Joël Chaskalovic and Franck Assous*

On generalized binomial laws to evaluate finite element accuracy: preliminary probabilistic results for adaptive mesh refinement

https://doi.org/10.1515/jnma-2019-0001
Received January 03, 2019; revised December 25, 2019; accepted February 24, 2020

Abstract: The aim of this paper is to provide new perspectives on the relative finite elements accuracy. Starting from a geometrical interpretation of the error estimate which can be deduced from Bramble–Hilbert lemma, we derive a probability law that evaluates the relative accuracy, considered as a random variable, between two finite elements P_k and P_m, $k < m$. We extend this probability law to get a cumulated probabilistic law for two main applications. The first one concerns a family of meshes, the second one is dedicated to a sequence of simplexes constituting a given mesh. Both of these applications could be considered as a first step toward application for adaptive mesh refinement with probabilistic methods.

Keywords: error estimates, probability, finite elements, Bramble–Hilbert lemma, mesh refinement

Classification: 65Gxx

The past decades have seen the development of finite element error estimates due to their influence on improving both accuracy and reliability in scientific computing.

In the *a priori* error estimates, an unknown (in most of the cases) constant is involved which depends, among others, on the basis functions of the considered finite element and on a given semi-norm of the exact solution one wants to approximate. Moreover, error estimates are only upper bounds of the approximation error yielding that its exact value is unknown.

This was the starting point that motivated us to consider the approximation error as a random variable to therefore derive a probability law of the relative accuracy between two Lagrange finite elements P_k and P_m, $k < m$ (see [4, 5]). A first study inspired from the same philosophy was already proposed in [1].

Our aim here is to generalize the results obtained in [4] for a given mesh $\mathcal{T}_h$ defined by a fixed mesh size h (corresponding to the largest diameter of all the simplexes of $\mathcal{T}_h$). We will extend them either to a family of meshes or, for a given mesh, to the sequence of the local elements that constitute it. As we will see, to proceed it, we will need to distinguish between the local and the global accuracy of two given Lagrange finite elements P_k and P_m. Both of these points of view will be introduced as potential applications for adaptive mesh refinement.

The paper is organized as follows. We recall in Section 1 the mathematical problem we consider, the Bramble–Hilbert lemma and the resulting error estimate that allowed us to introduce a probability law for the relative global accuracy between two finite elements P_k and P_m. Extension to a family of meshes is addressed in Section 2. In Section 3, one proposes a generalization that describes the relative local accuracy between two finite elements in a sequence of simplexes belonging to a given mesh. Then, the cumulated probabilistic law is derived either for a family of meshes or for a sequence of simplexes. Concluding remarks follow.

Joël Chaskalovic, D'Alembert, Sorbonne University, Paris, France.
***Corresponding author: Franck Assous,** Department of Mathematics, Ariel University, 40700 Ariel, Israël.
Email: franckassous55@gmail.com

1 Comparing global finite elements accuracy by a probabilistic approach

Let Ω be an open bounded and non empty subset of $\mathbb{R}^n$ and Γ its boundary which we assumed to be C^1-piecewise. Let also u be the solution to the second order elliptic variational formulation:

$$\textbf{(VP)} \quad \begin{cases} \text{Find } u \in V \text{ solution to:} \\ a(u,v) = l(v) \quad \forall v \in V \end{cases} \tag{1.1}$$

where V is a given Hilbert space endowed with the norm $\|\cdot\|_V$, $a(\cdot,\cdot)$ is a bilinear, continuous, and V-elliptic form defined on $V \times V$, and $l(\cdot)$ a linear continuous form defined on V. Classically, variational problem **(VP)** has one, and only one, solution $u \in V$ (see, e.g., [2]). In the sequel and for simplicity, we will restrict ourselves to the case where V is a usual Sobolev space of distributions.

Let us also consider the approximation u_h of u, solution to the approximate variational formulation:

$$\textbf{(VP)}_h \quad \begin{cases} \text{Find } u_h \in V_h \text{ solution to:} \\ a(u_h,v_h) = l(v_h) \quad \forall v_h \in V_h \end{cases} \tag{1.2}$$

where V_h is a given finite-dimensional subset of V.

To state a corollary of Bramble–Hilbert's lemma (cf. [6, 7]), we assume that Ω is exactly covered by a mesh $\mathcal{T}_h$ composed of N_K simplexes K_μ, $1 \leqslant \mu \leqslant N_K$, which respect the classical rules of regular discretization (see, e.g., [2] for the bidimensional case and [7] in $\mathbb{R}^n$). Moreover, we denote by $P_k(K_\mu)$ the space of polynomial functions defined on a given simplex K_μ of degree less than or equal to k, $k \geqslant 1$.

Then, we have the following well-known result [7].

Lemma 1.1. *Suppose that there exists an integer $k \geqslant 1$ such that the approximation u_h of V_h is a continuous piecewise function composed of polynomials which belong to $P_k(K_\mu)$, $1 \leqslant \mu \leqslant N_K$. Then, u_h converges to u in $H^1(\Omega)$:*

$$\lim_{h \to 0} \|u_h - u\|_{1,\Omega} = 0. \tag{1.3}$$

Moreover, if the exact solution u belongs to $H^{k+1}(\Omega)$, we have the following error estimate:

$$\|u_h - u\|_{1,\Omega} \leqslant \mathscr{C}_k\, h^k\, |u|_{k+1,\Omega} \tag{1.4}$$

where $\mathscr{C}_k$ is a positive constant independent of h, $\|\cdot\|_{1,\Omega}$ is the classical norm in $H^1(\Omega)$ and $|\cdot|_{k+1,\Omega}$ is the semi-norm in $H^{k+1}(\Omega)$.

In the sequel, we remind the probability law we derived in [4] which allowed us to evaluate the relative global accuracy, measured in H^1-norm, between two Lagrange finite elements.

We consider two families of Lagrange finite elements P_k and P_m corresponding to a set of values $(k, m) \in \mathbb{N}^2$ such that $0 < k < m$.

