

## Research Article

Joël Chaskalovic\* and Franck Assous

# A New Probabilistic Interpretation of the Bramble–Hilbert Lemma

<https://doi.org/10.1515/cmam-2018-0270>

Received October 25, 2018; revised November 12, 2018; accepted December 20, 2018

**Abstract:** The aim of this paper is to provide new perspectives on relative finite element accuracy which is usually based on the asymptotic speed of convergence comparison when the mesh size  $h$  goes to zero. Starting from a geometrical reading of the error estimate due to the Bramble–Hilbert lemma, we derive two probability distributions that estimate the relative accuracy, considered as a random variable, between two Lagrange finite elements  $P_k$  and  $P_m$  ( $k < m$ ). We establish mathematical properties of these probabilistic distributions and we get new insights which, among others, show that  $P_k$  or  $P_m$  is more likely accurate than the other, depending on the value of the mesh size  $h$ .

**Keywords:** Error Estimates, Finite Elements, Bramble–Hilbert Lemma, Probability

**MSC 2010:** 65N15, 65N30, 65N75

## 1 Introduction

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. However, in these error estimates, an unknown 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 the precise value of the approximation error is generally unknown. Due to quantitative uncertainties which are generated in the process of the mesh generator and, as a consequence, in the corresponding approximation too, it gave us the idea of considering the approximation error as a random variable. Therefore, we were able to evaluate the probability of the difference between two approximation errors corresponding to two different finite elements, and then, we got a probabilistic way to compare the relative accuracy between these two finite elements.

The paper is organized as follows. We recall in Section 2 the mathematical problem we consider and a corollary of the Bramble–Hilbert lemma to propose a geometrical interpretation of the error estimate which appears in this lemma. In Section 3 we derive two probability distributions to interpret and estimate the relative accuracy, considered as a random variable, between two Lagrange finite elements  $P_k$  and  $P_m$  ( $k < m$ ). Several mathematical properties of these probabilistic distributions are established in Section 4. Concluding remarks follow.

---

\*Corresponding author: Joël Chaskalovic, D'Alembert, Sorbonne University, Paris, France, e-mail: jch1826@gmail.com.

<https://orcid.org/0000-0003-1263-5313>

Franck Assous, Department of Mathematics, Ariel University, 40700 Ariel, Israel, e-mail: franckassous55@gmail.com

## 2 The Problem Model and a Geometrical Interpretation of an Error Estimate

Let  $\Omega$  be an open bounded, and nonempty subset of  $\mathbb{R}^n$  and  $\Gamma$  its boundary which we assumed to be  $C^1$ -piecewise, and let  $u$  be the solution to the second order elliptic variational formulation:

$$\text{Find } u \in V \text{ solution to } a(u, v) = l(v) \text{ for all } v \in V, \quad (\mathbf{VP})$$

where  $V$  is a given Hilbert space endowed with a 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 solution  $u \in V$  (see for example [4]). In this paper and for simplicity, we will restrict ourselves to the case where  $V$  is a usual Sobolev space of distributions. Let us also consider an approximation  $u_h$  of  $u$ , solution to the approximate variational formulation:

$$\text{Find } u_h \in V_h \text{ solution to } a(u_h, v_h) = l(v_h) \text{ for all } v_h \in V_h, \quad (\mathbf{VP})_h$$

where  $V_h$  is a given finite-dimensional subset of  $V$ . To state a corollary of Bramble–Hilbert’s lemma and a corresponding error estimate, we follow [6] or [5], and we assume that  $\Omega$  is exactly recovered by a mesh  $\mathcal{T}_h$  composed by  $N_K$   $n$ -simplices  $K_\mu$  ( $1 \leq \mu \leq N_K$ ), which respect classical rules of regular discretization (see for example [4] for the bidimensional case and [6] in  $\mathbb{R}^n$ ). Moreover, we denote by  $P_k(K_\mu)$  the space of polynomial functions defined on a given  $n$ -simplex  $K_\mu$  of degree less than or equal to  $k$  ( $k \geq 1$ ). Then we have the following result:

