Basis of *Best-Estimate Plus Uncertainties (BEPU)* Methodologies in Deterministic Safety Analysis of nuclear plants. Statistical methods in the uncertainty analysis

CSN

Colección
Otros Documentos
46.2024

CSN CONSEJO DE SEGURIDAD NUCLEAR

Basis of *Best-Estimate Plus Uncertainties (BEPU)* Methodologies in Deterministic Safety Analysis of nuclear plants.

Statistical methods in the uncertainty analysis

Rafael Mendizábal Sanz

Área de Ingeniería del Combustible Nuclear – Nuclear Fuel Engineering Branch.
Consejo de Seguridad Nuclear

Madrid, 2024

CSN CONSEJO DE SEGURIDAD NUCLEAR

Colección Otros Documentos
ODE-04.26

Publicado y distribuido por:
Servicio de Publicaciones
Consejo de Seguridad Nuclear
Pedro Justo Dorado Dellmans, 11. 28040 - Madrid
http://www.csn.es
peticiones@csn.es

Maquetación e Impresión:
Grupo Editorial Cinca, S.A.
General Ibáñez Ibero, 5-A, 28003 Madrid
Tel.: 91 553 22 72
grupoeditorial@edicionescinca.com
www.edicionescinca.com

Depósito legal: M-25771-2024

Index

Index of figures

Index of tables

Foreword

Foreword

The present book is devoted to describing the main basis of the BEPU methodologies used for deterministic safety analysis of nuclear plants. It is the product of many years of work in the Consejo de Seguridad Nuclear (CSN), the Spain's nuclear regulatory authority, in the evaluation and licensing of BEPU methodologies for accident analysis in Spanish nuclear plants. Such experience has crystalized in a number on scientific articles, presentations in congresses and meetings and lectures given in courses, which constitute a main source for the book.

The text tackles in detail the basis of the uncertainty analysis of calculated quantities and the inclusion of the uncertainty in the acceptance criteria established for such quantities. Then it refers to the statistical methods devoted to the testing of the fulfilment of the acceptance criteria. All these topics are treated in a very general fashion, so that the book can be interesting for practitioners on uncertainty analysis in different scientific and engineering fields. Anyway, the exposition never loses sight of the nuclear DSA, and therefore the statistical methods are those applied to a greater or lesser degree in such field. In this sense, the nonparametric Wilks' method and connected methods based on order statistics receive an especial attention. Methods not based on random sampling (e.g. experimental design methods, deterministic sampling or the law of linear error propagation) are not covered in the text.

A necessary companion to uncertainty analysis of safety outputs is sensitivity analysis, to help in the selection of the inputs that propagate most of the uncertainty to the output. The book focuses on the uncertainty analysis and does not address in depth the sensibility analysis.

The document is written from a regulatory standpoint, giving special relevance to the classification and description of methods. It somehow differs from other Statistics texts, where the methods are compared in detail, especially from the point of view of efficiency. Obviously, an accurate and precise knowledge of safety quantities calculated in accident analyses is desired. But the regulator may also consider the advantage of more conservative (i.e. less efficient) methods, assuming that they add safety margin to the results.

The book has a first part (chapters 1 to 3) describing the basis of uncertainty analysis of calculated quantities, the formulation of acceptance criteria for uncertain quantities and the statistical methods applied in testing the fulfilment of such criteria. This part gives the basis to understand the BEPU calculations in DSA of nuclear plants.

The second part of the book (chapters 4 and 5) refers to two additional topics in BEPU studies. One is the definition of safety margin (SM), a concept very used in the nuclear realm but without a universal definition. From the BEPU acceptance criteria a probabilistic definition of safety margin can be derived, introducing a special type of metrics in the space of safety quantities which proves very fruitful.

Chapter 5 deals with the validation of DSA methodologies, which are the calculational frameworks to calculate safety quantities in DBS. Methodologies must be validated against real data (e.g. experimental data). Validation and licensing are, in a sense, symmetric processes. Validation compares calculated values with real values, and licensing compares calculated values and regulatory limits. Therefore, BEPU tools could be (and should be) applied to validation of methodologies. As shown in chapter 5, interesting conclusions can be drawn from the probabilistic definition of safety margin.

I. Models, calculations, uncertainty

I. MODELS, CALCULATIONS, UNCERTAINTY

1.1. Calculations and models

Calculations are essential in Science and Engineering. Technological systems are designed and studied by means of calculations. The importance of calculus is so substantial that it is often considered one of the basic pillars of science, along with theory and experimentation [78].

To make calculations we need computational models, implemented in *codes*. Science works by elucidating reality and building and validating models of it. Computational models are numerical representations of real systems. They are used to make calculations, also termed simulations or predictions.

It is important to know the reliability of the results of calculations, their fidelity to reality, which depends mainly on two elements:

- The quality of computational models, related to their ability to predict real phenomena.
- The knowledge of quantities that are input to the calculation. A poor knowledge of them implies a low reliability of the results.

These two elements introduce uncertainty in the results of the calculations, as we describe below.

1.2. Uncertainty

Our knowledge of physical reality through models is inevitably afflicted by uncertainty. In general, we do not know the true value of a physical quantity. In some cases, this is due to inherent variability of the quantity; in other cases, it is due to imperfect knowledge of the quantity. Furthermore, models are imperfect because they are simplified versions of reality. Hence, we do not have full certainty that our simulations faithfully represent physical reality. We describe this fact saying that our calculated results are uncertain or have uncertainty. In the scientific and engineering realm, the term "uncertainty" connects with concepts such as lack of information, imperfect knowledge, lack of predictive capability, variability of magnitudes…

Definitions of uncertainty can be found in the scientific realm. A well-known and widely accepted definition of measurement uncertainty says… *parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand* [5]. Based on this definition, we can define the uncertainty of a numerical quantity as a mathematical representation of the dispersion (or multiplicity) of values that can reasonably be

attributed to it. The adopted mathematical representation must have a translation to numbers: uncertainty is defined mathematically and then is calculated numerically.

As stated, there are two main causes of the uncertainty of a quantity:

- *Variability.* The quantity may exhibit heterogeneity, and this fact impedes an exact prediction. For instance, the quantity may have fluctuations in time and/or space; or it may be an attribute of a manufactured item, so that it exhibits the uncertainty due to the manufacturing process.
- *Ignorance.* It is the lack of information about the value of the quantity.

In a sense, uncertainty can be regarded as a characteristic of the couple magnitude / observer [64]. The variability produces an uncertainty attributable to the magnitude, which is termed *aleatory* (or aleatoric, or random). On the other hand, ignorance is attributable to the observer, and produces an uncertainty termed *epistemic.* Epistemic uncertainty may be regarded as subjective, when it refers to the ignorance of an individual. When it represents the "collective ignorfance" (e.g. of a scientific community) about an issue, it has a more objective character.

Aleatory and *epistemic* are the two classical categories of uncertainty. The distinction between them, aside from the cause, is the reducibility. Epistemic uncertainty of a quantity is reduced when more information is obtained about it, while aleatory uncertainty is irreducible. Sometimes, epistemic uncertainty can be regarded as irreducible, when obtaining new information is very expensive or laborious (e.g. when it needs making new and costly experiments). Conversely, as we will show, sometimes aleatory uncertainties are treated as epistemic, when the interest is focused on the range of values of quantities rather than on their distribution.

Ultimately, uncertainty represents the lack of reliability in the value of the quantity, due to variability and/or to ignorance of the observer or user.

It is useful to recall the possible sources of information about the value of a physical quantity:

- Measurement.
- Control (when the value of the quantity is being controlled).
- Manufacturing specifications, when the quantity corresponds to a manufactured item.
- Calculation from values of other quantities, by means of predictive models.
- Expert opinion.
- …

Each one of these sources has an associated uncertainty. The uncertainty associated to measurement and control comes from the limitation of physical devices. On the other hand, the uncertainty associated to a calculated quantity derives from the imperfection of the prediction model and from the fact that the inputs to the calculation are commonly uncertain quantities as well.

In this document we focus on the uncertainty of quantities calculated with predictive models. We next discuss about the sources of such uncertainty.

1.3. Models and the black-box approach

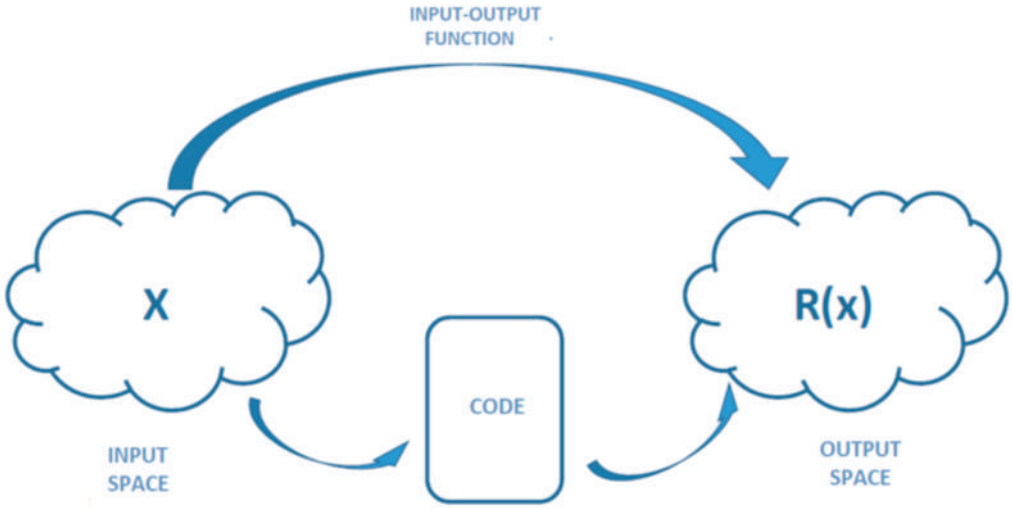
Models collect and structure our knowledge of physical reality and enable us to make predictions about it. Models are constructed in successive stages, each one starting from the product of the previous one.

1) Conceptual (i.e. qualitative) model
2) Mathematical model
3) Numerical model
4) Computational model (i.e. code)

Computational models transform input variables into output variables. The process of transforming by the model a given input in an output is a calculation or simulation. The set of input variables needed for a calculation of the model is often called the *input deck.* The set where input variables take values is the *input range or input space.* The set where output variables take values is the *output range or output space.*

Every computational model or code defines a mapping or transformation from the input space to the output space (*input-output transformation*). Every point in the input space represents a possible input deck of the model / code. The code transforms the input deck into a point in the output space (figure 1.1).

Figure 1.1. Computational code and input-output function. Adapted from [69]

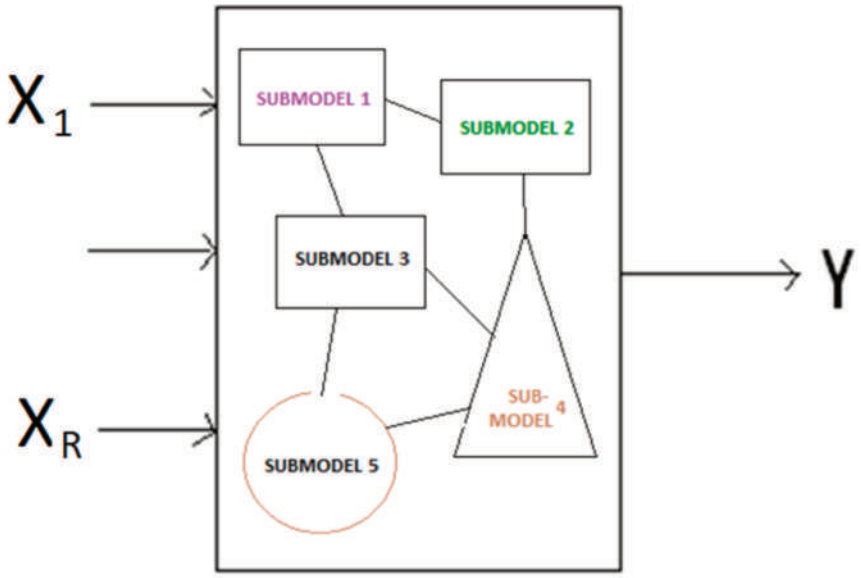


Deterministic models and codes are those that always produce the same output from a specified input deck. This repeatability property may seem trivial, but it is not so, because there are models that include random numbers generators, so that different runs of the code with the same inputs may lead to different choice of random numbers and thus different results.

The input-output transformation defined by a deterministic model is, mathematically, a function, because the image of a point in the input range is a single point in output range. When the sole information considered about a model is its associated input-output function, we are making the so-called "black-box approach" [64]. The adoption of such approach does not mean a real ignorance about the form of the model (i.e. the mathematical structure, equations or associated algorithms); simply, such information is not used (or is minimally used) in the analysis of outputs. In this case, the action of the model is regarded in an empirical fashion, via input-output computational experiments.

Codes are typically composed by interconnected submodels (figure 1.2). A submodel performs a part of the task of the complete code. In fact, the complete model is a hierarchical structure of submodels.

Figure 1.2. Computational code and submodels



Every model or submodel has input variables, output variables and parameters. Model parameters are numerical quantities included in the formulation of the model that are neither input nor output variables. If the value of any model parameter is changed, we consider that the model changes, even if its functional form is unchanged. But model parameters can be regarded, in a wide sense, as a special type of input parameters.

Suppose a computational model, producing a transformation input-output:

$$Y = M(X,C) \tag{1.1}$$

$\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{C}$ are, in general, multidimensional magnitudes. $\mathbf{Y}$ groups the safety outputs of interest. For calculations of Nuclear Safety (Deterministic Safety Analysis) the outputs of interest are calculated safety quantities defined by the regulator.

In (1.1) the input variables are represented by $\mathbf{X}$ and $\mathbf{C}$. $\mathbf{C}$ groups the model *calibration parameters* (also termed *free parameters*), a special type of model parameters that are used for calibration of the model i.e. for the adjustment of the model to real data.

On the other hand, $\mathbf{X}$ groups the remaining input variables (initial and boundary conditions, geometric and topological parameters, and other model parameters).

M is the input-output transformation defined by the model. For deterministic models, **M** is a function. As previously said, the "black-box approach" assimilates the model to its associated transformation **M**. In the sequel, the same letter (**M**) will be used indistinctly to symbolize the model and its associated transformation.

The term "input uncertainty" refers to the uncertainty of the input variables (**X**, **C**).

Expression (1.1) symbolizes the calculation of **Y** with a code. But how the calculated **Y** compares with the real value of **Y**? We complete (1.1) as

$$Y_{true} = M(X,C) + D(X) \tag{1.2}$$

$\mathbf{Y}_{\mathbf{true}}$ is the value of **Y** that would be produced in reality, while **Y** is the corresponding value calculated in a simulation of reality with the model **M**. The difference between the real and the calculated value of **Y** is **D(X)**, termed the *model bias* (also termed *model error,* or *model inadequacy*, or *model discrepancy*...) [49].

1.4. Propagation of uncertainty through models

Computational models (codes) operate with inputs and produce results (outputs). If we make mathematical operations with uncertain inputs we obtain uncertain results; the latter inherit their uncertainty from the former. It is said that uncertainty is *propagated* through a code calculation, from inputs to outputs. Therefore, the propagation explains the uncertainty of **Y** in (1.1). A part of the output uncertainty is propagated from the calibration parameters **C**; we can associate such uncertainty to the imperfection of models (they are simplified mathematical descriptions of reality). A perfect model would not need to be calibrated.

On the other hand, expression (1.2) shows that $\mathbf{Y}_{\mathbf{true}}$ has the uncertainty of the calculated **Y** combined with the uncertainty associated to the model bias. Indeed, model bias is an imperfectly known quantity (otherwise, adding **D** to **Y**, we would produce a perfect model), and hence has epistemic uncertainty. Therefore, the uncertainty derived from the imperfection of models has two components: one is propagated from calibration parameters to the calculated value; the other, produced by the model inadequacy, affects the true and unknown value of **Y**.

In summary, models and their input parameters are the two basic sources of uncertainty affecting our knowledge derived from calculation of the physical reality.

But this is not the complete picture yet. There is still another source of uncertainty that we next describe.

1.5. Second-level uncertainty

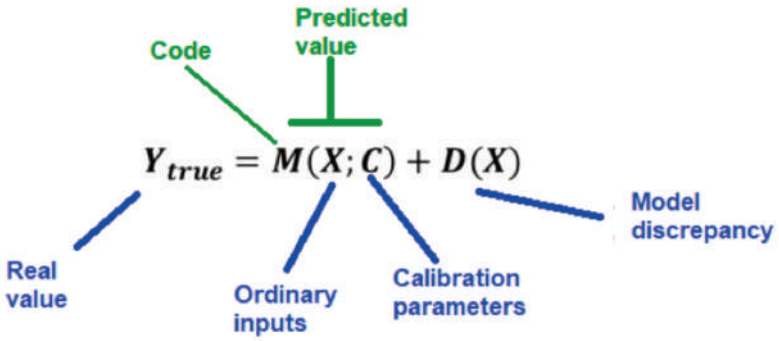
We have just defined uncertainty as a mathematical description of random variability and/or of ignorance about numerical quantities, although we have not specified yet how uncertainty is quantified.

In fact, uncertainty must be obtained from information and data about the quantity. Uncertainty must be modelled and calculated, and this process introduces a second layer of uncertainty. Calculated uncertainty may itself be an uncertain quantity. For the uncertainty of the calculated uncertainty, we use the term *second-level uncertainty*, and also *metauncertainty*. It has epistemic nature, because it reduces when the information about the quantity increases.

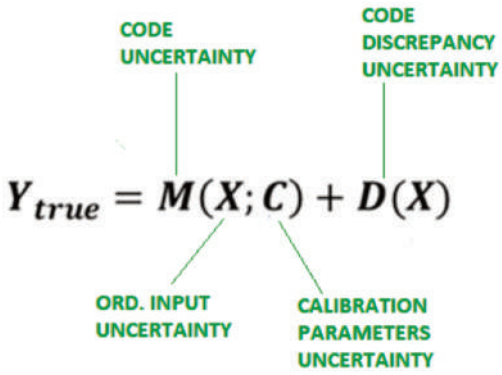
This metauncertainty may be regarded as arising from the imperfect knowledge about the input-output function associated to the model, and thus about the uncertainty propagation. Even if the formulation of the model (i.e. their equations or algorithms) is perfectly known, we have a limited information about the input-output function, commonly limited to a discrete, finite set of input decks and the corresponding outputs, and this does not enable an accurate knowledge of the function. This fact has inspired the term *code uncertainty* used by some authors [49] for this second-level uncertainty.

In a sense, the inclusion of second-level uncertainty completes the picture about the sources of uncertainty in calculations (Fig. 1.3)

Figure 1.3. Two visions of the sources of uncertainty implied in calculated quantities



(a)



(b)

1.6. Modelling and describing uncertainty

Aleatory uncertainty is modelled by means of probability theory (PT). In fact, PT was created to deal with random events.

Several mathematical structures can be used for representing epistemic uncertainty. Following [40] we can mention four theories:

I. Interval analysis
II. Fuzzy sets theory and possibility theory
III. Dempster-Shafer Evidence theory
IV. Probability theory

In this sequence, the complexity of representation grows from I to IV. Each theory in the sequence represents the uncertainty with a more complex structure than the previous. For instance, when the sole information about a scalar quantity is the range of values it takes, its uncertainty is represented by an interval (with no "inner structure"). For multidimensional quantities, the generalization of the interval is a region (e.g. a Cartesian product of intervals).

Theories II, III and IV represent an increasing *structure* of the uncertainty. It seems logical that an important point for assigning representation to the epistemic uncertainty is the amount of information available about the quantity.

There are experts who argue that poorer information about a quantity requires a simpler representation of uncertainty. This means, for instance, that the probability theory should not be applied when the information is scarce. In an antagonistic position are other experts, who believe that the PT can always represent the uncertainty, no matter how much information exists.

In this book we will focus on the modelling of uncertainty via interval analysis and probability theory, because they are the common choices in the Nuclear Safety realm.

The uncertainty is propagated through a model calculation under the form chosen to represent it. The probabilistic uncertainty propagates and gives rise to probability distributions of the results; the uncertainty modelled via intervals / regions propagates and produce also intervals / regions of the results.

It is appropriate to highlight the difference between two common models to represent a scalar uncertain variable: an interval and a uniform distribution. Both seem very similar, because the uniform distribution is defined in an interval. But they are completely different models. An interval conveys the sole information that the value of the variable in inside it, and no "inner structure" in the interval is specified. On the other hand, a uniform distribution also indicates that the value of the variable is inside the interval, but gives an additional information, namely that all the values in the interval are equiprobable. Therefore, the uniform distribution gives far more information than the interval.

Anyway, as we previously said, there are experts who, even when the sole information about a quantity is an interval, choose to assign it a uniform probability distribution, based e.g. on the principle of maximum entropy [119].

1.7. Probabilistic modelling of uncertainty

In the probabilistic framework, uncertain quantities are modelled as random variables. The uncertainty of a quantity is identified with its probability distribution (PD). More precisely, *certainty* (the antonym of uncertainty) is identified with probability.

For continuous random variables, probability distributions are expressed via two well-known numerical functions: the cumulative distribution function (CDF) and the probability density function (PDF).

Given a continuous scalar random variable V, the cumulative distribution function of V is defined as:

$$F_V(v) \equiv PR_V(V \leq v) \tag{1.3}$$

The probability density function (PDF) is the derivative (when it exists) of the CDF.

The concepts of CDF and PDF are extended to multidimensional continuous random variables.

A parameter of a PD is any numerical quantity that summarizes or describes an aspect of the distribution. There is a more restricted definition, as any numerical quantity (and different to the independent and the dependent variables) which is contained in the mathematical formula of the CDF and PDF of a PD. A parametric family of PD is the set of distributions having the same functional form but differing in the value of the parameters. There are numerous well-known parametric families, which can be found in textbooks of Probability Theory and Statistics. If we know the parametric family of a random variable, and the exact value of their parameters, then we know exactly the PD and thus the uncertainty of the variable.