The two corresponding inequalities given by (1.4), assuming that the solution u to **(VP)** belongs to $H^{m+1}(\Omega)$, are, respectively, written as:

$$\|u_h^{(k)} - u\|_{1,\Omega} \leqslant \mathscr{C}_k h^k\, |u|_{k+1,\Omega} \tag{1.5}$$

$$\|u_h^{(m)} - u\|_{1,\Omega} \leqslant \mathscr{C}_m h^m\, |u|_{m+1,\Omega} \tag{1.6}$$

where $u_h^{(k)}$ and $u_h^{(m)}$, respectively, denotes the P_k and P_m Lagrange finite element approximations of u, solution to $\textbf{(VP)}_h$.

In what follows, for simplicity, we set $C_k \equiv \mathscr{C}_k\, |u|_{k+1,\Omega}$ and $C_m \equiv \mathscr{C}_m\, |u|_{m+1,\Omega}$. Therefore, inequalities (1.5) and (1.6) become

$$\|u_h^{(k)} - u\|_{1,\Omega} \leqslant C_k h^k \tag{1.7}$$

$$\|u_h^{(m)} - u\|_{1,\Omega} \leqslant C_m h^m \tag{1.8}$$

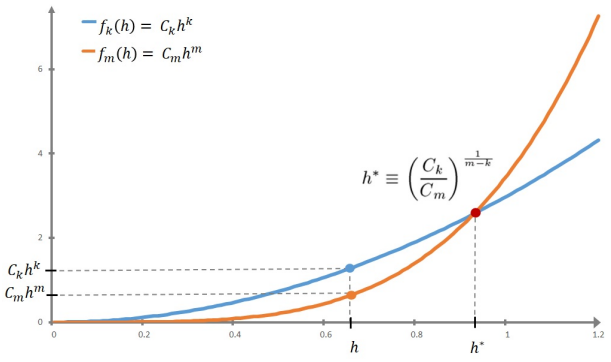

Fig. 1: Curves $f_k(h)$ and $f_m(h)$.

which show that the two polynomial curves defined by $f_k(h) \equiv C_k h^k$ and $f_m(h) \equiv C_m h^m$ are the upper bounds of the possible values for the two norms $\|u_h^{(k)} - u\|_{1,\Omega}$ and $\|u_h^{(m)} - u\|_{1,\Omega}$. More precisely, inequality (1.7) (resp. (1.8)) indicates that the norm $\|u_h^{(k)} - u\|_{1,\Omega}$ (resp. $\|u_h^{(m)} - u\|_{1,\Omega}$) is below the curve $f_k(h)$ (resp. $f_m(h)$), see Fig. 1.

As we are interested in comparing the relative positions of these two curves, we introduce their intersection point h^* defined by

$$h^* \equiv \left(\frac{C_k}{C_m}\right)^{1/(m-k)} = \left(\frac{\mathscr{C}_k\,|u|_{k+1,\Omega}}{\mathscr{C}_m\,|u|_{m+1,\Omega}}\right)^{1/(m-k)}. \tag{1.9}$$

Now, as often in numerical analysis, there is no *a priori* information to surely or better specify the distance between $\|u_h^{(k)} - u\|_{1,\Omega}$ (resp. $\|u_h^{(m)} - u\|_{1,\Omega}$) and the curve f_k or its precise value in the interval $[0, C_k h^k]$ due to (1.7) (resp. the curve f_m and the interval $[0, C_m h^m]$ due to (1.8)).

Indeed, this situation is the consequence of two main ingredients:

1. The solution u of **(VP)** is unknown, which motivates the use of a P_k finite element method to build an approximation $u_h^{(k)}$;
2. The way the mesh generator processes the mesh is *a priori* random which leads to a corresponding random approximation $u_h^{(k)}$.

It is the reason why we treat the possible values of the norm $\|u_h^{(k)} - u\|_{1,\Omega}$ as a random variable defined as follows.

Let us consider an experiment where the constitution of a random grid $\mathcal{T}_h$ and the corresponding random approximation $u_h^{(k)}$ are involved. Therefore, the approximation error $\|u_h^{(k)} - u\|_{1,\Omega}$ can also be viewed as a random variable, defined by the following probabilistic framework:

- A *random trial* corresponds to the grid constitution and the associated approximation $u_h^{(k)}$;
- The probability space Ω contains therefore all the possible results for a given random trial, namely, for all the possible grids that the mesh generator may processed, or equivalently, for all the corresponding approximations $u_h^{(k)}$.

Then, for a fixed value of k, we define the random variable $X^{(k)}$ by

$$\begin{aligned} &X^{(k)} : \ \Omega \to [0, C_k h^k] \\ &\boldsymbol{\omega} \equiv u_h^{(k)} \mapsto X^{(k)}(\boldsymbol{\omega}) = X^{(k)}(u_h^{(k)}) = \|u_h^{(k)} - u\|_{1,\Omega}. \end{aligned} \tag{1.10}$$

In the sequel, for simplicity, we will set:

$$iX^{(k)}(u_h^{(k)}) \equiv X^{(k)}(h).$$

Now, regarding the absence of information concerning the more likely or less likely values of the norm $\|u_h^{(k)} - u\|_{1,\Omega}$ in the interval $[0, C_k h^k]$, we will assume that the random variable $X^{(k)}(h)$ has a uniform distribution on the interval $[0, C_k h^k]$ in the following meaning:

$$\forall\,(\alpha,\beta) \in \mathbb{R}_+^2, \quad 0 \leqslant \alpha < \beta \leqslant C_k h^k : \ \text{Prob}\left\{X^{(k)}(h) \in [\alpha,\beta]\right\} = \frac{\beta - \alpha}{C_k h^k}. \tag{1.11}$$

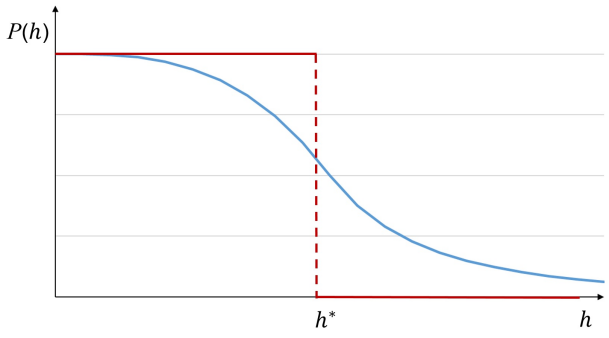

Fig. 2: Case $m - k \neq 1$: shape of the sigmoid distribution (1.14), $P(h) \equiv \text{Prob}\{X^{(m)}(h) \leqslant X^{(k)}(h)\}$.

