Article

Generalized Beta Prime Distribution Applied to Finite Element Error Approximation

Joël Chaskalovic [1,*] and **Franck Assous** [2]

[1] Jean Le Rond d'Alembert Institute, Sorbonne University, 4 Place Jussieu, CEDEX 05, 75252 Paris, France
[2] Department of Mathematics, Ariel University, Ariel 40700, Israel; franckassous55@gmail.com
[*] Correspondence: jch1826@gmail.com

Abstract: In this paper, we propose a new family of probability laws based on the Generalized Beta Prime distribution to evaluate the relative accuracy between two Lagrange finite elements P_{k_1} and P_{k_2}, $(k_1 < k_2)$. Usually, the relative finite element accuracy is based on the comparison of the asymptotic speed of convergence, when the mesh size h goes to zero. The new probability laws we propose here highlight that there exists, depending on h, cases where the P_{k_1} finite element is more likely accurate than the P_{k_2} element. To confirm this assertion, we highlight, using numerical examples, the quality of the fit between the statistical frequencies and the corresponding probabilities, as determined by the probability law. This illustrates that, when h goes away from zero, a finite element P_{k_1} may produce more precise results than a finite element P_{k_2}, since the probability of the event "P_{k_1} *is more accurate than* P_{k_2}" becomes greater than 0.5. In these cases, finite element P_{k_2} is more likely overqualified.

Keywords: error estimates; finite elements; bramble-hilbert lemma; probabilistic numerics

1. Introduction

Recently, we proposed a new vision to consider error estimates applied to finite element approximations (see [1–3]).

Indeed, we followed the emerging field *called probabilistic numerics* whose founding principles and aims consist of providing a statistical treatment of the errors and/or approximations that are made en route to the output of a deterministic numerical method, particularly the discretized solution of a partial differential equation (see, for example, [4–6]).

Then, each method that belongs to the field of probabilistic numerics introduces a probability measure on the approximated solution of a given classical numerical method.

One of the main applications of probabilistic numerics leads to the possibility for quantifying uncertainties contained in numerical approximations due to several causes, like numerical errors, randomness of the mesh generator, and so on.

In the same spirit, we introduced a probability measure of the classical error estimate considered as a random variable to derive two probability laws. This allowed us to model the relative accuracy between two Lagrange finite elements P_{k_1} and P_{k_2}, $(k_1 < k_2)$, analyzed as a random variable.

This new point of view enabled us to complete the classical results, which are limited to the convergence analysis. Indeed, the relative accuracy between two finite elements P_{k_1} and P_{k_2}, $(k_1 < k_2)$ is usually obtained by comparing the asymptotic rate of convergence when the mesh size h goes to zero. This enables one to conclude that the P_{k_2} finite element is more accurate than P_{k_1}, since h^{k_2} goes faster to zero than h^{k_1}.

The probability laws we derived confirmed our intuition (see also our previous approaches in [7]), that when h is set to a fixed value, greater than an undetermined threshold, the asymptotic behavior recalled above does not necessarily apply. Indeed, in the error estimates, the upper bound of the approximation error is an unknown constant that depends,

among others, on a given semi-norm of the unknown exact solution. Hence, the numerical comparison of error approximations between P_{k_1} and P_{k_2} cannot be achieved.

As a consequence, determining the smallest of the two concerned approximation errors remains an open question.

Starting from there, we first consider a given approximation error as a positive number whose position is unknown within an interval defined by the unknown upper bound of the corresponding error estimate. Then, we consider this position as the result of a random variable. Indeed, the approximation error depends on the approximation, hence, on quantitative uncertainties which are related to the mesh generator process.

In addition to the new insights we obtained from these probability laws, we also implemented practical cases [8], which enabled us to evaluate the quality of the fit between these probability laws and the corresponding statistical frequencies. We show that these two probability laws globally behave like the corresponding statistical frequencies. However, the fit was sometimes not precise enough. Next, we identified the reasons for this lack of accuracy: it basically originates in the excessive rigidity of the probabilistic assumptions we made to derive these laws.

Thus, we developed a new generation of probabilistic models based on the Generalized Beta Prime distribution (GBP). These models enabled us to derive, under probabilistic acceptable hypotheses, the probability law of the relative accuracy between two Lagrange finite elements.

In this paper, we introduce the probabilistic framework built to obtain this law; we justify its use and show that it fits well with several examples. This confirms, in a significant manner, the relevance of considering error estimates as random variables in a suitable probabilistic environment.

The paper is organized as follows. In Section 2, we recall the mathematical problem we consider and a corollary of the Bramble-Hilbert lemma, from which the previous probabilistic laws [8] have been derived. In Section 3, we show a typical numerical result we obtained combining numerical statistics and these probability laws. Then, in Section 4, we derive the new probability law that enable one to evaluate the relative error accuracy between two finite elements P_{k_1} and P_{k_2}, $(k_1 < k_2)$ relying on the GBP distribution. Finally, in Section 5, we show, using several examples, the high quality of the fit between the GBP probabilistic law and the corresponding statistical frequencies. Concluding remarks follow.

2. Abstract Problem and Finite Element Error Estimates

We consider an open-bounded and non-empty subset Ω of $\mathbb{R}^n$, and denote by Γ its boundary, assumed to be C^1-piecewise. We also introduce a Hilbert space V endowed with a norm $\|.\|_V$, and a bilinear, continuous and $V-$elliptic form $a(\cdot, \cdot)$ defined on $V \times V$. Finally, $l(\cdot)$ denotes a linear continuous form defined on V.

Let $u \in V$ be the unique solution to the second order elliptic variational formulation:

$$\begin{cases} \text{Find } u \in V \text{ solution to:} \\ a(u,v) = l(v), \quad \forall v \in V. \end{cases} \tag{1}$$

In this paper, we restrict ourselves to the simple case where V is the usual Sobolev space of distributions $H^1(\Omega)$. More general cases can be found in [2].

Now, for a given $h > 0$, let us introduce V_h, a finite-dimensional subset of V, and consider $u_h \in V_h$ an approximation of u, the solution to the approximate variational formulation:

$$\begin{cases} \text{Find } u_h \in V_h \text{ solution to:} \\ a(u_h, v_h) = l(v_h), \quad \forall v_h \in V_h. \end{cases} \tag{2}$$

In what follows, we are interested in evaluating error estimates for the finite element method. Hence, we first assume that domain Ω is exactly covered by a regular $\mathcal{T}_h$ mesh composed by N_s n-simplexes $K_j, (1 \leq j \leq N_s)$, which respects the classical rules for regular discretization (see, for example, ref. [9] for the bidimensional case or [10,11] in $\mathbb{R}^n$). We

also denote by $P_k(K_j)$ the space of polynomial functions defined on a given n-simplex K_j of degree less than or equal to k, $(k \geq 1)$.