There are nonparametric ways of defining a PD; for instance, graphical or numerical (table of values) expressions of the CDF and/or the PDF.

As previously mentioned, the probabilistic representation is the natural one for aleatory uncertainty. But it can also be applied to epistemic uncertainties.

1.8. Modelling uncertainty with intervals / regions

In this case, the uncertain variable is described by means of a numerical region in its range. For scalar variables, the uncertainty is typically represented by numerical intervals; for multidimensional variables, regions are used. In many cases such regions are Cartesian products of intervals.

This modelling of uncertainty is much less sophisticated than the probabilistic modelling, because all the information given is the region where values of the variable lie, and no "inner structure" of the information inside the region is provided.

The use of intervals/ regions is typical of epistemic uncertainties. For aleatory uncertainties, the probabilistic modelling is preferred. Nevertheless, when the information about the aleatory quantity

is scarce, uniform distributions are sometimes applied. And other possibility is using intervals or regions. In such case, some aleatory uncertainties are regarded as epistemic.

In Nuclear Safety calculations, the so-called *conservative approach* is based on a modelling of uncertainty via intervals, and often using one-sided intervals that overpredict uncertainty. Safety results have a "conservative direction", meaning that when their value progress in such direction, the severity of the calculated scenario grows. For instance, a scalar safety quantity Y having an upper regulatory acceptance limit L is such that growing values of Y imply growing severity of the scenario. The input parameters which are most influential in a safety result usually have a monotonic relation with the result, and this fact induces also a conservative direction in the input parameter. For instance, if X is an influential input on Y and they have a direct relation (i.e. increasing X implies increasing Y), it is clear that X has the same conservative direction than Y. In other words, if Y has an upper (resp. lower) safety limit, the conservative direction is that of growing (resp. decreasing) X. If Y and X have an inverse relation, they have opposite conservative directions. In summary, conservative methodologies model the uncertainty of the most influential inputs via pessimistic one-sided intervals i.e. assigning a pessimistic (or even very pessimistic) value to the parameter.

Notice that, in conservative methodologies, all the uncertainties are regarded as epistemic. Even if some of the input uncertainties are clearly aleatory and it is possible to assign them probability distributions, the analyst chooses to ignore such information and simply model them with an interval or region.

The uncertainty modelled with this type of pessimistic one-sided intervals is propagated through the calculations and give rise to pessimistic one-sided intervals for the safety results. One is tempted to think that the propagation is very simple and a single run of the model, with all the inputs in their pessimistic assignments, is enough to obtain the conservative value of the safety results, which are then compared to the regulatory limits. This approach often works well. But the input-output function of the model can be highly nonlinear and, in some cases, setting all inputs in their worst values does not produce the worst value of the output. In fact, the propagation of interval uncertainty can be regarded as an optimization process (or, rather, a "pessimization" process).

For conservative calculations, the goal of the propagation is finding values of outputs that, with a high certainty, are more severe than those produced in a hypothetical real scenario.

1.9. The uncertainty of the probabilistic uncertainty

It is not common a perfect knowledge of the uncertainty of a random variable. The knowledge of the variable is often limited to its range of definition and a set of discrete values (e.g. those contained

in a random sample). In other cases, there are reasons to suppose that the variable belongs to a specific parametric family, but there is not enough information to accurately estimate the value of the distribution parameters.

In this setting of partial, incomplete information, the PD of the variable cannot exactly be inferred. In other words, the PD has epistemic uncertainty, so that the uncertainty of the variable is itself uncertain.

Previously, we have referred to this "uncertainty of the uncertainty" as *metauncertainty* or second-level uncertainty. It is a usual companion to every uncertainty and decreases when the information on the variable grows. If, for instance, we know a very large random sample of the variable, second-level uncertainty becomes negligible.

The concept of "uncertainty of the uncertainty" is well known in the statistical realm, because Statistics is concerned with estimation of probability distributions from limited information. There are two well-known frameworks in Statistics: frequentist and Bayesian.

In frequentist Statistics, the parameters of the PD to estimate are regarded as fixed but unknown values i.e. they have epistemic uncertainty. From a random sample of a variable, statistical estimators can be constructed from sampled values for the parameters of the distribution, so that they are also random variables. The uncertainty of the estimators is epistemic and is described by the term "statistical confidence". This is the metauncertainty we are dealing with, and it decreases when sample sizes grow. A well-known form of expressing the uncertainty of a parameter is a *confidence interval*.

Let $\mathbf{V}$ be a continuous random variable, $\boldsymbol{\theta}$ a distribution parameter of $\mathbf{V}$, and $\mathbf{S}$ a random sample of $\mathbf{V}$. If C is a real number in the interval (0, 1), a confidence region of level C for the parameter $\boldsymbol{\theta}$ is a region RC($\mathbf{S}$), dependent of $\mathbf{S}$, contained in the range of $\boldsymbol{\theta}$ and fulfilling:

$$PR_{\mathbf{S}} \{\boldsymbol{\theta} \in RC(\mathbf{S})\} \geq C \tag{1.4}$$

C is termed the level of confidence.

The subindex $\mathbf{S}$ indicates the sampling distribution, which represents a "statistical uncertainty" termed "statistical confidence".

It is frequent the use of an alternative and more restricted definition, where the inequality of (1.4) is replaced by an equality:

$$PR_S\{h \in RC(S)\} = C \tag{1.5}$$

On the other hand, in Bayesian Statistics, parameters of distributions are considered random variables. The probability is regarded as a degree of belief. The uncertainty about the parameters before using the sampled values is described by a probability distribution termed the "prior distribution" or simply "the prior". The information contained in the sample is combined with the prior via the Bayes' theorem, to produce the "posterior distribution". Posterior distributions enable setting "credible intervals" (i.e. the Bayesian equivalent to confidence intervals) to describe the uncertainty of parameters.

Therefore, when the main information about the random variable is a random sample, obtained e.g. from a set of measurements or a Monte Carlo calculation, the second-level uncertainty is basically the "statistical uncertainty". Statistical estimates are uncertain quantities, having statistical uncertainty, which decreases when the sample size grows, and vanishes in the limit of infinite size. For this reason, the statistical uncertainty is associated to sampling and to the finite size of samples.

We identify the uncertainty of a random variable with its CDF. From limited information, we can estimate the CDF, which is a function and a curve. The simplest estimate of the CDF of a scalar random variable V, constructed from a SRS of V, S=(V_1…,V_N), is the so-called *empirical CDF* (ECDF), defined as:

$$F_{V,n}(v) \equiv \frac{1}{N}\sum_{i=1}^{N} 1(V_i \leq v) \tag{1.6}$$

(1.6) is a stepwise and right-continuous function (An alternative definition of the ECDF is obtained replacing in (1.6) the sign ≤ by <. It is a left-continuous function). There is in (1.6) a sum of Bernoulli variables (i.e. random variables which can take only values 0 and 1). For that reason, the product of N and $F_{V,n}(v)$ is a binomial random variable, the number of times V_i is not higher than v in N trials.

There are statistical techniques for setting confidence bands around the ECDF. These confidence bands are of two types: pointwise and simultaneous (or joint) bands (figure 1.5). Pointwise bands give, for each value v of the variable V, a confidence interval of level C for PR(V≤v) (i.e. the value of CDF on point v). On the other hand, joint or simultaneous bands are such that they enclose the complete CDF curve with a confidence not less than C.

Second-level uncertainty is epistemic, and there are other ways of modelling it (aside from the statistical one). The simplest one is interval analysis. Suppose that we know the parametric family of the variable distribution, and, additionally, we have an uncertainty interval for each distribution parameter. For every combination of values in the intervals we have a different CDF of the variable. The envelope of all these individual CDFs represents the uncertainty about the "true" CDF i.e. the second-level uncertainty. Such an envelope of CDF is termed a P-box (figure 1.4).

A way for considering the metauncertainty of a quantity is an "inflation of the uncertainty". In the probabilistic framework, this means an "extended" probability distribution for the uncertain quantity. An input parameter to a safety calculation has usually a conservative direction, such that a progress in this direction produce more severe results. In this case, the extension of the PD should be performed in the conservative direction. If the severity increases (resp. decreases) when the variable grows, and we have confidence bands or a P-box for the CDF, the lower (resp. upper) bound of the band or the P-box could be used to represent the uncertainty of the variable. The rationale of the procedure is clear: if C1 and C2 are two CDF and C1<C2 in all the range of the variable V, then C1 gives more weight than C2 to large values of V.

In the case of the confidence bands, either pointwise or joint bands could be used, with a high enough confidence level.

Figure 1.4. An interval between two cumulative distribution functions (P-Box). Taken from Wikipedia.

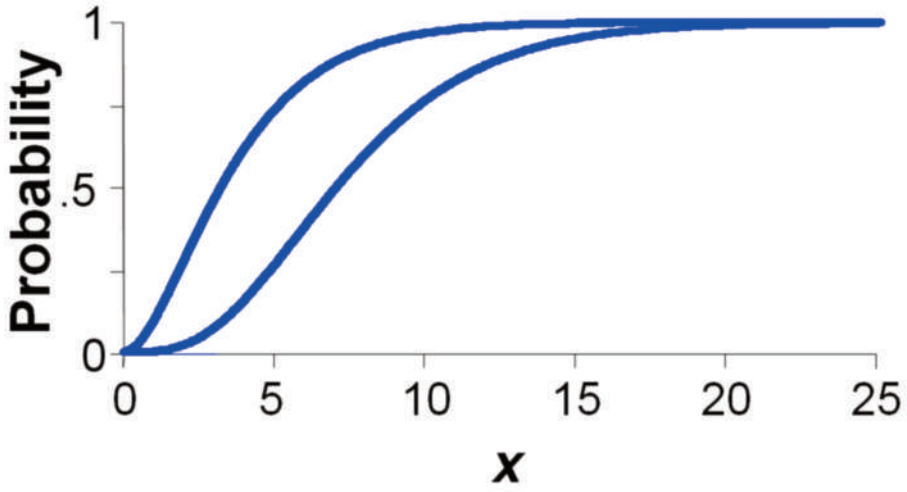
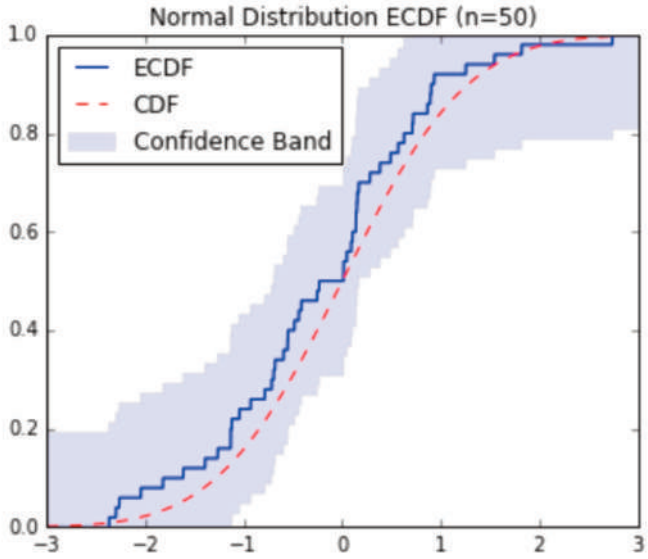


Figure 1.5. Empirical CDF (blue) with true CDF (red) and confidence band. Taken from https://bjlkeng.io/posts/the-empirical-distribution-function/



1.10. Descriptors of the uncertainty

In the probabilistic framework, the uncertainty of a quantity is described by a PD with epistemic uncertainty. But there are numerical quantities (parameters), typically constructed from random samples, that can give partial information about the probability distribution, namely.

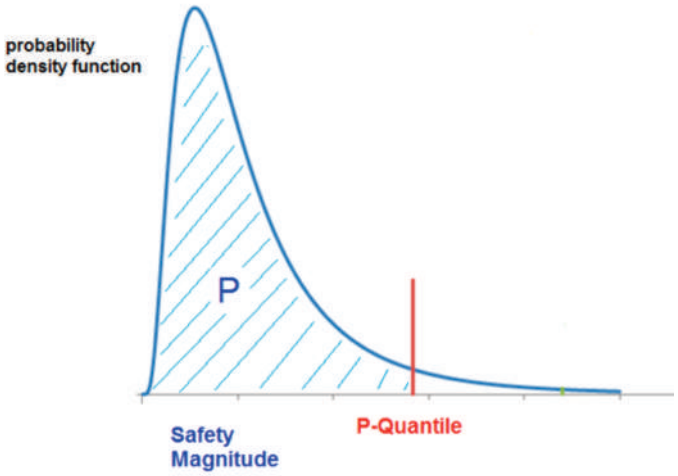
- Moments and central moments of the sample (e.g. mean, standard deviation, asymmetry coefficient, kurtosis coefficient) which are estimates of the corresponding population moments.
- Sample quantiles, which can be regarded as point estimates of quantiles of the random variable (also termed population quantiles). Given a number β in the interval (0, 1), the quantile of order β of a continuous random variable V is defined as (fig. 1.6)

$$V_\beta \equiv \inf\{v | F_V(v) \geq \beta\} \tag{1.7}$$

Likewise, there are regions in the range of the variable that provide partial information about the PD. They are termed *statistical regions.*

We call the previous elements *descriptors of the uncertainty.* In the next section we refer to statistical regions.

Figure 1.6. Quantile of a scalar and continuous random variable.



1.11. Statistical regions

A statistical region is a **statistical procedure** to construct, from a random sample of a random variable **V**, a region (i.e. subset) in the range of **V** or in the range of a distribution parameter. The region obtained from a specific random sample is also termed, by extension, statistical region. Therefore, in this document the term *statistical region* denotes, depending on the context, either the procedure of construction or a realisation of such procedure.

There are three basic types of statistical regions: confidence, prediction and tolerance regions [38, 52, 113]. Confidence regions have been already defined in section 1.9.

Let **V** a continuous random variable and **S** a random sample of **V**.

If p is a real number in the interval (0, 1), a prediction region of level p for **V** is a region **RP(S)** contained in the range of **V** and dependent of **S**, which satisfies the condition.

$$PR_{V,S}\{V \in RP(S)\} \geq p \tag{1.8}$$

In other words, a prediction region is a subset in the range of values of **V**, established from a random sample, that will contain a single new observation of **V** with a probability not less than a prespecified value.

If β and γ are real numbers in the interval (0, 1), a tolerance region of level (β, γ) for V is a region $RT(S)$ in the range of V, which depends on the sample S and fulfils:

$$PR_S\{PR_V\{V \in RT(S)\} \geq \beta\} \geq \gamma \tag{1.9}$$

β is termed the coverage level or probability level, and γ is termed the confidence level. The pair (β, γ) is termed the tolerance level of the region.

The subindex of probabilities in definitions from (1.8) and (1.9) indicates the probability distribution they are linked to. The subindex S indicates the sampling distribution, which represents a "statistical uncertainty" also termed "statistical confidence"

It is frequent the use of alternative, more restricted definitions to (1.8) and (1.9), where the inequality in confidence is replaced by an equality [52]:

$$PR_{V,S}\{V \in RP(S)\} = \beta \tag{1.10}$$

$$PR_S\{PR_V\{V \in RT(S)\} \geq \beta\} = \gamma \tag{1.11}$$

Many methods exist to construct statistical regions from simple random samples (SRS), i.e. when the values of the sample are i.i.d. (independent and identically distributed).

Confidence regions are descriptors of the epistemic uncertainty about probability distributions. When the sample size grows, such uncertainty decreases. In the limit of infinite size, the uncertainty vanishes, and the confidence region collapses to a point. It is desirable that such point coincides with the true value of the parameter. Confidence regions can be constructed from estimators of the parameter and their uncertainty. An estimator is termed *consistent* when it tends asymptotically to the true value of the parameter.

Prediction and tolerance regions are descriptors of the uncertainty of random variables. They model an uncertainty that has an aleatory part and an epistemic part. The former is expressed by the probability distribution of the variable; the latter is the metauncertainty, represented by the statistical confidence. In prediction regions, both uncertainties are combined, while in tolerance regions they are considered separately. The left-hand side of (1.9 and (1.11) is termed a *second order probability*.

When the sample size tends to infinity, both prediction and tolerance regions tend to fixed regions in the range of the random variable, having a probability content (i.e. coverage) equal to the prediction level and the coverage level, respectively. Such asymptotic regions depend on the procedure of construction of the statistical region and are descriptors of the aleatory uncertainty of the random variable.

The simplest statistical regions are intervals: confidence intervals (for scalar parameters) and prediction and tolerance intervals (for scalar random variables). All these intervals can be either one-sided or two-sided.

For two-sided intervals, both endpoints are functions of the sample S. For one-sided intervals, only one endpoint is function of S, the other one being the extreme value of the variable or parameter range. Endpoints which are functions of S are termed *statistical limits* (upper or lower, depending on the case). So, there are confidence, prediction and tolerance limits.

1.11.1. Tolerance regions and intervals

Tolerance regions [37, 38, 52, 113] are statistical descriptors of the uncertainty of random variables. As previously said, there are several definitions of tolerance region. In a wide sense, a tolerance region of level (β, γ) for a real continuous random variable $\mathbf{V}$ is a procedure that assigns to any random sample $\mathbf{S}=\{V_1,\ldots,V_N\}$ of $\mathbf{V}$ a region $\mathbf{RT(S)}$ in the range on $\mathbf{V}$ fulfilling (1.9). In a reduced sense, a tolerance region is any realization of the mentioned procedure. We adopt the formal definition (1.9), though some authors prefer the more restrictive version (1.11). Both definitions contain a second-order probability, involving two nested probabilities. The innermost probability is termed the *coverage* of the region, and it is a function of the sample $\mathbf{S}$:

$$\mu_V\big(\boldsymbol{RT}(\boldsymbol{S})\big) \equiv PR_V\{\boldsymbol{V} \in \boldsymbol{RT}(\boldsymbol{S})\} \tag{1.12}$$

In probability theory, the right-hand side of (1.12) is termed the measure (according to the variable $\mathbf{V}$) of the set $\mathbf{RT(S)}$. The coverage is a random variable with a probability distribution induced by the sampling. The definition (1.9) transforms to:

$$PR_S\{\mu_V\big(\boldsymbol{RT}(\boldsymbol{S})\big) \geq \beta\} \geq \gamma \tag{1.13}$$

from which a more compact form is derived

$$\mu_V\left(\boldsymbol{RT}(\boldsymbol{S})\right)_{1-\gamma} \geq \beta \tag{1.14}$$

based on the quantile of order 1-γ of the coverage.

The standard value of the tolerance level is γ=β=0.95. We speak of *(95, 95) intervals or 95/95 intervals.*

For a scalar random variable V, the simplest tolerance regions are *tolerance intervals*, which are either one-sided or two-sided.

One-sided tolerance intervals have only one endpoint which is a statistic (i.e. a numerical function of the sample), while the other endpoint is a fixed value, in fact an extreme value of the V range. When the upper (resp. lower) endpoint is a function of the sample, it is termed tolerance upper limit (resp. tolerance lower limit).

A statistic UT(S) is an upper tolerance limit (UTL) with level (β, γ) of V when

$$PR_S\{PR_V\{V \leq UT(S)\} \geq \beta\} \geq \gamma \tag{1.15}$$

equivalent to

$$PR_S\{V_\beta \leq UT(S)\} \geq \gamma \tag{1.16}$$

V_β is the β-quantile of V. Hence UT(S) is an upper confidence limit of the β-quantile of V with confidence level γ.

Similarly, LT(S) is a lower tolerance level (LTL) of level (β, γ) of V when

$$PR_S\{PR_V\{V \geq LT(S)\} \geq \beta\} \geq \gamma \tag{1.17}$$

equivalent to

$$PR_S\{V_{1-\beta} \geq LT(S)\} \geq \gamma \tag{1.18}$$

meaning that LT(S) is a lower confidence limit of the (1-β)-quantile of V with confidence level γ.

(L(S), U(S)) is a two-sided tolerance interval of level (β, γ) for V when

$$PR_S\{PR_V\{L(S) \leq V \leq U(S)\} \geq \beta\} \geq \gamma \tag{1.19}$$

It is interesting to notice that, if there is a procedure to obtain an upper tolerance limit of V from a sample S, then a lower tolerance limit is obtained using the same procedure. Indeed, if we call W= -V, the random sample of V, S=(V_1,…,V_N) transforms to a random sample of W, S*=(-V_1,…,-V_N). Then, an upper tolerance limit is obtained for W, UT(S*), and

$$W \leq UT(S^*) \iff V \geq -UT(S^*) \tag{1.20}$$

Therefore, -UT(S*) is a lower tolerance limit for V, with the same tolerance level than UT(S).

An important property of tolerance regions and intervals is their conservativeness (or conservatism) degree, that will be discussed in chapter 4.

1.11.2. Equal-tailed tolerance intervals

There is a special type of two-sided tolerance intervals, termed *equal-tailed tolerance intervals,* (ETTI) or *centred tolerance intervals* [52, 70]. Given the scalar random variable V and a simple random sample S of V, an equal-tailed tolerance interval of level (P, C) is an interval [L, U] (L and U being functions of the sample S) such that.

$$PR_S\left\{L(S) < V_{\frac{1-P}{2}} \quad AND \quad U(S) > V_{\frac{1+P}{2}}\right\} \geq C \tag{1.21}$$

i.e. the interval bounds simultaneously the quantiles of order (1+P)/2 and (1-P)/2 with a joint confidence level C. Notice that these quantiles have complementary orders, and that they encompass a probability P, thus showing that the equal-tailed interval is a special case of a (P, C)-level tolerance interval. The converse is not true; there are "standard" tolerance intervals (as defined in (1.19)) which are not centred.

The left-hand side of (1.21) is the probability of conjunction of two events, which is lower or equal than the probability of each event. Therefore, if (1.21) is fulfilled, the two conditions:

$$PR_S\left\{L(S) < V_{\frac{1-P}{2}}\right\} \geq C \tag{1.22}$$

$$PR_S\left\{U(S) > V_{\frac{1+P}{2}}\right\} \geq C \tag{1.23}$$

are satisfied as well. (1.22) indicates that L is a lower tolerance limit for V with level ((1+P)/2, C). Namely, L is a lower confidence limit (with level C) of the quantile of order (1-P)/2 of V. And (1.23) indicates that U is an upper tolerance limit for V with level ((1+P)/2, C). Namely, U is an upper confidence limit (with level C) of the quantile of order (1+P)/2 of V.