**Lemma 2.1.** *Suppose that there exists an integer  $k \geq 1$  such that the approximation  $u_h$  of  $V_h$  is a continuous piecewise function composed by polynomials which belong to  $P_k(K_\mu)$  ( $1 \leq \mu \leq N_K$ ). Then  $u_h$  converges to  $u$  in  $H^1(\Omega)$ :*

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

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

$$\|u_h - u\|_{1,\Omega} \leq \mathcal{C}_k h^k |u|_{k+1,\Omega}, \quad (2.1)$$

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

Let us now 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 (2.1), assuming that the solution  $u$  to **(VP)** belongs to  $H^{m+1}(\Omega)$ , are

$$\|u_h^{(k)} - u\|_{1,\Omega} \leq \mathcal{C}_k h^k |u|_{k+1,\Omega}, \quad (2.2)$$

$$\|u_h^{(m)} - u\|_{1,\Omega} \leq \mathcal{C}_m h^m |u|_{m+1,\Omega}, \quad (2.3)$$

where  $u_h^{(k)}$  and  $u_h^{(m)}$  denotes the  $P_k$  and  $P_m$  Lagrange finite element approximations of  $u$ , respectively. Now, if one considers a given mesh for the finite element of  $P_m$  which would contains those of  $P_k$ , then, for the particular class of problems where **(VP)** is equivalent to a minimization formulation **(MP)** (see for example [4]), one can show that the approximation error of  $P_m$  is always lower than those of  $P_k$ , and  $P_m$  is more accurate than  $P_k$  for all values of the mesh size  $h$  corresponding to the largest diameter in the mesh  $\mathcal{T}_h$ . Then, for a given mesh size value of  $h$ , we consider two independent meshes for  $P_k$  and  $P_m$  built by a mesh generator. So, usually, to compare the relative accuracy between these two finite elements, one asymptotically considers inequalities (2.2) and (2.3) to conclude that, when  $h$  goes to zero,  $P_m$  finite element is more accurate than  $P_k$ , as  $h^m$  goes faster to zero than  $h^k$ . However, for any application  $h$  has a static fixed value and this way of comparison is not valid anymore. Therefore, our point of view will be to determine the relative accuracy between two finite elements  $P_k$  and  $P_m$  ( $k < m$ ), for any given value of  $h$  for which two independent meshes have to be considered. To this end, let us set

$$C_k = \mathcal{C}_k |u|_{k+1,\Omega} \quad \text{and} \quad C_m = \mathcal{C}_m |u|_{m+1,\Omega}.$$

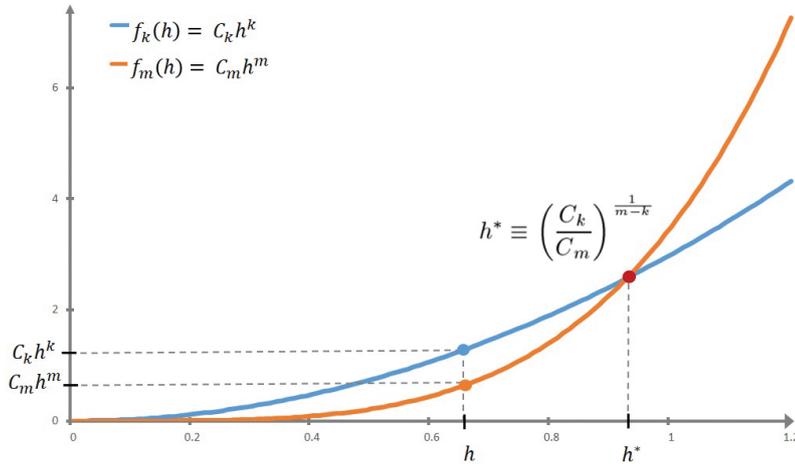


Figure 1: Curves  $f_k$  and  $f_m$  and existence domain of  $\|u_h^{(i)} - u\|_{1,\Omega}$ ,  $i = k$  or  $i = m$ .