Equation (1.11) means that if one slides the interval $[\alpha, \beta]$ anywhere in $[0, C_k h^k]$, the probability of the event $\left\{X^{(k)}(h) \in [\alpha, \beta]\right\}$ does not depend on where the interval $[\alpha, \beta]$ is in $[0, C_k h^k]$; this is the property of uniformity of the random variable $X^{(k)}$. This assumption could be relaxed if we were able to get a complementary information about the more likely position of $X^{(k)}(h)$ in the interval $[0, C_k h^k]$. This could be possible either from numerical experiments, or based on new theoretical results regarding error estimates.

We are now able to evaluate the probability of the event

$$\left\{\|u_h^{(m)} - u\|_{1,\Omega} \leqslant \|u_h^{(k)} - u\|_{1,\Omega}\right\} \equiv \left\{X^{(m)}(h) \leqslant X^{(k)}(h)\right\} \tag{1.12}$$

to estimate the relative global accuracy between two finite elements of order k and m, $k < m$.

Let us first start by defining the relative global accuracy between two Lagrange finite elements P_k and P_m, $k < m$.

Definition 1.1. Let P_k and P_m, $k < m$, be two Lagrange finite elements. Then, we will say that P_m is *globally* more accurate than P_k if

$$\|u_h^{(m)} - u\|_{1,\Omega} \leqslant \|u_h^{(k)} - u\|_{1,\Omega}. \tag{1.13}$$

We will recall now the probabilistic law established in [3, 5] to get an estimate on the relative global accuracy between two Lagrange finite elements P_k and P_m, $k < m$, for a fixed mesh size h.

Theorem 1.1. *Let $u \in H^{k+1}(\Omega)$ be the solution to* (1.1)*, and $u_h^{(i)}$, $i = k$ or $i = m$, $k < m$, the two corresponding Lagrange finite element P_i approximation solutions to* (1.2)*. We assume the two corresponding random variables $X^{(i)}(h)$, $i = k$ or $i = m$, defined by* (1.10) *are uniformly distributed on $[0, C_i h^i]$, where C_i are defined by* (1.7)–(1.8)*. Then, the probability such that P_m is* globally *more accurate than P_k, as introduced in* (1.1)*, is given by:*

$$\text{Prob}\left\{X^{(m)}(h) \leqslant X^{(k)}(h)\right\} = \begin{cases} 1 - \dfrac{1}{2}\left(\dfrac{h}{h^*}\right)^{m-k}, & 0 < h \leqslant h^* \\ \dfrac{1}{2}\left(\dfrac{h^*}{h}\right)^{m-k}, & h \geqslant h^*. \end{cases} \tag{1.14}$$

Remark 1.1. Immediate interpretations of the non linear probability law (1.14) are available (see also Fig. 2):

1. P_m finite element is not only *asymptotically* more accurate than P_k as $k < m$, when h goes to 0, as usually considered and as a consequence of error estimate (1.4). Indeed, for all $h \leqslant h^*$, the probability of P_m is more accurate than P_k belongs to $[0.5, 1]$. It means that P_m is *more likely* accurate than P_k for all of these values of h, and not only for arbitrarily small values of h.
2. On the contrary when $k < m$, the finite element P_k is more likely accurate when $h > h^*$. This new point of view allows us to recommend that, for specific situations like for adaptive mesh refinement, this finite element would be more appropriated, as long as one would be able to detect that $h > h^*$.

The next section is devoted to a possible application of the probabilistic law (1.14) to a family of meshes, for example, in the process of mesh refinement.

2 Extension to a family of meshes

The aim of this section is to extend to a family of meshes the previous results we recalled for a given mesh $\mathcal{T}_h$. For this purpose, we introduce a family of N regular meshes $(\mathcal{T}^{(n)})_{n=1,N}$ made of simplexes, each mesh $\mathcal{T}^{(n)}$ being characterized by its mesh size h_n. Let us also consider two Lagrange finite elements P_k and P_m, $k < m$. Therefore, for each mesh $\mathcal{T}^{(n)}$, we can write the corresponding probability law (1.14) for the event $\{X^{(m)}(h_n) \leqslant X^{(k)}(h_n)\}$, associated to the given mesh size h_n.

Our aim is now to evaluate the probability such that exactly n_e meshes, $n_e = 0, 1, \ldots, N$, satisfies P_m *is more accurate than* P_k. To this end, let us introduce the sequence of N independent Bernoulli random variables $(Y_n)_{n=1,N}$ defined by:

$$Y_n = \begin{cases} 1 & \text{if } X^{(m)}(h_n) \leqslant X^{(k)}(h_n) \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

and also the random variable S_N determined by

$$S_N = \sum_{n=1}^{N} Y_n. \tag{2.2}$$

As each Bernoulli variable Y_n indicates if P_m is more accurate than P_k or not on the corresponding mesh $\mathcal{T}^{(n)}$, S_N describes the number of meshes among the all N meshes such that P_m is more accurate than P_k.

Remark 2.1. For any mesh $\mathcal{T}^{(p)}$ and $\mathcal{T}^{(q)}$ belonging to $(\mathcal{T}^{(n)})_{n=1,N}$, characterized by their mesh size h_p and h_q, the knowledge of the event 'P_m is more accurate than P_k' on $\mathcal{T}^{(p)}$ does not enable us to conclude anything on $\mathcal{T}^{(q)}$. Hence, the N random Bernoulli variables Y_n, $1 \leqslant n \leqslant N$, are considered as independent.

It comes out from the nature of the event 'P_m is more accurate than P_k'. Indeed, for an unknown exact solution u and for the corresponding two approximations $u_{h_p}^{(i)}$ and $u_{h_q}^{(i)}$, $i = k$ or $i = m$, one cannot link the value Y_p and Y_q, associated to a mesh size h_p and h_q.

Therefore, we have the following result.