A more interesting result follows. A sufficient condition for the construction of an ETTI can be obtained by simply applying Bonferroni's inequality to (1.21):

$$\begin{aligned} PR_S\left\{L(S) < V_{\frac{1-P}{2}} \quad AND \quad U(S) > V_{\frac{1+P}{2}}\right\} \\ \geq PR_S\left\{L(S) < V_{\frac{1-P}{2}}\right\} + PR_S\left\{U(S) > V_{\frac{1+P}{2}}\right\} - 1 \end{aligned} \tag{1.21 bis}$$

Suppose that L(S) (resp. U(S)) is a lower (resp. upper) tolerance limit of level; then, the two probabilities on the right-hand side of (1.21bis) are higher or equal than (1+C)/2, and (1.21bis) implies the inequality (1.21).

Therefore, a sufficient condition for [L, U] being an ETTI with level (P, C) is that L (resp. U) is a lower (resp. upper) tolerance limit of level ((1+P)/2, (1+C)/2).

The procedures to calculate the two limits L(S) and U(S) are related; knowing one means knowing the other. As in section 1.11.1, let us call W=-V. Then, it is not difficult to prove that

$$W_{\frac{1+P}{2}} = -V_{\frac{1-P}{2}} \tag{1.24}$$

So that

$$-L(S) > W_{\frac{1+P}{2}} \Leftrightarrow \quad L(S) < V_{\frac{1-P}{2}} \tag{1.25}$$

And, finally,

$$L(S) = -U(S^*) \tag{1.26}$$

where S* is the random sample formed by the elements of S changed sign. The reciprocal relation (obtained by interchanging L and U) also holds.

To sum up, an equal-tailed tolerance interval with level (P, C) is also a standard two-sided tolerance interval with level (P, C), and the two endpoints are also one-sided tolerance limits (one upper and one lower) with level ((1+P)/2, C).

1.12. Propagation of uncertainty in BEPU calculations

Uncertainties are classified as aleatory or epistemic. But, rather than this classification, it is important how it is represented. In this sense, uncertain inputs can be classified as:

- *Inputs represented as aleatory,* modelled with probability distributions. Often input parameters with epistemic uncertainty are treated as aleatory, and endowed with a PD.
- *Inputs represented as epistemic* and modelled with intervals. Usually, they are one-sided intervals, but sometimes two-sided intervals are used.

These uncertainties propagate through the code calculation. Probabilistic input uncertainties propagate and give rise to probabilistic uncertainties of the calculation outputs. We say that input PD combine to produce output PD. Sometimes it is said that output PD are a "convolution" of input PD.

On the other hand, interval / region uncertainties also propagate through calculations and give rise to intervals / regions of the outputs. The combination of interval uncertainties is a process of optimization, where the upper and lower values of an output calculation are obtained.

When a calculation involves the two types of input uncertainties (PDs and intervals), how are the uncertainties propagated and combined? Suppose a simplest setting, with only two scalar inputs

to a calculation. One of them, X1, is modelled with a probability distribution. The other one, X2, is modelled with an interval. If we fix X2 to a value s of the interval, and propagate the uncertainty of X1 to the output Y, we obtain a CDF F(x1, s). If s varies through the whole interval, we obtain a set of CDF curves parametrized by s. This bundle produces a p-box, as defined in section 1.9, where we discussed how the uncertainty of a random variable is modelled when the distribution parameters are uncertain and represented by intervals. This result can be generalized, stating that the propagation of input uncertainties described by probability distributions and by intervals give rise to scalar outputs described by p-boxes.

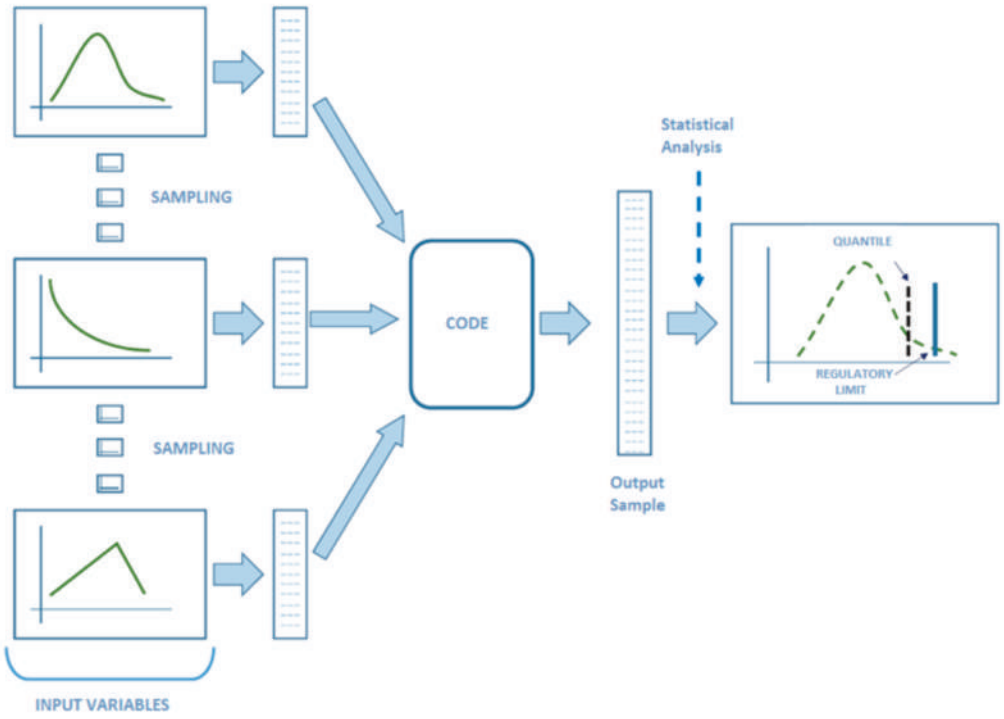
How is the propagation of uncertainties through a calculation performed in practice? For probabilistic uncertainties, the standard procedure is the Monte Carlo technique [48], with the following steps:

1. Uncertain inputs are identified. The inputs that propagate most of the uncertainty of the safety results are selected. Sensitivity analyses are very useful in this task [84]. It should be proved that the rejection of less important inputs does not imply a significant underprediction of the propagated uncertainty.
2. Uncertainty is assigned to the selected inputs, in the form of probability distributions. Dependencies among inputs must be modelled, via e.g. conditional probability distributions.
3. Random samples are obtained from the probability distributions of the inputs, using random (or pseudorandom) number generators. The sampling must take into account dependencies among inputs previously described. The result is a random sample of inputs decks to the model.
4. The model is run for each random input deck. The result of these runs is a random sample of calculation results.
5. The uncertainty of the calculation results is estimated, by means of statistical methods.

Consequently, Monte Carlo techniques imply the use of random sampling procedures. The simplest and most used procedure is the simple random sampling (SRS), when all the sampled values are mutually independent (It is often said that the sampled variables are independent and identically distributed (i.i.d.)). The use of pseudorandom numbers breaks this assumption of independency, but such numbers are extensively used in the generation of samples, because they emulate adequately the SRS. Aside from that, a very good property of the pseudorandom number generators is that a complete sequence of numbers can be reproduced, simply by repeating the numerical seed. This property is useful on studying the impact of the change of some parameter on the safety results.

Notice that previous step 4 requires most of the computational effort, especially for long-running code calculations.

Figure 1.7. Monte Carlo analysis of a safety calculation. Adapted from [69].



As we have previously stated, the information about the input-ouput function ($\mathbf{M}$ in (1.1)) associated to a code is incomplete. When Monte Carlo techniques are used for propagation, the information provided about $\mathbf{M}$ is a random sample of input decks and the corresponding random sample of output values. In the black-box approach this is practically our sole information on $\mathbf{M}$. From the finite random samples, statistical techniques estimate characteristics of the probability distributions of the outputs. The finiteness implies:

1. an imperfect knowledge of the propagation of uncertainty through the calculations, and, at the same time
2. an incomplete information about the probability distribution of the outputs, which are estimated with statistical techniques.

The second-level uncertainty produced by the finite size of Monte Carlo samples is termed "code uncertainty" by some authors [49] based on previous point 1; and receives the classical name "statistical confidence" or (simply "confidence"), based on point 2.

The code uncertainty is null or negligible only in very especial cases:

- When the model is very simple, so that the associated input-output function has an analytical, closed form. Then, the propagation of uncertainty may be analytically calculated.
- When the model is cheap to run, both in computer time and resources. Then, the model can be run a lot of times and thus be exhaustively sampled. In this case, the input-output function may be known very accurately.

Summarizing, when we study a code output via a Monte Carlo analysis, the uncertainty of the output has two components:

- Computational uncertainty, which is the uncertainty propagated from inputs and models.
- Code uncertainty, due to the ignorance about the propagation process and the finite size of the MC sample.

The propagation of uncertainties defined by intervals produces output intervals. In practice, Monte Carlo techniques could be used to perform the propagation, assigning to each input a uniform distribution in its interval. Clearly, this procedure is less efficient than optimization techniques. But, if the computational cost is affordable, the Monte Carlo size can be very large, and the results very precise.

Finally, we tackle the procedures of uncertainty propagation when some uncertain inputs are modelled as random variables and others are modelled with intervals. Rather than describing specific methods, we will refer to the usual procedure in nuclear DSA, with two steps:

- The probabilistic inputs are fixed on their nominal or mean values, and then an optimization procedure is applied for obtaining the values of interval-modelled inputs that produce the most severe safety outputs.
- The interval-modelled inputs are fixed in such optimized values, and then the propagation of uncertainty of the probabilistic inputs is performed (e.g. via Monte Carlo).

This is the procedure typically applied in LOCA/ECCS analyses performed with BEPU methodologies, where several important input parameters are modelled by intervals: the break position in the primary system, the break size, the discharge coefficient in the break…These parameters are "optimized" under the criterion of maximizing the peak cladding temperature (PCT). Indeed, the search of the "worst" break size, location and discharge coefficient is a regulatory requirement [16]. And in such optimization process, other inputs defined as discrete variables are used as well. A good

example is the worst single failure criterion, that requires postulating in the scenario the single failure of safety systems that maximizes the peak cladding temperature. There are methodologies termed "Extended BEPU" (E-BEPU) that treat these uncertainties about the safety systems as probabilistic, rather than looking for the worst configuration [23, 24].

Another category of inputs represented by one-sided intervals (i.e. conservative values) are the operational parameters limited by the Plant Technical Specifications. The safety calculations must prove that the allowed operation of the NPP (rather than the "real" operation is safe). Therefore, operational parameters should be fixed in their TS limits or, alternatively, should be modelled by intervals containig the TS limit or probability distributions giving significant probability to the range around the TS limit [122].

II. Introduction to BEPU methodologies in deterministic safety analysis

II. INTRODUCTION TO BEPU METHODOLOGIES IN DETERMINISTIC SAFETY ANALYSIS

2.1. Safety calculations: conservative and realistic calculations

Safety calculations are referred to potentially dangerous facilities, industries or activities. The goal of such calculations may be the design of safety equipment or systems, or the estimation of the risk associated to these types of systems. In safety calculations, the consideration of uncertainties is essential.

Nuclear Power Plants are designed by means of a framework termed Deterministic Safety Analysis (DSA). The basic goal of the method is proving that the design of the NPP, including their safety systems, can face, without a significant damage, a representative set of accidents and transients of the plant.

DSA, also named "accident analysis" has the following steps:

- A set of categories of accidental scenarios is established. Typically, each category groups scenarios with similar frequencies (here *frequency* is defined as probability per unit time).
- For each category, a few representative scenarios are selected, such that their consequences are more severe than all accidents in the category. They are commonly termed *design basis accidents and transients*. In the sequel we term them as *design basis scenarios* (DBS).
- Each DBS is simulated with adequate computational models (codes), calculating some safety quantities (representative of the scenario severity).
- It is checked if these calculated safety quantities fulfil regulatory acceptance criteria. If they do, the design is considered as safe. If any calculated safety quantity does not fulfil its criterion, the design must be improved till full satisfaction of the criteria is obtained.

It must be highlighted that the regulator establishes the main elements of DSA, namely the accident categories, the DBS, the calculated safety quantities for each DBS and their acceptance criteria. Likewise, the regulator commonly gives guidance about the capabilities of the codes to be used in the DBS simulation.

On the other hand, it is the licensee who chooses or develops the codes, performs the simulations and checks the fulfilment of the regulatory acceptance criteria.

A DSA methodology (sometimes termed *Evaluation Model*) [98] is a complete procedure to conduct the analysis of a specified type (or types) of design basis scenarios. A methodology includes:

- The computational predictive models (codes) used in the simulation of the DBS.
- The assumptions used in the calculations.
- The procedures to perform uncertainty and sensitivity analysis of the safety outputs.
- Ancillary tools as user guidelines, additional software, etc.

Ideally, every methodology should give the same result (independently of the user) in the analysis of the specified DBS. In real world, different users may obtain different outcomes. But methodologies should be robust, aimed at minimizing the "user effect".

Globally speaking, methodologies can be either conservative or realistic (also termed BEPU, *Best Estimate Plus Uncertainty*). Conservative methodologies are based on a pessimistic choice of models, assumptions and input values. The term pessimistic is interpreted as "producing more severe consequences (from the safety point of view) than expected in a real scenario". The value of an input variable (resp. a model) is more pessimistic than other value (resp. other model) when the former produce more severe results than the latter.

It is important to remark than "conservative analysis" is not synonym of "no uncertainty analysis". The conservatively calculated value of a safety quantity includes its uncertainty, and the conservative calculation implements an uncertainty analysis performed with intervals. Such intervals are usually selected in a pessimistic or even very pessimistic fashion. The propagation of the input uncertainties implies an optimization process, looking for the "worst" (in terms of consequences) safety results compatible with the input intervals. In most cases, the worst results correspond to a single calculation with all the influential inputs in their worst values. But, as previously said, this fact is not true in some cases, due to strong nonlinear effects in the code. The adoption of ultraconservative values for the outputs attempts to counteract these possible nonlinear effects. The ultimate goal is to ensure, with a very high certainty, the conservativeness of the calculated safety results.

Two elements are needed to perform a conservative calculation:

- The important input parameters and submodels in the calculation i.e. those with a significant impact on the safety results. Sensitivity analysis techniques can be used in this task.
- The uncertainty range and the conservative subrange of input and submodel parameters.

With such knowledge, it is possible to assign conservative enough values to the important input parameters.

Historically, in the realm of nuclear accident analyses, conservative methodologies were needed as long as the phenomenology of design basis scenarios was not known in depth, so that predictive models were simple and sometimes very empirical.

As the experimental and theoretical knowledge of the phenomena and the computational capability increased, the door began to open to the use of realistic, mechanistic models (termed *best estimate models*) in accident analysis. Methodologies based on calculations with this type of codes and supplemented with detailed and realistic uncertainty analyses of the safety results, are termed BEPU (*Best Estimate Plus Uncertainty*). They constitute an essential advance compared to the conservative approach. Most existing BEPU methodologies are probabilistic, and they are also known as *statistical methodologies*.

BEPU methodologies require the assignment of uncertainty to the important input parameters, in the form of probability distributions. Then, the uncertainty propagation to the outputs is commonly performed with Monte Carlo techniques. It is a far more complex and expensive process than in the case of conservative calculations, because many code runs are needed.

It is good to remark that BEPU methodologies typically have some (slight) degree of conservativeness, either in models or assumptions or values of input parameters. A good example is the spatial discretization (nodalization), that introduces errors in the results which are not perfectly known, and thus produces some uncertainty. It should be very difficult to model this uncertainty with a probability distribution. Rather, the common procedure is the adoption of a nodalization which introduces a moderate conservative bias in the results. Sometimes, BEPU is interpreted as *Better Estimate Plus Uncertainty* to account for these conservative choices.

2.2. Regulatory acceptance criteria

2.2.1. Regulatory acceptance criteria for conservative methodologies

It is very common, in Science and Engineering, that acceptance criteria are imposed on results of calculations. That is, the goal of the calculation is to prove that some physical quantities meet predetermined criteria. For instance, design calculations are aimed at proving that design criteria are satisfied. Basically, acceptance criteria are single or multiple mathematical inequalities, involving calculated quantities and acceptance limits.

For nuclear DSA calculations, acceptance criteria require that some safety outputs (representing the severity of the scenario) are below adequate limits. The goal is to maintain a proper margin to damage thresholds of the system.

Acceptance criteria are different depending on the conservative or realistic character of the calculation. In the former case, conservative values are calculated, and they include their uncertainty. In the realistic case, values are calculated with uncertainties, typically using probabilistic methods. In

such case, when we skip from conservative to realistic calculation, the acceptance criteria change from simple inequalities to probabilistic statements.

Regulatory Acceptance Criteria (RAC) are conditions that the safety results of DSA must fulfill. The nuclear regulatory authority defines the categories of design basis scenarios (DBS), and the safety quantities that must be calculated for each DBS and the RAC they must satisfy. The fulfillment of the RAC implies that the DBS do not produce unacceptable damage.

It seems logical that the RAC must take into account the calculational uncertainty of the safety outputs, and how this uncertainty is modelled.

Let us consider the simplest case, when we have a single safety quantity, Y, which is scalar and continuous. If Y increases when the severity of the scenario grows, the regulator will set an upper acceptance limit L on the calculated Y. When a conservative methodology is used, the RAC takes the form:

$$Y_C < L \tag{2.1}$$

Subindex C means that Y is conservatively calculated. It is assumed that Y_C is higher than Y_{real} with very high certainty, where Y_{real} would be the value of Y produced in a "real" scenario.

We define the *acceptance region* for Y as the region in the range of Y fulfilling the restriction of the RAC; in this case, the region is the one-sided interval Y<L, and we speak of *acceptance interval*. If, on the other hand, higher values of Y imply lower severity of the scenario, the regulator sets a lower acceptance limit on Y, and (2.1) takes the form:

$$Y_C > L \tag{2.2}$$

and the acceptance region is the one-sided interval Y>L. Any criterion of the form (2.2) can be recast in the form (2.1), with a simple change of variable. For instance, we can define Y* = H-Y and L*=H-L, where H is a value such that Y* and L* are positive values (i.e. H is higher than any value in the range of Y including the limit L). Then (2.2) transforms to $Y_C^* < L^*$. Therefore, for a scalar quantity Y we may consider only the case (2.1), with an upper acceptance limit, without loss of generality.

In some cases, rather than a single scalar quantity, there are multiple scalar safety quantities (defined by the regulator) for a specific DBS. In this case, all the safety quantities can be grouped in a multidimensional continuous quantity $\mathbf{Y} = (Y_1,\ldots,Y_D)$, where Y_j, j=1,…,D are real and continuous

scalar quantities. The regulator establishes the RAC setting restrictions on the components of **Y**, thus defining an "acceptance region" in the range of **Y**. The RAC for conservative methodologies is a generalization of (2.1):

$$Y_C \in R_A \tag{2.3}$$

where C again indicates "conservatively calculated" and $\mathbf{R_A}$ is the multidimensional acceptance region.

Typically, criterion (2.3) is a conjunction of individual criteria, one for each component of **Y**:

$$Y_{i,C} \in R_{A,i} \qquad i = 1,\ldots,\mathrm{D} \tag{2.4}$$

$R_{A,i}$ is the acceptance interval of the scalar safety output Y_i. In most cases. the acceptance intervals are one-sided, and the multidimensional acceptance region is a Cartesian product of one-sided intervals.

The joint criterion is fulfilled if and only if the **D** individual criteria are met for the analysed scenario.

Another way of writing (2.3) and (2.4) is

$$\bigwedge_{i=1}^{D} (Y_{i,C} \in R_{A,i}) \tag{2.5}$$

where we have introduced the symbol Λ, meaning logical conjunction i.e. simultaneous fulfilment.

If the joint criterion is true, then each one of the individual criteria is true as well. And only if all the individual criteria are satisfied, the joint criterion is fulfilled as well.

The scalar acceptance regions are typically regions below an acceptance limit. As previously discussed, we can assume without loss of generality that all the safety limits are upper, so that (2.4) is written

$$Y_{i,C} < L_i \qquad i = 1,\ldots,\mathrm{D} \tag{2.6}$$

Or, equivalently

$$\bigwedge_{i=1}^{D}(Y_{i,C} < L_i) \tag{2.7}$$

where the quantities $Y_{i,C}$ and L_i are positive. In this case, the acceptance region of the multidimensional criterion is

$$\boldsymbol{R_A} = \left\{(Y_1, \ldots, Y_D) \middle| \bigwedge_{i=1}^{D}(Y_i < L_i)\right\} \tag{2.8}$$

2.2.2. Reducing a multidimensional acceptance criterion to a scalar one

A multidimensional acceptance criterion as (2.6) or (2.7) can be transformed to a scalar acceptance criterion [64, 67]. The procedure is a simple change of variable. In the remainder of the section, we will drop the subindex C, to simplify the notation.

Other form of the multidimensional criterion (2.6) - (2.7) is

$$\begin{array}{l} Y_1 < L_1 \\ \ldots\ldots\ldots\ldots \\ Y_D < L_D \end{array} \tag{2.9}$$

which is a conjunction of D individual criteria. Without any loss of generality, we can assume that the Y_j and L_j are positive quantities.

We define the scalar quantity

$$W \equiv max\left(\frac{Y_1}{L_1}, \ldots, \frac{Y_D}{L_D}\right) \tag{2.10}$$

It is easy to see that (2.9) is equivalent to W < 1. The variable change (2.10) is a very simple way to transform the multidimensional RAC into a scalar one. Furthermore, though the Y_j may have different physical dimensions, W is nondimensional.

It is clear that, if W>1, there are one or more quantities Y_j that have surpassed their regulatory limits. Conversely, if any Y_j surpass its limit L_j, then W>1. Hence W can be regarded as a measure of two things: the severity of the analyzed scenario and the conservativeness of the **Y** calculated for that scenario. For a fixed scenario (i.e. a specific DBS) W is a measure of conservativeness of **Y**.