Therefore, instead of (2.2) and (2.3), we consider in the sequel the two next inequalities:

$$\|u_h^{(k)} - u\|_{1,\Omega} \leq C_k h^k, \tag{2.4}$$

$$\|u_h^{(m)} - u\|_{1,\Omega} \leq C_m h^m. \tag{2.5}$$

Let us remark that inequalities (2.4) and (2.5) show that the two polynomial curves defined by  $f_k(h) \equiv C_k h^k$  and  $f_m(h) \equiv C_m h^m$  play a critical role regarding the values of the two norms  $\|u_h^{(k)} - u\|_{1,\Omega}$  and  $\|u_h^{(m)} - u\|_{1,\Omega}$ . More precisely, these inequalities indicate that the norm  $\|u_h^{(k)} - u\|_{1,\Omega}$  (respectively  $\|u_h^{(m)} - u\|_{1,\Omega}$ ) is below the curve  $f_k(h)$  (respectively below the curve  $f_m(h)$ ) (see Figure 1). As we are interested in comparing the relative positions of these curves, we introduce their intersection point  $h^*$  defined by

$$h^* \equiv \left(\frac{C_k}{C_m}\right)^{\frac{1}{m-k}}. \tag{2.6}$$

Now, as often in numerical analysis, there is no *a priori* information to surely or better specify the relative distance between  $\|u_h^{(k)} - u\|_{1,\Omega}$  (respectively  $\|u_h^{(m)} - u\|_{1,\Omega}$ ), and the curve  $f_k$  or its precise value in the interval  $[0, C_k h^k]$  (respectively the curve  $f_m$  and the interval  $[0, C_m h^m]$ ). Moreover, we have to deal with finite element methods that return quantitative uncertainties in their calculations. This mainly comes from the way the mesh grid generator will process the mesh to compute the approximation  $u_h^{(k)}$ , leading to a partial noncontrol of the mesh, even for a given maximum mesh size. As a consequence, the corresponding grid is *a priori* random, and the corresponding approximation  $u_h^{(k)}$  too. For all of these reasons, we motivate that a probabilistic approach can provide a coherent framework for modeling quantitative uncertainties in finite element approximations. This is the purpose of the following section where we will establish two probability distributions which will allowed us to estimate the relative accuracy between two Lagrange finite elements.

### 3 The Two Probabilistic Models for Relative Finite Elements Accuracy

In this section, we will introduce a convenient probabilistic framework to consider the possible values of the norm  $\|u_h^{(k)} - u\|_{1,\Omega}$  as a random variable defined as follows:

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

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

$$X^{(k)} : \Omega \rightarrow [0, C_k h^k], \quad \omega \equiv u_h^{(k)} \mapsto X^{(k)}(\omega) = X^{(k)}(u_h^{(k)}) = \|u_h^{(k)} - u\|_{1,\Omega}. \quad (3.1)$$

In the sequel, for simplicity, we will set  $X^{(k)}(u_h^{(k)}) \equiv X^{(k)}(h)$ .

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  $[0, C_k h^k]$ . So, our interest is to evaluate the probability of the event

$$\{\|u_h^{(m)} - u\|_{1,\Omega} \leq \|u_h^{(k)} - u\|_{1,\Omega}\} \equiv \{X^{(m)}(h) \leq X^{(k)}(h)\},$$

which will allow us to estimate the relative accuracy between two finite elements of order  $k$  and  $m$  ( $k < m$ ). To proceed it, let us now introduce the two random events  $A$  and  $B$  as follows:

$$A \equiv \{\|u_h^{(m)} - u\|_{1,\Omega} \leq \|u_h^{(k)} - u\|_{1,\Omega}\}, \quad (3.2)$$

$$B \equiv \{\|u_h^{(k)} - u\|_{1,\Omega} \in [C_m h^m, C_k h^k]\}. \quad (3.3)$$

Then we have the following lemma:

**Lemma 3.1.** *Let  $A$  and  $B$  be the events defined by (3.2) and (3.3). Then we have*

$$\text{Prob}\{A\} = \frac{\text{Prob}\{B\}}{\text{Prob}\{B/A\}} \quad \text{for all } h < h^*. \quad (3.4)$$

