Noname manuscript No.
(will be inserted by the editor)

A Chernov bound for robust tolerance design and application

Ambre Diet[1,2] · Nicolas Couellan[2,3] · Xavier Gendre[2,4] · Julien Martin[1]

Received: date / Accepted: date

Abstract Within an industrial manufacturing process, tolerancing is a key player. The dimensions uncertainties management starts during the design phase, with an assessment on variability of parts not yet produced. For one assembly step, knowledge can be gained from the tolerance range required for the parts involved. In order to assess output uncertainty of this assembly in a reliable way, this paper presents an approach based on the deviation of the sum of uniform distributions. As traditional approaches based on Hoeffding inequalities do not give accurate results when the deviation considered is small, an improved upper bound is proposed. Then, the impact of the stack chain geometry on the bound definition is discussed. Finally, an application of the proposed approach in tolerance design of an aircraft sub-assembly is detailed. The main interest of the technique compared to existing methodologies is the management of the confidence level and the emphasis of the explicit role of the balance within the stack chain.

Keywords Design · Deviation · Manufacturing · Quality · Sum of uniform distributions · Tolerance

Ambre Diet
E-mail: ambre.diet@airbus.com
Nicolas Couellan
E-mail: nicolas.couellan@recherche.enac.fr
Xavier Gendre
E-mail: xavier.gendre@isae-supaero.fr
Julien Martin
E-mail: julien.ju.martin@airbus.com
[1] Tolerancing department, Airbus Operations S.A.S, 316 route de Bayonne, 31060 Toulouse, France.
[2] Institut de Mathématiques de Toulouse UMR 5219, Université de Toulouse, 31062 Toulouse, France.
[3] ENAC, Université de Toulouse, 7 avenue Édouard Belin, 31400 Toulouse, France.
[4] ISAE-SUPAERO, Université de Toulouse, 10 Avenue Édouard Belin, 31055 Toulouse, France.

1 Introduction

The management of dimensions uncertainties is a key player in the manufacturing process of various industrial sectors such as transportation (automotive, aeronautics, . . .) or household appliances industry.

Dimensions may have some deviation from the designed value without significant impact on the quality and functional requirements of the final product. Tolerance intervals are defined according to engineering knowledge and scientific analysis in order to determine these acceptable variations. A deviation out of the determined tolerance bounds is considered non-compliant and imply an action such as an investigation or a modification in the process or the design.

The perfect balance between functional requirements and process capability has to be found so that the specified tolerance interval is the most accurate possible. If the tolerance is too tight, the process might not have the capability to manufacture it and either there will be many rejected items or some costly improvement will be needed to produce compliant items. Otherwise, a too wide tolerance will lead to non-conformity with functional requirements of the final product and may lower the product performance. As there are often several steps in a manufacturing process, the propagation of uncertainty has also to be taken into account to specify the tolerance interval of following assembly steps.

In this paper, the focus is on the tolerances allocation during the design phase in which tolerancing activity does not only aim at anticipating the margins of uncertainty but also help in predicting their effects on the various assembly steps. These involve different physical characteristics of parts, such as part length, hole position, pin, . . . , called features. All tolerancing issues and notations are detailed in the engineering drawing and related documentation pratices [1] and [2].

In our case, there are no available dimensions measurements because the focus is on tolerance allocation in the design phase of a product prior production. Considering one specific assembly stage, one of the main concern is to assess the variability of an output feature of the assembly knowing the tolerance range of the input features. In Figure 1 which is a simple example in two dimensions, input features are the lengths of different parts and the output feature is the total length of interest in this assembly.

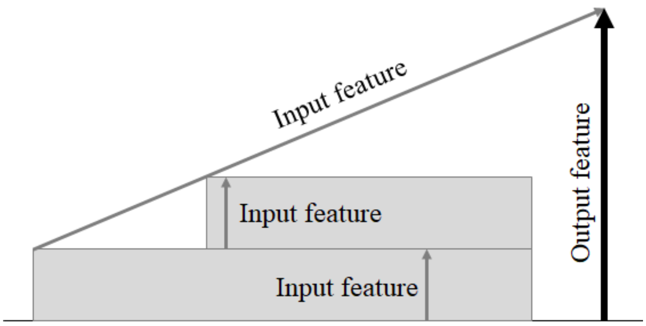

Fig. 1 An assembly example : inputs and output features identification.

In the design phase, both inputs and output tolerance intervals are assumed to be centered around the nominal dimension. To determine the variation around this nominal dimension, there are two main methods detailed in [3] : Worst Case and statistical approaches. These methods propose different ways to define an output tolerance range based on inputs tolerances. Some other approaches based on sampling, fuzzy arithmetic or analytical procedures are reviewed in [4].

Worst Case approach is to consider all assembly parts delivered at their worst acceptable value (assembly output tolerance equals the sum of the input tolerances). Statistical approach, also called RSS approach (square Root of the Sum of Squares), gives a result assuming all input features are normally distributed (assembly output tolerance equals the square root of the sum of squares of input features tolerances). Statistical result gives a much tighter tolerance range result than the worst case approach, but it does not hedge against the case where input are not reasonably close to their nominal value. To find a balance between this two approaches, Bender [5] proposed to multiply the statistical result by an empiric coefficient of 1.5 to obtain an inflated statistical result which is supposed to give a result tighter than the worst case but more conservative than RSS. However, this technique does not apply if number of assembly inputs is low as it gives a wider result than the worst case approach. Several other statistical methodologies have been studied to obtain the best trade-off between worst case and statistical approaches. For instance, Skowronski and Turner [9] proposed a method relying on Monte Carlo techniques. Choi *et al.* [10] studied an approach based on Taguchi's method requiring the definition a quadratic loss function. The tolerance allocation problem is formulated as a minimization of the sum of machining cost and quality loss. Manufacturing cost considerations for tolerance allocation is beyond the scope of this article. Pillet *et al.* [11] proposed to consider weighted inertial tolerancing. Inertial tolerancing works with mean square deviation (inertia) of the output feature as limit instead of considering a tolerance interval. Then, they applied a weighting system based on the number of assembly inputs to obtain a reasonable tolerance result. An other approach has been studied in [12] by taking an interest in the meaning of the conformity. Instead of limiting the assembly output variability, they propose a formal definition of statistical conformity that does not apply individually to a part but to a part population.