Let us first briefly recall the result of [11], which is the starting point of our study. Let $\|.\|_1$ be the classical norm in $H^1(\Omega)$ and $|.|_{k+1}$ the semi-norm in $H^{k+1}(\Omega)$, and h the mesh size, namely the largest diameter of the elements of mesh $\mathcal{T}_h$. We then have:

Lemma 1. *Suppose that there exists an integer $k \geq 1$ such that approximation u_h of V_h is a continuous piecewise function composed by polynomials that belong to $P_k(K_j)$, $(1 \leq j \leq N_s)$.*

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

$$\|u_h - u\|_1 \leq \mathscr{C}_k h^k |u|_{k+1},\tag{3}$$

where $\mathscr{C}_k$ is a positive constant independent of h.

Let us now consider two families of Lagrange finite elements P_{k_1} and P_{k_2} corresponding to a set of values $(k_1, k_2) \in \mathbb{N}^2$ such that $0 < k_1 < k_2$.

The two corresponding inequalities given by (3), assuming that the solution u to (1) belongs to $H^{k_2+1}(\Omega)$, are:

$$\|u_h^{(k_1)} - u\|_{1,\Omega} \quad \leq \quad \mathscr{C}_{k_1} h^{k_1} |u|_{k_1+1,\Omega},\tag{4}$$

$$\|u_h^{(k_2)} - u\|_{1,\Omega} \quad \leq \quad \mathscr{C}_{k_2} h^{k_2} |u|_{k_2+1,\Omega},\tag{5}$$

where $u_h^{(k_1)}$ and $u_h^{(k_2)}$, respectively, denote the P_{k_1} and P_{k_2} Lagrange finite element approximations of u.

Now, if we consider a given mesh for the finite element of P_{k_2}, which contains that of P_{k_1}, then, for the particular class of problems where the variational formulation (1) is equivalent to a minimization formulation (see for example [9]), one can show that the approximation error of P_{k_2} is always lower than that of P_{k_1}, and that P_{k_2} is more accurate than P_{k_1} for all values of mesh size h.

Next, for a given mesh size value of h, we consider two independent meshes for P_{k_1} and P_{k_2} built by a mesh generator (see Figure 1).
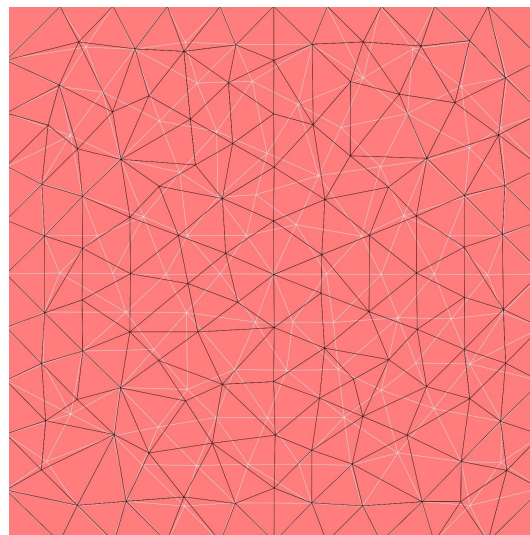

Figure 1. Superposition of two independent meshes generated with the same mesh size h.

Usually, in order to compare the relative accuracy between these two finite elements, one asymptotically considers inequalities (4) and (5) and concludes that, when h goes to zero, P_{k_2} is more accurate than P_{k_1}, since h^{k_2} goes faster to zero than h^{k_1}.

However, for a given application, h has a fixed value and this method of comparison is no more valid. Therefore, we propose a solution to determine the relative accuracy between the finite elements P_{k_1} and P_{k_2}, $(k_1 < k_2)$, for a given value of h, for which two independent meshes are considered.

To this end, let us set:

$$\beta_{k_1} = \mathscr{C}_{k_1} h^{k_1} |u|_{k_1+1,\Omega} \text{ and } \beta_{k_2} = \mathscr{C}_{k_2} h^{k_2} |u|_{k_2+1,\Omega}, \tag{6}$$

that allows us to rewrite (4) and (5) as:

$$\|u_h^{(k_1)} - u\|_{1,\Omega} \leq \beta_{k_1}, \tag{7}$$

$$\|u_h^{(k_2)} - u\|_{1,\Omega} \leq \beta_{k_2}. \tag{8}$$

Now, as explained in [3], there is no a priori available information to surely or better specify the relative position of the approximation errors $\|u_h^{(k_1)} - u\|_{1,\Omega}$ and $\|u_h^{(k_2)} - u\|_{1,\Omega}$, which respectively belong to the interval $[0, \beta_{k_1}]$ and $[0, \beta_{k_2}]$, where β_{k_i} are unknown since $\mathscr{C}_{k_i}$ and $|u|_{k_i+1,\Omega}$ are unknown quantities.

In [3], we explained why quantitative uncertainties have to be handled using finite element computations. This originates in the way the mesh grid generator builds the mesh used to compute the approximation $u_h^{(k_i)}$, $(i = 1, 2)$; this leads to a partial non-control of the mesh, even for a given maximum mesh size. Indeed, most meshes are generated by a mesh generator based on a Delaunay-Voronoi method. As a consequence, the corresponding point generation, and/or the treatment of ambiguities inevitably contains some randomness. Therefore, the generated grid is a priori random, and so is the corresponding approximation $u_h^{(k_i)}$ $(i = 1, 2)$.

For these reasons, we introduced in [3] a probabilistic framework that enables one to consider the possible values of the norm $\|u_h^{(k)} - u\|_{1,\Omega}$ as a random variable, which is defined as follows:

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

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

$$X^{(k)}: \quad \mathbf{\Omega} \quad \rightarrow \quad [0, \beta_k] \tag{9}$$

$$\boldsymbol{\omega} \equiv u_h^{(k)} \quad \mapsto \quad X^{(k)}(\boldsymbol{\omega}) = X^{(k)}(u_h^{(k)}) = \|u_h^{(k)} - u\|_{1,\Omega}. \tag{10}$$

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

Therefore, our goal is now to evaluate the probability of the event:

$$\left\{ \|u_h^{(k_2)} - u\|_{1,\Omega} \leq \|u_h^{(k_1)} - u\|_{1,\Omega} \right\} \equiv \left\{ X^{(k_2)}(h) \leq X^{(k_1)}(h) \right\}, \tag{11}$$

which will enable us to estimate the more likely accuracy of the two finite elements P_{k_1} and P_{k_2}, $(k_1 < k_2)$.