The fact that a multidimensional safety quantity can be transformed to a scalar one is important, because in many cases we will focus on the scalar problem without loss of generality.

(2.10) is not the unique transformation from the multidimensional variable to a scalar one. Another valid transformation could be:

$$W \equiv max\left(\frac{Y_1}{w_1 L_1}, \ldots, \frac{Y_D}{w_D L_D}\right) \tag{2.11}$$

where the w_j are numbers in the interval (0, 1). Again, W<1 is equivalent to (2.9). w_j is a multiplier less than one, reducing somehow the regulatory limit L_j. Therefore, when (2.11) is used, W<1 is a stricter criterion than when (2.10) is used. (2.11) reduces to (2.10) when all the w_j are equal to 1. In both cases, all the individual RAC must be fulfilled, but different importance (given by the value of $1/w_j$) may be assigned to each individual criterion.

The value of W depends on the factors w_j used. We conclude that there are simple variable changes (2.11) to transform a multidimensional safety quantity **Y** in a scalar quantity W that is regarded as a measure of the conservativeness of **Y**. The multidimensional RAC transforms to W<1.

To sum up, the multidimensional case with RAC (2.9) can always be reduced to a scalar problem with a simple RAC. In many aspects, we can focus on the scalar case without loss of generality.

2.2.3. Regulatory acceptance criteria with uncertainty

If, instead of a conservative methodology, we use a BEPU methodology, the safety output **Y** is calculated as an uncertain quantity, and we should modify (2.3) as follows:

$$Y \in R_A \quad \textit{with high enough certainty} \tag{2.12}$$

The subscript C has been removed, because Y is no longer calculated as a conservative quantity, but rather as a realistic uncertain quantity.

Most BEPU methodologies are probabilistic and consider **Y** as a random variable. For them, the measure of certainty is probability, and criterion (2.12) is formulated as

$$PR_{\mathbf{Y}}\{\mathbf{Y} \in \mathbf{R_A}\} \geq P_0 \tag{2.13}$$

Criterion (2.13) is probabilistic and states that **Y** must be inside the acceptance region with a high enough probability, because P_0 is a high value of probability (i.e. close to one) established by the regulator. We term it the regulatory *probability level.* The standard value is 0.95.

Subindex **Y** indicates that the probability on the left-hand side of (2.13) is calculated with the probability distribution of **Y**. We term the condition inside the probability i.e. $\mathbf{Y} \in \mathbf{R_A}$ the *restriction* of the criterion (2.13).

The simplest regulatory acceptance criterion occurs when **Y** is a scalar quantity Y. Let us suppose that the RAC consists in setting an upper regulatory limit L on Y. The expression (2.13) of the RAC with uncertainty reduces to

$$PR_Y\{Y < L\} \geq P_0 \tag{2.14}$$

2.2.4. The acceptance criteria in terms of limit exceedance probability

The criterion (2.13) can be rewritten as

$$P_{exc}(\mathbf{Y}) < 1 - P_0 \tag{2.15}$$

where

$$P_{exc}(\mathbf{Y}) \equiv PR_{\mathbf{Y}}\{\mathbf{Y} \notin \mathbf{R_A}\} \tag{2.16}$$

is the probability of **Y** being outside the acceptance region i.e exceeding some regulatory limit. We term it *limit exceedance probability*.

In the scalar case, the definition is

$$P_{exc}(Y) \equiv PR_Y\{Y \geq L\} \tag{2.17}$$

i.e. the probability of exceeding limit L. The criterion is

$$P_{exc}(Y) < 1 - P_0 \tag{2.18}$$

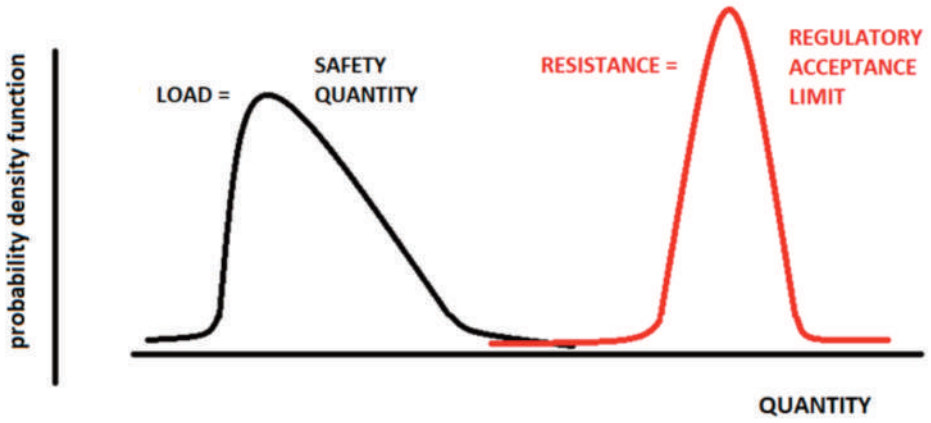
One minus $\mathbf{P_{exc}}$ is the *acceptance probability* of the safety quantity, which is the left-hand side of (2.13) and (2.14).

This acceptance probability can be interpreted from several standpoints, and consequently have different synonyms. We next mention 5 interpretations of this probability.

1. **A degree of fulfilment** of the restriction inside the RAC. (2.13) could be read as "the degree of fulfilment of $Y \in R_A$ is higher or equal than P_0".
2. **A measure of the acceptance region**, because, mathematically, probability is a measure.
3. **A safety margin**. Figure 2.3 shows that the acceptance probability is the area enclosed by the PDF of a scalar Y and the limit L; it is evident that can be interpreted as a "distance" from the PDF to L. In chapter 4 we will discuss in depth the concept of *safety margin*, interpreted as a metrics in the range of safety quantities. The margin between the calculated safety quantity and the regulatory acceptance limit is sometimes termed the *licensing margin*. Hence criterion (2.14) requires a high enough value of the licensing margin. This interpretation can be extended to the multidimensional case; the probability $PR_{\mathbf{Y}}\{\mathbf{Y} \in R_A\}$ can be regarded as a margin from $\mathbf{Y}$ to the boundary of the multidimensional acceptance region.
4. **One minus a failure probability**. The calculation of Y can be regarded as a Reliability problem. Reliability is a branch of science and engineering dealing with failures of physical objects and systems when they are submitted to mechanical loads. Typically, an object fails when the load applied on it is higher than its strength or resistance. In this sense, we can regard the quantity Y as a "load" on the nuclear plant caused by an accident, and the limit L as the "resistance" (fig. 2.1). Then, the exceedance probability can be viewed as the probability of the load overcoming the resistance i.e. the probability of failure. This interpretation of accident analysis in terms of Reliability can be very enlightening and, as we will see, opens the door to the application of techniques from the Reliability field to the BEPU field.
5. T**he parameter of a Bernoulli experiment**. When a Monte Carlo technique is applied to propagate the uncertainty through a code, a random sample of input decks is obtained and fed to the model, producing a random sample of outputs. Each code run can be regarded as a computational Bernoulli experiment, where a random value of the safety output is obtained and regarded as *success*, when $\mathbf{Y} \in \mathbf{R_A}$ or *failure*, when $\mathbf{Y} \notin \mathbf{R_A}$. The complete Monte Carlo is viewed as a *binomial experiment,* where N trials are performed and an outcome of S successes and F=N-S failures is obtained. In this sense, the acceptance probability can be considered a success probability in a computational experiment.

Interpretations 4 and 5 are jointly addressed in section 3.3.1.2, in the description of methods to verify the fulfilment of a RAC.

Figure 2.1. Analogy of nuclear safety calculation with Reliability concepts. Adapted from [64].



Notice that the acceptance criterion with uncertainty, as stated in this section, contains two regulatory limits, namely:

- The limit on **Y**, defined by the boundaires of the acceptance region R_A (and L in the scalar case).
- The limit on the exceedance probability ($1-P_0$), whose standard value is 0.05

2.2.5. The acceptance criteria in terms of quantiles

There are more avatars of the RAC with uncertainty. Expression (2.14) can be written in another form, using the concept of quantile of a random variable. The quantile of order λ of a continuous random variable Y, where λ is a real number in the interval (0, 1), is the value Y_λ such that the probability of Y being below Y_λ is λ.

$$Y_\lambda \equiv inf\{y|F_Y(y) \geq \lambda\} \tag{2.19}$$

If the CDF of Y is continuous and strictly increasing, its inverse function exists, and (2.19) simplifies to

$$Y_\lambda \equiv F_Y^{-1}(\lambda) \tag{2.20}$$

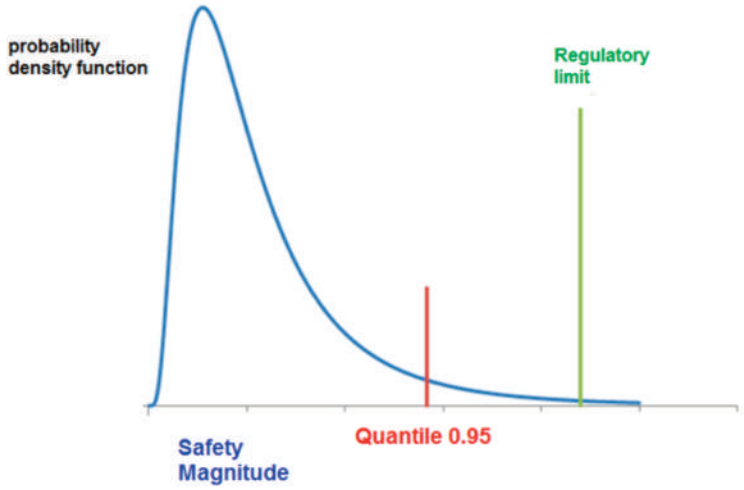
Obviously, the higher the order, the higher the quantile.

In terms of quantiles, (2.14) is rewritten as

$$Y_{P_0} < L \tag{2.21}$$

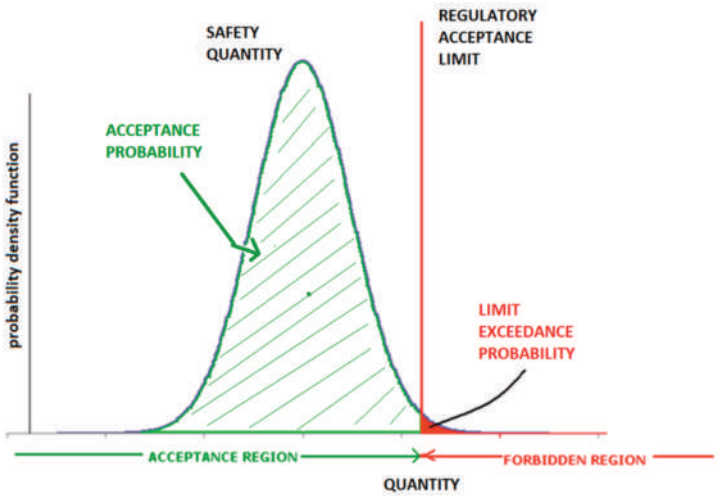
simply stating that the quantile of order P_0 of Y must be less than the regulatory limit L. A comparison of this expression with (2.1), corresponding to a conservative calculation, is interesting. Shifting from a conservative to a BEPU methodology implies replacing a conservatively calculated Y (with a conservatism that is not quantified, but is assumed "sufficient") by a high-order quantile of Y, the order being established by the regulator (Fig. 2.2). In other words, in the transition from conservative to BEPU analysis, a quantile of very high (but unknown) order is replaced by a quantile of high and prespecified order.

Figure 2.2. The 0.95-quantile of a safety quantity when the acceptance criterion is fulfilled.



Consequently, in the scalar case, the acceptance criterion can be written in several forms. Two parameters can be regarded as figures of merit of the analysis: the quantile or order P_0 of Y and the exceedance probability of Y (fig. 2.3). Contrary to the quantile-based form of the criterion, the form based on exceedance probability is directly generalizable to the multidimensional case, as we have previously seen.

Figure 2.3. Limit exceedance probability



2.2.6. Adding the second-level uncertainty: BEPU regulatory acceptance criteria

The RAC becomes probabilistic when the calculational uncertainty, probabilistically modelled, is introduced. Then the RAC takes the forms shown in section 2.2.4. However, we know that there is still more uncertainty to be introduced into the RAC. There is the second-level uncertainty (also termed metauncertainty or code uncertainty), produced by the imperfect knowledge of the propagation through the model and the finiteness of the Monte Carlo sample. It in epistemic and becomes negligible when the sample is very large.

The imperfect knowledge of the probability distribution of Y implies an imperfect knowledge of the exceedance probability of Y and (in the scalar case) of the P_0-quantile of Y. Hence these parameters have second-level uncertainty, that can be modelled in different ways, for instance via intervals or probability distributions. When we have a Monte Carlo sample of Y, statistical methods can be applied to the estimation of the probability distribution, and the estimates have a statistical uncertainty, termed "statistical confidence" (the term we encompass both the frequentist and the Bayesian viewpoint).

Therefore, the second-level uncertainty, in the form of statistical confidence, should be added to the criteria of section 2.2.4. There are two basic ways to do it, that we next describe [64].

Firstly, the two uncertainties can be juxtaposed, meaning that a second layer of uncertainty is applied e.g. to (2.13), producing

$$PR_{\mathbf{S}}\{PR_{\mathbf{Y}}\{\mathbf{Y} \in \boldsymbol{R}_{\boldsymbol{A}}\} \geq P_0\} \geq C_0 \tag{2.22}$$

This form contains the two uncertainties but does not combine them. The left-hand side of (2.22) is a second-order probability (i.e. two nested probabilities). The inner part refers to the probability distribution of Y. The outer part refers to the second-level uncertainty i.e. the ignorance about the mentioned probability distribution. The subindex $\mathbf{S}$ can also be regarded as meaning "sampling", because usually the uncertainty propagation is performed by means of a *pure* (also termed *crude*) Monte Carlo, where a simple random sample (SRS) of the safety output is obtained. In such case, the exceedance probabilities and quantiles have sampling distributions. C_0 is termed *confidence level,* and has a value close to 1 defined by the regulator.

Other possibility, alternative to (2.22), is the combination of the two uncertainties, producing

$$PR_{\boldsymbol{Y},\boldsymbol{S}}\{\boldsymbol{Y} \in \boldsymbol{R_A}\} \geq P_0^* \tag{2.23}$$

where the regulator would establish a combined probability level P_0^*, close to 1 and not necessarily equal to P_0 in (2.22). The probability on the left-hand side of (2.23) is the expected value of the acceptance probability of **Y**, and therefore (2.23) is rewritten as

$$E_{\boldsymbol{S}}[PR_{\boldsymbol{Y}}\{Y \in \boldsymbol{R_A}\}] \geq P_0^* \tag{2.24}$$

The separation of uncertainties in form (2.22) is present in the concept of tolerance region, while the combination of uncertainties in form (2.23) is present in the concept of prediction region. Both concepts are exposed (as descriptors of uncertainty) in section 1.11.

In terms of the acceptance probability, (2.22) is written

$$PR_{\boldsymbol{S}}\{P_{ACC}(\mathbf{Y}) \geq P_0\} \geq C_0 \tag{2.25}$$

And, more compactly

$$P_{ACC}(\boldsymbol{Y})_{1-C_0} \geq P_0 \tag{2.26}$$

which is a condition on the quantile of order 1-C_0 of the acceptance probability. And the combined form of (2.24) is expressed as

$$E_S[P_{ACC}(\mathrm{Y})] \geq P_0^* \tag{2.27}$$

Comparing (2.26) and (2.27) it is clear that, if $P_0^*=P_0$, criterion (2.26) is stricter than (2.27), because low-order quantiles are lower than the expected value, except for very unusual distributions.

When sample sizes increase, the statistical uncertainty reduces, and the left-hand sides of both criteria are closer to each other. In the limit of infinite samples, both criteria reduce to a common form.

In conclusion, and from the regulatory standpoint, it makes sense to choose the separated form (2.22) instead of the combined one (2.23), mainly because the former is, in general, more conservative than the latter. Therefore, we consider that the **BEPU regulatory acceptance criterion** is, in the scalar case

$$PR_S\{PR_Y\{Y < L\} \geq P_0\} \geq C_0 \tag{2.28}$$

And, in the multidimensional case

$$PR_S\{PR_Y\{Y \in R_A\} \geq P_0\} \geq C_0 \tag{2.29}$$

As previously said, C_0 is a high value of confidence (close to 1), established by the regulator (with standard value equal to 0.95) and termed the *regulatory confidence level.* The pair (P_0, C_0) is termed the *regulatory tolerance level,* with a standard value equal to (0.95, 0.95). With such level, the acceptance criteria (2.27) and (2.28) are qualified as *95/95 criteria or (95, 95) criteria.*

(2.28) and (2.29) are rather complex expressions, with two nested probabilities. Here reading is simpler than writing; e.g. (2.28) reads "Y is less than L with a probability not less than P_0 and a statistical confidence not less than C_0".

The scalar criterion (2.28) can be simplified using the P_0-quantile of Y

$$PR_Y\{Y_{P_0} < L\} \geq C_0 \tag{2.30}$$

An statistician would translate (2.30) saying that "an upper confidence limit of level C_0 of the P_0 –quantile of Y must be lower than L". We saw in section 1.11.1 that a confidence limit of a quantile is termed a *tolerance limit.* Then, (2.30) requires that an upper tolerance limit with level (P_0, C_0) of Y must be lower than L symbolized as:

$$Y_{(P_0,C_0)} < L \tag{2.31}$$

In terms of the limit exceedance probability, (2.28) and (2.29) are rewritten:

$$PR_S\{P_{exc}(Y) < 1 - P_0\} \geq C_0 \tag{2.32}$$

and

$$PR_S\{P_{exc}(Y) < 1 - P_0\} \geq C_0 \tag{2.33}$$

stating that the exceedance probability must be less than 1-P_0 with a confidence not less than C_0. A more compact form is obtained in terms of a quantile of the exceedance probability.

$$P_{exc}(Y)_{C_0} < 1 - P_0 \tag{2.33bis}$$

(2.28) to (2.33bis) are all valid forms of the RAC with second-level uncertainty. They represent the regulatory acceptance criterion for BEPU methodologies.

2.2.7. Evolving from conservative to BEPU RAC

In the historical evolution from conservative to BEPU methodologies, regulatory acceptance criteria have increasing complexity, due to the introduction of two levels of uncertainty. Figures 2.4 and 2.5 illustrate this evolution when the Monte Carlo technique is used in the propagation of uncertainty through the calculation. Figure 2.4 displays the addition of two layers of probability to the former simple criterion for a scalar Y. When the sample size tends to infinity, the second-level uncertainty tends to zero and the form simplifies from (2.28) to (2.14), as shown in the figure. This is the case when the code runs are cheap (in computation time and other resources) and can be massively performed. The simplified form (2.14) would be also appropriate in the special (and unusual) case when the model is so simple that the propagation of uncertainties can be calculated very accurately e.g. in a closed form.

Figure 2.4. The evolution of the regulatory acceptance criterion for a scalar quantity: from conservative to BEPU methodologies

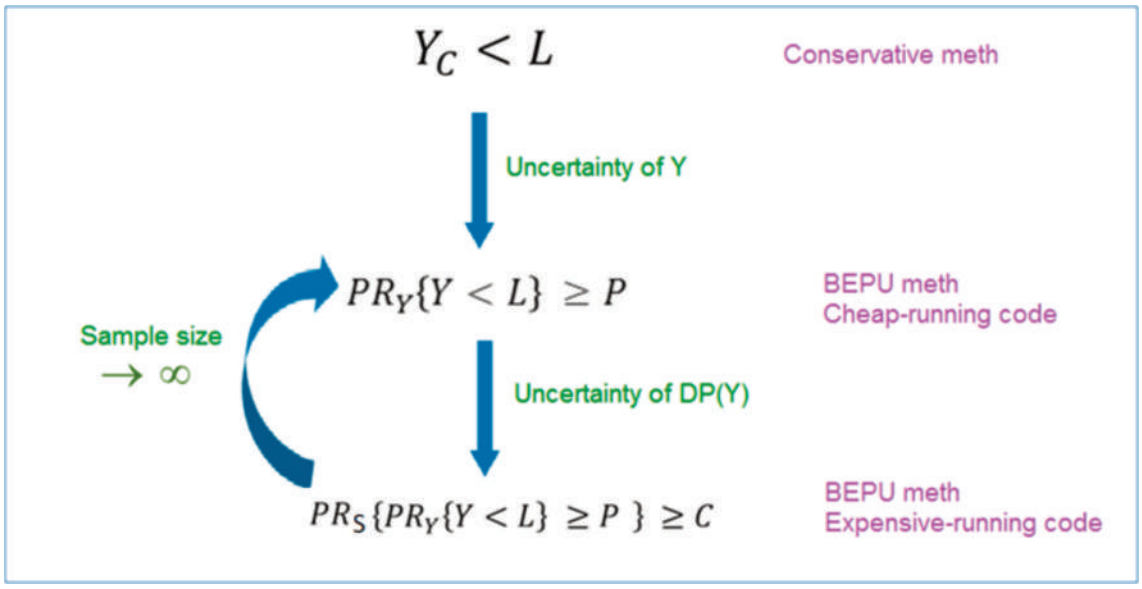
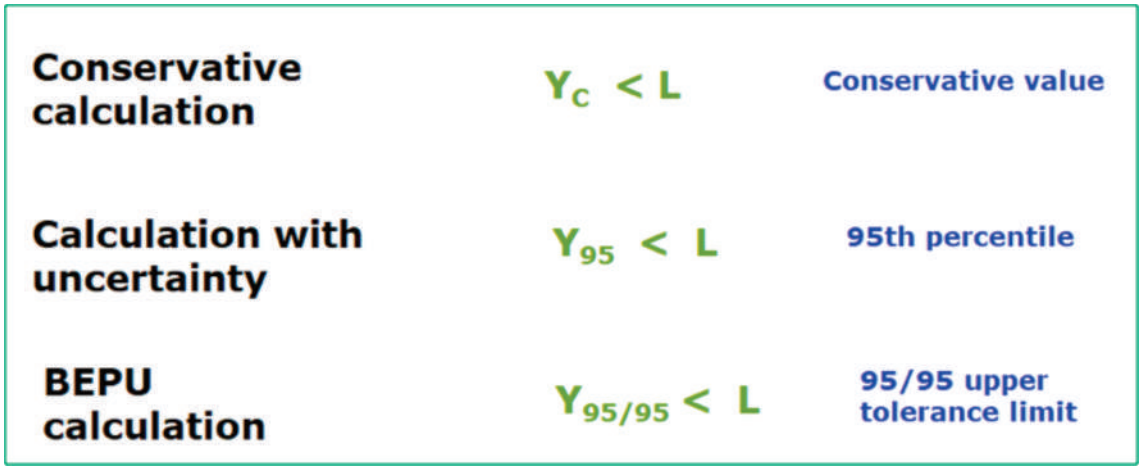


Figure 2.5 shows the same evolution, with a more compact notation, stating that the quantity compared with the limit L is, respectively, a conservative value of Y, a quantile of Y, and an upper tolerance level of Y.