Note that tolerance intervals are highly related to the assemblies processes capabilities. Even if suppliers process capability indicators should be monitored as detailed in [13], the normality of features distributions can not always be verified.

One of the objective of the tolerancing is to assess the same confidence in a tolerance interval whatever the distribution of inputs are, as long as these inputs are delivered within the claimed tolerance range. Indeed, suppliers of parts receive a nominal value and two dimension limits. They are also required to follow a target distribution, however checking this compliance is difficult in practice. At the design stage, it is impossible to characterize the features distributions from measurement data. Uniform distribution is a better option to hedge against less favorable distributions of suppliers values.

Knowing lower and upper limits, the less informative distribution is the uniform distribution. It means results obtained with this assumption still stands for alternative distribution provided that distribution support is finite. If the support is not finite, as is the case for Gaussian distribution, uniform assumption is still a good candidate because this is a conservative approach.

A mathematical tool is proposed to define an accurate assembly output tolerance range considering uniform input features and taking into account the stack chain inputs structure. Indeed, result on output tolerance is highly dependant on how balanced is a stack chain. A balanced stack chain means that all contributors have the same impact on the output. Conversely, the predominance of one contributor in the assembly leads to an unbalanced stack chain. The aim is to present an analytical result that links stack chain inputs structure and output tolerance range. This kind of outcomes could be obtained from Monte Carlo simulation but such a procedure is not analytical and does not provide information on the link between inputs and output tolerances.

The paper is organized as follows: statistical framework is introduced in Section 2, main results are presented in Section 3: First part is devoted to traditional approach on deviations and following parts detail improvement on the upper bound accuracy and balance term introduction. In Section 4, a simulation study is carried out in order to represent and compare our results. Finally, an example on airframe assembly with real inputs data is performed.

2 Statistical framework

Consider a set of input features $X_1, \ldots, X_n \in \mathbb{R}$ and an output feature $Y \in \mathbb{R}$. All input features are assumed to be independent random variables for the reason that assembly parts are supposed to be separately produced. The main interest here is in the variability of the output Y and especially in a way to define a tolerance range for this feature.

From an engineering perspective, input features $X_1, \ldots, X_n \in \mathbb{R}$ shall correspond to the stack chain contributors of an assembly such as parts dimensions, while the output feature Y relates to the top level requirement.

Each input feature is assumed to be centered around a nominal dimension and has its own variability characterized by its tolerance range $[-v_i, v_i], \forall i \in \{1, \ldots, n\}$ where $v_1, \ldots, v_n > 0$ are the tolerance bounds. This variability reflects the uncertainty linked to the process (temperature, control plan, ground motion, delivery types, ...).

In order to discuss about the feature Y, assembly step must be modeled to represent the link between inputs and output of the assembly. for isoconstrained mechanisms, a common approach in tolerancing is the linear coefficient model (see [6]). If the variations are supposed to be small around the nominal dimension, the linear approach is appropriate. More elaborated models than linear one could be considered, such as studied in [7] where assembly geometry is taken into account. In this paper, the framework is to work with a linear model on centered tolerances.

Based on the knowledge of inputs tolerances and influence coefficients on the output, output result is seen as a linear combination of all inputs weighted by known influence coefficients (previously determined with a 3D CAD tool and only linked to the assembly geometry). Let denote $\alpha_1, \ldots, \alpha_n$ the coefficients for a linear tolerance model and input dimension features $Z_1, \ldots, Z_n$, then

$$Y = \sum_{i=1}^{n} \alpha_i Z_i.$$

For ease of notations, the weighted features denoted by $\alpha_1 Z_1, \ldots, \alpha_n Z_n$ are directly treated. These input features are denoted already multiplied by their respective influence coefficients by $X_1, \ldots, X_n$. In this formalism,

$$Y = \sum_{i=1}^{n} X_i.$$

For a given confidence level ρ, the aim is to determine the associated tolerance interval $[-t, t]$ for the output feature Y, verifying

$$\mathbb{P}(|Y| \geqslant t) \leqslant \rho.$$

The tolerance interval is determined based on the distribution of the output feature which depends on input features distributions.

A popular practice is to consider all features as Gaussian which leads to a Gaussian output feature. By applying the commonly used 6σ methodology, the confidence level is $\rho = 0.0027$. In Gaussian framework, the result is $t = 3\sigma_Y$ with σ_Y the standard deviation of the feature Y.

Within the Gaussian framework, the 6σ methodology gives the standard deviation of each input feature : $v_i/3, \forall i \in 1, \ldots, n$. As input features are assumed independent, the standard deviation of the output feature in the Gaussian case is $\frac{1}{3}\sqrt{\sum_{i=1}^{n} v_i^2}$. Again, the 6σ methodology leads to the interval $[-T_{RSS}, T_{RSS}]$ for the output feature tolerance, where $T_{RSS} = \sqrt{\sum_{i=1}^{n} v_i^2}$. This tolerance interval is commonly called the statistical result or RSS (Root Sum of Squared) result by the tolerancing community. However, as tolerance allocation is considered in the design phase, Gaussian assumption can not be verified from measurement data on features. Only tolerances bounds of input features $v_1, \ldots, v_n$ are available.

Input features are considered as uniform random variables, since it is the least informative available distribution given our knowledge about inputs. The purpose is to characterize the deviation of the sum of uniform independent random variables. Killmann and Von Collani [14] studied the distribution of the sum of uniform features. Their idea was to explicitly calculate density of the sum but such a closed form is numerically intractable and therefore not suited to our context.

Note that our objective is to focus on the quantile of the distribution of Y ensuring a given probability ρ to be out of tolerance. This probability value is fixed in our framework and the tolerancing problem uniformly according to ρ is not planned to be addressed here. This is the point of view of the field of *optimal transport* as developed in [15] but controlling distribution tails leads to poor results in practice for reasonable values of ρ.