Now, regarding the absence of information concerning the more or less likely values of the norm $\|u_h^{(k_i)} - u\|_{1,\Omega}$, $(i = 1, 2)$ in the interval $[0, \beta_{k_i}]$, $(i = 1, 2)$, we assumed in [1], that each random variable $X^{(k_i)}$, $(i = 1, 2)$ had a uniform distribution on the interval $[0, \beta_{k_i}]$, and that they were also independent.

This probabilistic framework enabled us to obtain, in a more general context (see Theorem 3.1 in [1]), the density of the probability of the random variable Z defined by

$Z = X^{(k_2)} - X^{(k_1)}$ and, consequently, $\mathcal{P}_{k_1,k_2} \equiv Prob\left\{ X^{(k_2)} \leq X^{(k_1)} \right\}$. This probability corresponds to the value of the entire cumulative distribution function $F_Z(z)$ at $z = 0$, defined by:

$$F_Z(z) = \int_{-\infty}^{z} f_Z(z)dz , \tag{12}$$

with $f_Z(z)$ denoting the density function of the random variable Z.

For the purpose of the present work, by performing elementary adaptations of Theorem 3.3 in [1], we get the corresponding probability law given by:

$$\mathcal{P}_{k_1,k_2}(h) = \left| \begin{array}{ll} 1 - \dfrac{1}{2}\left(\dfrac{h}{h^*_{k_1,k_2}}\right)^{k_2-k_1} & \text{if} \quad 0 \leq h \leq h^*_{k_1,k_2}, \quad (13) \\[3mm] \dfrac{1}{2}\left(\dfrac{h^*_{k_1,k_2}}{h}\right)^{k_2-k_1} & \text{if} \quad h \geq h^*_{k_1,k_2}, \quad (14) \end{array} \right.$$

where $h^*_{k_1,k_2}$ is defined by:

$$h^*_{k_1,k_2} \equiv \left(\frac{\mathscr{C}_{k_1}|u|_{k_1+1,\Omega}}{\mathscr{C}_{k_2}|u|_{k_2+1,\Omega}}\right)^{\frac{1}{k_2-k_1}}. \tag{15}$$

The shape of this law is similar to a "sigmoid" curve, as one can see in Figure 2.

We already remarked in [3] that the probability law (13) and (14) can asymptotically—when $k_2 - k_1$ goes to infinity—lead to the limit situation we defined as the "two-steps" model (see Figure 2).

Nevertheless, we also proved in [3] that under suitable probabilistic assumptions, one can directly derive this "two-steps" probabilistic law for all non-zero integers k_1 and k_2 to finally obtain the following probability law:

$$\mathcal{P}_{k_1,k_2}(h) = \left| \begin{array}{ll} 1 & \text{if } 0 < h < h^*_{k_1,k_2}, \\[2mm] 0 & \text{if } h > h^*_{k_1,k_2}. \end{array} \right. \tag{16}$$

The next section is dedicated to the analysis of the fit between statistical data and the above probability laws (13), (14) and (16).

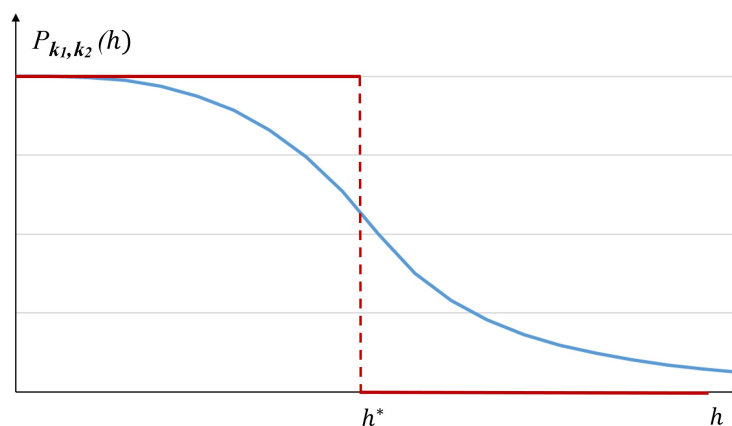**Figure 2.** General shape of probabilistic law (13) and (14) together with the limit Heaviside case (16).

3. Numerical Statistics and Probability-Law Comparison

Here, we compare the theoretical probability law (13) and (14) and the statistics obtained from two examples of approximate variational formulations (2).

As shown in [8], the "two-steps" law (16) provides good-quality fits in several numerical cases. However, this law is a bit rough and cannot really follow the variations of the convexity of the statistical data curve (see below). For this reason, we also tested the accuracy of the fit with the "sigmoid" law given by (13) and (14).

In order to numerically check the accuracy of the fit between the probability law (13) and (15) and the corresponding statistical data produced by numerical simulations, we considered two finite elements P_{k_1} and P_{k_2} ($k_1 < k_2$).

One of the main ingredients of the method is the computation of $h^*_{k_1,k_2}$, as defined by (15), which is evaluated using a *maximum likelihood estimator*; see for instance [12].

In our case, this principle is applied as follows: for a given finite element P_{k_1}, we consider N different meshes with the same (maximum) mesh size h. Then, we compute:

$$\max_{N,h} \frac{\|u_h^{(k_1)} - u\|_1}{h^{k_1}} \tag{17}$$

which constitutes, using estimate (4), the maximum likelihood estimator for $\mathscr{C}_{k_1}|u|_{k_1+1}$. Indeed, due to inequality (4), quantity $\frac{X^{k_1}(h)}{h^{k_1}}$ is also a uniform random variable whose support is $[0, \mathscr{C}_{k_1}|u|_{k_1+1}]$.

Then, one can show [13] that for a given uniform random variable Y whose support is $[0, \theta]$, θ being an unknown real parameter, the maximum likelihood estimator $\hat{\theta}$ is given by:

$$\hat{\theta} = \max(Y_1, \ldots, Y_N),$$

where $(Y_1, \ldots, Y_N)$ is a sample built with independent and identically distributed random variables $(Y_i)_{i=1,N}$, with the same distribution as Y.

In our case, this implies that (17) is the maximum likelihood estimator for $\mathscr{C}k_1|u|_{k_1+1}$, since N and h each take a finite number of values.

Doing the same for another finite element P_{k_2}, ($k_1 < k_2$), we obtain that the estimator for $h^*_{k_1,k_2}$, denoted $\widehat{h^*_{k_1,k_2}}$, is defined by:

$$\widehat{h^*_{k_1,k_2}} = \left(\frac{\max\limits_{N,h} \dfrac{\|u_h^{(k_1)} - u\|_1}{h^{k_1}}}{\max\limits_{N,h} \dfrac{\|u_h^{(k_2)} - u\|_1}{h^{k_2}}} \right)^{1/k_2 - k_1} \tag{18}$$

Then, one can easily compute the probability law given by (13) and (14), when one replaces $h^*_{k_1,k_2}$ by its estimator $\widehat{h^*_{k_1,k_2}}$.