Proposition 2.1. The distribution of probabilities corresponding to the *exact* number of meshes satisfying P_m is more accurate than P_k is given by

$$\text{Prob}\{S_N = 0\} = (1 - \mathcal{P}(h_1)) \cdots (1 - \mathcal{P}(h_N)) \tag{2.3}$$

$$\text{Prob}\{S_N = n_e\} = \sum_{\substack{(i_1,\ldots,i_N)\in\{1,\ldots,N\} \\ i_l \neq i_q,\ l\neq q}} \mathcal{P}(h_{i_1}) \cdots \mathcal{P}(h_{i_{n_e}}) \cdots (1 - \mathcal{P}(h_{i_{n_e+1}})) \cdots (1 - \mathcal{P}(h_{i_N})), \quad 1 \leqslant n_e \leqslant N-1 \tag{2.4}$$

$$\text{Prob}\{S_N = N\} = \mathcal{P}(h_1) \cdots \mathcal{P}(h_N) \tag{2.5}$$

where the quantities $\mathcal{P}(h_{i_j}) \equiv \text{Prob}\{X^{(m)}(h_{i_j}) \leqslant X^{(k)}(h_{i_j})\}$ are given by the probability distribution (1.14) of Theorem 1.1.

In what follows, the sequence of values $(\text{Prob}\{S_N = n_e\})_{n_e=0,N}$ will be called the exact probabilistic distribution of the relative *global accuracy*.

Proof. To establish formulas (2.3)–(2.5), it is sufficient to notice that N random Bernoulli variables $(Y_n)_{n=1,N}$ defined by (2.1) are independent.

Indeed, (2.3)–(2.5) is a direct generalization of the binomial law for the variable S_N we would have to consider if all the Bernoulli variables $(Y_n)_{n=1,N}$ had been defined by the same probability p, given by

$$p \equiv \text{Prob}\{Y_n = 1\} = \text{Prob}\{X^{(m)}(h_n) \leqslant X^{(k)}(h_n)\} \quad \forall n = 1 \text{ to } N.$$

This completes the proof. □

Remark 2.2. From Proposition 2.1 one can also get the *cumulated* distribution of the number of meshes for which P_m *is more accurate than* P_k. It corresponds to the probability such that at least n meshes, $n = 1, \ldots, N$, are such that P_m *is more accurate than* P_k.

Using the definition of the random variable S_N given by (2.2), this cumulated distribution corresponds to the probabilities defined by

$$\forall n_e = 1, \dots, N: \quad \text{Prob}\{S_N \geqslant n_e\} = \sum_{j=n_e}^{N} \text{Prob}\{S_N = j\}. \tag{2.6}$$

Formula (2.6) is not easy to express explicitly, due to formulas (2.3)–(2.5). However, in the next proposition, we will prove a recurrence relation which allows us to determine an algorithm to compute the exact probabilistic distribution as well as the cumulated one.

To this end, let p_N denote the probability defined by $p_N \equiv \text{Prob}\{Y_N = 1\}$. Then, we have the following result.

Proposition 2.2. Let $(\mathcal{T}^{(n)})_{n=1,N}$ be a family of N regular meshes composed of simplexes, each mesh $\mathcal{T}^{(n)}$ being characterized by its mesh size h_n. Then, we have

$$\forall\, n_e = 1, N: \quad \text{Prob}\{S_N = n_e\} = p_N \,\text{Prob}\{S_{N-1} = n_e - 1\} + (1 - p_N)\,\text{Prob}\{S_{N-1} = n_e\}. \tag{2.7}$$

Proof. Formula (2.7) corresponds to the decomposition of the event $\{S_N = n\}$ into two independent events

$$\text{event}\,(\{S_{N-1} = n_e - 1\} \cap \{Y_N = 1\}) \;\text{ and }\; \text{event}\,(\{S_{N-1} = n_e\} \cap \{Y_N = 0\}).$$

Then, (2.7) results from classical probabilistic property. □

Remark 2.3. Proposition 2.2 enables us to process the computation of the exact probabilistic law which corresponds to $(\text{Prob}\{S_N = n_e\})_{n_e=0,N}$, and therefore, the cumulated one too.

Indeed, first of all, let us notice that

$$\forall n = 1, N: \quad \text{Prob}\{S_n = n\} = \text{Prob}\left\{\sum_{l=1}^{n} Y_l = n\right\} = \prod_{l=1}^{n} \text{Prob}\{Y_l = 1\} = p_1 \dots p_n \tag{2.8}$$

as the Bernoulli variables $(Y_l)_{l=1,n}$, $\forall\, n = 1, N$, are independent.

Moreover, with the same arguments, we also notice that

$$\forall n = 1, N: \text{Prob}\{S_n = 0\} = (1 - p_1) \dots (1 - p_n). \tag{2.9}$$

Relations (2.8)–(2.9) directly give: $\text{Prob}\{S_n = 0\}$ and $\text{Prob}\{S_n = n\}$ $\forall n = 1, N$. Therefore, by the help of the recurrence relation (2.7) and relations (2.8)–(2.9), one can compute step by step the exact probabilistic law $(\text{Prob}\{S_N = n_e\})_{n_e=0,N}$ as follows:

$$\begin{aligned}
&\text{Step 1:} && \text{Prob}\{S_2 = 1\} \text{ by (2.7)}\\
&\text{Step 2:} && \text{Prob}\{S_3 = 1\},\ \text{Prob}\{S_3 = 2\} \text{ by (2.7)}\\
&&& \dots\\
&\text{Step N:} && \text{Prob}\{S_N = 1\},\ \dots,\ \text{Prob}\{S_N = N - 1\} \text{ by (2.7)}
\end{aligned}$$

Remark 2.4. The cumulated probabilistic distribution is then a direct consequence of Proposition 2.2 due to formula (2.6). Finally, in the particular case corresponding to $n = 1$, we can derive an explicit expression of $\text{Prob}\{S_N \geqslant 1\}$. This is the purpose of the next proposition.