*Proof.* Let us use the following splitting:

$$\text{Prob}\{A\} = \text{Prob}\{A \cap B\} + \text{Prob}\{A \cap \bar{B}\}, \quad (3.5)$$

where  $\bar{B}$  denotes the opposite event of  $B$ . Now, by the definition of the conditional probability we have

$$\text{Prob}\{A \cap B\} = \text{Prob}\{A/B\} \cdot \text{Prob}\{B\} = \text{Prob}\{B\},$$

since the probabilistic interpretation of the Bramble–Hilbert lemma in the case  $h < h^*$  corresponds to

$$\text{Prob}\{A/B\} = 1.$$

Then equation (3.5) can be written as

$$\text{Prob}\{A\} = \text{Prob}\{B\} + \text{Prob}\{A \cap \bar{B}\}, \quad (3.6)$$

which can be transformed by the help of the conditional probability as follows:

$$\text{Prob}\{A\} = \text{Prob}\{B\} + \text{Prob}\{\bar{B}/A\} \cdot \text{Prob}\{A\},$$

or equivalently,

$$\text{Prob}\{A\} = \frac{\text{Prob}\{B\}}{1 - \text{Prob}\{\bar{B}/A\}} = \frac{\text{Prob}\{B\}}{\text{Prob}\{B/A\}}, \quad (3.7)$$

which corresponds to (3.4).  $\square$

Then we have two options regarding the nature of the dependency between the events  $A$  and  $B$  which will lead us to get two different distribution laws of probabilities of the event  $\{X^{(m)}(h) \leq X^{(k)}(h)\}$ . The next two subsections are devoted to the dependency modeling between  $A$  and  $B$ .

### 3.1 The Two Steps Model

The first case we will consider states that since, a priori, no information is available in numerical analysis to consider any kind of dependency between the events  $A$  and  $B$ , we assume in this subsection that these events are independent.

**Corollary 3.2.** *Let  $A$  and  $B$  be the two events defined by (3.2) and (3.3) and let us assume they are independent. Then the probability distribution of the event  $\{X^{(m)}(h) \leq X^{(k)}(h)\}$  is given by*

$$\text{Prob}\{X^{(m)}(h) \leq X^{(k)}(h)\} = \begin{cases} 1 & \text{if } 0 < h < h^*, \\ 0 & \text{if } h > h^*. \end{cases} \quad (3.8)$$

*Proof.* As the events  $A$  and  $B$  are supposed independent, we have

$$\text{Prob}\{A/B\} = \text{Prob}\{A\}.$$

As a consequence, by Lemma 3.1 equation (3.7) gives after simplification

$$\text{Prob}\{A\} = 1 \quad \text{for all } h < h^*.$$

With the same kind of arguments, when  $h > h^*$  we get

$$\text{Prob}\{A\} = 0 \quad \text{for all } h > h^*. \quad \square$$

Let us now examine the main properties of probabilistic distribution (3.8):

- For any  $h$  smaller than  $h^*$ ,  $P_m$  finite element is not only asymptotically better than  $P_k$  finite element as  $h$  becomes small, but they are *almost surely* more accurate for all these values of  $h$  such that  $h < h^*$ .
- For any  $h$  greater than  $h^*$ ,  $P_k$  finite element becomes *almost surely* more accurate than  $P_m$  finite element, even if  $k < m$ .

This last feature upsets the widespread idea regarding the relative accuracy between  $P_k$  and  $P_m$  ( $k < m$ ) finite elements. It clearly indicates that there exist cases where  $P_m$  finite elements *surely* must be overqualified and a significant reduction of implementation and execution cost can be obtained without a loss of accuracy. Furthermore, one may expect to get a probabilistic distribution where more variations would appear, as it is in this two steps model, between the probability of the event  $\{X^{(m)}(h) \leq X^{(k)}(h)\}$  and the mesh size  $h$ . It is certainly due to the assumption we considered regarding the independency between the events  $A$  and  $B$ . The purpose of the next subsection we will be devoted to relax this assumption by directly computing the probability of the event  $\{X^{(m)}(h) \leq X^{(k)}(h)\}$ .