Finally, in order to numerically check the validity of this probability law, we now compare the probabilities computed by (13), (14) and (18), with the corresponding statistical frequencies computed on the N meshes, for each fixed value h of the mesh size, as follows.

We considered for the two finite elements P_{k_1} and P_{k_2} ($k_1 < k_2$), the same number N of different meshes with the same (maximum) mesh size h. From there, we computed the approximate solution $u_h^{(k_1)}$ and $u_h^{(k_2)}$, and tested if $\|u_h^{(k_2)} - u\|_1 \leq \|u_h^{(k_1)} - u\|_1$. Then, we repeated the same process for different values of h, either lower or greater than $\widehat{h^*_{k_1,k_2}}$. This gives, as a function of h, the percentage of cases where the approximation error of P_{k_2} is lower than the approximation error of P_{k_1}.

In all cases, we used package FreeFem++ [14] to compute the P_{k_1} and P_{k_2} finite element approximations. Basically, the mesh generator of FreeFem++ is based on the Delaunay-Voronoi algorithm, and allows us to fix the maximum mesh size h for N different meshes of the same domain Ω. Since FreeFEM++ is based on updated and modern methods of meshing, we are confident that using another mesh generator will lead to equivalent results.

Now, in order to motivate the new probability law of the next section, we recall typical results obtained in [8]. To this end, let us look at the numerical tests we implemented to analyze the relative accuracy between two finite elements.

Hence, we consider the following classical elliptic problem, with obvious notations:

Find $u \in H^1(\Omega)$ solution to:

$$\begin{cases} -\Delta u = f \text{ in } \Omega, \\ \quad u = g \text{ on } \partial\Omega, \end{cases} \tag{19}$$

where, for simplicity, Ω is the open unit square in $\mathbb{R}^2$: $\Omega =]0,1[\times]0,1[$. The associated variational formulation, which is analogous to (1), can readily be derived. According to the choice of f and g, we considered, as examples, a *stiff* problem, where the solution exhibits rapid variations or, alternatively, a *very smooth* problem.

More precisely, when the right-hand side $f(x,y)$ is the Runge-like function, (see [15] or [16]), given by:

$$f(x,y) = \frac{1}{1 + \alpha x^2} \frac{1}{1 + \alpha y^2},$$

α being a real parameter, we derived in [8], the exact solution of (19) according to a suitable choice of the Dirichlet condition, parametered by g.

Then, considering 100 meshes for each value of h and $\alpha = 3000$ in the Runge function, we plot in Figure 3 the statistical frequencies corresponding to "P_3 more accurate that P_1" (blue line), together with the corresponding "sigmoid" probability law (orange line).
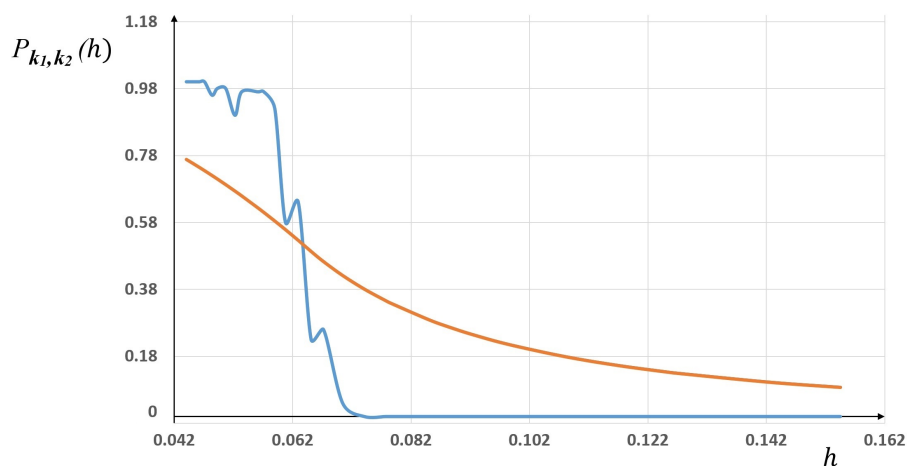

Figure 3. P_1 versus P_3 for the Runge function with $\alpha = 3000$. Comparison between the statistical frequencies (blue) and the "Sigmoid law" (orange).

As one can see, the fit between the statistics and the probability law is not satisfactory. The same gaps were observed for other simulations with different sets of parameter values, and for several pairs of finite elements P_{k_1} and P_{k_2}.

We now illustrate the probabilistic law for a very smooth solution to the variational problem (1). To build such a case, we choose the right-hand side f of problem (19) as:

$f(x,y) = 2\pi^2 \sin(\pi x)\cos(\pi y)$, so that $u(x,y) = \sin(\pi x)\cos(\pi y)$ is the exact solution of the problem, provided that the Dirichlet boundary condition g is taken as the trace of $u(x,y)$ on the boundary $\partial\Omega$ (for more details, see [8]).

As previously, we first compute $\widehat{h^*_{k_1,k_2}}$ defined by (15), then we compute the probabilistic model recalled above. After that, we compare these results to the statistical frequencies.

As an example, we consider the finite elements P_2 and P_3, the corresponding results being depicted in Figure 4.

As one can see, the comparison with the sigmoid law in (13) and (14) gives only a global trend and is not accurate.

Here, as for the previous Runge example, this lack of accuracy will be improved by the new probabilistic laws derived in the next section. We will show how to enrich the "sigmoid" model, which actually only depends on one parameter, namely, $h^*_{k_1,k_2}$.
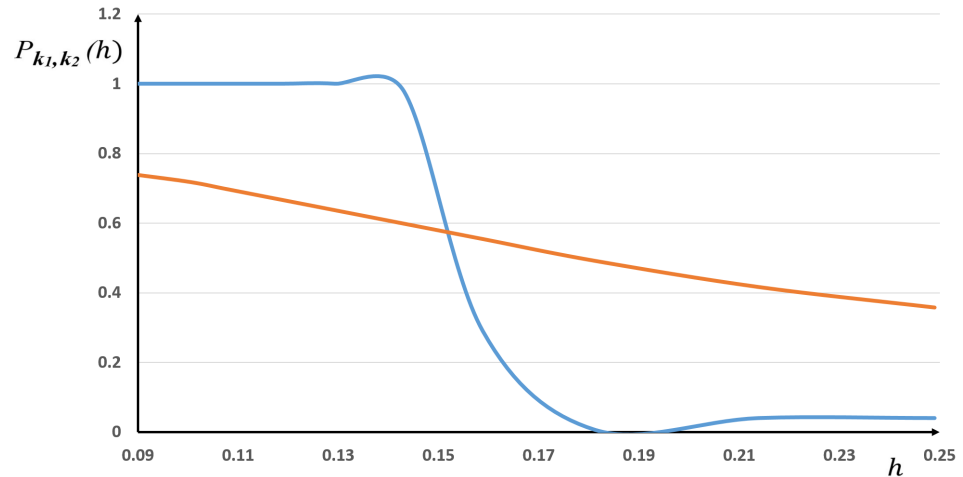

Figure 4. P_2 versus P_3 for the smooth case. Comparison between the statistical frequencies (blue) and the two probabilistic "Sigmoid" laws (orange).