In the uniform case, input features standard deviations are now $v_i/\sqrt{3}, \forall i \in 1,\ldots,n$, and the standard deviation of the output feature is $\frac{1}{\sqrt{3}}\sqrt{\sum_{i=1}^{n} v_i^2}$. In this case, the 6σ methodology is applied with standard deviations of uniform distributions and the output feature tolerance interval would be $[-\sqrt{3}\times T_{RSS}, \sqrt{3}\times T_{RSS}]$.

The coefficient $\sqrt{3}$ is an accurate coverage factor on the statistical result if the $v_1,\ldots,v_n$ are all equal but it does not address the case where they are unbalanced. Yet, if one of the feature predominates over others, for a same confidence level the output tolerance interval should be tighter, as shown in the Figure 2.

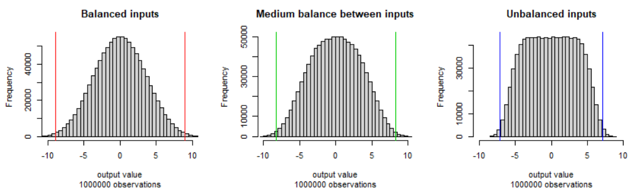

Fig. 2 Output feature distribution for several balance ratio of inputs.

The aim of our approach is to introduce a shape coefficient in order to correct the RSS interval result assuming the input distributions are uniforms. This balance indicator aims to determine how inputs contribution to the output variation is distributed. Indeed, this coefficient will depend on how unbalanced input features are. It also depends on the selected confidence level ρ. The value $\sqrt{3}$ for this shape coefficient means that $v_1 = v_2 = \cdots = v_n$. The more input features are unbalanced, the lower the form coefficient value is.

Next, the focus will be on the role of this coefficient and its impact on the probability ρ.

3 Main results

If Gaussian independent input features are considered with a tolerance interval $[-v_i, v_i]$, $\forall i \in \{1,\ldots,n\}$, the associated standard deviation from the 6σ methodology is $v_i/3$, $\forall i \in \{1,\ldots,n\}$. If the features are denoted N_i and $N_i \sim \mathcal{N}(0, v_i/3)$, then $\forall i \in 1,\ldots,n$, the standard Gaussian deviation inequality gives

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} N_i\right| \geqslant t\right) \leqslant 2\exp\left(-\frac{t^2}{2\sum_{i=1}^{n}\left(\frac{v_i}{3}\right)^2}\right)$$

and then

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} N_i\right| \geqslant \frac{1}{3}\sqrt{2log\left(\frac{2}{\rho}\right)\sum_{i=1}^{n} v_i^2}\right) \leqslant \rho$$

that is equivalent to

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} N_i\right| \geqslant l_\rho \times T_{RSS}\right) \leqslant \rho$$

with

$$l_\rho = \frac{1}{3}\sqrt{2log\left(\frac{2}{\rho}\right)}.$$

For fixed ρ and independent uniform input features $U_i \sim \mathcal{U}([-v_i, v_i])$, $\forall i \in 1,\ldots,n$, our aim is to determine f such that

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} U_i\right| \geqslant f \times l_\rho \times T_{RSS}\right) \leqslant \rho. \quad (1)$$

This f coefficient aims to correct the RSS value obtained in the 6σ Gaussian case, together with the fixed value l_ρ which manages the confidence level ρ.

3.1 Hoeffding approach for the deviation of a sum of bounded random variables

Traditional approaches based on deviations are related to the Hoeffding inequality which provides an upper bound on the probability that the sum of bounded independent random variables deviates more than a certain amount.

As detailed in [16] and [17], this inequality applied to the sum of uniform independent random variables $Y = \sum_{i=1}^{n} X_i$ gives a non asymptotic upper bound for the probability of deviation. This result is summarized in the Proposition 1.

Proposition 1. *Let $v_1, \dots, v_n > 0$, if $X_1, \dots, X_n$ are independent random variables such that*

$$\forall i \in \{1, \dots, n\},\ |X_i| \leqslant v_i,\ a.s.$$

then,

$$\forall t > 0, \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geqslant t\right) \leqslant 2\exp\left(-\frac{t^2}{2\sum_{i=1}^{n} v_i^2}\right).$$

Proof. See Section 2.6 in [17]. □

Setting $t = \sqrt{2\log\left(\frac{2}{\rho}\right)\sum_{i=1}^{n} v_i^2}$ leads to

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant \sqrt{2\log\left(\frac{2}{\rho}\right)\sum_{i=1}^{n} v_i^2}\right) \leqslant \rho,$$

and with $f = 3$ in order to match the expression (1). Now, it becomes

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant 3 \times l_\rho \times T_{RSS}\right) \leqslant \rho.$$

Hoeffding approach only takes into account the fact that random variables are bounded. However, here the information that features are uniform random variables is also available. This information will be used to find a tighter upper bound for the deviation of a sum of uniform random variables. As a result, a lower value for the coefficient f is obtained.

3.2 Chernov approach to improve the bound for a sum of uniform random variables

As it has just been mentioned, the Hoeffding approach is solely based on the support of the distribution involved in the deviation inequality. To improve such an upper bound, this distribution has to be considered more carefully. To this end, the well-known Cramér-Chernov bounding method is available. Such an approach is based on the following inequality derived from Markov's inequality and valid for any real random variable W and any real $\lambda > 0$,

$$\forall t > 0,\ \mathbb{P}\left(W - \mathbb{E}[W] \geqslant t\right) \leqslant e^{-\lambda t}\mathbb{E}\left[e^{\lambda(W-\mathbb{E}[W])}\right].$$

Thus, the method consists in optimizing this upper bound with respect to $\lambda > 0$ in order to exhibit a sharper deviation inequality.

