ un sous-espace dense de $L^2(m)$ et Γ un opérateur bilinéaire symétrique positif de $\mathbb{D} \times \mathbb{D}$ dans $L^1(m)$ vérifiant

Email addresses: bouleau@enpc.fr (Nicolas Bouleau), chorro@cermics.enpc.fr (Christophe Chorro).

1) le calcul fonctionnel de classe $C^1 \cap Lip$: si $U = (U_1, \dots, U_n) \in \mathbb{D}^n$ et $F \in C^1(\mathbb{R}^n) \cap Lip = \{C^1$ et lipschitzienne$\}$ alors $F(U_1, \dots, U_n) \in \mathbb{D}$ et $\Gamma[F(U_1, \dots, U_n)] = \sum_{i,j=1}^{n} F'_i(U)F'_j(U)\Gamma[U_i, U_j]$,

2) $1 \in \mathbb{D}$, la forme $\mathcal{E}[F,G] = \frac{1}{2}\int \Gamma[F,G]dm$ est fermée au sens où $\mathbb{D}$ est complet pour la norme $\| \,.\, \|_{\mathcal{E}} = (\| \,.\, \|^2_{L^2(m)} + \mathcal{E}[.])^{\frac{1}{2}}$.

Cette notion, dérivée de la théorie des formes de Dirichlet ([1], [6]), s'est avérée bien adaptée pour traduire les phénomènes de propagation des erreurs dans les modèles physiques et financiers ([2], [3], [4]) car les structures d'erreur ont la propriété de se transporter simplement par image (définition 3.2), de permettre les opérations de produit (définition 5.1) et d'autoriser les calculs sous hypothèses lipschitziennes. Une structure d'erreur étant un espace de probabilité muni d'un opérateur gérant la précision, il est naturel de chercher à étendre l'identification statistique des lois de probabilité au cas des structures d'erreur. Ainsi, si θ est un paramètre inconnu à valeurs dans un ouvert Θ de $\mathbb{R}^d$, nous voulons mettre en place expérimentalement une structure d'erreur dans laquelle l'opérateur Γ exprime la précision de la connaissance que l'on a sur θ avec les moyens statistiques employés. Lorsque l'on estime θ à l'aide d'un modèle paramétrique régulier, l'inégalité de Cramer-Rao nous conduit à poser l'identification fondamentale suivante : $\underline{\underline{\Gamma}}[Id] = J^{-1}$ où J est la matrice d'information de Fisher. Ceci est bien naturel, Fisher lui-même présentait J comme une précision (intrinsic accuracy) sur le paramètre (cf [5]). Le but de cette note est de tester la robustesse de cette identification. En ce qui concerne les changements de variables injectifs réguliers (définition 3.1) on peut voir que la reparamétrisation induite sur le modèle se traduit par la notion de structure d'erreur image grâce aux propriétés analytiques de la matrice J (proposition 3.3). Dans le problème de l'estimation directe de la grandeur $\psi(\theta)$ lorsque ψ n'est pas injective nous obtenons des résultats asymptotiques commodes pour les applications (proposition 4.1). Enfin, lorsque l'on s'intéresse à l'estimation de grandeurs produit, où les composantes sont estimées à l'aide d'expériences indépendantes, le phénomène naturel de sommation des erreurs s'exprime dans la notion de structure d'erreur produit (proposition 5.2) et ouvre la voie à une généralisation aux grandeurs infini-dimensionnelles.