### 3.2 The “Sigmoid” Model

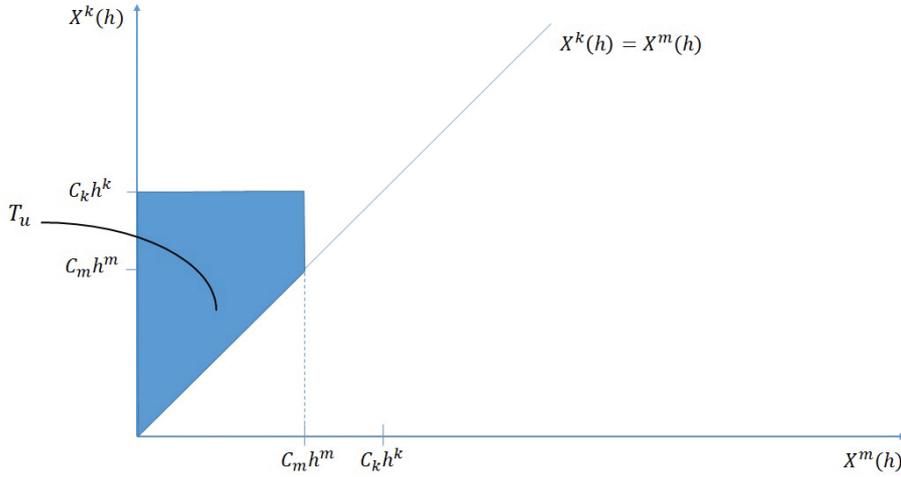
To avoid the hypothesis of independency between the events  $A$  and  $B$  defined by (3.2) and (3.3), we will directly evaluate the probability of the event  $A$  without considering anymore the splitting we wrote in formula (3.6). However, we will assume that the two random variables  $X^{(i)}(h)$  ( $i = k$  or  $i = m$ ) defined by (3.1) are independent and uniformly distributed on  $[0, C_i h^i]$  ( $i = k$  or  $i = m$ ). This is the aim of the following theorem.

**Theorem 3.3.** *Let  $u$  be the solution to the second order variational elliptic problem (VP) and  $u_h^{(i)}$  ( $i = k$  or  $i = m$ ,  $k < m$ ), the two corresponding Lagrange finite element  $P_i$  approximations, solution to the approximated formulation (VP) <sub>$h$</sub> . We assume the two corresponding random variables  $X^{(i)}(h)$  ( $i = k$  or  $i = m$ ) defined by (3.1) are independent and uniformly distributed on  $[0, C_i h^i]$ , where  $C_i$  are defined by (2.4)–(2.5). Then the probability of the event  $\{X^{(m)}(h) \leq X^{(k)}(h)\}$  is given by*

$$\text{Prob}\{X^{(m)}(h) \leq X^{(k)}(h)\} = \begin{cases} 1 - \frac{1}{2} \left( \frac{h}{h^*} \right)^{m-k} & \text{if } 0 < h \leq h^*, \\ \frac{1}{2} \left( \frac{h^*}{h} \right)^{m-k} & \text{if } h \geq h^*. \end{cases} \quad (3.9)$$

*Proof.* Let us first consider a fixed value of  $h$  such that  $h < h^*$ . In this case,  $f_m(h) < f_k(h)$ , or in other words,  $0 < C_m h^m < C_k h^k$  and due to the Bramble–Hilbert lemma (see Figure 1), one must deal with the following inequalities:

$$X^{(k)}(h) \leq C_k h^k \quad \text{and} \quad X^{(m)}(h) \leq C_m h^m < C_k h^k. \quad (3.10)$$



**Figure 2:** Area corresponding to  $X^{(m)}(h) \leq X^{(k)}(h)$ .

Then, to compute the probability such that  $X^{(m)}(h) \leq X^{(k)}(h)$ , we consider inequalities (3.10) in the plane  $(0; X^{(m)}(h), X^{(k)}(h))$  (see Figure 2) in which the two random variables belong to the rectangle  $R_t$  defined on  $[0, C_m h^m] \times [0, C_k h^k]$ . Our purpose is to characterize the points in  $R_t$  that satisfy  $X^{(m)}(h) \leq X^{(k)}(h)$ . Obviously, it only concerns the points which are above the bisector  $X^{(k)}(h) = X^{(m)}(h)$ , namely the points which belong to the trapezium  $T_u$  (see Figure 2) whose surface is given by