Figure 2.5. The form of regulatory acceptance criteria (green), in a compact notation, assuming the standard tolerance level 95/95. In blue, the quantity to compare with the regulatory limit L.



There is a basic difference between the verification of criteria (2.21) and (2.31). In the first case, the figure of merit (FOM) to compare with the limit L is a quantile of Y, which is a fixed quantity only depending on the code and the input probability distributions. In the second case, the FOM is a tolerance level, which is a random variable depending on the code, the input PD and also on the random sample used. If the sample changes, the FOM changes.

2.2.8. The multidimensional case: joint versus individual fulfilment of acceptance criteria

When Y is a multidimensional quantity, $Y=(Y_1 \ldots, Y_D)$, the acceptance criterion without uncertainty

$$Y_C \in R_A \tag{2.34}$$

is typically equivalent to

$$Y_{i,C} \in R_{A,i} \qquad i = 1,\ldots,D \tag{2.35}$$

Expression (2.35) is understood as a conjunction (i.e. simultaneous fulfilment) of the D individual criteria. We write the logical statement

$$(\boldsymbol{Y} \in \boldsymbol{R_A}) \Leftrightarrow \bigwedge_{i=1}^{D}(Y_i \in R_{A,i}) \tag{2.36}$$

where the subscript C has been dropped. The joint criterion is fulfilled if and only if the individual criteria are fulfilled.

Using (2.36), the criterion (2.22) is rewritten

$$PR_S\left\{PR_{\boldsymbol{Y}}\left\{\bigwedge_{i=1}^{D}(Y_i \in R_{A,i})\right\} \geq P_0\right\} \geq C_0 \tag{2.37}$$

For every component Y_j of Y

$$\boldsymbol{Y} \in \boldsymbol{R_A} \Rightarrow Y_j \in R_{A,j} \tag{2.38}$$

From logical statements we can derive probability statements. When statement A implies statement B, the probability of A is less or equal than the probability of B. Then, from (2.38):

$$PR_{\boldsymbol{Y}}(\boldsymbol{Y} \in \boldsymbol{R_A}) \leq PR_{Y_j}(Y_j \in R_{A,j}) \tag{2.39}$$

and

$$PR_{\boldsymbol{Y}}(\boldsymbol{Y} \in \boldsymbol{R_A}) \geq P_0 \Rightarrow PR_{Y_j}(Y_j \in R_{A,j}) \geq P_0 \tag{2.40}$$

So, the fulfilment of the joint criterion with uncertainty and to the level P_0 implies the fulfilment of each individual criterion with uncertainty and to the same level.

Finally, adding the second-order uncertainty to (2.40):

$$PR_S\{PR_{\boldsymbol{Y}}(\boldsymbol{Y} \in \boldsymbol{R_A}) \geq P_0\} \leq PR_S\{PR_{Y_j}(Y_j \in R_{A,j}) \geq P_0\} \tag{2.41}$$

so that, for every j=1,…,D

$$PR_S\{PR_{\boldsymbol{Y}}(\boldsymbol{Y} \in \boldsymbol{R_A}) \geq P_0\} \geq C_0 \quad \Rightarrow \quad PR_S\left\{PR_{Y_j}(Y_j \in R_{A,j}) \geq P_0\right\} \geq C_0 \tag{2.42}$$

Therefore, the fulfilment of the joint BEPU criterion to a level (P_0, C_0) implies the fulfilment of every individual criterion to the same level. This is an obvious result, because the joint criterion is stronger than each individual criterion. The converse is not true. The fulfilment of all the individual criteria does not ensure the satisfaction of the joint criterion to the same level.

In conclusion, it is safer, from the regulatory point of view, to require the joint fulfillment than the individual one to the same level.

In Annex II we prove that a sufficient condition for the fulfilment of the joint criterion is the fulfilment of the individual criteria to an increased level (P_0^*, C_0^*), where

$$P_0^* \equiv 1 - \frac{1 - P_0}{D} \quad ; \quad C_0^* \equiv 1 - \frac{1 - C_0}{D} \tag{2.43}$$

In the case that the individual criteria are tested independently, the probability in (2.37) is factorised

$$PR\left\{\prod_{j=1}^{D} PR\{Y_i \in R_{Ai}\} \geq P_0\right\} \geq C_0 \tag{2.44}$$

and in Annex II it is also proved that a sufficient condition for the fulfilment of the joint criterion is the fulfilment of the individual criteria to an increased level (P_0^*, C_0^*), where

$$P_0^* \geq \sqrt[D]{P_0} \qquad C_0^* \geq \sqrt[D]{C_0} \tag{2.45}$$

When P_0 and C_0 tend to 1, the lower bounds in (2.45) tend to those in (2.43).

2.2.9. An example of BEPU acceptance criteria: LOCA/ECCS analysis according to 10 CFR 50.46

Our previous discussion about the BEPU acceptance criteria in DSA is very well exemplified in the regulation of LOCA/ECCS analyses in USA. Every LWR plant must perform (as a part of the Deterministic Safety Analysis) an analysis of design basis accidents representing a loss of coolant of the primary system. Such accidents, termed LOCA, are major challenges to the safety system of the plant, particularly to Emergency Core Cooling System (ECCS). The LOCA/ECCS analysis is so important that it is regulated in the US Code of Federal Regulations, in the section 10 CRF 50.46 [16], which was formerly released in 1974. At that time, the knowledge about the LOCA thermohydraulic phenomena was still limited. So, it was decided to require very conservative LOCA analyses, following very strict prescriptions described in the Appendix K to Title 10 CRF 50 [17]. Additionally, the regulatory acceptance criteria were stated in [16], and main safety quantities to be calculated were defined: peak cladding temperature (PCT), cladding maximum local oxidation (MLO), and core-wide oxidation (CWO, that quantifies the hydrogen generation during the LOCA).

After a period of experimental, theoretical and calculational development, the USNRC decided to revise 10 CFR 50.46. The new version was released in 1989 and gave the licensees the option to perform realistic LOCA/ECCS analyses, based on best estimate models and supplemented with an uncertainty analysis of the safety quantities. The restrictions of the conservative RAC were maintained for realistic calculations, but the fulfilment was exacted "with a high level of probability". This requirement implied a probabilistic modelling of the uncertainty. Nevertheless, some uncertain variables were excluded of such treatment. The most noted was the break size: the rule requires analysing the plausible range of break sizes and locations. And this amounts to modelling the uncertainty of the break size via an interval, rather than with a probability distribution.

The 1989 issuance of 10 CFR 50.46 had some shortcomings about uncertainty issues:

- The "high level of probability" was not quantified.
- It was not clear whether the "high level of probability" referred specifically to the individual fulfilment or to the joint fulfilment of acceptance criteria.
- Nothing was said about second-level uncertainty.

To complement the revised ECCS rule, USNRC released in 1989 the Regulatory Guide 1.157 [96], that includes the requirements for models and assumptions in a BEPU LOCA/ECCS analysis. Furthermore, it includes guidelines for carrying out the uncertainty analysis and fixes 0.95 as the acceptable level for the probability of fulfilling the acceptance criteria. But the guide barely refers to the second-level uncertainty, and does not establish a regulatory level for it.

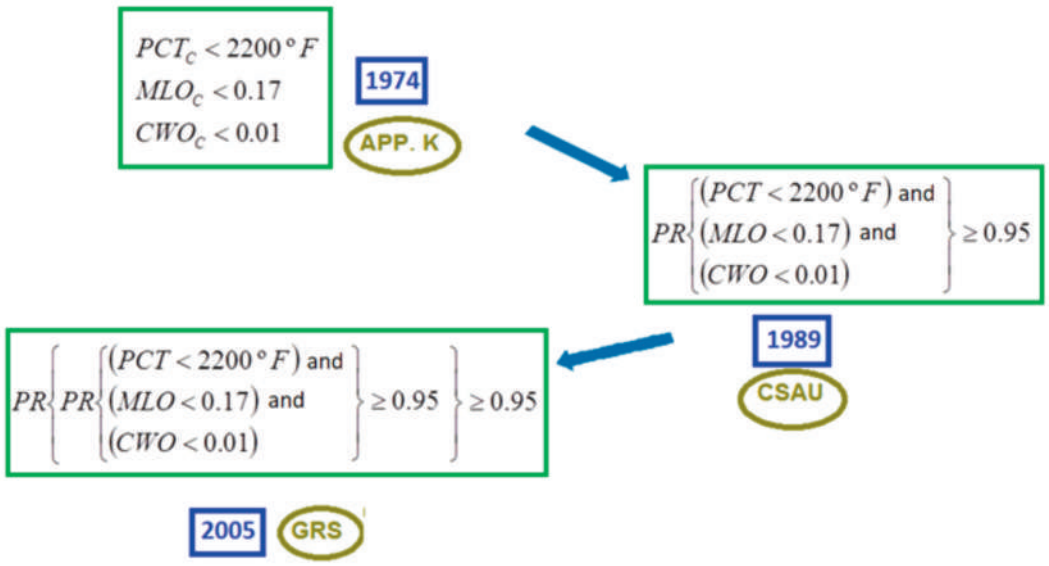
Another landmark in the revision of the ECCS rule was the publication, also in 1989, of CSAU (*Code Scaling, Applicability and Uncertainty Evaluation Methodology*) [93], which was the first BEPU methodology. The CSAU framework has been adopted for many BEPU methodologies. It was generalized to other types of transients and accidents in NPP with the EMDAP (*Evaluation Model Development and Assessment Process*) developed by the USNRC technical staff and presented in the Regulatory Guide 1.203 [98].

The first application of CSAU used a metamodel to propagate the uncertainty (see section 3.3.3.2). The former BEPU methodologies used similar procedures and in general did not consider the second-level uncertainty (or maybe considered it as combined with the first-level uncertainty).

About a decade after CSAU, the German company GRS (*Gesellschaft für Anlagen- und Reaktorsicherheit*) developed a methodology [32] that followed the CSAU scheme, but changed completely the method of propagation. GRS used the original code (instead of a metamodel) to propagate the uncertainty, and applied Wilks' method (see section 3.3.1.1.1) to calculate tolerance limits. Therefore, they took into account the metauncertainty due to the finite size of the simple random sample and used a confidence level of 95%. After the boom of Wilks' method, regulators started to require explicitly a specific tolerance level for the fulfilment of acceptance criteria in DSA.

Figure 2.6 illustrates the evolution of LOCA/ECCS RAC.

Figure 2.6. Regulatory Acceptance Criteria for LOCA/ECCS analyses. Evolution from conservative to BEPU methodologies.



III. Verifying the fulfilment of BEPU acceptance criteria: statistical methods

III. VERIFYING THE FULFILMENT OF BEPU ACCEPTANCE CRITERIA: STATISTICAL METHODS

3.1. Testing the fulfillment of a RAC: a statistical problem

The last step of the analysis of design basis scenarios in DSA is the verification of fulfilment of the regulatory acceptance criteria (RAC). If that is verified, the design is considered as safe. If there is any RAC which is not met, the safety design must be modified to achieve compliance with the criterion. The BEPU acceptance criteria are probabilistic statements, representing requirements on probability distributions. Therefore, statistical techniques are needed to verify if the criteria are satisfied. In fact, when the uncertainty is modelled probabilistically, it must be calculated statistically. Testing the fulfilment of a BEPU criterion means entering the realm of statistical inference, which is classically divided into two fields: estimation and hypothesis test. These two fields are much related; if we know techniques to estimate parameters of random variables, we can test hypotheses about those parameters. The BEPU criteria can be regarded as hypotheses about random variables (safety quantities) [34]. The statistical test should have as null or default hypothesis the conservative one i.e. that "the criterion is not met". It is the alternative hypothesis ("the criterion is met") the one that must be proved.

In chapter 2 we show how the scalar RAC may be expressed in terms either of quantiles or of limit exceedance probabilities, and how this fact can be extended to multidimensional RAC. Because confidence limits of quantiles are tolerance limits, it is clear that statistical methods for estimation of quantiles and probabilities and for construction of tolerance regions / intervals are needed for our purpose.

Estimators have statistical uncertainty, representing the second-level uncertainty in the RAC. The standard deviation of an estimator is termed *standard error.* The precision of the estimator is related to the inverse of its variance. Estimator precision and efficiency are almost equivalent terms. In this sense, it is important to distinguish two meanings of the term "conservative" in the present context. In Statistics, the term *conservative* is often used as antonym of *efficient*. That is, an estimator is more conservative than other when the former is less precise than the latter. However, in Nuclear Safety, the term *conservative* means *severe* (in the sense of damage or nearness to damage). If the scalar safety quantity Y has an upper regulatory limit, a quantile estimator of Y is more conservative when it produces higher values. And the same can be said of exceedance probability estimators. One of such estimators may be very conservative (because it has a large bias) and, at the same time, it may be very precise (low standard error).

Typically, estimators are based on random samples of the safety quantities, and the larger the sample size, the more precise is the estimator.

NPP licensees must test the fulfilment of the RAC in DSA. With such goal, they choose a statistical method among those accepted by the regulator, who has the responsibility of assessing the analyses. In this setting, the properties of efficiency and conservativeness are very important. And other features of the methods are important too:

- Ease of use
- Number of assumptions of the method: more assumptions may imply greater difficulty in the justification of the use of the method.

The estimation methods can be classified according to several features:

- *Parametric vs. nonparametric methods.* Parametric methods assume that the random variable under study belongs to a specific parametric family. Nonparametric methods do not rely on such assumptions, but on much weaker hypotheses e.g. the continuity of the CDF or the PDF.

- *Frequentist vs. Bayesian methods*. Frequentist methods regard the parameters to estimate as fixed but unknown quantities. They can implement *a priori* information e.g. in the form of parametric distributions. They provide estimators (with known distribution) or confidence regions for the unknown parameters. On the other hand, Bayesian methods regard the parameters to estimate as random variables. They implement *a priori* information in the form of *prior distributions* of the parameters. Bayes' theorem combines prior information with sample information, and produce the so-called *posterior distribution* of the parameters, which allows e.g. setting credible intervals for them.

- *Exact vs approximate methods*.

Reference [29] can give an idea of the wide variety of analyses in Nuclear Safety where BEPU methods have been applied.

In the remainder of this chapter, we will describe and analyze statistical methods to test the fulfilment of the RAC. We will distinguish the RAC with or without second-level uncertainty; the scalar from the multidimensional RAC and the methods based on quantiles and tolerance regions from those based on exceedance probabilities.

3.2. The verification of criteria without second-level uncertainty

When the code runs are very cheap (in terms of calculation time and resources) so that running it massively has an affordable cost, the second-level uncertainty can in practice be cancelled. The RAC has then a simple form, and the verification of the fulfilment is equally simple. A "method of probability regions" should be applied, as next described.

A P_0-probability region for $\mathbf{Y}$ is a region $\mathbf{RP}$ in the range of $\mathbf{Y}$ such that the probability of $\mathbf{Y}$ being inside $\mathbf{RP}$ is at least P_0. If a P_0-probability region $\mathbf{RP}$ is obtained for $\mathbf{Y}$ and $\mathbf{RP}$ is inside the acceptance region of the RAC, then the RAC is fulfilled. When the probability distribution of the safety output is accurately known, it is very simple to calculate a probability region of this type.

For scalar safety quantities, a corresponding "method of probability intervals" is applied. If a P_0-probability interval is constructed for Y, and such interval is contained in the acceptance region of Y, the RAC is satisfied. Any type of interval (two-sided or one-sided) is valid for the method. Nonetheless, if Y has a one-sided acceptance interval, the most efficient choice is a one-sided probability interval. Specifically, if Y has an upper acceptance limit L, the figure of merit to calculate and to compare with L is the quantile of order P_0 of Y.

Other possibility is the use of the limit exceedance probability for the safety quantity ($\mathbf{Y}$ or Y), that can be calculated accurately and compared to the regulatory limit $1\text{-}P_0$.

3.3. The verification of criteria with second-level uncertainty

3.3.1. The scalar case

When Y is a scalar continuous safety quantity, the BEPU criterion including second-level uncertainty can be formulated in terms either of a quantile of Y or of the limit exceedance probability of Y. Accordingly, the statistical methods to verify the criteria can be categorized as follows:

- Methods of quantile estimation, that we term *Q-methods.*
- Methods of exceedance probability, that we term *P-methods.*

The terms "Q-methods" and "P-methods" are taken from [34]. Ultimately, both categories solve the same problem. Indeed, there is a property of duality that we formulate as follows: from a given Q-method we can derive a P-method, and vice versa, and each method is termed *dual* of the other [52].

3.3.1.1. Methods of quantile estimation (Q-methods)

A first category refers to methods of estimation of the P_0-quantile of Y. There is a complete field of statistical estimation devoted to these methods, described as "quantile estimation" and "tolerance intervals calculation", because a one-sided tolerance limit is a confidence limit of a quantile.

Specifically, an upper (resp. lower) tolerance limit with level (P_0, C_0) of Y is an upper (resp. lower) confidence limit with level C_0 of the P_0-quantile (resp. the (1-P_0)-quantile) of Y.

Therefore, quantile estimation and tolerance limit setting are basically the same thing.

We know that the scalar acceptance criteria in the nuclear DSA are one-sided i.e. they are defined by upper or lower safety limits, and the multidimensional RAC are simply conjunctions of one-sided scalar criteria.

So, for scalar safety quantities, the "regulatory acceptance regions" are one-sided intervals in the range of the quantities. The scalar Q-methods in BEPU analysis can be generically termed *methods of tolerance intervals,* because they are based on the following result: if a random sample S of Y is obtained, and a tolerance interval with at least the regulatory level (P_0, C_0) is constructed from S, and such interval is contained in the acceptance interval of Y, then the BEPU RAC is fulfilled. Any type of tolerance interval (be it one-sided, standard two-sided or equal-tailed) with the regulatory tolerance level is adequate to apply these methods. Nevertheless, it seems logical that, for a one-sided acceptance interval, the most efficient tolerance interval (TI) should be one-sided too. The endpoint of a one-sided TI is a tolerance limit, which is a confidence limit of a quantile of prespecified order of the random variable.

In the specific case of an upper acceptance limit L, if we calculate an upper tolerance limit (UTL), with level not less than the regulatory level, the RAC is equivalent to the condition that the UTL is below L.

It is important to emphasize the difference between the methods used for the RAC with and without second-level uncertainty. If there is no metauncertainty, we use probability regions, that basically depend only on the probability distribution of the variable and on the coverage level P_0. On the other hand, when the second-level uncertainty is not negligible, we use tolerance regions, that depend on the probability distribution of the variable and the tolerance level (P_0, C_0), but depend on the random sample of the variable too. Hence, there is a random component on the tolerance region that does not exist in the probability regions. For this reason, it is important to make sure that the random sample is truly obtained in a random fashion.

Many methods of quantile estimation and tolerance intervals construction exist. There is one overwhelmingly used in the field of nuclear thermal-hydraulics and accident analysis: the nonparametric Wilks' method, which will be simply term Wilks' method throughout this document. We next make a detailed description of this method. After that, other procedures of quantile estimation, presently less used than Wilks' method, are mentioned.

3.3.1.1.1. The nonparametric Wilks' method

Samuel S. Wilks (1906-1964) was a Texan mathematician, one of the most renowned statisticians in the first half of 20th century. During his career, he made significant contributions to mathematical statistics, mainly in the fields of multivariate statistics, regression, theory of likelihood ratio tests, etc.

Figure 3.1. Samuel S. Wilks.

One of his main contributions was developed in response to a requirement formulated by Walter A. Shewhart (a father of statistical quality control in manufacturing processes). Shewhart created the concept of "statistical tolerance interval", as a statistical counterpart to the tolerance intervals used in quality control to accept / reject manufactured items [90]. The problem of how to construct tolerance intervals with a specified level of coverage and confidence was brilliantly solved by Wilks in 1941 [106], when he proved that order statistics could be used as distribution-free tolerance limits.

The emergence of Wilks' method attracted the interest of other American mathematicians to the field of tolerance regions. In 1943, Abraham Wald published an extension of Wilks' method to multidimensional variables [102]. Shortly after, John Tukey and Henry Scheffé studied the conditions of applicability of the method [87].

Wilks' nonparametric method is a universally known and applied technique. Born in the field of quality control, it found a widespread use on uncertainty analysis of calculations. At the end of the 1980s, the German company GRS started to apply it to BEPU analysis in nuclear thermohydraulics [32]. From then on, the method (sometimes known as "GRS method") started to be massively applied in uncertainty analyses of nuclear safety calculations, based on simple random samples of the calculation safety results. Former BEPU methods were based on the use of metamodels to propagate uncertainty (see section 3.3.3.2) , following the pioneering CSAU methodology [93]. The advent of Wilks' method turned upside down the situation, because it was adopted by a majority of BEPU methodologies. For instance, Westinghouse changed its BEPU methodology for LOCA/ECCS analysis in PWR from one based in metamodels [112] to one based on Wilks' method [28]

The term "nonparametric" refers to the fact that tolerance intervals generated with Wilks' method are distribution-free i.e. the procedure to obtain them is the same whatever the distribution of the analysed random variable is.