The following proposition is stated and proved:

Proposition 2. *Let $v_1, \dots, v_n > 0$, if $X_1, \dots, X_n$ are independent random variables such that*

$$\forall i \in \{1, \dots, n\},\ X_i \sim \mathcal{U}([-v_i, v_i]),$$

then,

$$\forall t > 0, \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geqslant t\right) \leqslant 2\inf_{\lambda>0}\left\{\exp\left(\phi(\lambda, t)\right)\right\}$$

where the function ϕ is defined for any $\lambda, t > 0$ by

$$\phi(\lambda, t) = \sum_{i=1}^{n} \log\left(\frac{e^{\lambda v_i} - e^{-\lambda v_i}}{2\lambda v_i}\right) - \lambda t. \quad (2)$$

Proof. Let $\lambda > 0$, applying Markov inequality to the positive random variable $\exp\left(\lambda\sum_{i=1}^{n} X_i\right)$ gives the following upper bound on the probability that the sum of uniform independent random variables deviates more than $t > 0$

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geqslant t\right) \leqslant \frac{\mathbb{E}[e^{\lambda\sum_{i=1}^{n} X_i}]}{e^{\lambda t}}.$$

Thus, using the symmetry of the uniform distribution, for any $t > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geqslant t\right) \leqslant 2\exp\left(\phi(\lambda, t)\right).$$

The upper bound is valid for any value of $\lambda > 0$ and the announced result follows by taking the infimum according to λ.

□

The optimization of ϕ function with respect to λ is highly related to the input features balance. Next, a way to characterize this dependency is detailed.

3.3 Dependency on the features balance

The aim of this section is to provide details on how to determine the value of λ which is obtained from a function minimization in the upper bound previously presented.

A concept of balance between input features is introduced. This balance represents the discrepancy between the uniform distributions parameters : if all uniform random variables have the same parameters, it means a perfect balance between tolerance bounds. Otherwise, one of the random variables within the sum may have a much larger support set than others and it leads to imbalance between tolerance bounds.

The upper bound result from Proposition (2) is taken. The idea is to bound from above this result by introducing a specific term that identifies the influence of the balance within $v_1, \ldots, v_n$.

The focus is on the sum of logarithms in the function ϕ given in (2) that can be rewritten as

$$\sum_{i=1}^{n} \log\left(\frac{e^{\lambda v_i} - e^{-\lambda v_i}}{2\lambda v_i}\right) \tag{3}$$

$$= \lambda \sum_{i=1}^{n} v_i + \sum_{i=1}^{n} \log\left(\frac{1 - e^{-2\lambda v_i}}{2\lambda v_i}\right)$$

$$= n \log\left(\frac{1 - e^{-2\lambda \bar{v}}}{2\lambda \bar{v}}\right) + S_\lambda \tag{4}$$

with S_λ defined as follows

$$S_\lambda = \sum_{i=1}^{n} \left(\log\left(\frac{1 - e^{-2\lambda v_i}}{2\lambda v_i}\right) - \log\left(\frac{1 - e^{-2\lambda \bar{v}}}{2\lambda \bar{v}}\right)\right).$$

The term S_λ quantifies the imbalance between uniform distributions parameters. In the next two propositions, results are proposed about the upper bound on the probability that the sum of uniform independent random variables deviates from its expected value.

Proposition 3. *Let $v_1, \ldots, v_n > 0$ and the mean $\bar{v}$ be defined as*

$$\bar{v} = \frac{1}{n} \sum_{i=1}^{n} v_i.$$

If $X_1, \ldots, X_n$ are independent random variables such that $\forall i \in \{1, \ldots, n\}$, $X_i \sim \mathcal{U}([-v_i, v_i])$ then,

$$\forall t > 0,\ \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geqslant t\right) \leqslant 2 \exp\left(\psi(\lambda_0, t)\right).$$

where for any $\lambda, t > 0$

$$\psi(\lambda, t) = -\lambda t + \lambda n \bar{v} + n \log\left(\frac{1 - e^{-2\lambda \bar{v}}}{2\lambda \bar{v}}\right) + \lambda \sum_{i=1}^{n} |v_i - \bar{v}|$$

and λ_0 is such that

$$\frac{\partial \psi(\lambda_0, t)}{\partial \lambda} = 0$$

For a set of tolerance bounds $v_1, \ldots, v_n > 0$ and a fixed probability ρ, t verifies $\psi(\lambda_0, t) = \rho$. The value of interest t is obtained by inversion with respect to t of the function ψ. With this expression, the balance within $v_1, \ldots, v_n$ appears via $\sum_{i=1}^{n} |v_i - \bar{v}|$. Indeed, this term is large for unbalanced values $v_1, \ldots, v_n$ and small otherwise. Next, Propostion 3 is proven.

Proof. Let function h be defined as

$$\forall x > 0, \quad h(x) = \log\left(\frac{1 - e^{-x}}{x}\right).$$

This function is $\frac{1}{2}$-Lipschitz continuous (proof is postponed in the appendix A) and therefore

$$\forall x, y > 0, \quad |h(x) - h(y)| \leqslant \frac{1}{2} |x - y|. \tag{5}$$

This inequality is applied for $x = 2\lambda v_i\ \forall i \in \{1, \ldots, n\}$ and for $y = 2\lambda \bar{v}$ and sum the terms to obtain

$$S_\lambda \leqslant \lambda \sum_{i=1}^{n} |v_i - \bar{v}|.$$

The announced result follows from this upper bound on S_λ in the equation (4). □

In the previous proposition, the balance ratio of the v_i was quantified through the absolute values $|v_i - \bar{v}|$. It is natural to consider also the variance to this end and this is the purpose of the next proposition.

Proposition 4. *Let $v_1, \ldots, v_n > 0$ and the mean $\bar{v}$ and the variance $\operatorname{Var}(v)$ be defined as*

$$\bar{v} = \frac{1}{n} \sum_{i=1}^{n} v_i \quad \textit{and} \quad \operatorname{Var}(v) = \frac{1}{n} \sum_{i=1}^{n} (v_i - \bar{v})^2.$$

If $X_1, \ldots, X_n$ are independent random variables such that $\forall i \in \{1, \ldots, n\}$, $X_i \sim \mathcal{U}([-v_i, v_i])$ then,

$$\forall t > 0,\ \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geqslant t\right) \leqslant 2 \exp\left(\tilde{\psi}(\lambda_0, t)\right).$$

where for any $\lambda, t > 0$

$$\tilde{\psi}(\lambda, t) = -\lambda t + \lambda n \bar{v} + n \log\left(\frac{1 - e^{-2\lambda \bar{v}}}{2\lambda \bar{v}}\right) + \frac{n\lambda^2 \operatorname{Var}(v)}{2}$$

and λ_0 is such that

$$\frac{\partial \tilde{\psi}(\lambda_0, t)}{\partial \lambda} = 0$$

As in Proposition 3, tolerance bounds $v_1, \ldots, v_n > 0$ and a fixed probability ρ lead to a value t obtained by inversion with respect to t of the function $\tilde{\psi}$. The proof of Proposition 4 is as follows.