$$S(T_u) = C_m h^m (C_k h^k - C_m h^m) + \frac{C_m^2 h^{2m}}{2},$$

while the total surface of the rectangle  $R_t$  is equal to  $C_m C_k h^{m+k}$ . As we assume that the two random variables  $X^{(k)}(h)$  and  $X^{(m)}(h)$  are independent and uniformly distributed, the probability  $\text{Prob}\{X^{(m)}(h) \leq X^{(k)}(h)\}$  corresponds to the ratio between the two surfaces of  $T_u$  and  $R_t$  and we have

$$\begin{aligned} \text{Prob}\{X^{(m)}(h) \leq X^{(k)}(h)\} &= \frac{S(T_u)}{S(R_t)} = \frac{C_m h^m (C_k h^k - C_m h^m) + \frac{1}{2} C_m^2 h^{2m}}{C_m C_k h^{m+k}}, \\ &= 1 - \frac{1}{2} \frac{C_m}{C_k} h^{m-k}. \end{aligned}$$

Using the definition (2.6) of  $h^*$ , we get

$$\text{Prob}\{X^{(m)}(h) \leq X^{(k)}(h)\} = 1 - \frac{1}{2} \frac{C_m}{C_k} h^{m-k} = 1 - \frac{1}{2} \left( \frac{h}{h^*} \right)^{m-k} \quad \text{for all } h < h^*.$$

Let us consider now the second case where  $h > h^*$ . The curve  $f_m(h) = C_m h^m$  is above the curve  $f_k(h) = C_k h^k$  and by the same arguments we used above, one must deal with the following inequalities:

$$X^{(m)}(h) \leq C_m h^m \quad \text{and} \quad X^{(k)}(h) \leq C_k h^k < C_m h^m.$$

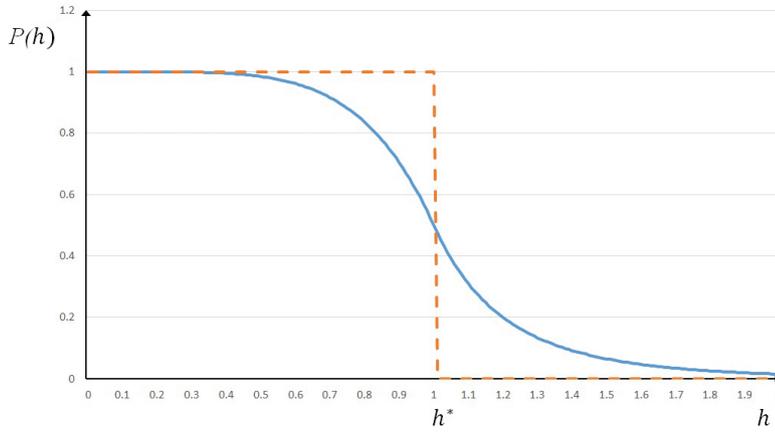
Then, if we change the role between  $k$  and  $m$ , we can directly write

$$\begin{aligned} \text{Prob}\{X^{(k)}(h) \leq X^{(m)}(h)\} &= \frac{C_k h^k (C_m h^m - C_k h^k) + \frac{1}{2} C_k^2 h^{2k}}{C_m C_k h^{m+k}} \\ &= 1 - \frac{1}{2} \frac{C_k}{C_m} h^{k-m}. \end{aligned}$$

Hence, the probability of the complementary event  $X^{(m)}(h) \leq X^{(k)}(h)$  which interests us is given by

$$\begin{aligned} \text{Prob}\{X^{(m)}(h) \leq X^{(k)}(h)\} &= 1 - \text{Prob}\{X^{(k)}(h) \leq X^{(m)}(h)\} \\ &= \frac{1}{2} \frac{C_k}{C_m} \cdot \frac{1}{h^{m-k}} = \frac{1}{2} \left( \frac{h^*}{h} \right)^{m-k}, \end{aligned}$$

where we used the definition (2.6) of  $h^*$ . □



**Figure 3:** Case  $m - k \neq 1$ : shape of the sigmoid distribution (3.9) and the two steps corresponding one (3.8), ( $P(h) \equiv \text{Prob}\{X^{(m)}(h) \leq X^{(k)}(h)\}$ ).

The global shapes of the two probabilistic distributions (3.8) and (3.9) are plotted in Figure 3 and particular features of (3.9) are described in the next section.