4. A New Probability Law Based on GBP

This section is devoted to the new probabilistic law we derived in order to evaluate the relative error accuracy between two finite elements P_{k_1} and P_{k_2}, $(k_1 < k_2)$. To motivate this new approach, let us make some remarks:

1. In the previous works that enabled us to derive the "Sigmoid" probability law, we had made some assumptions (i.e., uniform distribution and independency). Since we now aim at obtaining a better fit between the probabilistic law and the statistical data, we will now relax the hypothesis of uniformity we had made on the densities of the random variables $X^{(k_i)}(h)$, $(i = 1, 2)$, (see [1–3]).

2. In order to obtain more degrees of freedom for the shape of these densities, let us first consider the random variable $Z = X^{(k_2)} - X^{(k_1)}$. Our goal is to obtain, for the cumulative distribution function F_Z defined by (12) at point $z = 0$, a curve whose shape looks like a "Sigmoid". We thus have to enrich our modeling process by adding more degrees of freedom, the "Sigmoid" probability law (13) and (14) containing only one parameter $h^*_{k_1,k_2}$. For this reason, we now consider a density f_Z for the random variable Z, whose associated cumulative distribution function F_Z at point $z = 0$ will include two exogenous parameters; these will be statistically estimated.

Keeping in mind these remarks, we begin by introducing the probability density function f_X of the normalized Beta random variable X (see [17]), defined by:

$$\forall x \in [0,1] : f_X(x; p, q) \equiv \frac{x^{p-1}(1-x)^{q-1}}{\displaystyle\int_0^1 u^{p-1}(1-u)^{q-1}du} = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}, \qquad (20)$$

with p and q, two parameters of shape that belong to $\mathbb{R}^*_+$, and $\Gamma(.)$, the classical Gamma function.

Among the numerous properties of the Beta distribution, let us mention the one that particularly motivates us into considering it. Depending on the two parameters p and q, the shapes of the corresponding cumulative distributions are very rich. In particular, they include the shape of the "Sigmoid" curve we are looking for to fit the statistics we get in this numerical example.

However, one cannot directly apply the Beta distribution to get the probability law we are looking for. More specifically, two properties have to be taken into account:

1. The support of the Beta density f_X of the random variable X, denoted $Supp\,X$, is included in $[0, 1]$. Nevertheless, that of the random variable Z is $[-\beta_{k_1}, \beta_{k_2}]$, since $Z = X^{(k_2)} - X^{(k_1)}$ and $Supp\,X^{(k_i)} \subset [0, \beta_{k_i}], (i = 1, 2)$. This will make us use a suitable transformation of the density f_X to guarantee the correct support of the density f_Z of Z.

2. We are looking for a probability law for the event $\left\{ X^{(k_2)} \leq X^{(k_1)} \right\}$ as a function of h, that belongs to $[0, +\infty[$. Consequently, we have to apply another transformation to the density f_Z to guarantee this property for the support of h.

To achieve these transformations, we establish the following results:

Lemma 2. *Let Z be the random variable defined by $Z \equiv X^{(k_2)} - X^{(k_1)}$, where $X^{(k_i)}(i = 1, 2)$ are defined by (10). Let $X \hookrightarrow B(p, q)$ be the Beta distribution parameterized by two given shape parameters $(p, q) \in \mathbb{R}_+^{*2}$. Assume that Z is given by:*

$$Z = -\beta_{k_1} + (\beta_{k_1} + \beta_{k_2})X. \tag{21}$$

Then, the probability density function f_Z of Z can be written as:

$$f_Z(z) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{\beta_{k_1}^{p-1}\beta_{k_2}^{q-1}}{(\beta_{k_1} + \beta_{k_2})^{p+q-1}} \left(1 + \frac{z}{\beta_{k_1}}\right)^{p-1}\left(1 - \frac{z}{\beta_{k_2}}\right)^{q-1} \mathbb{1}_{[-\beta_{k_1}, \beta_{k_2}]}(z), \tag{22}$$

where $\mathbb{1}_{[-\beta_{k_1}, \beta_{k_2}]}$ is the indicator function of the interval $[-\beta_{k_1}, \beta_{k_2}]$.

Proof. As we already noticed, the support of f_Z is clearly within $[-\beta_{k_1}, \beta_{k_2}]$ as soon as those of $X^{(k_i)}$ are in $[0, \beta_{k_i}]$.

Now, let us evaluate the cumulative distribution function $F_Z(z)$ of the random variable Z as defined by (21):

$$F_Z(z) \quad = \quad Prob\left\{ Z \leq z \right\} = Prob\left\{ -\beta_{k_1} + (\beta_{k_1} + \beta_{k_2})X \leq z \right\} \tag{23}$$

$$= \quad Prob\left\{ X \leq \frac{z + \beta_{k_1}}{\beta_{k_1} + \beta_{k_2}} \right\} = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^{\frac{z+\beta_{k_1}}{\beta_{k_1}+\beta_{k_2}}} u^{p-1}(1-u)^{q-1}du. \tag{24}$$

Then, by deriving $F_Z(z)$ as given by (24), we obtain the expression (22) of the density f_Z. □

Remark 1. *Let us now give an interpretation of this result. As mentioned above, in our previous works (see, for example, [1,3]), we assumed the random variables $X^{(k_i)}$ to be uniformly distributed on their support $[0, \beta_{k_i}]$. This enabled us to derive the density f_Z of the random variable Z (see Theorem 3.1 in [1]).*

Here, Lemma 2 can be interpreted regarding the new assumption implicitly made on the random variables $X^{(k_i)}$. Indeed, if we rewrite Z as:

$$Z = \beta_{k_2}X - (\beta_{k_1} - \beta_{k_1}X),$$

by setting:

$$X^{(k_1)} \equiv \beta_{k_1} - \beta_{k_1}X \text{ and } X^{(k_2)} = \beta_{k_2}X, \tag{25}$$

we observe that the support of $X^{(k_i)}$ belongs to $[0, \beta_{k_i}]$, on the one hand, and that the difference $X^{(k_2)} - X^{(k_1)}$ is equal to Z, on the other hand.