Proposition 2.3. Let N be the total number of meshes which belong to a family of regular meshes $(\mathcal{T}^{(n)})_{n=1,N}$ made of simplexes, and h_n denote the mesh size of a given mesh $\mathcal{T}^{(n)}$. We assume that $N = N_1 + N_2$, where N_1 is the number of meshes such that $h_n \leqslant h^*$ and N_2 those such that $h_n > h^*$. Then, we have

$$\text{Prob}\{S_N \geqslant 1\} = 1 - \frac{1}{2^{N_1}} \left[\frac{h_1 \dots h_{N_1}}{h^{*N_1}}\right]^{m-k} \left[1 - \frac{1}{2}\left(\frac{h^*}{h_{N_1+1}}\right)^{m-k}\right] \dots \left[1 - \frac{1}{2}\left(\frac{h^*}{h_N}\right)^{m-k}\right]. \tag{2.10}$$

Proof. By definition of the opposite event of $\{S_N \geqslant 1\}$, we can directly write

$$\text{Prob}\{S_N \geqslant 1\} = 1 - \text{Prob}\{S_N = 0\} = 1 - (1 - \mathcal{P}(h_1)) \cdots (1 - \mathcal{P}(h_N)) \tag{2.11}$$

where expression (2.3) was used.

So, taking into account the probability law (1.14) of Theorem 1.1 we can make explicit the probability such that on N meshes, at least one mesh satisfies P_m *is more accurate than* P_k. Indeed, we conclude by using the definitions of N_1 and N_2 and the corresponding expressions in the probabilistic law (1.14) to get (2.10). □

Remark 2.5. As we recall before, when $k < m$, it is often sought that the P_m finite element is more accurate than the P_k one. However, the probability given by (2.10) shows that even the event 'There is a least one mesh among N meshes such that P_m is more accurate than P_k' is not a sure event. Nevertheless, one can prove that this event is asymptotically sure. It is the purpose of the next proposition based on the following lemma.

Lemma 2.1. *Let β be a real number such that $0 < \beta < 1$. Let p be a given integer and $(x_n)_{n\geqslant p}$ be a sequence of real numbers satisfying $0 < x_n \leqslant \beta \ \ \forall\, n \geqslant p$. Then, the sequence product $(\Pi_N)_{N\in\mathbb{N}}$ defines by $\Pi_N \equiv \prod_{n=p}^{N} x_n$ converges to 0 when N goes to $+\infty$.*

Proof. As the sequence $(x_n)_{n\geqslant p}$ belongs to the interval $]0, \beta]$ such that $0 < \beta < 1$, we can write

$$0 < \frac{\Pi_{N+1}}{\Pi_N} = x_{N+1} \leqslant \beta < 1. \tag{2.12}$$

Therefore, Π_N is a decreasing sequence, bounded from below by 0, so it converges.

To compute the limit of the sequence Π_N let us consider the inequalities

$$0 < x_p \dots x_N \leqslant \left(\max_{p\leqslant n\leqslant N} x_n \right)^{N-p+1} \leqslant \beta^{N-p+1} < 1 \tag{2.13}$$

since we assume that β belongs to $]0, 1[$.

As a consequence, the sequence $\left(\beta^{N-p+1}\right)_{N\in\mathbb{N}}$ goes to 0 when N goes to infinity, and due to squeeze theorem [8], the sequence Π_N too. □

Now we can formulate the following proposition.

Proposition 2.4. Let $(\mathcal{T}^{(n)})_{n=1,N}$ be a given family of regular meshes and $(h_n)_{n=1,N}$ the corresponding sequence of mesh sizes. We assume that there exists $h_{\max} \in R_+^*$ such that

$$\forall n > 0 : \ \ h_n \leqslant h_{\max} \ \text{ such that } \ 0 < \mathcal{P}(h_{\max}) < 1. \tag{2.14}$$

Let also S_N be the random variable introduced in (2.2). Then, we have

$$\lim_{N\to+\infty} \text{Prob}\,\{S_N \geqslant 1\} = \lim_{N\to+\infty} \sum_{n_e=1}^{N} \text{Prob}\,\{S_N = n_e\} = 1. \tag{2.15}$$

Proof. We consider the expression of Prob $\{S_N \geqslant 1\}$ given by (2.11) and introduce the sequence $(x_n)_{n=1,N}$ defined by $x_n \equiv 1 - \mathcal{P}(h_n)$.

Then, one can write Prob $\{S_N \geqslant 1\}$ as follows

$$\text{Prob}\,\{S_N \geqslant 1\} = 1 - \prod_{n=1}^{N} x_n. \tag{2.16}$$

Now, one can check from formula (1.14) that

$$x_n = 1 - \mathcal{P}(h_n) \in]0, 1[\quad \forall\, n = 1, N$$

since $\mathcal{P}(h_n) \neq 0$ and $\mathcal{P}(h_n) \neq 1 \ \ \forall\, h_n$, $0 < h_n \leqslant h_{\max}$, as we assume (2.14).

Consequently, as the probabilistic law $\mathcal{P}(h)$ defined by (1.14) is decreasing on $\mathbb{R}_+$, we have

$$\mathcal{P}(h_n) \geqslant \mathcal{P}(h_{\max}) \quad \forall n = 1, N \tag{2.17}$$

and finally, there exists a real number $\beta \equiv 1 - \mathcal{P}(h_{\max})$, $0 < \beta < 1$, such that

$$x_n = 1 - \mathcal{P}(h_n) \leqslant \beta < 1 \quad \forall\, n = 1, N.$$

Therefore, from Lemma 2.1 we obtain the claimed asymptotic behavior of Prob $\{S_N \geqslant 1\}$. □

Remark 2.6. We notice that in Proposition 2.4 the assumption (2.14) corresponds to any concrete application where only finite mesh sizes are considered.

The idea of Proposition 2.4 is that, for a large number of meshes, the probability to get at least one mesh such that P_m *is more accurate than* P_k goes to 1 with the number of meshes. This property could be taken into account for the adaptive mesh refinement, where a family of meshes is built to handle the large variations of the approximate solution.

In other words, this result indicates that for a large number of meshes, one has to consider that P_m finite elements will surely be more accurate, at least on one mesh. However, it is not a sufficient indication to motivate the implementation of these finite elements which are more expensive than the P_k finite elements when $k < m$.

Furthermore, to apply this approach to adaptive mesh refinement, one needs to transfer the results of this section which concern *global* behaviors of a family of meshes to get *local* information to refine a given mesh, mainly depending on the local gradient of the approximated solution. This is the purpose of the next section.