1. Introduction

An error structure is a term $(W, \mathcal{W}, m, \mathbb{D}, \Gamma)$ where $(W, \mathcal{W}, m)$ is a probability space, $\mathbb{D}$ is a dense vector subspace of $L^2(m)$ and Γ is a positive symmetric bilinear map from $\mathbb{D} \times \mathbb{D}$ into $L^1(m)$ fulfilling:

1) the functional calculus of class $C^1 \cap Lip$ i.e. if $U = (U_1, \dots, U_n) \in \mathbb{D}^n$ and $F \in C^1(\mathbb{R}^n) \cap Lip = \{C^1$ and Lipschitz$\}$ then $F(U_1, \dots, U_n) \in \mathbb{D}$ and $\Gamma[F(U_1, \dots, U_n)] = \sum_{i,j=1}^{n} F'_i(U)F'_j(U)\Gamma[U_i, U_j]$,

2) $1 \in \mathbb{D}$, the bilinear form $\mathcal{E}[F,G] = \frac{1}{2}\int \Gamma[F,G]dm$ is closed i.e. $\mathbb{D}$ is complete under the norm $\| \,.\, \|_{\mathcal{E}} = (\| \,.\, \|^2_{L^2(m)} + \mathcal{E}[.])^{\frac{1}{2}}$. We always write $\Gamma[F]$ for $\Gamma[F,F]$ and $\mathcal{E}[F]$ for $\mathcal{E}[F,F]$.

This notion is derived from the theory of Dirichlet forms ([1] Ch.1, [6]). It is a natural extension of the classical Gauss approach ([3]) and seems to be a good way for studying the propagation of errors and the sensitivity to changes of parameters in physical and financial models ([3], [4]). For $U = (U_1, \dots, U_n) \in \mathbb{D}^n$ the intuitive meaning of the matrix $\underline{\underline{\Gamma}}[U] = [\Gamma[U_i, U_j]]_{1 \le i,j \le n}$ is the variance-covariance of errors on U ([4] Ch.1). From the hypotheses mentioned above $\mathcal{E}$ is a local Dirichlet form and Γ its associated squared field operator. The operations of taking images by mapping and making countable products naturally provide error structures on spaces of stochastic processes ([1] Ch.2, [3], [4] Ch.6). Since a probability space $(W, \mathcal{W}, m)$ can be known thanks to statistical experiments, we raise the problem of the empirical identification of an error structure. In the same way as the σ-additivity of m on $\mathcal{W}$ could not result from experiments but is a fundamental mathematical hypothesis, our error structure will have to verify the closedness property 2) (This cannot be deduced from observations). Thus let θ be a parameter taking its values in an open set $\Theta \subset \mathbb{R}^d$. It is frequently useful to treat θ as the realisation of a random variable $V : (\Omega, \mathcal{A}, \mathbb{P}) \to \Theta$ with a known distribution ρ chosen by combining experience with convenience ([8] p.225). Let X be a random variable defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in a mesurable space $(E, \mathcal{F})$. We suppose that the conditional law of X, given $V = \theta$, is P_θ. Classically, to estimate θ we may use the statistical model $(P_\theta)_{\theta\in\Theta}$ generated by the observation of X. Here we want to equip Θ with an error structure

$$S^V = (\Theta, \mathcal{B}(\Theta), \rho, \mathbb{D}^V, \Gamma^V) \qquad (1)$$

where Γ^V will express the precision of our knowledge on θ. Our approach is to consider Γ^V as the inverse of the Fisher matrix which is an accuracy measure for regular statistical models ([5]). We will study the behaviour of this identification through changes of variables and products to show its remarkable stability.

2. The Cramer-Rao Inequality (C.R.I.) and the Fundamental Identification (F.I.)

We suppose that $(P_\theta)_{\theta\in\Theta}$ satisfies the conditions of regular models ([7] p.65):

(a) The measures P_θ are absolutely continuous with respect to the σ-finite measure μ and $\frac{dP_\theta}{d\mu} = f(.,\theta) > 0$.

(b) $\theta \to f(x,\theta)$ is continuous for μ-almost all x.

(c) We set $g(x,\theta) = \sqrt{f(x,\theta)}$. The function $\theta \to g(.,\theta)$ is differentiable in $L^2(\mu)$ with derivative $\phi : E \times \Theta \to \mathbb{R}^d$; thus the positive semidefinite matrix $J(\theta) = 4\int \phi(x,\theta)\phi(x,\theta)^t d\mu(x)$ is defined as the Fisher information matrix of our model.