Proof. The Lipschitz continuity of h ensures the following inequality (see for example Lemma 1.2.3 in [18] for a proof of this result):

$$\forall x, y > 0, \quad |h(x) - h(y)| \leqslant \frac{L}{2} \|x - y\|^2 .$$

This result applied to $x = 2\lambda v_i \; \forall i \in \{1, \ldots, n\}$ and for $y = 2\lambda \bar{v}$ gives

$$\forall \lambda, v_1, \ldots, v_n > 0, \quad |h(2\lambda v_i) - h(2\lambda \bar{v})| \leqslant \frac{\lambda^2}{2} \|v_i - \bar{v}\|^2$$

and finally, since $S_\lambda = \sum_{i=1}^n (h(2\lambda v_i) - h(2\lambda \bar{v}))$,

$$S_\lambda \leqslant \frac{n\lambda^2 \operatorname{Var}(v)}{2}.$$

The announced result follows from this upper bound on S_λ in the equation (4). □

4 Applications

The first part of this section will describe how our upper bound behaves on different stack chains obtained from simulations. The second part will focus on a practical study on an industrial example of tolerance definition within an aircraft assembly.

4.1 Simulations

4.1.1 Tolerance design on an assembly example

First step is to simulate stack chains. Stack chains represented by features X_1, X_2, X_3, X_4, X_5 are randomly generated with a number of inputs $n = 5$. Their tolerance intervals values v_1, v_2, v_3, v_4, v_5 are also randomly generated between 1 and 5. The aim is to assess on the output tolerance variability via tolerance intervals to be defined, and accepting an out-of-tolerance rate ρ. A stack chain is generated with tolerance inputs intervals bounds and traditional output results as following:

Table 1 Example of a stack chain characterization.

X_1	X_2	X_3	X_4	X_5	RSS	WC
$v_1 = 5$	$v_2 = 4$	$v_3 = 3$	$v_4 = 2$	$v_5 = 1$	7.4	15

The first approach is based on Monte Carlo methods. $N = 10^5$ observations are generated from $n = 5$ uniform distributions and are summed. The probability to be out of a given tolerance interval can therefore be asymptotically estimated and considered as a near theoretical result. The two following methods give an output interval bound according to the two approaches proposed in this paper. The provided upper bound depends on the selected confidence level ρ. This level is the probability for the output feature to be out of the designed output tolerance interval. The higher the confidence level, the wider the tolerance interval is. Indeed, if more values to be out of tolerance are allowed, the output tolerance interval should be broader. Figure 3 illustrates the results obtained from Monte Carlo draws and from the methods detailed in this paper. Among the three methods, Figure 3 shows that the Lipschitz and Quadratic approaches give a looser upper bound than the Chernov method.

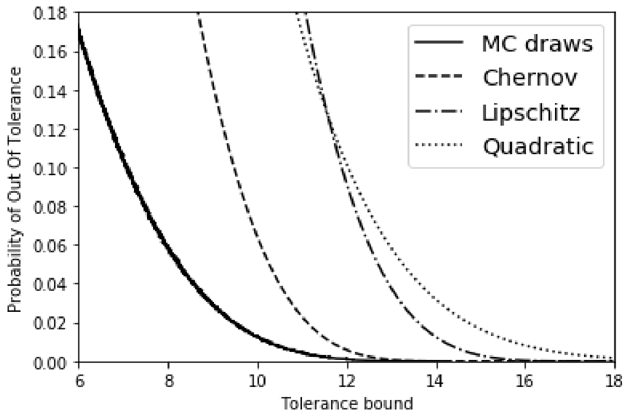

Fig. 3 Example of the behavior of the probability to be out of tolerance with respect to the value of the output tolerance bound for the discussed approaches.

The benefits of the methods proposed in this article is that they do not require Monte Carlo draws, nor asymptotic estimation of the probability to be out of tolerance. Indeed, the provided bounds offer theoretical non asymptotic guarantees and eliminate any risk of rare events that Monte Carlo methods would not generate. Moreover, for large assemblies, the number of Monte Carlo draws needed to obtain a sufficiently sharp result would grow with the number of input features in the assembly. The discussed formula are closed and directly usable in practice and cheaper to compute than Monte Carlo simulations.

4.1.2 Influence of the assembly geometry

In order to represent the balance within a stack chain, previous sections introduced the following term

$$S_\lambda = \sum_{i=1}^{n} \left(\log \left(\frac{1 - e^{-2\lambda v_i}}{2\lambda v_i} \right) - \log \left(\frac{1 - e^{-2\lambda \bar{v}}}{2\lambda \bar{v}} \right) \right).$$

In particular, taking arbitrarily parameter $\lambda = 1$ leads to

$$S_1 = \sum_{i=1}^{n} \left(\log \left(\frac{1 - e^{-2v_i}}{2v_i} \right) - \log \left(\frac{1 - e^{-2\bar{v}}}{2\bar{v}} \right) \right).$$

This quantity can be used as an indicator of the balance of the stack chain. Indeed, the more balanced the stack chain is, the lower the value is and vice versa.

As mentioned in the previous part, one of our main issue is to take into account the traditional RSS result and the balance of the stack chain. This explains why hereafter the choice is to display the coefficient f with respect to some balance indicator such as S_1 or other dispersal measures within input features.

First, in Figure 4, the results are showed for the coefficient f obtained from a Monte Carlo simulation of uniform distributions with 2×10^5 drawn observations. Next, the coefficient f is displayed according to the Chernov methodology as detailed in Proposition 2. Finally, it shows that boundings by Lipschitz and Quadratic approaches directly depend on some balance factor.