3 From a global to a local accuracy comparison of finite elements

3.1 A local probability law to compare accuracy of finite elements

In the previous sections, we described a probabilistic approach to estimate for a given mesh $\mathcal{T}_h$ or for a sequence of meshes $(\mathcal{T}^{(n)})_{n=1,N}$, the relative global accuracy between two Lagrange finite elements P_k and P_m, $k < m$. We obtained laws of probability that depend on the mesh size h, namely the size of the largest diameter in the mesh. Accordingly, the results we got are *global* and do not explicitly consider any local behavior that could play an important role, particularly when one considers adaptive mesh refinement.

Our purpose is now to derive a *local* comparison tool between two Lagrange finite elements. This approach will be possible if one recalls that the error estimate (1.4), deduced from Bramble Hilbert's lemma, is elaborated with two main ingredients. The first one is Cea's lemma and the second one is the interpolation error (see [7]).

So, for any simplex K belonging to a regular mesh $\mathcal{T}_h$, let us introduce $\Pi_K^{(k)}$, the local K-Lagrange interpolation operator of degree k, that defines the local interpolation by the help of polynomials of degree lower than or equal to k on K.

Then, one can write the global interpolation error $\|u - \Pi_h u\|_{1,\Omega}$ as follows:

$$\|u - \Pi_h u\|_{1,\Omega} = \left(\sum_{K \in \mathcal{T}_h} \|u - \Pi_K^{(k)} u\|_{1,K}^2 \right)^{1/2} \tag{3.1}$$

where Π_h denotes the global Lagrange interpolation operator on the mesh $\mathcal{T}_h$.

So, under the conditions of Lemma 1.1, for each $K \in \mathcal{T}_h$, to obtain the next local estimate, we follow P. A. Raviart and J. M. Thomas, (see [7]), to get

$$\|u - \Pi_K^{(k)} u\|_{1,K} \leqslant \mathscr{C}_k^{'} \, h_K^k \, |u|_{k+1,K} \leqslant \mathscr{C}_k^{'} \, h_K^k \, |u|_{k+1,\Omega} \tag{3.2}$$

where h_K denotes the diameter of the simplex K and $\mathscr{C}_k^{'}$ a positive constant which does not depend on K and h_K, but depends on the reference element defining the Lagrange P_k finite element [7].

Then, the quantity $\mathscr{C}_k^{'} \, |u|_{k+1,\Omega}$ does not depend on K either. This independency will be further crucial when we will extend our results to the mesh refinement process.

Moreover, as a consequence of Céa's lemma, we can also consider the following estimate:

$$\|u - u_h\|_{1,\Omega} \leqslant \frac{M}{\alpha} \|u - \Pi_h u\|_{1,\Omega} \tag{3.3}$$

where M is the continuity constant and α the ellipticity constant of the bilinear form $a(\cdot, \cdot)$.

Now, due to (3.1) and (3.3), we highlight that the accuracy of a given finite element P_k can be *locally* characterized by estimate (3.2).

As a consequence, we define the relative *local accuracy* between two finite elements P_k and P_m as the relative local interpolation accuracy on a given simplex K as follows.

Definition 3.1. Let P_k and P_m, $k < m$, be two Lagrange finite elements and K a given simplex which belongs to $\mathcal{T}_h$. We will say that P_m is *locally* more accurate than P_k on K if

$$\|u - \Pi_K^{(m)} u\|_{1,K} \leqslant \|u - \Pi_K^{(k)} u\|_{1,K}. \tag{3.4}$$

Therefore, if we assume that the exact solution u of the variational formulation **(VP)** belongs to $H^{m+1}(\Omega)$, we can write inequality (3.2) for both of the $\Pi_K^{(i)}$ local operators, $i = k$ or $i = m$, and we have

$$\|u - \Pi_K^{(k)} u\|_{1,K} \leqslant \mathscr{C}_k^{'} h_K^k |u|_{k+1,\Omega} \tag{3.5}$$

$$\|u - \Pi_K^{(m)} u\|_{1,K} \leqslant \mathscr{C}_m^{'} h_K^m |u|_{m+1,\Omega}\,. \tag{3.6}$$

Setting $C_k^{'} \equiv \mathscr{C}_k^{'} |u|_{k+1,\Omega}$ and $C_m^{'} \equiv \mathscr{C}_m^{'} |u|_{m+1,\Omega}$, inequalities (3.5) and (3.6) become

$$\|u - \Pi_K^{(k)} u\|_{1,K} \leqslant C_k^{'} h_K^k \tag{3.7}$$

$$\|u - \Pi_K^{(m)} u\|_{1,K} \leqslant C_m^{'} h_K^m \tag{3.8}$$

which are the twins of inequalities (1.7) and (1.8).

The main difference between (1.7)–(1.8) and (3.7)–(3.8) is the meaning of h. Here, in (3.7)–(3.8), h_K denotes the local diameter of the simplex K, whereas in (1.7)–(1.8), h is the involved maximum mesh size of $\mathcal{T}_h$.

As a consequence, if we introduce the random variables $X_K^{(i)}(h_K)$, $i = k$ or $i = m$ and $k < m$, defined by

$$X_K^{(i)}(h_K) \equiv \|u - \Pi_K^{(i)} u\|_{1,K} \tag{3.9}$$

we can directly get the probability of the event $\{X_K^{(m)}(h) \leqslant X_K^{(k)}(h)\}$ corresponding to *P_m is locally more accurate than P_k* on K specified in Definition 3.1 by adapting formulas (1.14) as follows.

Corollary 3.1. Let K be a given simplex of diameter h_K belonging to a given regular mesh $\mathcal{T}_h$. Let $u \in H^{k+1}(\Omega)$ be the solution to (1.1) and $u_h^{(i)}$, $i = k$ or $i = m$, $k < m$, the corresponding Lagrange finite element P_i approximation solutions to (1.2).