In other words, the choice to write Z given by (21) as a dimensional Beta distribution on $[-\beta_{k_1}, \beta_{k_2}]$ (which corresponds to altering the location and scale of the standard Beta distribution), accordingly modifies the hypothesis of uniformity of the two variables $X^{(k_i)}$ to the one of a non-dimensional Beta distribution.

We are now capable to derive the new probability distribution $\mathcal{P}_{k_1,k_2}(h)$ in order to evaluate which, between two Lagrange finite elements P_{k_1} and P_{k_2}, $(k_1 < k_2)$, is the more likely accurate. This is the purpose of the following Theorem.

Theorem 1. *Let $X^{(k_i)}(i = 1, 2)$, be the two random non-dimensional Beta distribution variables defined by (25) and Z the corresponding random variable defined by (21).*

Then, $\mathcal{P}_{k_1,k_2}(h)$ is the cumulative distribution function of a Generalized Beta Prime random variable H, whose density of probability f_H, is defined by four parameters $(p, q, k_2 - k_1, h^*_{k_1,k_2})$, and we have:

$$\mathcal{P}_{k_1,k_2}(h) = Prob\{H \geq h\} = \int_h^{+\infty} f_H(s; q, p, k_2 - k_1, h^*_{k_1,k_2})\, ds, \tag{26}$$

where:

$$f_H(s; q, p, k_2 - k_1, h^*_{k_1,k_2}) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \frac{(k_2-k_1)}{h^*_{k_1,k_2}} \left(\frac{s}{h^*_{k_1,k_2}}\right)^{q(k_2-k_1)-1} \left[1 + \left(\frac{s}{h^*_{k_1,k_2}}\right)^{k_2-k_1}\right]^{-p-q}. \tag{27}$$

Proof. On the one hand, we are looking for a random variable H whose $[0, +\infty[$, which is associated to $\mathcal{P}_{k_1,k_2}(h)$ equal to $F_Z(0)$. On the other hand, (24) was derived by the help of the dimensionless Beta distribution $B(p, q)$, whose support belongs to $[0, 1]$. Hence, we set the following change of variable in (24):

$$s = \frac{u}{1-u},$$

and we get:

$$F_Z(z) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^{\frac{z+\beta_{k_1}}{\beta_{k_2}-z}} s^{p-1}(1+s)^{-p-q}\, ds. \tag{28}$$

Now, taking $z = 0$ in (28), and using $\beta_{k_i}(i = 1, 2)$, given by (6) and $h^*_{k_1,k_2}$ by (15), we obtain:

$$F_Z(0) = \mathcal{P}_{k_1,k_2}(h) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^{\frac{\beta_{k_1}}{\beta_{k_2}}} s^{p-1}(1+s)^{-p-q}\, ds, \tag{29}$$

$$= \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^{(h^*_{k_1,k_2}/h)^{k_2-k_1}} s^{p-1}(1+s)^{-p-q}\, ds. \tag{30}$$

A last change of variables, $t = \dfrac{1}{s}$, in (30) leads to:

$$F_Z(0) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_{(h/h^*_{k_1,k_2})^{k_2-k_1}}^{+\infty} t^{q-1}(1+t)^{-p-q}\, dt. \tag{31}$$

Finally, by considering the probability of the complementary event considered in (31) as a function of h (namely, $1 - \mathcal{P}_{k_1,k_2}(h)$), and after derivation with respect to h, we get the density f_H defined in (27), which is a GBP probability density function parameterized by $(q, p, k_2 - k_1, h^*_{k_1,k_2})$, see for example [17].

In other words, we have obtained:

$$Prob\left\{X^{(k_1)} \leq X^{(k_2)}\right\} = 1 - Prob\left\{X^{(k_2)} \leq X^{(k_1)}\right\}(h) = \int_o^h f_H(s; q, p, k_2 - k_1, h^*_{k_1,k_2})ds, \tag{32}$$

and $\mathcal{P}_{k_1,k_2}(h)$ in (26) is deduced by complementarity. □

Remark 2.

1. *From (26) and (27), or equivalently from (30), we can obtain the asymptotic behavior of the probability of the event $\left\{ X^{(k_1)} \leq X^{(k_2)} \right\}$ as the mesh size h goes to 0.*

   Clearly, $Prob\left\{ X^{(k_1)} \leq X^{(k_2)} \right\}$ goes to 0 with h. In other words, we have found with the new probabilistic law the usual result, which claims that the finite element P_{k_2} is more accurate than P_{k_1}, since from (4) and (5), h^{k_2} goes faster to zero than h^{k_1} when $k_1 < k_2$.

   Here, by (30), the same property is expressed in terms of probability, namely: the event $\left\{ X^{(k_1)} \leq X^{(k_2)} \right\}$ occurs almost never, or equivalently, the event $\left\{ X^{(k_2)} \leq X^{(k_1)} \right\}$ is an almost surely one, since its probability is equal to 1.

2. *From (26), using the positivity of the density f_H, we conclude that $\mathcal{P}_{k_1,k_2}(h)$ is a decreasing function of h. This property was already observed with the "Sigmoid" model in [1,3] where uniformity of the random variables $X^{(k_i)}$, $(i = 1, 2)$, was assumed.*

   In other words, this property claims that the event $\left\{ P_{k_2} \text{ is more accurate than } P_{k_1} \right\}$ is true only for small h, and the more h increases the less likely this property is. Moreover, when h becomes large, asymptotically the event $\left\{ P_{k_2} \text{ is more accurate than } P_{k_1} \right\}$ occurs almost never.

5. Numerical Comparisons

This section is devoted to the numerical comparison of the GBP probabilistic law (26), as derived in Theorem 1, with the statistical frequencies, computed as explained in Section 3.

Concerning the mesh generation, as explained above, we have considered N different meshes with the same (maximum) mesh size h. In these conditions, if we denote by $\mathcal{T}_h$ one of the N meshes, each other mesh is constructed as a random perturbation of $\mathcal{T}_h$ obtained by moving and/or adding internal vertices, in such a way that h remains constant, the number of points on the boundary being constant.

Let us also explain how we proceeded to determine the four parameters $(q, p, k_2 - k_1, h^*_{k_1,k_2})$ of density f_H. First of all, we considered and fixed two finite elements, for example P_1 and P_3, which corresponds to $k_2 - k_1 = 2$.

Regarding the three other parameters, p, q and $h^*_{k_1,k_2}$, we have implemented a least squares optimization (see for instance [18]) to minimize the norm L^2 of the difference between the statistical frequencies and the corresponding values, computed by the GBP model. This was computed using the Excel solver.