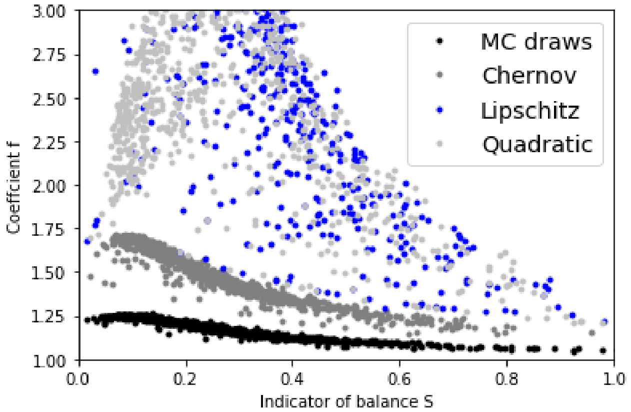

Fig. 4 Link between the coefficient f and the balance factor S_1 with parameter $\rho = 0.05$.

An almost linear behavior of the result with respect to the balance factor S_1 is observed for the Monte Carlo approach and for the Chernov methodology. This factor S_1 seems to be a relevant indicator to characterize f coefficient. As expected and due to the upper bounds defined in these methods, both Lipschitz and Quadratic approaches give results much more conservative. Still the Quadratic approach is more accurate for small values of S_1. This is explained by the fact that the result with Quadratic approach in Proposition 4 takes into account the variance of input feature bounds. For a small S_1, input features are balanced and variance is a more regular control quantity for the structure of the stack chain than the sum of absolute deviations around the mean $\bar{v}$ introduced in Proposition 3.

4.2 Case study

In this part, the focus is on industrial practices at Airbus. First, the example of an assembly from an aircraft is taken and results from the methodology proposed in this article are showed. Then, the common process of tolerance definition at Airbus is detailed and explanation about how it is related to the approaches presented in the paper are provided. Finally, all stack chains are represented in a real aeronautical product perimeter according to the balance factor.

4.2.1 An assembly example

The assembly in Figure 5 is related to a genreic frame misalignment for an Airbus aircraft.

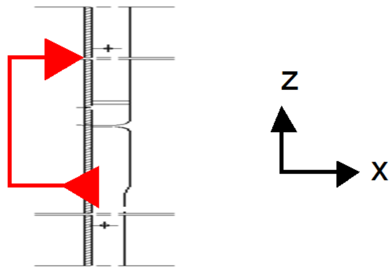

Fig. 5 Example of vertical frame misalignment with respect to the last rigid point.

Table 2 gives the stack chain data of this requirement. Tolerance bounds value have been modified.

Table 2 Stack chain of the top level requirement: frame misalignment - last rigid point.

Name of the contributor	**Tolerance interval**
Frame 1	±1
Frame 2	±0.5
Process tolerance	±0.25
Process tolerance	±0.23
Process tolerance	±0.2
Process tolerance	±0.2
Process tolerance	±0.15
Process tolerance	±0.13
Process tolerance	±0.1
Process tolerance	±0.09

It involves 10 input features in the assembly and tolerance data are scaled and unit free. Table 3 provides traditional tolerancing worst case and RSS results. The application of the different methods proposed in this paper gives the results depicted in Figure 6.

Table 3 Result for the top level requirement: frame misalignement - last rigid point.

Worst Case result	RSS result
±2.85	±1.23

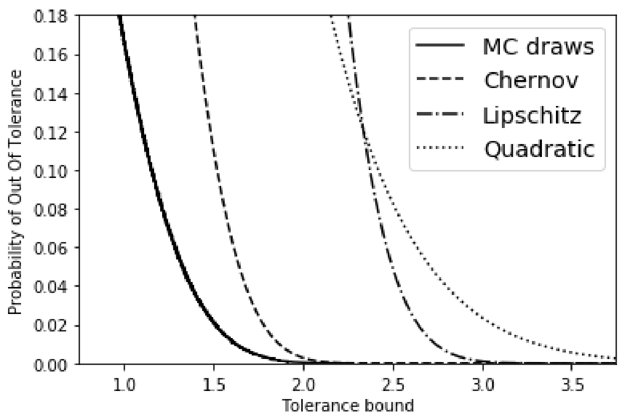

Fig. 6 Real case study of a bound value according to the confidence level.

Results for the real case study are very similar to simulated data conclusion. Indeed, Lipschitz and Quadratic analytical approaches still do not give a sharp bound result compared to the Monte Carlo and Chernov approaches. In practice, the more accurate method in order to design an output tolerance should be the Chernov one. Note that the values for the probability of Out Of Tolerance probability considered are generally the lower percentages on the y-axis as the aim in practice is to limit the scrap rate and in Airbus practices, the reference value is often 0.27%. This reference value is displayed in the Figure 7 which focuses on the percentage below 1% for Out Of Tolerance probability. The limits defined by RSS and WC methods applied to the assembly example are also displayed.

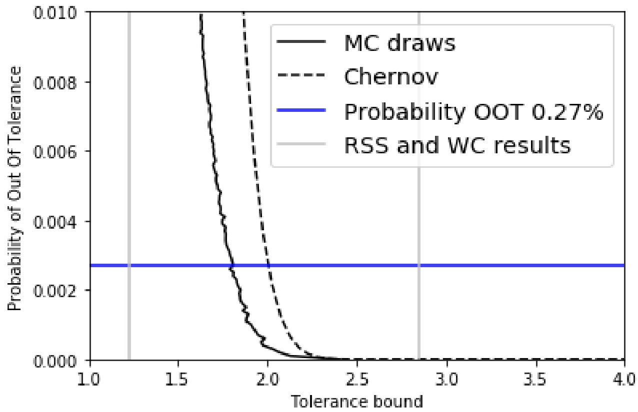

Fig. 7 Real case study of a bound value according to the confidence level zoomed on relevant values for Airbus.