(d) $\theta \to \phi(.,\theta)$ is continuous in $L^2(\mu)$.

The following result is called the Cramer-Rao inequality (**C.R.I.**) and gives a bound of estimation for the quadratic risk.

Theorem 2.1 ([7] p.73) *Let $\psi : \mathbb{R}^d \to \mathbb{R}^m$ be differentiable and $(P_\theta)_{\theta\in\Theta}$ be a regular model with $\forall\theta \in \Theta$ $det(J(\theta)) \neq 0$. If $T(X)$ is an unbiased estimator of $\psi(\theta)$ such that $\mathbb{E}[T(X)^2 \mid V = \theta]$ is locally bounded in θ then*

$$\mathbb{E}[(T(X) - \psi(\theta))(T(X) - \psi(\theta))^t \mid V = \theta] \geq \psi'(\theta)J^{-1}(\theta)\psi'(\theta)^t.$$

We suppose up to the end that J is regular. Let us look at the error structure (1) we want to determine. If the components of identity are in $\mathbb{D}^V$, according to the functional calculus, we have for $F \in Lip^1(\Theta) = \{F \in C^1(\Theta)$ and Lipschitz$\}$, $\Gamma^V[F] = (\nabla F)^t \underline{\underline{\Gamma}}^V[Id](\nabla F)$. The **C.R.I.** leads us to state the following fundamental identification:

$$\underline{\underline{\Gamma}}^V[Id] = J^{-1}. \qquad \textbf{(F.I.)}$$

As well as the statistical identification of a probability space presupposes the σ-additivity of the measure, we want to determine an error structure deriving from experiment in which $\mathcal{E}^V$ is a closed form. Thus, we suppose from now on the existence of a dense vector subspace of $L^2(\rho)$ denoted by $\mathbb{D}^V$ and of an operateur Γ^V fulfilling conditions 1) and 2) such that $Lip^1(\Theta) \subset \mathbb{D}^V$ and , according to the **F.I.**, for all F in $Lip^1(\Theta)$ $\Gamma^V[F] = F'J^{-1}(F')^t$. This hypothesis dictates conditions on ρ and J^{-1} and is linked to the notion of closability of operators ([6] p.42). Moreover, as $\mathbb{D}^V$ may not be uniquely defined, we take it minimal for inclusion, this implies the density of $Lip^1(\Theta)$ in $\mathbb{D}^V$ for the norm $\|\ .\ \|_{\mathcal{E}^V}$.

We want now to test the robustness of the **F.I.** by comparing its properties with the well-known behaviour of the Fisher information in the classical framework of parametric estimation.

3. Changes of variables: the injective case

We are going to show the stability of the **F.I.** for regular changes of variables.

Definition 3.1 *We suppose that $\psi : \Theta \to \mathbb{R}^d$ is injective of class $\mathcal{C}^1 \cap Lip$. This change of variables is said to be regular if $det(\psi'(x)) \neq 0$ for all x.*

There are two ways to equip $\psi(\Theta)$ with an error structure. A first one, since ψ is injective, is to estimate $\psi(\theta)$ using the regular model $(P_{\psi^{-1}(a)})_{a\in\psi(\Theta)}$: if we denote by $J^{\psi(V)}$ the Fisher matrix of this model, according to the **F.I.**, we can define $\Gamma^{\psi(V)}$ on $Lip^1(\psi(\Theta))$ by $\Gamma^{\psi(V)}[F] = (\nabla F)^t(J^{\psi(V)})^{-1}(\nabla F)$ where $J^{\psi(V)} = [\psi'(\psi^{-1})^{-1}]^t\ [J(\psi^{-1})]\ [\psi'(\psi^{-1})^{-1}]$. Since ψ is regular and $\mathcal{E}^V$ closed we can show that the form $\mathcal{E}^{\psi(V)}$ is closable and we note $\mathbb{D}^{\psi(V)}$ the domain of its smallest closed extension. Thus, $S^{\psi(V)}$ is defined as the error structure on $\psi(\Theta)$ associated to the direct estimation of $\psi(\theta)$. A second way, as for probability spaces, is to take the image of an error structure by a mapping ([1] p.186) this gives another error structure on $\psi(\Theta)$:

Definition 3.2 *Let $S = (W, \mathcal{W}, m, \mathbb{D}, \Gamma)$ be an error structure and $Y : W \to \mathbb{R}^d \in \mathbb{D}^d$ such that $Y(\mathcal{W})$ is an open set of $\mathbb{R}^d$. Let us define $\widetilde{\mathbb{D}_Y} = \{f \in L^2(Y_*m) \mid f(Y) \in \mathbb{D}\}$ and for $f \in \widetilde{\mathbb{D}_Y}$, $\widetilde{\Gamma_Y}[f](x) = \mathbb{E}_m[\Gamma[f(Y)] \mid Y = x]$.*

If we denote by $\mathbb{D}_Y$ the closure of $Lip^1(Y(\mathcal{W}))$ in $(\widetilde{\mathbb{D}_Y}, \|\,.\,\|_{\widetilde{\mathcal{E}_Y}})$ and by Γ_Y the restriction of $\widetilde{\Gamma_Y}$ to $\mathbb{D}_Y$ then $\psi_*S = (Y(\mathcal{W}), \mathcal{B}(Y(\mathcal{W})), Y_*m, \mathbb{D}_Y, \Gamma_Y)$ is an error structure called the image structure of S by Y.

Then we have the following expected property, the two errors structures coincide:

Proposition 3.3 *The fundamental identification is preserved by the transformation ψ. In other terms:*

$$\psi_*S^V = S^{\psi(V)}.$$

Proof: Using the definition 3.3, we have for $F \in Lip^1(\psi(\Theta))$, $\Gamma^V_\psi[F](a) = \mathbb{E}_\rho[\nabla(F(\psi))^t\ J^{-1}\ \nabla(F(\psi)) \mid \psi = a]$. Since ψ is injective $\Gamma^V_\psi[F](a) = (\nabla_a F)^t[\psi'(\psi^{-1}(a))][J^{-1}(\psi^{-1}(a))][\psi'(\psi^{-1}(a))]^t\nabla_a F$. Computing $J^{\psi(V)}$ we deduce the coincidence of error operators on $Lip^1(\psi(\Theta))$ and we conclude thanks to the density of $Lip^1(\psi(\Theta))$ and the closedness of the two error structures. ■

Remark 1 i) The condition of regularity of ψ can be omitted introducing a notion of infinite information.
ii) The last proposition may be linked to the sufficiency principle since P_θ depends on θ only through ψ in the injective case.

4. The non-injective case

We are now in a special situation: we have put in correspondence an error structure and a parametric model thanks to the Fisher information. But on one side non-injective changes of variables are allowed and on the other side they meet difficulties. We take benefit of this remark to propose a new asymptotic bound for the estimation of a parameter in this case which is directly linked with the notion of error structure. Here we suppose that ψ is a function in $Lip^1(\Theta)$ not necessarily injective but such that $\psi(\Theta)$ is an open set of $\mathbb{R}^d$ (in order to apply definition 3.2 with $Y = \psi$). In this section, $(X_1, \dots, X_n)$ will be, given $V = \theta$, a $n-$sample of P_θ. To estimate $\psi(\theta)$ the reparameterisation introduced in the previous section is meaningless. Nevertheless it is natural to use, for all n, the model generated by the observation of $(X_1,\dots,X_n)$ given $\psi(V)=a$. Now let us illustrate the relevance of the **F.I.** showing that Γ_ψ^V is the corner stone of asymptotic results in our new problem of estimation.