## 4 Properties of the sigmoid probability distribution

We give now the main properties of the sigmoid probability distribution given by (3.9). To this end, we will denote by  $\mathcal{P}(h)$  the probability defined by

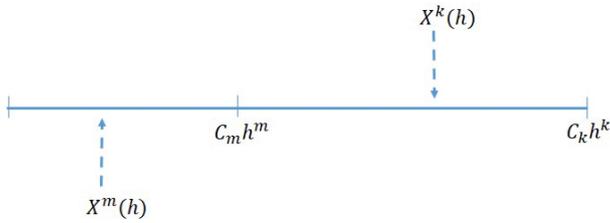
$$\mathcal{P}(h) \equiv \text{Prob}\{X^{(m)}(h) \leq X^{(k)}(h)\}.$$

**(1)** The first feature we observe concerns the global shape of  $P(h)$  together with (3.9), drawn in Figure 3 for  $m - k \neq 1$ , which looks like a kind of sigmoid roughly approximated by a stepwise function given by (3.8) from Lemma 3.1. In this way, we achieve our objective to relax the dependency assumption between the events  $A$  and  $B$ . As a consequence, nonlinearity appears in the relation described by (3.9) between the probability of the event “ $P_m$  finite element is more accurate than  $P_k$  finite element” and the mesh size  $h$ .

**(2) Behavior of  $\mathcal{P}(h)$  in the neighborhood of  $0^+$ .** Directly, we get

$$\lim_{h \rightarrow 0^+} \mathcal{P}(h) = \lim_{h \rightarrow 0^+} \text{Prob}\{X^{(m)}(h) \leq X^{(k)}(h)\} = 1, \quad (4.1)$$

which corresponds to the classical understanding of the error estimate in (2.1) which derives from the Bramble–Hilbert lemma, namely asymptotically when the maximum of the mesh size  $h$  goes to zero. Indeed, in these cases  $h$  “is sufficiently small”, and despite the unknown values of the constants  $C_k$  and  $C_m$  which appear in (2.4) and (2.5), one concludes as expected that the finite element  $P_m$  is more accurate than the finite element  $P_k$ , if  $k < m$ . But, the question is to determine what does it mean when  $h$  “is sufficiently small”. We will partially discuss about this in the next point regarding the behavior of  $\mathcal{P}(h)$  at the neighborhood of  $h^*$  given by (2.6). From a probabilistic point of view the result (4.1) is also intuitive because, when  $h$  goes to  $0^+$ , the quantity  $C_m h^m$  goes to 0 faster than  $C_k h^k$  ( $k < m$ ). Depicting the relative position of  $X^{(m)}(h)$  and  $X^{(k)}(h)$  in a one-dimensional way (see Figure 4), it is clear that the probability of the event  $\{X^{(m)} \leq X^{(k)}\}$  goes to 1 when  $h$  goes to zero, as  $X^{(m)} \leq C_m h^m$  due to the Bramble–Hilbert lemma. However, the interest of any probability distribution is to get additional information concerning the relative accuracy between two given finite elements, not only when  $h$  goes to zero, as we will see further. Here, we just mentioned that we find again the well known conclusion to compare two finite elements when the mesh size is arbitrarily small. Indeed, finite element  $P_m$  is not only *asymptotically* more accurate than  $P_k$  as  $k < m$ . Indeed, for all  $h \leq h^*$ , the probability for  $P_m$  to be more accurate than  $P_k$  is between 0.5 to 1. It means that  $P_m$  is *more likely*



**Figure 4:** Relative one-dimensional position between  $X^{(m)}$  and  $X^{(k)}$  ( $h < h^*$ ).

accurate than  $P_k$  for all of these values of  $h$ . We also notice that we have not anymore the event “ $P_m$  is more accurate than  $P_k$ ” as an almost sure event as we got in subsection 3.1 with the law (3.8). This is because we dropped the hypothesis of dependency between the events  $A$  and  $B$  which leads to a more general and realistic probabilistic distribution.