The nonparametric Wilks' method

In the sequel we will describe Wilks' nonparametric method, established by S.S. Wilks [106-108] and modified in its theoretical basis and extended by J. Tukey and H. Scheffé [87, 94, 95].

Let Y be a scalar and continuous random variable with a continuous cumulative distribution function (CDF), termed as F_Y. Let $S=(Y_1,\ldots,Y_N)$ be a simple random sampling (SRS) of Y, with size N. We term the *ordered sample* corresponding to S the N-tuple obtained by sorting the sample S from the minimum to the maximum element of S.

$$S_N^* \equiv (Y_{1:N}, \ldots, Y_{N:N})$$

$$Y_{1:N} < \cdots < Y_{i:N} < Y_{i+1:N} < \ldots < Y_{N:N} \tag{3.1}$$

It is observed that, in (3.1), the sorted elements have been reindexed. The element $Y_{i:N}$ is the i-th smaller element, and the (N-i+1)-th larger element of the sample S. It is clear that every element $Y_{i:N}$, i=1,…,N is a function of the sample S; if the sample changes, the elements may change. For this reason, the element $Y_{i:N}$ is termed *order statistic* of ordinal i. This definition makes sense, because (in the statistical jargon) a *statistic* is defined as a numerical function of a random sample.

A synonym of order statistic, sometimes used, is *ordered statistic.* The acronym OS represents both terms.

Some OS deserve special names. Clearly, $Y_{1:N}$ is the *sample minimum* and $Y_{N:N}$ is the *sample maximum*. $Y_{2:N}$ and $Y_{N-1:N}$ are sometimes termed the *sample second minimum and sample second máximum,* respectively, and so on [10]. When N is an odd number N=2k+1, the order statistic $Y_{k+1:N}$ is termed the *median* of the sample. When N is even, N=2k, the median is no longer an order statistic, but rather the half-sum of the OS $Y_{k:N}$ and $Y_{k+1:N}$.

Order statistics are very connected to sample quantiles. Let λ be a number in the interval (0,1). The λ-quantile of the sample has not an undisputed definition. It may be the OS with ordinal equal to the integer part of $\lambda \cdot N$, or the integer part of $\lambda \cdot (N+1)$; in some cases, it is defined as a convex combination of order statistics, as we will later see.

This relation between OS and sample quantiles was a major inspiration for Wilks' method. In fact, Wilks proved that order statistics can be used as tolerance limits i.e. confidence limits of quantiles. More specifically, OS can be used as endpoints of tolerance intervals.

Let $Y_{r:N}$ and $Y_{s:N}$, r<s, two order statistics of the same sample. It is clear that $Y_{r:N} < Y_{s:N}$. We define the coverage of the interval ($Y_{r:N}$, $Y_{s:N}$) as

$$\mu_Y(Y_{r:N}, Y_{s:N}) \equiv PR_Y\{Y_{r:N} < Y < Y_{s:N}\} \equiv F_Y(Y_{s:N}) - F_Y(Y_{r:N}) \tag{3.2}$$

Therefore, given two OS obtained from a random sample S, the coverage of the interval they span is the probability that a new "extraction" of Y, independent of S, falls inside the interval. We employ the symbol μ_Y, meaning that the coverage is a mathematical measure of the interval. $Y_{r:N}$ and $Y_{s:N}$ are functions of the sample, and thus the coverage of the interval they encompass is also a function of the sample i.e. a random variable with sampling distribution.

The basis of Wilks' method is the following theorem, that we give without demonstration [106]. Given a continuous scalar random variable, Y, with continuous CDF, the coverage of the interval ($Y_{r:N}$, $Y_{s:N}$) is a random variable which follows a distribution beta with parameters s-r and N-s+r+1.

The beta(R, T) distribution, where R and T are real positive numbers, has the PDF

$$f_{beta(R,T)}(x) = \frac{x^{R-1}(1-x)^{T-1}}{B(R,T)} \qquad 0 \le x \le 1 \tag{3.3}$$

B(R, T) is the so-called beta function with parameters R and T

$$B(R,T) \equiv \int_0^1 x^{R-1}(1-x)^{T-1}\,dx \tag{3.4}$$

that can be expressed in terms of the *gamma function*

$$B(R,T) = \frac{\Gamma(R)\Gamma(T)}{\Gamma(R+T)} \tag{3.5}$$

The CDF of the beta(R, T) distribution is

$$F_{beta(R,T)}(x) \equiv \frac{1}{B(R,T)}\int_0^x z^{R-1}(1-z)^{T-1}\,dz \equiv I_x(R,T) \tag{3.6}$$

$I_x(R, T)$ is the regularized incomplete beta function.

It must be highlighted that the distribution of the coverage of $(Y_{r:N}, Y_{s:N})$ is always beta(s-r, n-s+r+1) whatever the "parent" distribution F_Y is. This is the reason why Wilks' method is distribution-free. In fact, the term "nonparametric", applied to Wilks' method, means (in this context) "distribution-free".

A corollary of the previous theorem is that the variable $F(Y_{r:N})$ for r=1,…,N follows a beta(r, N-r+1) distribution. It is a random variable obtained by the application of the CDF of Y to the OS obtained from the sample S.

The previous theorem, obtained by Wilks, enables the use of OS as endpoints of tolerance intervals. For one-sided intervals, there is only one endpoint function of the sample, which is named *tolerance limit.* Depending on the case, it can be either an upper or a lower tolerance limit.

Given two real numbers β and γ in the interval (0, 1), the condition for $(Y_{r:N}, Y_{s:N})$ being a tolerance interval with level (β,γ) of Y is

$$PR_S\{PR_Y\{Y_{r:N} < Y < Y_{s:N}\} \geq \beta\} \geq \gamma \tag{3.7}$$

Using the previous Wilks' theorem, (3.7) transforms into

$$PR_S\{beta(s - r, N - s + r + 1) \geq \beta\} \geq \gamma \tag{3.8}$$

We term (3.8) the *two-sided Wilks' formula.* It is an inequality relating ordinals r and s, sample size N and tolerance level (β,γ). (3.8) is easily rewritten as:

$$F_{beta(s-r,\ N-s+r+1)}\ (\beta)\ \leq 1 - \gamma \tag{3.9}$$

The function on the left hand side of (3.9) is the CDF of the beta(s-r, N-s+r+1) distribution, which is related with the incomplete beta function:

$$F_{beta(s-r,\ N-s+r+1)}\ (\beta) = I_{\beta}\ (s - r,\ N - s + r + 1) \tag{3.10}$$

Thus, an alternative form to (3.9) is

$$1 - I_{\beta}\ (s - r,\ N - s + r + 1)\ \geq \gamma \tag{3.11}$$

Also (3.9) can be cast in terms of quantiles of the beta distribution:

$$beta(\ s - r,\ N - s + r + 1)_{1-\gamma}\ \geq\ \beta \tag{3.12}$$

stating that the quantile of order 1- γ of the given beta distribution must be not lower than β.

Using the functional form of the beta distribution, we obtain from (3.9) the integral version of two-sided Wilks' formula

$$\frac{\Gamma(N+1)}{\Gamma(s-r)\Gamma(N-s+r+1)} \int_0^{\beta} x^{s-r-1}(1-x)^{N-s+r}\, dx \leq 1-\gamma \tag{3.13}$$

Successive integration by parts of the left-hand side of (3.13) yields the summation form of two-sided Wilks' formula:

$$\sum_{i=s-r}^{N} \binom{N}{i} \beta^i (1-\beta)^{N-i} \leq 1-\gamma \tag{3.14}$$

Inequalities (3.9) and (3.11) to (3.14) are all valid forms of *two-sided Wilks' formula*, adequate for two-sided tolerance intervals.

If we fix the tolerance level (β,γ) and the ordinals r and s (possibly as functions of N) then it can be proved that there is a minimum integer value of N satisfying Wilks' formula. This value is here termed the *minimum sample size:*

$$N_{min}\ (\beta,\gamma,s,r) \equiv\ \text{Min}\ \{N \in Z^{+} \mid beta(s - r,\ N - s + r + 1)_{1-\gamma} \geq \beta\} \tag{3.15}$$

A very usual choice of the ordinals is r < (N+1)/2 and s=N-r+1. In such case, the ordinals r and s are termed conjugate, and (3.8) transforms to

$$PR_S \{\text{beta}(N - 2r + 1, 2r) \geq \beta \} \geq \gamma \tag{3.16}$$

or, in terms of the incomplete beta function

$$1 - I_\beta (N - 2r + 1, 2r) \geq \gamma \tag{3.17}$$

In this case we term r the *rank* of the TI. In Table 3.1, the minimum sample size obtained from (3.15) is given for Wilks' (95, 95) two-sided TI encompassed by conjugate OS. Clearly N_{min} is a growing function of the rank r. We also term these values *critical sample sizes*.

Table 3.1. Order statistics and minimum sample sizes for Wilks' two-sided 95/95 tolerance intervals, using conjugate OS as endpoints.

Rank r	N_{min} for Wilks' two-sided 95/95 tolerance intervals	OS for Wilks' two-sided 95/95 tolerance intervals
1	93	(1, 93)
2	153	(2, 152)
3	208	(3, 206)
4	260	(4, 257)
5	311	(5, 307)
6	361	(6, 356)
7	410	(7, 404)
8	458	(8, 451)
9	506	(9, 498)
10	554	(10, 545)
11	601	(11, 591)

(3.15) can be further minimized, in terms of ordinals r and s, producing what we term *minimum minimorum (mm)* sample size. Such value depends only on the tolerance level and is obtained when r=1 and s=N i.e. when the sample extremes are the interval endpoints. Then (3.16) reduces to

$$PR_S\{beta(N - 1,2) \geq \beta\} \geq \gamma \tag{3.18}$$

and

$$N_{mm}(\beta,\gamma) \equiv Min\{N \in Z^+ / beta(N - 1,2)_{1-\gamma} \geq \beta\} \tag{3.19}$$

For practitioners of Wilks' method, a well-known value is $N_{mm}(0.95, 0.95)=93$.

The one-sided case.

Wilks' method also enables setting one-sided tolerance intervals. Previous expressions for the two-sided case are directly applicable to the one-sided case, simply setting r=0 when $Y_{s:N}$ is used as upper tolerance limit and setting s=N+1 when $Y_{r:N}$ is used as lower tolerance limit.

The condition for the OS with ordinal s to be an upper tolerance limit with level (β,γ) of Y is

$$PR_S\{PR_Y\{Y < Y_{s:N}\} \geq \beta\} \geq \gamma \tag{3.20}$$

which can be written as

$$PR_S\{F_Y(Y_{s:N}) \geq \beta\} \geq \gamma \tag{3.21}$$

namely

$$PR_S\{beta(s, N - s + 1) \geq \beta\} \geq \gamma \tag{3.22}$$

(3.22) is one-sided Wilks' formula for an upper tolerance limit. It coincides with (3.8) when we set r=0. Once fixed the tolerance level and the ordinal s (as a function of N), there exists a minimum value of N fulfilling (3.22). Conversely, for a fixed value of N, there is a minimum integer value of s fulfilling (3.22). In terms of the quantiles of beta distribution, (3.22) transforms to

$$beta(s, N - s + 1)_{1-\gamma} \geq \beta \tag{3.23}$$

Hence, there are two versions of the Wilks' method and of the corresponding Wilks' formula: the two-sided one, with $1 \leq r < s \leq N$, and the one-sided one, with r=0 and $1 \leq s \leq N$ or with s=N+1 and $1 \leq r \leq N$.

The minimum sample size solving (3.23) is

$$N_{min}(\beta,\gamma,s,0) \equiv \text{Min}\{N \in Z^{+}/beta(s, N - s + 1)_{1-\gamma} \geq \beta\} \tag{3.24}$$

Table 3.2 shows the minimum sample sizes for s= N, N-1, N-2 and N-3 and tolerance level (95, 95). They are very well-known numbers (critical values) for the practitioners of Wilks' method.

Table 3.2. Minimum sample sizes obtained from (3.24) with tolerance level (95, 95).

s		N_{min}
N	Sample maximum	59
N-1	Sample second maximum	93
N-2	Sample third maximum	124
N-3	Sample fourth maximum	153

The size (3.24) can be further minimized with respect to the ordinal s. The *minimum minimorum* (*mm*) is obtained for s=N i.e. when the sample maximum is used as upper tolerance limit. In such case, (3.22) takes a very simple form

$$1 - \beta^{N} \geq \gamma \tag{3.25}$$

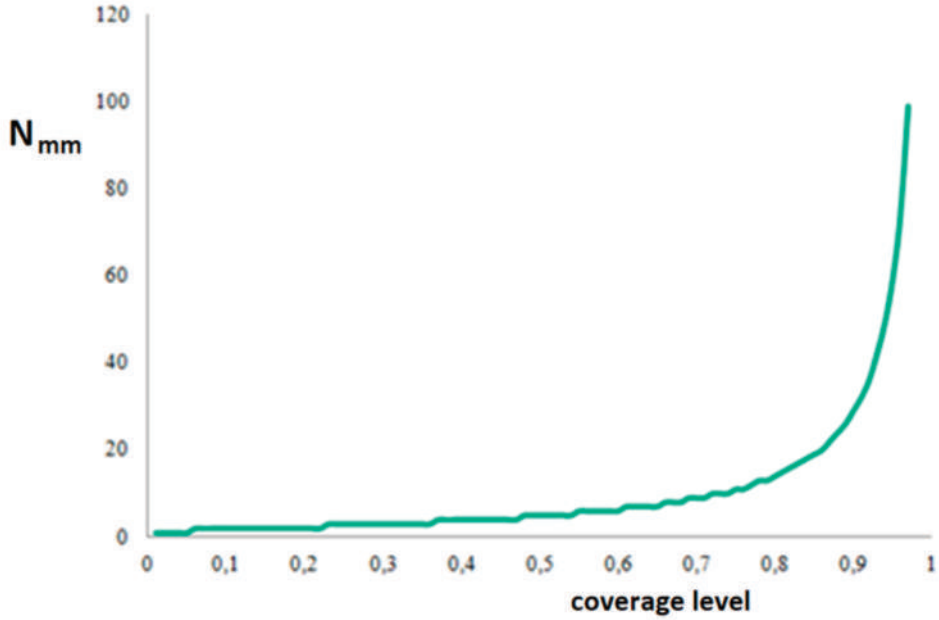
The *mm* size then derives from (3.25):

$$N_{mm} = \left\lceil \frac{log(1-\gamma)}{log\beta} \right\rceil \tag{3.26}$$

the partial bracket representing the ceiling function (i.e. integer value plus one).

Figure 3.2 plots N_{mm} versus the coverage level β for γ=0.95.

Figure 3.2. *mm* sample size as a function of the coverage level, for a confidence level of 0.95.



From now on, we will term (3.8) (and equivalent forms) the *two-sided Wilks' formula* and (3.22) (and equivalent forms) the *one-sided Wilks' formula*. It is easy to prove that, for a given tolerance level, the *mm* sample size derived from the two-sided Wilks' formula is higher than the one derived from one-sided Wilks' formula.

There is an interesting property of symmetry in Wilks' one-sided method, which is not usually mentioned. If the order statistic $Y_{r:N}$ is a lower tolerance limit of Y with level (P, C), with r< (N+1)/2, then the conjugate order statistic $Y_{N-r+1:N}$ is an upper tolerance limit with the same level, and vice versa. This result derives from the fact that, if V is a random variable with distribution beta(A, B), then 1-V follows a distribution beta(B, A). Consequently, Wilks' one-sided method allows setting pairs of *simultaneous tolerance limits* (one upper, one lower) except when N is an odd number and r=(N+1)/2 (and hence both limits coincide).

Any method of construction of upper tolerance limits (including Wilks' method) can also be used to calculate lower tolerance limits. Simply, it must be realized that a lower tolerance limit for Y is obtained via changing the sign of an upper tolerance limit (with the same level) of (-Y).

3.3.1.1.2. Two-sided confidence intervals for quantiles

Since its establishment, Wilks' method sparked the interest of many statisticians in the use of order statistics to set tolerance intervals [75, 76, 87, 94, 95, 102].

The method is used for setting the two basic types of tolerance intervals for a scalar continuous random variable: one-sided (bounded by upper or lower confidence limits for quantiles) and two-sided tolerance intervals. However there is a third possibility, combining characteristics of the previous two, namely setting a two-sided confidence interval on a specific quantile of the variable. For calculated safety outputs, this option is not particularly useful, because the option of a one-sided tolerance limit (in the conservative direction) is preferred. But we mention it here for the sake of completeness. The interval between two OS of Y, ($Y_{k1:N}$, $Y_{k2:N}$), k1<k2 is a confidence interval with level C of the Pth quantile of Y, Y_P, when the following condition is fulfilled:

$$I_P(k1, N - k1 + 1) - I_P(k2, N - k2 + 1) \geq C \tag{3.27}$$

where $I_P(A, B)$ is the incomplete beta function defined in (3.6). Typically, values of k1 and k2 are as close together as possible. Alternatively, the condition (3.27) can be written in terms of the binomial distribution.

3.3.1.1.3. Using Wilks' one-sided tolerance limits to obtain two-sided tolerance limits

According to our previous discussion, setting a two-sided tolerance interval via Wilks' method obviously requires the two-sided version of the method (i.e. using two order statistics as endpoints), and therefore the sample size needed is obtained from the two-sided Wilks' formula. But there is a procedure, proposed in [64, 67], to derive a two-sided tolerance interval from a one-sided tolerance limit of a transformed variable obtained from Wilks' method. This procedure enables to construct two-sided TI based on sample sizes derived from one-sided Wilks' formula.

Let us consider again the scalar continuous random variable Y with a continuous CDF. Let $Y_{1,...,}Y_N$ a simple random sample of Y, with size N. And, finally, let Y_0 be a fixed value in the range of Y, independent of the elements of S. We term Y_0 the *pivotal value*.

We define the following scalar variable:

$$\Delta \equiv |Y - Y_0| \tag{3.28}$$

The cumulative distribution function of Δ is

$$F_{\Delta}(\delta) = F_Y(Y_0 + \delta) - F_Y(Y_0 - \delta) \tag{3.29}$$

Since F_Y is continuous, so is F_{Δ}. Thus, Wilks' method is applicable to Δ in order to obtain an upper tolerance limit of level (P, C), Δ_U. It is straightforward to see that

$$\Delta<\Delta_U \iff Y_0 - \Delta_U < Y < Y_0 + \Delta_U \tag{3.30}$$

From (3.30) it is evident that, if Δ_U is an upper tolerance limit with level (P,C), then (Y_0-Δ_U, Y_0+Δ_U) is a two-sided tolerance interval for Y with the same level. The centre of the interval is the pivotal value Y_0. Therefore, to obtain a two-sided tolerance interval for Y it is enough to obtain a one-sided upper tolerance limit with the same level for the transformed variable Δ. If Wilks' method is used, the sample size needed is derived from the one-sided Wilks' formula. For instance, setting a two-sided (95, 95) tolerance interval with our method requires a *mm* value of only 59, plus an additional pivotal value Y_0, instead of the value 93 derived from the two-sided Wilks' formula.

It is worth to mention that the transformation (3.28) can be applied to any method of setting tolerance intervals (aside from Wilks'). The equivalence (3.30) allows the replacement of the two-sided problem by a one-sided one. Parametric methods can be applied when there is strong evidence that Δ follows the corresponding parametric distribution. Otherwise, distribution-free methods should be used.

It seems obvious that the described procedure is more efficient when the pivotal value Y_0 is closer to the central values of Y (mean, median, mode, that are unknown *a priori*). If Y is a random variable with a distribution basically unknown, it is difficult to obtain a central value from it. However, if Y is an output of a calculation with random inputs (with known distribution), the nominal value Y_0, resulting from a calculation with the inputs in "nominal" (i.e. central) values, can be in the central region of Y range. Only when the input-output function associated to the code is strongly nonlinear, the nominal value Y_0 can be far from the central region.

In relation with the use of Wilks' method to set the upper tolerance limit of Δ, the order statistics of Δ do not have a simple relation with those of Y. In the simplest case, when the maximum of a sample of 59 is used, it is clear that

$$\Delta_{59:59}=max(Y_{59:59} - Y_0,\ Y_0 - Y_{1:59}) \tag{3.31}$$

Using (3.31), the resulting tolerance interval is

$$\begin{aligned} &(2Y_0 - Y_{59:59}, Y_{59:59}) && when \ \frac{Y_{59:59} + Y_{1:59}}{2} \geq Y_0 \\ &(Y_{1:59},\ 2Y_0 - Y_{1:59}) && when \ \frac{Y_{59:59} + Y_{1:59}}{2} \leq Y_0 \\ &(Y_{1:59},\ Y_{59:59}) && when \ \frac{Y_{59:59} + Y_{1:59}}{2} = Y_0 \end{aligned} \tag{3.32}$$

We can compare two methods of setting a two-sided tolerance interval of level (P, C) for Y:

M1. The standard Wilks' method (described in section 3.3.1.1.1) take a simple random sample of Y with size given by the two-sided Wilks´ formula, and then use the interval encompassed by the sample minimum and maximum.

M2. Take a simple random sample of Y with size given by the one-sided Wilks' formula, use a pivotal value of Y and then construct the interval defined in the right-hand side of (3.30).

The method M2 is less expensive than M1 (it needs a lower computational effort) and has the advantage of producing an interval centred on the pivotal value. For instance, for a standard tolerance value (95, 95), method M1 requires a *mm* sample size of 93, while method M2 requires only a size 59 (plus the pivotal value). In general, M2 requires a computational effort N_1+1, N_1 being the sample size obtained from one-sided Wilks' formula.

In exchange, method M2 may be less efficient than M1, producing larger tolerance intervals. It is expected that the more centred the pivotal value Y_0 is in the distribution of Y, the more efficient will be method M2.