We assume the two corresponding random variables $X_K^{(i)}(h)$, $i = k$ or $i = m$, defined by (3.9) are uniformly distributed on $[0, C_i^{'} h_K^i]$, where $C_i^{'}$ are defined by (3.7)–(3.8). Then, the probability such that P_m is *locally* more accurate than P_k on K is given by:

$$\text{Prob}\left\{X_K^{(m)}(h) \leqslant X_K^{(k)}(h_K)\right\} = \begin{cases} 1 - \dfrac{1}{2}\left(\dfrac{h_K}{h^*}\right)^{m-k}, & 0 < h_K \leqslant h^* \\ \dfrac{1}{2}\left(\dfrac{h^*}{h_K}\right)^{m-k}, & h_K \geqslant h^* \end{cases} \tag{3.10}$$

where h^* is defined by (1.9), but where the constants $C_k^{'}$ and $C_m^{'}$ introduced in (3.7) and (3.8) replace C_k and C_m defined in (1.7) and (1.8).

Remark 3.1. We notice that the corresponding value of h^* does not depend on the simplex K as we consider in inequality (3.2) the semi-norm of u in $H^{k+1}(\Omega)$, on the one hand, and as the constant $\mathscr{C}_k^{'}$, (due to $\mathscr{C}_k$), does not depend on K too, as we mention above, on the other hand.

Therefore, formula (3.10) gives us an evaluation of the *local* accuracy comparison between two finite elements P_k and P_m based on the *local* comparison accuracy between the corresponding K-Lagrange interpolation errors of degrees k and m.

We are now in position to extend this local result to a sequence of simplexes which belongs to a fixed mesh and which are related to adaptive mesh refinement process. This is the purpose of the next subsection.

3.2 Toward applications for adaptive mesh refinement

We now consider a *given* mesh $\mathcal{T}_h$ composed of N simplexes K_μ whose diameters are denoted by $(h_\mu)_{\mu=1,N}$.

For each simplex K_μ, $1 \leqslant \mu \leqslant N$, we consider the probability law of the event 'P_m *is locally more accurate than* P_k on K_μ' which is given by (3.10). Now, as in Section 2, let us introduce the N Bernoulli random independent variables $(Y_\mu)_{\mu=1,N}$ defined by (2.1) where we replace $X^{(i)}(h)$, $i = k$ or $i = m$, by $X^{(i)}_{K_\mu}(h)$, and also the corresponding random variable S_N defined by (2.2).

Thanks to the total similitude between the mathematical formalism of Section 2 and the present one, we directly get the analogous formulas of (2.3)–(2.5) and (2.10) but with a total different meaning. Particularly, by adapting (2.10) of Proposition 2.3 to the current situation, we get the following result.

Proposition 3.1. Let us denote by N the total number of simplexes of a given mesh $\mathcal{T}_h$. Assuming that $N = N_1 + N_2$, where N_1 is the number of simplexes satisfying $h_\mu \leqslant h^*$ and N_2 the number of simplexes such that $h_\mu > h^*$. Then, we have

$$\text{Prob}\,\{S_N \geqslant 1\} = 1 - \frac{1}{2^{N_1}} \left[\frac{h_1 \dots h_{N_1}}{h^{*N_1}}\right]^{m-k} \left[1 - \frac{1}{2}\left(\frac{h^*}{h_{N_1+1}}\right)^{m-k}\right] \dots \left[1 - \frac{1}{2}\left(\frac{h^*}{h_N}\right)^{m-k}\right]. \tag{3.11}$$

Remark 3.2. We would like to highlight that even if the formalism is totally equivalent between this section and Section 2, one has to carefully distinguish the different meanings of the two situations.

Indeed, here the Bernoulli variables $(Y_\mu)_{\mu=1,N}$ determine, for a given elementary simplex K_μ of a given mesh, if P_m is *locally* more accurate than P_k on K_μ, while the Bernoulli variables introduced in the previous section characterize, for a given mesh belonging to a sequence of meshes, if P_m is *globally* more accurate than P_k on this mesh.

Proposition 3.1 shows again that, albeit if it is usually assumed that finite elements P_m are more accurate than P_k, $k < m$, formula (3.11) highlights that even *to get at least one simplex on* N *such that* P_m *is locally more accurate than* P_k is not a sure event, as its probability is different from one. Moreover, one can get here a result similar to Proposition 2.4, after modifying our framework to the adaptive mesh refinement process.

So, let us describe what happens for a given mesh $\mathcal{T}_h$, where adaptive mesh refinement is applied, to improve the computed solution in steep gradient areas. For our purpose, we distinguish in $\mathcal{T}_h$ two kinds of simplexes. Those who are not going to be changed, and those which will be refined. Let us denote by N', $N' \leqslant N$, the number of new simplexes created in the mesh by the refinement process.

We also introduce N_1, $N_1 \leqslant N'$, the number of new simplexes such that $h_\mu \leqslant h^*$ and $N_2 = N' - N_1$ the rest of new simplexes. Therefore, the random variable S_N defined by (2.2) becomes $S_{N'}$, and describes now the total number of new simplexes in the refinement process such that P_m is locally more accurate than P_k. We remark that these considerations make sense, since h^* does not depend on the simplexes K involved in the mesh $\mathcal{T}_h$ (see Remark 3.1).

Then, our interest is to determine the behavior of Prob $\{S_{N'} \geqslant 1\}$, (equivalently determined by (3.11) if one changes N by N'), when N' goes to infinity, namely, where the number of new simplexes N' concerned by the refinement process becomes large. This situation corresponds to the framework of adaptive mesh refinement where one is usually interested by locally refining the mesh: this is performed by identifying in the mesh the areas such that the gradient of the approximated solution is large.

Therefore, we are now in position to determine for a given set of new simplexes N' which becomes large, the asymptotic probability such that *there exists at least one new simplex where* P_m *is locally more accurate than* P_k.

This is the purpose of the following proposition which is the twin of Proposition 2.4.

Proposition 3.2. Let $\mathcal{T}_h$ be a given mesh composed of N simplexes and let N', $N' \leqslant N$, be the number of new simplexes obtained by mesh refinement. We also assume that there exists $h_{\max} \in \mathbb{R}_+^*$ which satisfies the equivalent condition of (2.14) for all the diameters h_μ, $\mu = 1, N'$, defining the new simplexes. Then we have

$$\lim_{N' \to +\infty} \text{Prob}\,\{S_{N'} \geqslant 1\} = 1. \tag{3.12}$$

This result is not surprising due to the total similarity between formulas (2.10) and (3.11), (if one replaces N by N' in (3.11)).