4.2.2 Industrial practices : Airbus

Among the methods used by Airbus to define a tolerance in the design phase, one of them is an approximation of Monte Carlo simulation data under uniform assumption for contributors distribution and a disproportion parameter. As for the Gaussian case, quantile at 0.27% are observed on Monte Carlo simulations and a linear regression with respect to the factor parameter is carried out to obtain the result. For a set of tolerance bound $v_1, \dots, v_n > 0$ for an input features balance ratio D, this rule gives an output feature tolerance interval $[-T_{Airbus}, T_{Airbus}]$ defined as :

$$T_{Airbus} = \beta \times (-0.56D + 1.04) \times T_{RSS} \quad (6)$$

with T_{RSS} as defined in previous parts,

$$\forall v_1, \dots, v_n > 0, \quad D = \frac{\max_i(v_i) - \bar{v}}{\sum_{i=1}^{n} v_i}$$

and $\beta > 0$ a coefficient. In practice and without claim of universality, Airbus industrial approach takes $\beta = 1.6$ as a relevant value for business and in order to ensure continuity with former practices.

This D factor measures how far from the mean is the main contributor of the stack chain and has the advantage of being understandable. This quantity is highly correlated to the term S_1 previously introduced as shown on the Figure 8 :

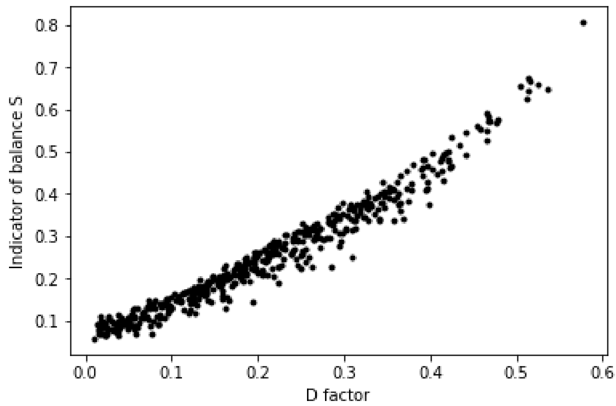

Fig. 8 Correlation between the term S_1 and the balance factor D.

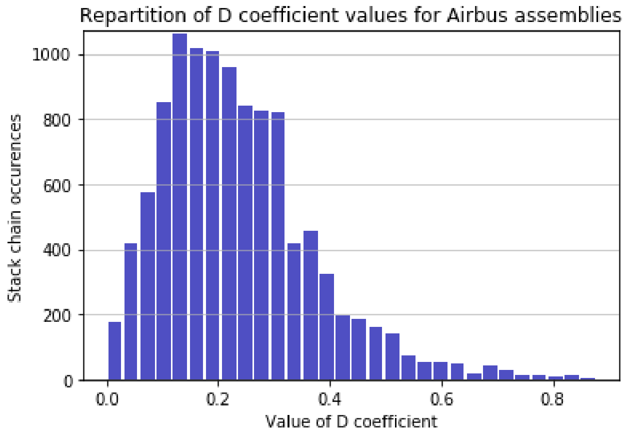

Fig. 10 Repartion of D values for Airbus assemblies.

With this definition, a high D still implies an unbalanced stack chain. Conversely, a small value of B means a balanced between stack chain inputs. Figure 9 shows a few examples of this factor with the respect to the stack chain structure.

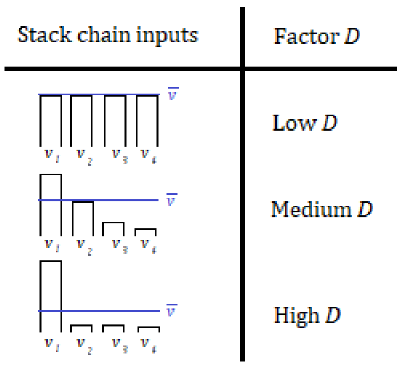

Fig. 9 Stack chain structure and balance factor D.

In practice, 70% of the stack chains have a D value between 0.10 and 0.36. The Figure 10 details the repartition of the coefficient value for Airbus stack chains.

In the industrial example of the frame misalignment. The probability to be out of tolerance for the output feature is set at $\rho = 0.0027$, which corresponds to the acceptable 0.27% out of the interval from the 6σ methodology. Table 4 summarizes the tolerance interval obtained for the frame misalignment. Three results are displayed : the Monte Carlo approach with 200000 drawn observations for each input feature, the Chernov approach proposed in this article with $\rho = 0.0027$ and the industrial practice presented in (6).

Table 4 Tolerance interval results according to the different approaches.

Method	Monte Carlo $\rho = 0.27\%$	Chernov $\rho = 0.27\%$	Industrial practice
Tolerance interval	± 3.56mm	± 4.01mm	± 3.53mm

For a level $\rho = 0.27\%$, the result from the industrial rule is very close to the value observed on Monte Carlo simulations. The Chernov method gives a more conservative result but ensures a precise probability ρ to be out of the interval for the output feature.

4.2.3 Performance of the different approaches on industrial cases

Focusing on a real sample of aeronautical assemblies, all stack chains have been analyzed in order to obtain the value of the f coefficient times l_ρ for $\rho = 0.27\%$, according to the different methodologies. The Airbus rule that can be used for decision helping is also displayed. This is an approximation for the selected confidence level.

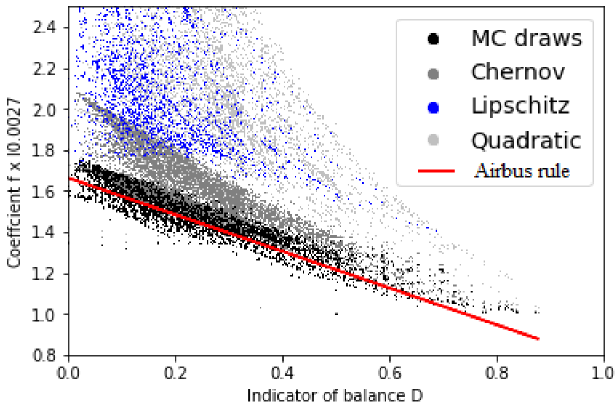

Fig. 11 Link between the term introduced and the balance factor D.

The same trend that for simulated data is retrieved: Better result for Chernov approach and linearity in D. For Monte Carlo simulations, a similar linear behaviour is observed with respect to the balance factor D. The airbus rule seems to be a good approximation for evaluate the coeficient f times l_ρ with a confidence level $\rho = 0.27\%$.