3.3.1.1.4. Using Wilks' one-sided tolerance limits to obtain multidimensional tolerance regions

Wilks' method is only applicable to scalar random variables. But it can also be applied to the construction of tolerance regions for multidimensional random variables, by means of a simple variable change. The sole requirement is (as in the method described in the previous section) the knowledge of a value in the range of the multidimensional random variable (pivotal value), independent from the simple random sample used. The method is described in section 3.3.2.2 and constitutes an alternative to the Wald extension of Wilks' method, described in the same section.

3.3.1.1.5. The merits of Wilks' method

Wilks' method has been extensively used in nuclear thermalhydraulics and DSA, since it began to be applied back in the 1980's. There are a number of merits that justify such success [66]:

- It is very simple to use.
- It is based on a minimum number of assumptions, because it is
    - Nonparametric (i.e. distribution-free), applicable to every scalar continuous random variable with continuous CDF. Scheffé and Tukey [87] even justified its application to any CDF.
    - Frequentist; it does not require the use of a prior distribution.
    - Exact: it is not based on any approximation.
- It can be applied to multidimensional random variables (including calculated time trends), through variable changes.
- It is conservative, compared to other methods of quantile estimation and tolerance intervals setting, meaning that the variance of the quantile estimators that uses (order statistics) is large compared to other estimation methods. This can be regarded as an advantage, from a regulatory point of view, because the larger dispersion in Wilks' estimators ultimately tends to give results nearer to the acceptance limits than other methods.

3.3.1.1.6. Equal-tailed tolerance intervals constructed with order statistics

Equal-tailed tolerance intervals, defined, in section 1.11.2, can be (similarly to Wilks' method) constructed with order statistics. We have previously remarked that every one-sided TI has a simultaneous and conjugate one-sided TI. This fact does not imply that a one-sided TI can be regarded as an ETTI. Rather than requiring two simultaneous tolerance limits, the ETTI requires a stronger condition, namely the simultaneous coverage (with a specific joint confidence) of two quantiles.

Following Wilks' method, we look for a nonparametric method for setting ETTI based on order statistics. We consider again the scalar, continuous random variable Y, having a continuous CDF, and a simple random sample S of size N of Y, $S=(Y_{1,\ldots},Y_N)$. An equal-tailed TI covers simultaneously both quantiles (of order (1+P)/2 and (1-P)/2, respectively) of Y, and so it seems adequate to use as endpoints of the interval two conjugate order statistics: $Y_{r:N}$ and $Y_{N-r+1:N}$ with r < (N+1)/2. The condition to be fulfilled is:

$$PR\left\{Y_{r:N} < Y_{\frac{1-P}{2}} \quad AND\ Y_{N-r+1:N} > Y_{\frac{1+P}{2}}\right\} \geq C \tag{3.33}$$

where Y_Q is the quantile of order Q of Y, 0<Q<1.

This type of ETTI has been used several times in the BEPU analysis realm [21, 39, 42] in place of the standard two-sided TI.

If [L, U] is an equal-tailed tolerance interval with level (P, C), then, simultaneously,

- [L,U] is a two-sided tolerance interval with level (P, C)
- L is a lower tolerance limit with level ((1-P)/2, C)
- U is an upper tolerance limit with level ((1+P)/2, C).

In Annex I it is proved that condition (3.33) is equivalent to

$$1 - I_{\frac{1+P}{2}}(N-r+1,r)\left[2 - \sum_{k=0}^{r-1}\binom{N}{k}\left(\frac{1-P}{2}\right)^k\left(\frac{1+P}{2}\right)^{N-k}\left(I_{\frac{1+P}{2}}(N-k-r+1,r)\right)\right] \geq C \quad (3.34)$$

where $I_\lambda(A, B)$ is the incomplete beta function with parameters A, B and λ. Inequality (3.34) relates the sample size (N), the rank (r), and the tolerance level (P, C). It is the counterpart for ETTI of Wilks' formula (3.11) for standard two-sided Wilks' TI constructed with conjugate order statistics of the variable. Given (P, C) and r, there is a minimum N fulfilling (3.34). This is the minimum sample size needed to establish the equal-tailed interval. It is evident that inequality (3.34) is far more complex than (3.11).

The simplest case is when r=1. Then (3.34) reduces to:

$$1 - I_{\frac{1+P}{2}}(N,1)\left[2 - \left(\frac{1+P}{2}\right)^N I_{\frac{1+P}{2}}(N,1)\right] \geq C \quad (3.35)$$

As an example, the minimum sample sizes for tolerance level (95, 95) and different choices of r (for conjugate OS), are given in Table 3.3, comparing the case of equal-tailed and standard two-sided Wilks' tolerance intervals.

In Table 3.3 it is observed that:

1. The minimum sample size needed to construct tolerance intervals using conjugate order statistics is higher for equal-tailed TI than for standard Wilks' two-sided TI.
2. In both cases, the minimum sample size grows with increasing rank of the OS used. The difference of the two sizes is also growing with the rank, while their quotient is decreasing.

It seems clear that, when N grows, both types of intervals tend to the same limit: the fixed interval encompassed by the quantiles of order (1-P)/2 and (1+P)/2. Therefore, it is not surprising that the quotient of the minimum sample sizes for the two schemes tends to 1 when N tends to infinity.

Table 3.3. Order statistics and minimum sample sizes for equal-tailed and standard Wilks' two-sided 95/95 tolerance intervals, using conjugate OS as endpoints. Taken from [70].

Rank r	N_{min} for equal-tailed 95/95 tolerance intervals	OS for equal-tailed 95/95 tolerance intervals	N_{min} for Std Wilks' two-sided 95/95 tolerance intervals	OS for Std Wilks' two-sided 95/95 tolerance intervals	Ratio (N_{min} ET)/(N_{min} STD)
1	146	(1, 146)	93	(1, 93)	1.5699
2	221	(2, 220)	153	(2, 152)	1.4444
3	287	(3, 285)	208	(3, 206)	1.3798
4	348	(4, 345)	260	(4, 257)	1.3385
5	407	(5, 403)	311	(5, 307)	1.3086
6	464	(6, 459)	361	(6, 356)	1.2853
7	519	(7, 513)	410	(7, 404)	1.2659
8	574	(8, 567)	458	(8, 451)	1.2533
9	627	(9, 619)	506	(9, 498)	1.2391
10	680	(10, 671)	554	(10, 545)	1.2274
11	732	(11, 722)	601	(11, 591)	1.2180

As stated in section 1.12, a sufficient condition for [L, U] being an ETTI with level (P, C) is that L (resp. U) is a lower (resp. upper) tolerance limit of level ((1+P)/2, (1+C)/2). The minimum sample sizes needed are obtained from one-sided Wilks' equation. In [125] it is shown that such minimum sample sizes, for ranks from 1 to 11 and P=C=0.95 are equal to those shown in Table 3.3.

In [125] it is also proved that a necessary condition that [L, U] is an ETTI with level (P, C) is that U is an UTL and L is a LTL, both with level ((1+P)/2, $C^{1/2}$). When C is near 1, this necessary condition and the previous sufficient condition tend to be the same.

The ETTIs constructed with conjugate order statistics can also be regarded as standard two-sided tolerance intervals with a tolerance level obtained from the two-sided Wilks' formula, given the rank and the sample size. For a fixed confidence level C, the value of P that solves the Wilks' formula is directly obtained from (3.12):

$$P_{ST} = {}_{1-C}beta(s - r, \quad N - s + r + 1) \tag{3.36}$$

The results of (3.34) for the different ETTI intervals are given in table 3.4.

All these intervals have level (95, 95) when regarded as ETTI. But, when regarded as standard two-sided tolerance intervals, with a confidence level of 0.95, they have coverage levels higher than 0.95 and decreasing with growing rank. The maximum coverage level is around 0.968 and corresponds to 1st rank.

Table 3.4. Coverage level of ETTI regarded as Wilks' two-sided tolerance intervals. Taken from [70].

Rank r	N_{min} for ETTI 95/95	OS for ETTI 95/95	Coverage level from (3.34) with C=0.95
1	146	(1, 146)	0.9679
2	221	(2, 220)	0.9653
3	287	(3, 285)	0.9637
4	348	(4, 345)	0.9626
5	407	(5, 403)	0.9617
6	464	(6, 459)	0.9611
7	519	(7, 513)	0.9605
8	574	(8, 567)	0.9600
9	627	(9, 619)	0.9596
10	680	(10, 671)	0.9593
11	732	(11, 722)	0.9589

3.3.1.1.7. Other Q-methods

Last point quoted in section 3.3.1.1.5 could also be regarded as a weakness of Wilks' method, meaning that it is less efficient than other methods. More efficient methods enable a better knowledge of the P_0-quantile of a safety quantity, which is also an advantage from the regulator's standpoint.

In fact, there are many procedures of quantile estimation alternative to Wilks' method. A large number of them has been studied and applied in simulation of accident scenarios in nuclear plants. But licensing methodologies and calculations are mostly based on Wilks´ method.

In the remainder of this section we will mention, and briefly describe, alternative methods of quantile estimation. We will distinguish nonparametric and parametric methods. Rather than trying an exhaustive enumeration and description we will mention those methods that could be more interesting in the field of nuclear BEPU calculations.

NONPARAMETRIC Q-METHODS

Aside from Wilks' method, there are other quantile estimation methods based on order statistics. We will focus on those producing unbiased and consistent quantile estimators. Consistency is the property of convergence in probability to the true value of the parameter as the sample size tends to infinity. From the distribution of the estimator, confidence limits are derived, in general making approximations. In contrast, we recall that Wilks' method is based on a direct and exact calculation of the confidence limit, as a biased quantile estimate.

There are many and diverse quantile estimators [18, 89, 114]. From them, confidence limits of the quantiles can be obtained.

Assuming that Y is a continuous random variable with continuous and strictly increasing CDF, the inverse function of the CDF exists (and it is also continuous and strictly increasing). It is termed the *quantile function* of Y. The quantile of order λ ($0<\lambda<1$) of Y is defined as

$$Y_\lambda = F_Y^{-1}(\lambda) = inf\{y : F_Y(y) \geq \lambda\} \tag{3.37}$$

Quantile estimators use the values of a random sample of the variable to produce an estimate of the quantile. A quantile estimator can be obtained by building up an approximate CDF of the variable, and using its inverse:

$$\hat{Y}_\lambda = \hat{F}_Y^{-1}(\lambda) \tag{3.38}$$

(The circumflex denotes an estimate).

The simplest estimate of the CDF, constructed from a SRS of Y, S=($Y_{1,...}$,Y_N), is the so-called *empirical CDF (ECDF)*, defined as:

$$F_{Y,n}(y) \equiv \frac{1}{N}\sum_{i=1}^{N} 1(Y_i \leq y) \tag{3.39}$$

(3.39) is a stepwise and right-continuous function. (An alternative, left-continuous definition of the ECDF is obtained replacing in (3.39) the sign ≤ by <)

From the ECDF, the *plug-in estimates* of quantiles are obtained:

$$\widehat{Z_\lambda} = inf\{z \in \mathbb{R} : F_{Z,n}(z) \geq \lambda\} \tag{3.40}$$

Since (3.39) is an stepwise function, the plug-in quantile estimators are order statistics.

$$\widehat{Y_\lambda} = Y_{\lceil n\lambda \rceil:n} \tag{3.41}$$

where the partial brackets ⌈ ⌉ denote the ceiling function (integer value plus one). There is not an undisputed or standard definition of *sample quantile*; rather, there are multiple definitions [45]. The simplest are single order statistics; others are simple interpolations i.e. convex combinations of two consecutive (adjacent) OS. Other estimators take advantage of the fact that $F(Y_{k:N})$ follows a beta distribution i.e. they have the same basis than Wilks' method. We next make a brief description of several of these methods.

1. Interpolation methods

Rather than the empirical CDF (3.39) and its inverse function (i.e. quantile function), which are stepwise functions, continuous estimates of the PDF can be used in the construction of quantile estimators. There are smoothing techniques [114] to do this. Local smoothing produce quantile estimators which are a convex combination (i.e. weighted mean) of two consecutive OS.

These confidence intervals for quantiles are interpolations between intervals based on OS. They produce more efficient (i.e. less conservative) intervals than the intervals based on OS. The basis of these methods is the fact that a Wilks' interval with level (P, C) has not, in general, a coverage exactly equal to P with a confidence exactly equal to C.

Beran and Hall (1993) [4] proposed to obtain more efficient tolerance intervals as simple linear interpolations between adjacent exact confidence intervals (CI). The interpolation is done in terms of the statistical confidence and reduces the coverage error to orders N^{-1}, while for classical nonparametric confidence intervals of quantiles the coverage errors are of order $N^{-½}$ for both one-sided and two-sided

intervals. The authors stated that simple linear interpolation always errs on the conservative side for CI of a quantile (i.e. intervals tend to overcover slightly). These results are valid for any smooth distribution of the observations.

Another method of interpolation of Wilks' intervals is due to Hutson (1999) [44]. It calculates nonparametric confidence intervals for quantiles using "fractional order statistics" (FOS), which is a generalization by Stigler (1977) [120] of the concept of OS. The U^{th} quantile of X is estimated as the fractional OS $X_{(N+1)U:N,}$ which, in general, cannot be calculated from the sample (because (N+1)U is, in general, a non-integer value). The FOS of a uniform distribution have beta distribution, just like the OS. Stigler [120] proved that, under moderate regularity conditions, the distribution of the FOS is well approximated by the distribution of linear combinations of OS. Specifically, the distribution of the FOS $Y_{(N+1)U:N,}$ is well approximated by the distribution of

$$(1-\varepsilon)\, Y_{\lfloor (N+1)U:N \rfloor} + \varepsilon\, Y_{\lfloor (N+1)U+1:N \rfloor} \tag{3.42}$$

when $0<\varepsilon<1$ and agrees in distribution when $\varepsilon=0$, where $\varepsilon=(N+1)U - \lfloor (N+1)U \rfloor$. $\lfloor\ \rfloor$ is the floor function (i.e. the "integer part"). Hutson proved that the discrepancy between both estimates of the quantile is $O(N^{-2})$, and it is zero when $\varepsilon=0$. Therefore, Hutson method makes a simple interpolation based on the coverage, while the interpolation of Beran & Hall is based on the statistical confidence.

More efficient estimators may be used in place of the linear interpolation estimator e.g. kernel estimators or quasi-quantiles.

An exact nonparametric upper confidence limit for the U^{th} quantile of Y is the FOS $Y_{(N+1)U2:N}$ such that:

$$PR\{Y_U \leq Y_{(N+1)U2:N}\} = 1 - I_U[(N+1)U2, (N+1)(1-U2)] \tag{3.43}$$

If a confidence level C is desired, the value of U2 is determined numerically by solving

$$I_U[(N+1)U2, (N+1)(1-U2)] = C \tag{3.44}$$

An approximate solution to (3.44) is

$$U2 = beta[(N+1)U, (N+1)(1-U)]_C \tag{3.45}$$

Young and Mathew (2014) proposed linear extrapolation of Wilks' intervals to obtain both one-sided and two-sided nonparametric tolerance intervals, when the sample size is below the minimum sample size derived from Wilks' formulas [111].

Reference [86] includes the application of Beran-Hall and Hutson methods to a BEPU LOCA/ECCS calculation in a PWR reactor.

2. Estimators based on smooth quantile function. L-estimates

Another option is a global smoothing, based on the construction of an empirical smooth quantile function (i.e. continuous and strictly growing). This procedure gives rise to L-estimators of quantile, which are convex combinations of all the sample OS. A specific type of L-estimators are kernel estimators, using a PDF as weight function. Contrary to the sample quantiles, L-estimators are not, in general, asymptotically normal. In return, they may be more efficient.

A basic shortcoming of L-estimators is that they are not equivariant under monotonic transformations of data, while the class of continuous and strictly increasing CDF is closed under such transformations. For this reason, the use of L-estimators is not recommended by some authors. In [115] it is stated that single OS (that do fulfil the mentioned property of equivariance) are the only "reasonable" estimators of quantiles.

3. Asymptotic estimators

When the size of the random sample tends to infinity, the distribution of the OS that are not extremes (i.e. minimum or maximum) tends to a normal one.

An estimator approximately unbiased and consistent of the quantile of order λ of Y is the OS:

$$Y_{[N\lambda]:N} \tag{3.46}$$

If the derivative of CDF (i.e. the PDF, f_Y) of Y exists, and $f_Y(Y_\lambda)>0$, it can be proved that its asymptotic distribution, as a function on N, is normal.

$$N\left(Y_\lambda\ ,\ \frac{\lambda(1-\lambda)}{N[f_Y(Y_\lambda)]^2}\right) \tag{3.47}$$

Therefore, the sample quantile (3.46) is asymptotically normal and unbiased, but it is not distribution-free, in the sense that its variance depends on the parent PDF. The asymptotic normality derives from the fact that the estimator admits a so-called Bahadur representation [3] and thus it follows the central limit theorem [34, 88].

The asymptotic normal distribution enables the calculation of approximate confidence intervals for quantiles centred on the sample quantiles. The problem is that the variance includes the value of the PDF on the exact quantile, obviously unknown. In [34] there is an asymptotic approach based on

order statistics and on the distribution (3.47) that does not need the knowledge of $f_Y(Y_\lambda)>0$. Serfling [88] demonstrates that the two mentioned asymptotic approaches produce confidence intervals that coincide asymptotically. Other methods have been proposed that use (3.47) and replace $f_Y(Y_\lambda)>0$ by a consistent estimator (i.e. one tending to it when $N \rightarrow \infty$).

For finite sample sizes, these asymptotically normal estimators are approximate. This can be a drawback for their possible use in licensing safety analyses.

4. Estimators based on alternative sampling procedures

The estimators hitherto described are mostly based on pure Monte Carlo analysis (i.e. simple random sampling). There are quantile estimators more efficient (i.e. with lower variance) based on other types of random sampling. Variance reduction techniques are applicable to Monte Carlo analyses [9], for instance:

- Importance sampling
- Control variates
- Controlled stratification

For some of them, the existence of a Bahadur representation has been proved, and thus the quantile estimator is asymptotically normal and unbiased [34].

In [34], this type of asymptotic estimators have been studied, and numerical experiments have proved them (in general) more efficient than Wilks' method.

5. Bootstrap estimators

To finish this section of nonparametric Q-methods, let us briefly mention bootstrap estimators [25, 116]. Existence of Bahadur representations has been proved for bootstrap quantile estimators [118]. Reference [86] describes the application of bootstrap techniques to a BEPU analysis of a PWR large break LOCA.

PARAMETRIC Q-METHODS

Parametric tolerance regions are constructed assuming that the random variable belongs to a specific parametric family [38, 91]. The method to construct the region is typically specific of the chosen family. The application of the methods requires the previous demonstration that the probability distribution of the variable truly fits the parametric family. Statistical goodness-of-fit (GoF) tests can be used with such purpose.

Once the adequacy of the parametric family is proved, the parametric methods are, in general, more efficient than the nonparametric ones. For this reason, the use of nonparametric methods must be adequately justified by providing strong evidence of the belonging to the parametric family. Typical GoF tests have as default hypothesis the belonging to the tested distribution. But, for our purpose, the default hypothesis should be the conservative one (i.e. the variable does not belong to the parametric family).

Even if the data about the random variable do not prove to belong to the specific family, the parametric method could be applied if it is proved that it produces more conservative results than nonparametric methods.

A general procedure to obtain a parametric frequentist method of quantile estimation has two steps:

1. Calculation of the maximum likelihood estimator of the unknown parameters of the parametric distribution.
2. Calculation of the quantile of the fitted distribution.

But there are also direct methods to obtain tolerance limits. About parametric tolerance regions, the best known are those based on the normal distribution [52]. The simplest ones, for scalar variables, are the so-called *normal intervals,* having the following endpoints:

- Upper endpoint (i.e. upper tolerance limit)

$$\bar{Y} + k \cdot S_Y \tag{3.48}$$

- Lower endpoint (i.e. lower tolerance limit)

$$\bar{Y} - k \cdot S_Y \tag{3.49}$$

where Y is a scalar normal variable and

$$\bar{Y} \equiv \frac{1}{N} \sum_{j=1}^{N} Y_j \qquad S_Y \equiv \sqrt{\frac{1}{N-1} \sum_{j=1}^{N} \left(Y_j - \bar{Y}\right)^2} \tag{3.50}$$

are, respectively, the mean and the standard deviation of the SRS of Y, and k are coefficients that we identify below. Once more, we distinguish between one-sided and two-sided tolerance intervals.

$$One - sided\ upper\ tolerance\ limit\ with\ level\ (\beta,\gamma): \quad \bar{Y} + k^{(1)} \cdot S_Y \tag{3.51}$$

$$One - sided\ lower\ tolerance\ limit\ with\ level\ (\beta,\gamma): \quad \bar{Y} - k^{(1)} \cdot S_Y \tag{3.52}$$

$$Two - sided\ tolerance\ interval\ with\ level\ (\beta,\gamma):$$

$$\left(\bar{Y} - k^{(2)} \cdot S_Y \ , \ \bar{Y} + k^{(2)} \cdot S_Y\right) \tag{3.53}$$

Coefficients $k^{(1)}$ and $k^{(2)}$ are termed *normal tolerance factors,* respectively one-sided and two-sided. They are functions of tolerance level and sample size [79]. For fixed level and size, two-sided factors are larger than one-sided ones.

When the sample size tends to infinity, coefficient $k^{(1)}$ tends to the quantile of order β of the normal standard distribution N(0,1); and coefficient $k^{(2)}$ tends to the quantile of order $(1+\beta)/2$ of N(0,1).

Coefficients $k^{(1)}$ are obtained exactly as quantiles of the *noncentral Student distribution* [113]:

$$k^{(1)} = \frac{1}{\sqrt{N}} \ T\left(N-1, Z_\beta \sqrt{N}\right)_\gamma \tag{3.54}$$

where (β,γ) is the tolerance level and $T(M, A)_q$ is the q^{th} quantile of the noncentral Student variable with noncentrality parameter A and M degrees of freedom. Z_β is the quantile of order β of the N(0, 1) random variable.

Coefficients $k^{(2)}$ do not have a closed expression as (3.54), but can be obtained in an exact fashion as functions of the tolerance level and the sample size.