**(3) Behavior of  $\mathcal{P}(h)$  in the neighborhood of  $h^*$ .** The probabilistic stepwise law (3.8) did not describe the case  $h$  equals  $h^*$ . However, here, the sigmoid probability distribution (3.9) can be extended by continuity to  $h = h^*$  as we simply have

$$\lim_{h \rightarrow h^{*-}} \mathcal{P}(h) = \lim_{h \rightarrow h^{*+}} \mathcal{P}(h) = \frac{1}{2},$$

and then, we extend  $\mathcal{P}(h)$  by continuity at  $h^*$  by setting

$$\mathcal{P}(h^*) = \text{Prob}\{X^{(m)}(h^*) \leq X^{(k)}(h^*)\} \equiv \frac{1}{2}.$$

This feature illustrates that when  $h = h^*$ ,  $C_k h^{*k} = C_m h^{*m}$ , and the two norms  $\|u_h^{(k)} - u\|_{1,\Omega}$  and  $\|u_h^{(m)} - u\|_{1,\Omega}$ , which measures each approximation error of the two corresponding Lagrange finite elements, are somewhere below the two curves (see Figure 1), or in other words, somewhere in the same interval as we here:  $[0, C_k h^{*k}] = [0, C_m h^{*m}]$ . Then the probability to get  $\{X^{(m)}(h^*) \leq X^{(k)}(h^*)\}$  is equal to 0.5. This new behavior claims that when  $h$  approaches the critical value  $h^*$  the event “ $P_m$  finite element is more accurate than  $P_k$  finite element” is equally likely to occur or not to occur. As a consequence the accuracy between the two finite element  $P_k$  and  $P_m$  is equivalent. It is clearly a new theoretical information because, as we mentioned above, the values of the two constants  $C_k$  and  $C_m$  are totally unknown. Indeed, we already suspected and pointed out by data mining techniques (see for example [1], [2] and [3]) that this situation would occur. Here, we complete this suspicion by a theoretical probabilistic framework.

**(4)** Despite the usual point of view which claims that  $P_m$  finite element are more accurate than  $P_k$  ones, we get here that  $P_k$  finite element is *more likely* accurate than  $P_m$  when  $h > h^*$ . This new point of view allows us to recommend that for specific situations, like for adaptive refinement meshes for example,  $P_k$  finite element would be locally more appropriated as long as one will be able to detect the case  $h > h^*$ .

## 5 Conclusions

In this paper we present a new way to investigate the relative accuracy between two finite elements. Indeed, leaving the classical asymptotic point of view usually considered to compare the speed of convergence for different approximation errors, we got new insights for understanding error estimates. The way we thought the error estimates is not restricted to the finite element method but can be extended to other approximation methods. Indeed, the underlying idea is that, given a class of numerical schemes and their corresponding error estimates, one is able to rank them, not only in terms of asymptotic speed of convergence as usual, but also by evaluating the *almost surely* most accurate. For instance, considering numerical schemes that approximate solution to ordinary differential equations, one could argue, why (or why not!) RK4 scheme would be implemented rather than another simplest one. More generally, based on such novel probabilistic interpretation of approximation errors, the classical error estimates (not only limited to ODEs or PDEs approximations)

have to overcome and rethink. For example, it can also be successful for numerical quadrature integrations. Indeed, in that case, the structure of the error estimate also enables to proceed to a similar geometrical interpretation as we proposed for finite element error estimates. As the central keys in our methodology are the ability to get a geometrical point of view of the error estimates, on the one hand, and the handling of the random nature of the approximation implementations, on the other hand, this could bring new insights to sort numerical quadrature formulas, by evaluating the *almost surely* most accurate.

**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] F. Assous and J. Chaskalovic, Error estimate evaluation in numerical approximations of partial differential equations: A pilot study using data mining methods, *C. R. Mecanique* **341** (2013), 304–313.
- [3] F. Assous and J. Chaskalovic, Indeterminate constants in numerical approximations of PDEs: A pilot study using data mining techniques, *J. Comput. Appl. Math* **270** (2014), 462–470.
- [4] J. Chaskalovic, *Mathematical and Numerical Methods for Partial Differential Equations*, Springer, Cham, 2013.
- [5] P. G. Ciarlet, Basic error estimates for elliptic problems, in: *Handbook of Numerical Analysis. Vol. II*, North Holland, Amsterdam (1991), 17–351.
- [6] P. A. Raviart and J. M. Thomas, *Introduction à l'analyse numérique des équations aux dérivées partielles*, Masson, Paris, 1982.