To simplify, let us consider that (P_θ) allows a unequivocal definition of the maximum likelihood estimator: $\forall n\in\mathbb{N}$, $\forall(x_1,\dots,x_n)\in(E)^n$, the equation $\sum\limits_{i=1}^{n}\frac{\partial}{\partial\theta}log f(x_i,\theta)=0$ has a unique solution denoted by $\hat{\theta_n}(x_1,\dots,x_n)$ which is a maximum for the function $\theta\to\prod\limits_{i=1}^{n}f(x_i,\theta)$. In this section we assume that Θ is a convex bounded subset of $\mathbb{R}$ (this is not very restrictive for the applications and could be easily extended to any finite dimension). We call r the Hellinger's distance and recall that $r(P_\theta,P_{\theta'})=\int(\sqrt{f(x,\theta)}-\sqrt{f(x,\theta')}\,)^2d\mu(x)$.

We are now able to prove the following result which provides a consistent estimator with convenient asymptotic properties:

Proposition 4.1 *a) If both conditions:* $0<\underset{\Theta}{inf}\ J(\theta)\le\underset{\Theta}{sup}\ J(\theta)<\infty$ *and* $\forall\theta$, $\forall\delta>0$, $\underset{u\in U_\theta,|u|>\delta}{inf}\ r(P_\theta,P_{\theta+u})>0$ *are fulfilled where* $U_\theta=\{u\in\mathbb{R}\mid\theta+u\in\Theta\}$, *then* $\forall a\in\psi(\Theta)$,

a_1) $\forall\ \varepsilon>0$,

$$\mathbb{E}[\mathbf{1}_{|\psi(\hat{\theta_n}(X_1,\dots,X_n))-a|>\varepsilon}\mid\psi(V)=a]\underset{n\to\infty}{\longrightarrow}0\,.$$

a_2) Given $\psi(V)=a$,

$$\sqrt{n}(\psi(\hat{\theta_n}(X_1,\dots,X_n))-a)\underset{\mathcal{L}(\mathbb{P})}{\longrightarrow}G_a$$

where G_a *is a random variable having the density* $g(x,a)=\mathbb{E}_\rho[\mathbf{1}_{\psi'\neq0}\frac{1}{\sqrt{2\pi\frac{\psi'^2}{J}}}e^{\frac{-x^2J}{2\psi'^2}}\mid\psi=a]$ *with respect to Lebesgue measure on* $\mathbb{R}$ *and consequently* $\Gamma_\psi^V[Id](a)$ *for variance.*

b) Suppose that $(P_\theta)_{\theta\in\Theta}$ *can be extended to a regular model on an open set* Θ' *such that* $\overline{\Theta}\subset\Theta'$. *Let the two conditions* $0<\underset{\Theta'}{inf}\ J(\theta)\le\underset{\Theta'}{sup}\ J(\theta)<\infty$ *and* $\forall\delta>0$ $\underset{\theta\in\Theta'}{inf}\ \underset{u\in U_\theta,|u|>\delta}{inf}\ r(P_\theta,P_{\theta+u})>0$ *be satisfied, then* $\forall a\in\psi(\Theta)$,

$$\mathbb{E}[n(\psi(\hat{\theta_n}(X_1,\dots,X_n))-a)^2\mid\psi(V)=a]\underset{n\to\infty}{\longrightarrow}\Gamma_\psi^V[Id](a).$$

Proof: a_1) Using [7] page 185, the conditions a) imply $\forall\theta\in\Theta,\ \forall\varepsilon>0$,

$$\mathbb{E}[\mathbf{1}_{|\hat{\theta_n}(X_1,\dots,X_n)-\theta|>\varepsilon}\mid V=\theta]\underset{n\to\infty}{\longrightarrow}0\,.$$

By Fubini's theorem

$$\mathbb{E}[\mathbf{1}_{|\psi(\hat{\theta_n}(X_1,\dots,X_n))-a|>\varepsilon}\mid\psi(V)=a]$$

is equal to

$$\mathbb{E}_\rho[\int\mathbf{1}_{|\psi(\hat{\theta_n}(x_1,\dots,x_n))-a|>\varepsilon}f(x_1,.)\dots f(x_n,.)d\mu(x_1)\dots d\mu(x_n)\mid\psi=a],$$

and we can conclude by dominated convergence theorem reminding that ψ is a Lipschitz function.