5 Conclusion

Robust approaches are proposed for tolerance definition in the design phase allowing the management of confidence level. From known input tolerance intervals of an assembly and for a selected confidence level, the output tolerance interval can be determined. The result will be robust against poor or unknown industrial capabilities because uniform distributions on tolerance intervals are assumed for input features.

The Chernov method is particularly accurate and gives an output tolerance interval result close to the reality, tight enough to be industrially relevant, and ensures also the selected probability as confidence level. A balance factor is also provided. This factor is strongly related to how tight an interval should be according to the disproportion of the stack chain. An almost linear behavior of the result is obtained from the Chernov methodology with respect to this balance factor.

Future directions of this work would be to consider more adversarial distributions for input features. For instance, bimodal distributions or truncated distributions could be studied in order to hedge against industrial practices with the induced bias of machine or thrust effect. Additionally, extension to this work in more general nonlinear settings for the model output Y as a function of the input features could be thought of. Indeed, if the model is nonlinear but the variance of each input component is finite, [19] suggests that one could also control the model output as proposed in our study.

A Proof of Lipschitz continuity of function h

The definition of Lipschitz continuity of a function $f : \mathbb{R} \to \mathbb{R}$ is recalled. Let $L > 0$, if the function is such that

$$\forall x, y \in \mathbb{R},\ |f(x) - f(y)| \leqslant L\,|x - y|$$

then, f is said to be Lipschitz continuous with constant L. In this appendix, the previously claimed Lipschitz continuity of function h defined by

$$\forall x > 0, \quad h(x) = \log\left(\frac{1 - e^{-x}}{x}\right)$$

is proven with constant $L = 1/2$.

The first and second derivative functions of h are easily obtained as

$$\forall x > 0, \quad h'(x) = \frac{e^{-x}}{1 - e^{-x}} - \frac{1}{x}$$

and

$$\forall x > 0, \quad h''(x) = \frac{1 + (1 + x^2)e^{-x}}{x^2(1 - e^{-x})^2}.$$

Since $h''(x) \geqslant 0$ for any $x > 0$, then h' is an non decreasing function. Moreover, $h'(x)$ tends to $-1/2$ when $x \to 0^+$ and to 0 when $x \to +\infty$. Thus, the conclusion is that $|h'(x)| \leqslant \frac{1}{2}$ and finally that h is Lipschitz continuous with $L = 1/2$ by a straightforward integration argument.

Acknowledgements Industrial application detailed in the article is the topic of a patent application FR 15508 (patent pending n°1912668).

This work was partly supported by the French *Agence Nationale de la Recherche* through a CIFRE contract 2017/1390 in partnership with Airbus.

References

1. ASME Y14.5M-1994, *Dimensioning and Tolerancing*, The American Society of Mechanical Engineers, 1994.
2. ISO 1101:2017, *Geometrical product specifications (GPS) — Geometrical tolerancing — Tolerances of form, orientation, location and run-out*, ISO/TC 213 - International Organization for Standardization, 2017.
3. K. W. Chase and W. H. Greenwood, *Design issues in mechanical tolerance analysis*, Manufacturing Review, 1988, 1(1), pp. 50-59.
4. B. Heling, T. Oberleiter, B. Schleich , K. Willner , S. Wartzack, *On the Selection of Sensitivity Analysis Methods in the Context of Tolerance Management*, ASME. J. Verif. Valid. Uncert. March 2019; 4(1): 011001. https://doi.org/10.1115/1.4043912.

5. A. Bender, *Benderizing tolerances-a simple practical probability method of handling tolerances for limit-stack-ups*, Graphic Science, 1962, vol. 17.
6. S. C. Liu, S. J. Hu, *Variation Simulation for Deformable Sheet Metal Assemblies Using Finite Element Methods.*, ASME. J. Manuf. Sci. Eng. August 1997, 119(3): 368–374. https://doi.org/10.1115/1.2831115.
7. Y. Wang, L. Li, N. W. Hartman, J. W. Sutherland, *Allocation of assembly tolerances to minimize costs*, CIRP Annals, Volume 68, Issue 1, 2019, Pages 13-16, ISSN 0007-8506, https://doi.org/10.1016/j.cirp.2019.04.027.
8. Greenwood, W. H., and Chase, K. W. (August 1, 1988). *Worst Case Tolerance Analysis with Nonlinear Problems.*, ASME. J. Eng. Ind. August 1988; 110(3): 232–235. https://doi.org/10.1115/1.3187874
9. V. J. Skowronski and J. U. Turner, *Using Monte-Carlo variance reduction in statistical tolerance synthesis*, Computer-Aided Design, 1997, vol. 29, no 1, p. 63-69.
10. H.-G. R.Choi, M.-H. Park and E. Salisbury, *Optimal tolerance allocation with loss functions*, Journal of Manufacturing Science and Engineering, 2000, vol. 122, no 3, p. 529-535.
11. M. Pillet, D. Duret and A. Sergent, *Weighted inertial tolerancing*, Quality Engineering, 2005, vol. 17, no 4, p. 687-693.
12. L. Leblond and M. Pillet, *Conformity and statistical tolerancing*, International Journal of Metrology and Quality Engineering 9, 2018: 1.
13. P. J. Drake, P. J. Drake and V. Srinivasan, *Dimensioning and tolerancing handbook*, New York : McGraw-Hill, 1999, ch. 8.
14. F. Killmann and E. Von Collani, *A note on the convolution of the uniform and related distributions and their use in quality control*, Economic Quality Control, 2001, vol. 16, no 1, p. 17-41.
15. C. Villani, *Topics in optimal transportation*, Issue 58 of Graduate studies in mathematics, ISSN 1065-7339, American Mathematical Soc., 2003.
16. W. Hoeffding, *Probability Inequalities for Sums of Bounded Random Variables*, Journal of the American Statistical Association, 58:301, 13-30, DOI: 10.1080/01621459.1963.10500830, 1963.
17. S. Boucheron, G. Lugosi and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*, Oxford university press, 2013.
18. Y. Nesterov, Yurii, *Introductory lectures on convex optimization: A basic course.* Vol. 87. Springer Science & Business Media, 2013.
19. A. Marchina. *Concentration inequalities for functions of independent random variables*, Ph.D. Thesis. Université Paris-Saclay, 2017.