Again, it shows that one has to carefully consider the relative local accuracy between P_m and P_k Lagrange finite elements, $k < m$, except if locally, the number of simplexes becomes very large.

However, for a fixed number N' of new simplexes, this phenomena is more pronounced depending on the minimum number of simplexes n'_e, $n'_e = 1, N'$, satisfying P_m would be more accurate than P_k. Unfortunately, deriving the analogous formulas of (2.3)–(2.5) and (2.6) for a sequence of simplexes, to explicit the probability of the event $\{S_{N'} \geqslant n'_e\}$ is inextricable.

So, one can compare for two different values of n'_e the behavior of the corresponding probabilities. This is the purpose of the following proposition.

Proposition 3.3. Let $n'_{e,1}$ and $n'_{e,2}$ be two integers such that $1 \leqslant n'_{e,1} < n'_{e,2} \leqslant N'$, and let $S_{N'}$ be the random variable equivalently defined as (2.2). Then, we have

$$\text{Prob}\left\{S_{N'} \geqslant n'_{e,2}\right\} \leqslant \text{Prob}\left\{S_{N'} \geqslant n'_{e,1}\right\}. \tag{3.13}$$

Proof. The proof results from the following identity

$$\text{Prob}\left\{S_{N'} \geqslant n'_{e,1}\right\} = \sum_{n'_e = n'_{e,1}}^{n'_{e,2}-1} \text{Prob}\left\{S_{N'} = n'_e\right\} + \text{Prob}\left\{S_{N'} \geqslant n'_{e,2}\right\} \tag{3.14}$$

and consequently, (3.13) holds. □

Therefore, Proposition 3.3 clearly indicates that the larger n'_e the less the probability such that at least n'_e simplexes satisfy P_m is locally more accurate than the P_k.

This points out that we cannot easily disqualify the P_k finite element in comparison with the P_m one, particularly for a mesh refinement process.

Remark 3.3. Notice that it is also possible to compute the exact distribution and the associated cumulated distribution, by applying the same principles as those applied to derive formula (2.7) of Proposition 2.2 and formulas (2.8) and (2.9).

4 Discussion and conclusion

In this paper, we proposed a new geometrical-probabilistic approach to evaluate the relative accuracy between two Lagrange finite elements P_k and P_m, $k < m$. Basically, we distinguished two cases: a global approach and a local one. Both cases are based on a probabilistic interpretation of the error estimate one can derive from Bramble–Hilbert lemma.

The global approach is based on the probabilistic law (1.14) we derived in [4], from which a first extension to a family of meshes is proposed. Regarding the local accuracy between two finite elements, we recall the two main components required to establish the *a priori* error estimate and we highlight that it is centrally based on the local interpolation error. This leads us to transpose our global analysis to the local one to get the corresponding probability distribution (3.10).

Afterwards, we extended our global and local results to two principal applications, both of them are concerned with adaptive mesh refinement. The global probability law has been used to describe the case of a family of meshes while the local probabilistic one helped us to treat a fixed mesh composed of a sequence of simplexes.

These results strengthen those we got in [4] and show that even if we consider a family of meshes (respectively, a sequence of simplexes), the event '*to get at least one mesh (respectively, at least one simplex) such that P_m is more globally (respectively, locally) accurate than P_k*' is not a sure event, (cf. Propositions 2.3, 3.1, and Proposition 3.3 for a more general case).

However, it is proved in Propositions 2.4 and 3.2 that, for a great number of meshes (respectively, a great number of simplexes), this event is asymptotically sure. We also proved a recurrence relation, see (2.2), in the case of a family of meshes, which can be adapted to the case of a sequence of simplexes. This enables to compute, for example, the probability such that *at least 50 percents of meshes (resp., at least 50 percents of simplexes) are such that P_m is more globally (resp., locally) accurate than P_k*.

Finally, we have to mention that for all concrete applications, one will have to precisely estimate the critical value h^*. Indeed, all the probabilistic laws we derived are based on formulas (1.14) and (3.10) which depend on h^*. Since h^* strongly depends on the semi-norm $H^{k+1}(\Omega)$ of the exact solution u to the variational problem, all the available techniques which belong to *a priori* estimate theory of solutions to partial differential equations will be involved to evaluate h^*.

We are presently working on numerical examples for which (nearly) everything could be computed. This allows us to make our results more concrete. Basically, the method is based on the determination of statistical estimators, that will lead us to numerically specify the value of h^*.

Acknowledgment: The authors want to warmly dedicate this research to pay homage to the memory of Professors André Avez and Gérard Tronel who largely promote the passion of research and teaching in mathematics.

References

[1] F. Assous and J. Chaskalovic, Data mining techniques for scientific computing: Application to asymptotic paraxial approximations to model ultra-relativistic particles, *J. Comput. Phys.* **230**, (2011), 4811–4827.
[2] J. Chaskalovic, Mathematical and numerical methods for partial differential equations, Springer Verlag, 2013.
[3] J. Chaskalovic and F. Assous, A new probabilistic interpretation of Bramble–Hilbert lemma, *Computational Methods in Applied Mathematics* **20** (2018), No. 1, 79–87.
[4] J. Chaskalovic and F. Assous, A new probabilistic interpretation of Bramble–Hilbert lemma, October 2018. *arXiv: 1803.09547* [math.NA].
[5] J. Chaskalovic and F. Assous, Asymptotic relation between probabilistic distributions which quantify the relative accuracy between two finite elements, October 2018. *arXiv: 1803.09552* [math.NA].
[6] P. G. Ciarlet, Basic error estimates for elliptic problems. In: *Handbook of Numerical Analysis, Vol. II* (Eds. P. G. Ciarlet and J. L. Lions), North Holland, Amsterdam, 1991.
[7] P. A. Raviart et J. M. Thomas, Introduction à l'analyse numérique des équations aux dérivées partielles, Masson, Paris, 1982.
[8] J. Stewart, Multivariable Calculus, Cengage Learning Publisher, 2008.