The proof of a_2) and b) is similar thanks to [7] page 185. ■

5. Product structure

The product of two error structures is defined as follows ([1] p.200)

Definition 5.1 *If* $S_i=(W_i,\mathcal{W}_i,m_i,\mathbb{D}_i,\Gamma_i)$ *(i=1,2) are two error structures, the product structure* $S_1\otimes$

S_2 is defined as the structure $(W_1 \times W_2, \mathcal{W}_1 \otimes \mathcal{W}_2, m_1 \otimes m_2, \mathbb{D}, \Gamma)$ whith $\Gamma[f](x,y) = \Gamma_1[f(.,y)](x) + \Gamma_2[f(x,.)](y)$ and where $\mathbb{D}$ is defined naturally (cf [1]).

Here we want to evaluate the parameter $\theta = (\theta_1, \theta_2)$ where θ_1 and θ_2 are supposed to be independent i.e. V_1 and V_2 are independent random variables. If we estimate the components of θ with two independent regular models we obtain the following proposition that expresses a summation of errors via the product structure notion:

Proposition 5.2 *1) $S^V = S^{V_1} \otimes S^{V_2}$*
2) If ψ_1 and ψ_2 are regular changes of variables then $(\psi_1, \psi_2)_* S^{(V_1,V_2)} = \psi_1\ {}_* S^{V_1} \otimes \psi_2\ {}_* S^{V_2}$.

Remark 2 We can notice that this property is intuitively linked with the additive property of the Fisher information for independent experiments. As we can make the product of many countably error structures ([1] p.203), we are able to obtain, working component per component, an empirical error calculus associated to the estimation of $\theta = (\theta_i)_{i \in \mathbb{N}}$.

6. Conclusion

Through statistical experiments, we have seen that the fundamental identification gives an error structure intrinsically linked to the observed physical phenomenon. The remarkable robustness of this identification, regarding injective changes of variables and products, yields a particularly fitting tool for finite dimensional estimation.

The existence of such an error structure built from the parametric model allows to propagate the accuracy through calculations performed with the parameter thanks to a coherent specific differential calculus (property 1) of Γ). Moreover error calculus provides a natural framework concerning the study of non-injective mapping. A possible extension will be to generalize such an experimental protocol when J is singular and also to explore more precisely the connections between Dirichlet forms and asymptotic statistics. Finally, we wonder whether the semi-parametric and non-parametric estimation theories [9] could lay the foundation of an infinite dimensional identification in order to get Γ on the Wiener space, using a direct functional reasoning instead of a component per component argument as above.

References

[1] N. Bouleau, F. Hirsch : *Dirichlet Forms and Analysis on Wiener Space*, de Gruyter, 1991.

[2] N. Bouleau : *Calcul d'Erreur Complet Lipschitzien et Formes de Dirichlet*, J. de Math. Pures et App., 80.9, 961, 2001.

[3] N. Bouleau : *Error Calculus and Path Sensitivity in Financial Models*, Math. Finance, vol 13, n°1, 115-134, 2003.

[4] N. Bouleau : *Error Calculus for Finance and Physics: The Language of Dirichlet Forms*, de Gruyter, 2003.

[5] R.A. Fisher : *Theory of Statistical Information*, Proc. Cambridge Philo. Soc. Vol XXII, Pt5, 1925.

[6] M. Fukushima : *Dirichlet Forms and Markov Process*, North-Holland, 1980.

[7] I.A. Ibragimov, R.Z. Has'minskii : *Statistical Estimation*, Springer, 1981.

[8] E.L Lehmann, G.Casella : *Theory of Point Estimation*, Springer Texts In Statistics, Springer-Verlag, Berlin, 1998.

[9] A.W Van der Vaart : *Asymptotic Statistics*, Cambridge series in Statistical and Probabilistic Mathematics, 1998.