The results we obtained are depicted in Figure 5 for $N = 500$ different meshes. As one can see, the least squares algorithm found the optimal parameters p, q and $h^*_{k_1,k_2}$ such that the fit between the statistical frequencies and the GBP law (26) is very satisfactory.
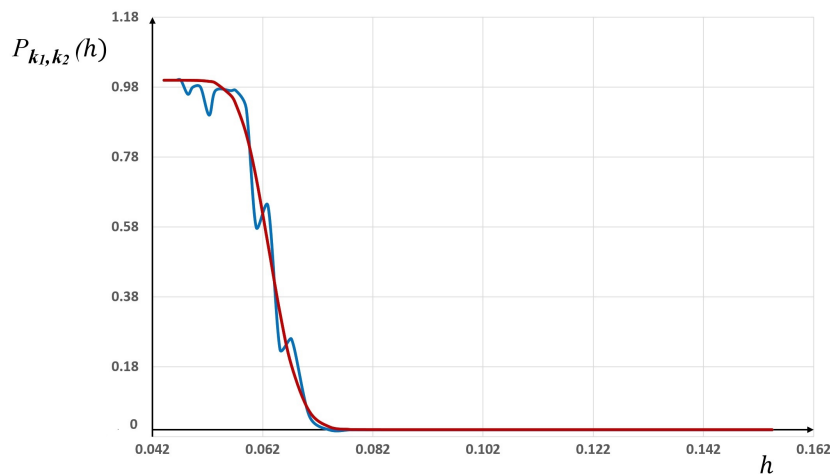

Figure 5. P_1 versus P_3 for the Runge function with $\alpha = 3000$. Comparison between the statistical frequencies (blue) and the GBP law (red).

This is clearly due to the the richness and flexibility of this distribution, which provides two supplementary degrees of freedom, p and q, with respect to the "Sigmoid" law. This also validates a posteriori the approach presented in Section 4, which describes the randomness values of the approximation errors $\|u_h^{(k_i)} - u\|_{1,\Omega}, (i = 1, 2)$, within their respective interval $[0, \beta_{k_i}], (i = 1, 2)$.

Another example that enables one to appreciate the accuracy of the fit may be obtained by comparing the results implemented with the two Lagrange finite elements P_1 and P_4, also for $N = 500$. In this case, as one can see in Figure 6, the "Sigmoid" law only describes the global trend of the statistical frequencies, whereas the GBP law perfectly fits the corresponding data.
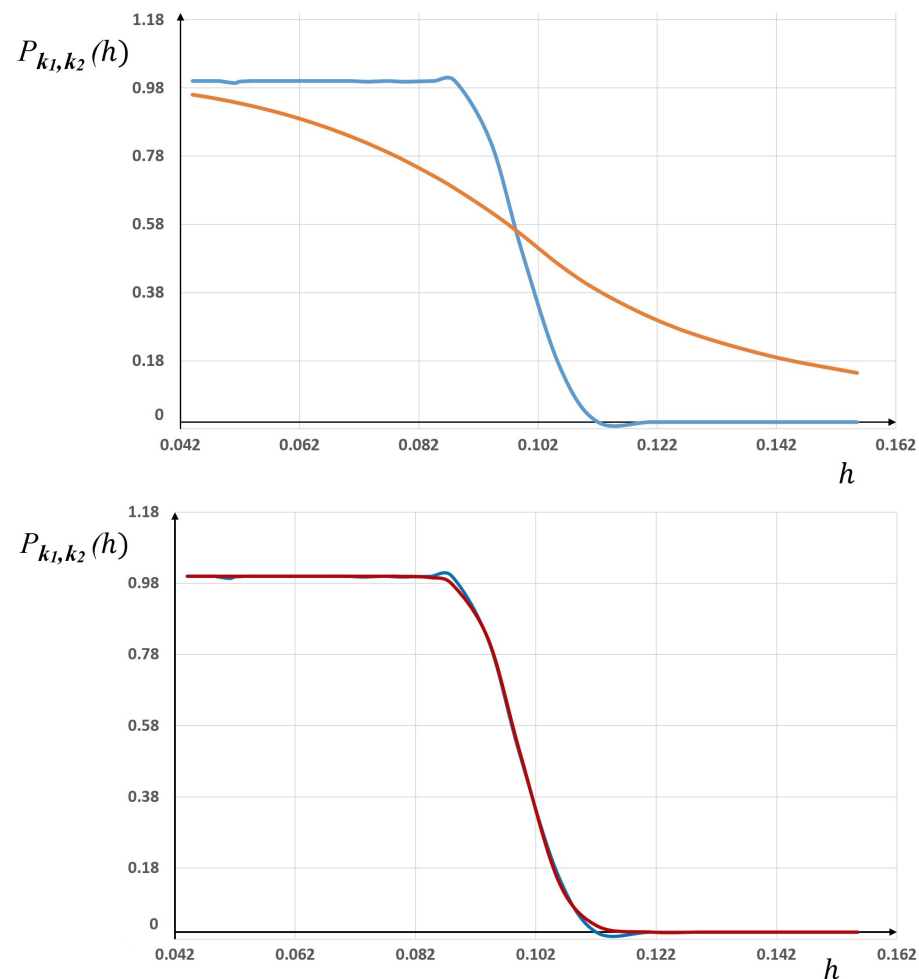

Figure 6. P_1 versus P_4. **Up**: comparison between the statistical frequencies (blue) and the Sigmoid (orange). **Down**: comparison between the statistical frequencies (blue) and the GBP law (red).

As a last illustration, let us present the results obtained when comparing the P_2 and P_3 finite elements. The difference in the quality of the fit between the "Sigmoid" law and the GBP law becomes even more important than in the previous case. Indeed, in this case, $k_2 - k_1$ is equal to one and the "Sigmoid" law becomes a linear function of h when $h \leq h_{k_1,k_2}^*$ in formula (13). In addition, the GBP law once again fit the statistical frequencies very well (see Figure 7).
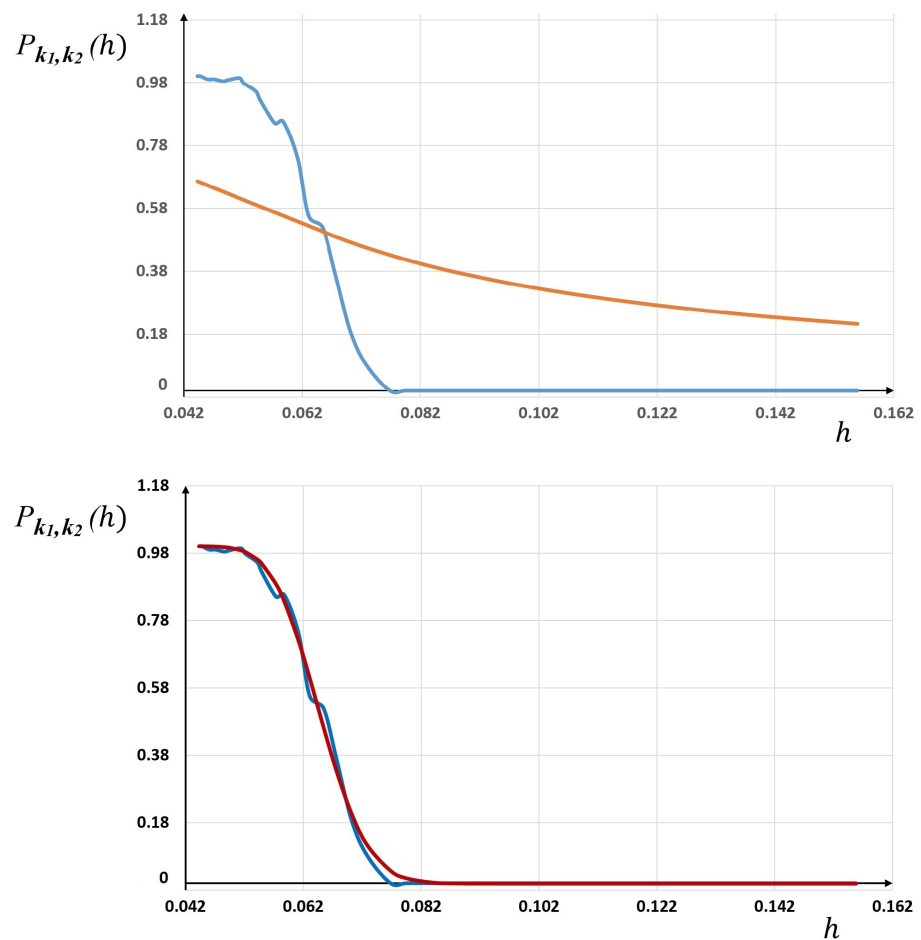
$P_{k_1,k_2}(h)$

$P_{k_1,k_2}(h)$

Figure 7. P_2 versus P_3. **Up**: comparison between the statistical frequencies (blue) and the Sigmoid (orange). **Down**: comparison between the statistical frequencies (blue) and the GBP law (red).

6. Conclusions

In this paper, we derived a new family of probabilistic laws to compare the accuracy between two Lagrange finite elements P_{k_1} and P_{k_2} ($k_1 < k_2$). Based on a new probabilistic approach introduced in [1], we extended these previous results to improve the fit between the statistical frequencies obtained by implementing practical cases and the corresponding probabilities.

We recall, in Section 3, the main problems we were confronted with regarding the accuracy of the previous probabilistic laws we derived in [1–3], by identifying the gaps we observed with the corresponding statistical frequencies.

Afterwards, in Section 4 we derived the new probabilistic law based on the Generalized Beta Prime law. In Section 5, we showed the very good fit we obtained between the numerical results and the new probability law. This corrected the main observed deficiencies described above with simpler laws.

Finally, the new probabilistic law, together with our statistical validation, confirmed the validity of considering the approximation errors as random variables defined in a well-adapted probabilistic framework. Of course, this probabilistic approach is not restricted to Lagrange P_k finite elements and can be extended to any other finite elements, like rectangular Q_k. It can also be generalized to any types of approximation in the following sense: given a class of numerical schemes and their corresponding approximation error estimates, we should be able to order them not only in terms of asymptotic rate of convergence but also by evaluating the most probably accurate one.

References

1. Chaskalovic, J. A probabilistic approach for solutions of determinist PDE's as well as their finite element approximations. *Axioms* **2021**, *10*, 349. [CrossRef]
2. Chaskalovic, J.; Assous, F. Explicit k-dependence for P_k finite elements in $W^{m,p}$ error estimates: Application to probabilistic laws for accuracy analysis. *Appl. Anal.* **2021**, *100*, 2825–2843. [CrossRef]
3. Chaskalovic, J.; Assous, F. A new mixed functional-probabilistic approach for finite element accuracy. *Comput. Methods Appl. Math.* **2020**, *20*, 799–813. [CrossRef]
4. Abdulle, A.; Garegnani, G. A probabilistic finite element method based on random meshes: A posteriori error estimators and Bayesian inverse problems. *Comput. Methods Appl. Mech. Eng.* **2021**, *384*, 113961. [CrossRef]
5. Hennig, P.; Osborne, M.A.; Girolami, M. Probabilistic numerics and uncertainty in computations. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2015**, *471*, 20150142. [CrossRef] [PubMed]
6. Oates, C.J.; Sullivan, T.J. A modern retrospective on probabilistic numerics. *Stat. Comput.* **2019**, *29*, 1335–1351. [CrossRef]
7. Assous, F.; Chaskalovic, J. Data mining techniques for scientific computing: Application to asymptotic paraxial approximations to model ultra-relativistic particles. *J. Comput. Phys.* **2011**, *230*, 4811–4827. [CrossRef]
8. Chaskalovic, J.; Assous, F. Numerical validation of probabilistic laws to evaluate finite element error estimates. *Math. Model. Anal.* **2021**, *26*, 684–695. [CrossRef]
9. Chaskalovic, J. *Mathematical and Numerical Methods for Partial Differential Equations*; Springer: Berlin/Heidelberg, Germany, 2013.
10. Ciarlet, P.G. *The Finite Element Method for Elliptic Problems*; Classics in Applied Mathematics; SIAM: Philadelphia, PA, USA, 2002.
11. Raviart, P.A.; Thomas, J.M.; Thomas, J.M. *Introduction à L'analyse Numérique des Équations aux Dérivées Partielles*; Elsevier Masson: Paris, France, 1982.
12. Rossi, R.J. *Mathematical Statistics: An Introduction to Likelihood Based Inference*; John Wiley and Sons: New York, NY, USA, 2018.
13. Jeune, M.L. *Théorie de L'estimation Paramétrique Ponctuelle*; Springer: Paris, France, 2010.
14. Hecht, F. New development in FreeFem++. *J. Numer. Math.* **2012**, *20*, 251–265. [CrossRef]
15. Runge, C. Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten. *Zeitschrift für Mathematik und Physik* **1901**, *46*, 224–243.
16. Epperson, J.F. On the Runge Example. *Am. Math. Mon.* **1987**, *94-4*, 329–341. [CrossRef]
17. Gavin E. *Crooks, Field Guide to Continuous Probability Distributions*; Berkeley Institute for Theoretical Science: Berkeley, CA, USA, 2019.
18. Lawson, C.L.; Hanson, R.J. *Solving Least Squares Problems*; SIAM in Applied Mathematics: Philadelphia, PA, USA, 1995.