There are random variables that can be transformed to normal variables via monotonic functions. Then, tolerance intervals for such variables are obtained by setting tolerance intervals for the normal variables and applying the inverse transformation to them.

Additionally, there are techniques to construct tolerance regions and intervals for non-normal parametric families of distributions [38, 110].

BAYESIAN Q-METHODS

Our previous discussion on Q-methods has focused on frequentist methods, that estimate the desired quantile of the safety output Y from the PD assigned to the uncertain inputs. A Bayesian perspective could be envisaged, when there is prior information about the distribution of Y, that can be combined with the distribution propagated from the inputs by means of Bayes' theorem. The prior information can come from analysis of the same or a similar scenario in the same or similar plants, or from related scenarios studied in tests in experimental facilities. Reference [82] describes a methodology implementing a Bayesian analysis of the distribution of safety outputs.

3.3.1.2. Methods based on estimation of exceedance probability (P-methods)

In the process of testing the fulfilment of regulatory acceptance criteria, P-methods are an alternative to Q-methods. P-methods are based on the estimation of the exceedance probability of regulatory limits (P_{exc}); the ultimate goal is proving that P_{exc} is below a regulatory threshold ($1\text{-}P_0$) with high enough confidence (C_0). The statistical procedure is to construct an upper confidence limit of level C_0 for P_{exc}, and then show that such confidence limit is below $1\text{-}P_0$.

Q-methods are clearly more used than P-methods in the BEPU realm. Anyway, some authors have focused the formulation of RAC in terms of exceedance (or acceptance) probabilities [103-105].

Given the continuous scaler random variable Y and the acceptance limit L, we define

$$EX(\mathrm{Y}) \equiv 1(Y > L) \tag{3.55}$$

EX is a function of the value Y, and is a discrete Boolean random variable (i.e. it can only take on two values, 0 and 1). The value of EX(Y) is 1 when y exceeds L, and 0 otherwise. Thus, EX(Y) is an indicator of Y exceeding L, and it follows a Bernoulli distribution, with parameter P_{exc}. Given a random sample of N values of Y, $(Y_{1,\ldots},Y_N)$ it can be transformed to a sample of N values of EX, $(EX(Y_1),\ldots, EX(Y_N))$, formed by zeros and ones. The number of ones in the sample is the number of exceedances, a random variable following a binomial distribution with parameters N (sample size) and P_{exc} (probability). The SRS of EX is regarded as the result of a *binomial experiment.* Each code run can be regarded as a *Bernoulli experiment*, with the result 1 when Y exceeds L and 0 otherwise.

Similarly to quantile estimation methods, there is a great variety of methods to estimate exceedance probabilities from random samples. Some methods only use the information given by the sample of EX(Y), while other use the sample of values of Y, and hence utilize more information in the inference.

Again, we can classify the methods as parametric and nonparametric, frequentist and Bayesian, and exact and approximate methods.

NONPARAMETRIC FREQUENTIST P-METHODS

Typically, these methods only need two integer values as inputs: the sample size N and the total number of exceedances in the sample, NE. In fact, these methods use the results of a binomial experiment (N trials, NE exceedances) to make an estimation of the exceedance probability. There are many different methods to estimate a so-called *binomial probability* [8].

The basis of the methods is the fact that NE follows a binomial distribution with parameters N (the sample size) and P_{exc} (the binomial probability). The probability of NE exceedances is:

$$PR(NE) = \binom{N}{NE} \; (P_{exc})^{NE} \, (1 - P_{exc})^{N-NE} \tag{3.56}$$

Given the sample size N, the right-hand side of (3.56), when regarded as function of the binomial parameter P_{exc}, is termed *likelihood*.

Among the nonparametric methods we give a special mention to the Clopper-Pearson interval [14]. It is the dual of Wilks' method (which is a Q-method). Both are:

- Nonparametric (i.e. distribution-free).
- Based on binomial distribution.
- Frequentist.
- Exact (i.e. make no approximation).
- Conservative, compared to alternative methods.

They are almost contemporary: Clopper-Pearson interval was published in 1938, while Wilks' method was published three years later (1941).

Clopper-Pearson intervals are confidence intervals of the binomial probability and can be either two-sided or one-sided. The interval endpoints are quantiles of a beta distribution. The Clopper-Pearson upper confidence limit of level C for the exceedance probability is:

$$L_{up} = beta(NE + 1, N - NE)_C \tag{3.57}$$

i.e. the quantile of order C of a beta random variable. Such quantile is not defined when N=NE, and we set L_{up}=1. In that case, the acceptance criterion is not met, no matter the value of N.

When NE=0 (i.e. zero exceedances in the sample), (3.57) reduces to

$$L_{up} = 1 - \sqrt[N]{1 - C} \tag{3.58}$$

Using (3.57), the regulatory acceptance criterion takes the form:

$$beta(NE + 1, N - NE)_C < 1 - P \tag{3.59}$$

provided that (P, C) is not less than the regulatory tolerance level. Criterion (3.59) coincides with one-sided Wilks' formula when the order statistic with ordinal N-NE is used as an upper tolerance limit of Y with level (P, C) and the upper regulatory limit is L. When NE=0, (3.58) holds and the lowest N fulfilling (3.59) to level (95, 95) is 59, a well-known figure in Wilks' theory.

This result is not surprising, in light of the duality of Clopper-Pearson and Wilks' methods. If the maximum of an SRS of size 59 is below the regulatory limit L, we conclude that the probability of exceeding such limit is below 0.05 with a 95% confidence.

There are many other nonparametric methods of estimation of a binomial probability [8], all of them having as input the numbers N and NE. Among these methods, Clopper-Pearson is qualified as conservative [8], similarly to Wilks' among the Q-methods. There are indeed P-methods more efficient than the Clopper-Pearson interval.

NONPARAMETRIC BAYESIAN P-METHODS

Nonparametric Bayesian P-methods regard the exceedance probability as a random variable. The information about P_{exc} prior to the knowledge of the sample is expressed as a prior probability distribution (simply termed *prior*). The posterior probability distribution (termed *posterior*) of P_{exc} is obtained by combining the prior and the information of the sample, using Bayes' theorem. Specifically, the posterior is the normalized product of the prior and the binomial likelihood (3.56).

Priors for P_{exc} can be either discrete or continuous, producing respectively discrete and continuous posteriors. A special case of continuous prior is the beta distribution. It is easily proved that, when the prior has beta distribution, the posterior is beta too; for this reason, it is said that beta and binomial distributions are *conjugate* [1]. Assuming that there are NE exceedances in a sample of

N, if the prior is beta(d,e), d and e being real numbers, then the posterior has the distribution beta(d+NE, e+N-NE) [1].

The frequentist confidence intervals are replaced, in Bayesian Statistics, by the so-called *credible intervals*, which contain, with a specified probability, the estimated parameter. Credible intervals (either one-sided or two-sided) are easily obtained from the posterior distribution.

For the exceedance probability, the upper credible limit with level C is

$$L_{up} = beta(d + NE,\ e + N - NE)_C \tag{3.60}$$

Hence the RAC takes the form

$$beta(d + NE,\ e + N - NE)_C < 1 - P \tag{3.61}$$

provided that P and C are not less than the regulatory levels P_0 and C_0.

Expressions (3.61) and (3.60) make evident that the first (resp. second) parameter of the beta distribution can be regarded as a "number of exceedances" (resp. non-exceedances) of the regulatory limit. The mean of the beta(d, e) variable is d/(d+e) and hence we can interpret the prior beta(d,e) as conveying the information of "d exceedances in a binomial experiment of d+e trials", even when d and/or e are non-integer values. And the distribution is updated (via Bayes' theorem) by adding NE to the number of exceedances and N to the number of trials. The prior distribution is based on information independent of the binomial experiment, and such information can include the results of previous experiments, or simply expert opinion.

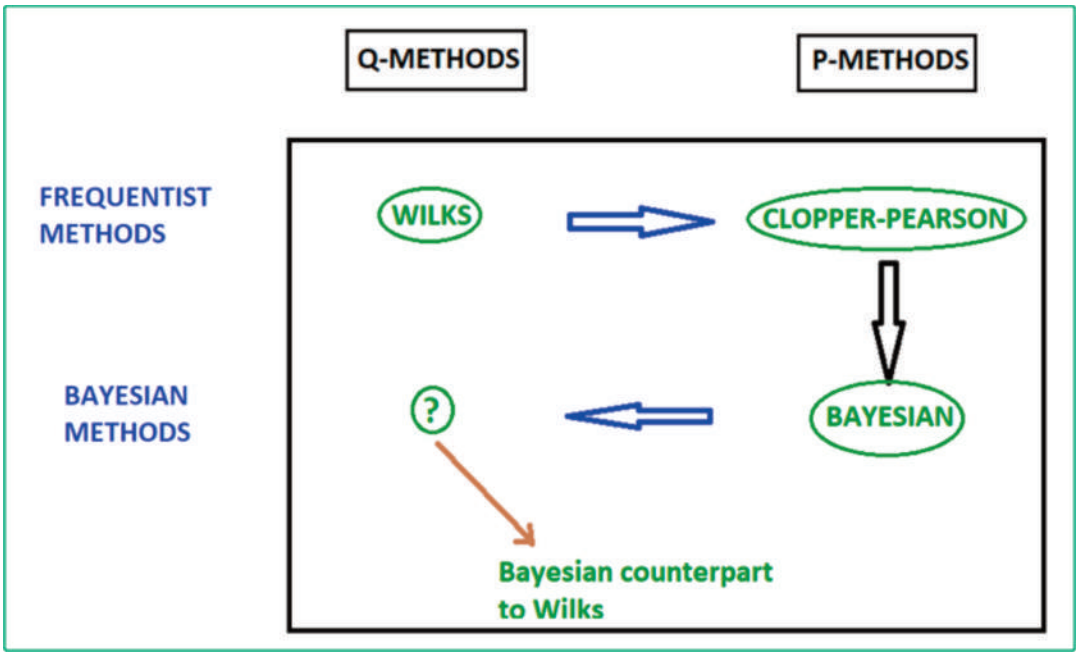
A field of Science and Engineering that is focused on the calculation of exceedance probabilities is Reliability. Failure probabilities of a system (equipment, component, structure…) are probabilities that the load on the system exceeds the system capacity [26]. Clopper-Pearson interval and Bayesian methods can be applied in Reliability [1]. Bayesian methods for calculation of system failure probability use priors derived from generic databases (i.e. data of failures of equal or similar systems). It is widely admitted the combination of such generic data with specific data obtained for the system under study.

In the case of nuclear DSA, the prior information could consist of previous analysis for the same (or very similar) plant design, operating conditions and accident scenario. Otherwise, the inclusion of the prior information could distort the results (though a distortion in the conservative direction could be admitted). The effort of justifying the prior choice could outweigh an eventual gain in margins.

There is a clear similarity between the upper confidence limits (3.57) (Clopper-Pearson) and (3.60) (Bayesian), based on quantiles of the beta distribution. The first one is a confidence limit, the second one a credible limit (i.e. a Bayesian confidence limit). Clopper-Pearson interval does not use prior information but coincides with the Bayesian limit when d=1 and e=0. So, Clopper-Pearson gives the same result as a Bayesian method with the improper prior "beta(1,0)" (it should be remembered that the parameters of the beta distribution must real positive numbers) termed "Wallis prior" by some authors [19]. It can be regarded as the prior information "1 exceedance in 1 trial". This equivalence is a hint of the well-known conservative character of Clopper-Pearson method [66, 105].

The result (3.60) is the Bayesian counterpart to Clopper-Pearson method, which is the dual of Wilks' method. So, we can regard (3.61) as a Bayesian counterpart to Wilks' formula when the analyzed output is scalar (fig. 3.3). Given the regulatory tolerance level, the number of limit exceedances and the prior parameters d and e, there is a minimum sample size N which satisfies (3.61). Depending on the prior information (d, e), this minimum N is higher or lower than the one derived for Wilks' method (19, 105).

Figure 3.3. Defining a Bayesian counterpart to Wilks' method via P-methods.



It may happen that a detailed prior information (e.g. in the form of binomial experiment results), even of the generic type, is not available, but there is a prior point estimate P* of the exceedance probability. The so-called *constrained non-informative priors* [1], where the value of the distribution mean is the unique information, are adequate in this case. They can be approximated by beta distributions e.g. a beta(d, e) prior such that d/(d+e)=P*.

When there is no *a priori* information about the exceedance probability, *non-informative priors* may be used [1]. Some examples are:

- **Uniform U[0,1] prior**, i.e. the beta(1,1) distribution, corresponding to "one exceedance in two trials". The resulting posterior is beta(NE+1, N–NE+1).
- **Jeffreys prior, beta(½, ½)**, capturing the information "half a exceedance in one trial". The resulting posterior is beta(NE+½, N–NE+½).

For a given binomial sample, these two priors give results less conservative than Clopper-Pearson.

Other priors, more conservative than the non-informative ones e.g. assigning an exceedance probability in the interval (0.5, 1), can be applied. The more conservative the prior, the higher the minimum sample size needed [19, 105].

Other possibility is to use non-conjugate priors i.e. different to beta priors. Their drawback is that they are not updated as easily as the conjugate ones, and the posterior distribution must be obtained with more complex procedures (numerical integration, rejection sampling, etc.). A well-known non-conjugate prior for probabilities is the truncated lognormal distribution.

When the *a priori* information is scanty, a procedure for the assignation of priors is the *Maximum Entropy Principle (MEP)*. This principle was established by E.T. Jaynes in 1957, and states that the probability distribution best representing the "state of knowledge" of a given quantity is the one maximizing the information entropy, subject to constraints that represent the information about the variable. In order to obtain the Maximum Entropy (ME) distribution, a problem of "maximization with constraints" must be solved.

MEP is a general procedure for assigning probability distributions to uncertain input parameters of BEPU calculations, whenever the information about them is limited to the value of the range and/or some moments of the distribution [82, 119].

The ME distribution of a binomial probability depends on the available information (i.e. the constraints), typically in the form of ranges and moments (e.g. mean, variance, …) of the distribution. In [1, 119], ME distributions are given for different types of information.

Concerning the prior updating, we should distinguish two concepts:

- When the new information is the result of a binomial experiment, the prior is updated (as we know) through Bayes' theorem producing the posterior.

- When the new information is only about prior distribution ranges or moments, the prior is updated to produce a "better" prior. For ME priors there is a specific update procedure: the so-called *Maximum Relative Entropy Principle* [119].

OTHER NONPARAMETRIC P-METHODS

Other nonparametric P-methods are based on the same procedure than many Q-methods: the construction of an estimate of the probability distribution of the safety quantity Y (typically as an estimated PDF). Rather than obtaining a quantile estimate (as Q-methods), some P-methods integrate an estimated density from L to the upper bound of the Y range, thus producing confidence intervals of the exceedance probability of Y. Anyway, it is important to recall that the regulatory limit L is commonly located in the upper tail of the probability distribution of the safety quantity, rather than on the central part. Thus, a good estimate of the exceedance probability requires an adequate estimate of the upper tail of the PDF. And this is not a very easy task when SRS is used, because most points of the sample of Y are located in the central range defined by the PDF.

Focusing on nonparametric or, at least, semiparametric estimates [121], we have kernel estimators of PDF, which are continuous functions of the form

$$\hat{f}(y) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{y - Y_i}{h}\right) \tag{3.62}$$

In (3.62) K is a function, named *kernel*, which integrates to 1 in its domain; usually, a parametric PDF is chosen as kernel (e.g. the standard normal PDF). So, (3.62) can be regarded as a semiparametric (rather than nonparametric) estimate of the PDF. h is termed *smoothing parameter*, and is related to the dispersion of the kernel. The tails of the distribution (3.62) are characterized by the choice of K. For instance, if the kernel is normal, the tails are normal too.

PARAMETRIC P-METHODS

Parametric methods assume that the random variable Y belongs to a specific parametric family. As previously stated, if the variable follows the parametric distribution, the parametric method typically produces results less conservative than nonparametric ones. Therefore, goodness-of-fit (GoF) tests should be applied, and the default hypothesis should be that Y does not follow the distribution. Nonparametric methods should be applied unless strong evidence exists that Y follows the parametric distribution. In [38] it is pointed out that the confidence limits of exceedance probabilities, when they are small, are not robust with respect to wrong distributional hypotheses.

Sometimes there are reasons to believe that Y follows a specific parametric distribution. In other cases, there is no hint about it, and a possibility is testing the compatibility of the data with different parametric distributions. When a distribution is found that fits adequately the data, the distribution parameters are estimated, and the PDF is integrated to find the limit exceedance probability. Since we need an upper confidence limit for P_{exc}, confidence intervals must be assigned to the distribution parameters. This can be done using the probability distributions of the parameter estimators or applying other statistical methods (e.g. bootstrap).

Reference [117] is an example of application of these techniques to the BEPU study of a PWR LOCA.

For some parametric families, there are direct methods to calculate confidence intervals of the exceedance probabilities, based on their dual methods of tolerance intervals.

When the scalar variable Y follows a normal distribution, there is a direct procedure based on the method of normal tolerance intervals, described in 3.3.1.1.7. The procedure consists in setting

$$L = \bar{Y} + k^{(1)} \cdot S_Y \tag{3.63}$$

i.e. the (β,γ)-upper tolerance limit for Y in (3.51) is equated to the regulatory limit. Recalling that the coefficient $k^{(1)}$ is obtained in (3.54) from the noncentral Student distribution, we have

$$\frac{L-\bar{Y}}{S_Y} = \frac{1}{\sqrt{N}}\ T\left(N-1, Z_\beta\sqrt{N}\right)_\gamma \tag{3.64}$$

An upper γ-confidence limit (with $\gamma \geq C_0$) of the exceedance probability is obtained as one minus the β value that solves equation (3.64). It is worth to notice that such confidence limit depends on the ratio

$$RI \equiv \frac{L-\bar{Y}}{S_Y} \tag{3.65}$$

which, in Reliability, is termed the *reliability index*. It is a normalized distance between the safety variable Y (regarded as a load) and the regulatory limit L (regarded as a resistance). The higher RI, the lower the exceedance probability, as it seems logical. This dependency is consistent with the

insights exposed in chapter 4, where we discuss the use of the acceptance probability (i.e. one minus the exceedance probability) as a safety margin, interpreted as a distance from the variable Y to the limit L.

The described method is also applicable when the safety quantity is not normally distributed, but can be transformed to a normal variable G through a growing monotonic function h:

$$G \equiv h(Y) \tag{3.66}$$

In that case, the exceedance probability is preserved by the transformation, and

$$PR_Y\{Y \geq L\} = PR_Z\{G \geq h(L)\} \tag{3.67}$$

and the previous method for normal variables is used to establish an upper confidence limit for the exceedance probability of G, which coincides with the exceedance probability of Y.

3.3.2. The multidimensional case

In this section, we address the problem of testing the fulfilment of regulatory acceptance criteria when they are multidimensional i.e. when there is a multiple criterion that is the conjunction of several scalar criteria. In this case, there is a multidimensional safety variable, whose components are the scalar safety quantities involved in the individual criteria. The scalar components are typically different scalar quantities (even with different physical units). In this sense, making the magnitudes nondimensional is a fundamental step.

There are two ways of proving the fulfilment of a multidimensional RAC. A possibility is proving the scalar individual criteria, but using an increased tolerance level. The other way is dealing directly with the multidimensional criterion.

3.3.2.1. Proving scalar individual criteria to prove the joint criterion

Scalar safety quantities involved in the RAC may be mutually dependent. We should distinguish two types of dependency:

- *Intrinsic dependency*, when there is a causal relation between the variables, or they are effects of a common cause. In both cases, it is possible to find correlations in the occurrence of the different variables.

- *Sampling dependency:* effective dependency ultimately is related to the sampling procedure. Values independently sampled (e.g. via simple random sampling) are statistically independent, even if they come from intrinsically dependent variables (and even if they come from the same random variable).

For instance, two safety quantities which are outputs of the same code calculation may have, in principle, some dependency, because they are produced by the same values of input variables. The strength of the relation will depend on the proximity of both variables; if there is a causal relation, the dependency can be very strong. Similar relations may arise when two quantities are outputs of different code calculations but being derived from inputs corresponding to the same plant state. But, ultimately, statistical dependency is a property of the sampling process. Values of two variables having intrinsic dependency are independent if they have been sampled independently.

Therefore, if two scalar components are outputs from the same calculations, or from different calculations derived from a common state plant, they will exhibit their intrinsic dependency. If they are output from different calculations with input decks which are statistically independent, they are statistically independent as well.

We must distinguish between the joint (multidimensional) acceptance criteria and the individual (scalar) criteria composing it. Furthermore, we must distinguish between the tolerance level of the joint RAC and the tolerance level of each individual criterion.

Let us suppose that there are D individual criteria, each one with its own tolerance level:

$$PR_{S_i}\left\{PR_{Y_i}\left\{Y_i \in R_{A,i}\right\} \geq P_i\right\} \geq C_i \quad , \quad i = 1, \ldots, D \tag{3.68}$$

Then (3.68) is a sufficient condition for the joint criterion:

$$PR\left\{PR\left\{\bigwedge_{i=1}^{D}\left(Y_i \in R_{A,i}\right)\right\} \geq P_0\right\} \geq C_0 \tag{3.69}$$

whenever the following inequalities hold

$$\sum_{i=1}^{D} P_i - D + 1 \geq P_0 \quad and \quad \sum_{i=1}^{D} C_i - D + 1 \geq C_0 \tag{3.70}$$

The proof of this statement is given in annex 2.

Conditions (3.70) can be rewritten as

$$\begin{aligned} \bar{P} &\geq 1 - \frac{1 - P_0}{D} \\ \bar{C} &\geq 1 - \frac{1 - C_0}{D} \end{aligned} \tag{3.71}$$

which are conditions on the arithmetic mean of the probability and confidence individual levels. The lower bounds in the right-hand side of (3.71) are higher than P_0 and C_0, respectively, and they increase when the dimension D grows.

If a homogeneous individual tolerance level is assumed i.e.

$$\begin{aligned} P_i &= P \quad i = 1, \dots, D \\ C_i &= C \quad i = 1, \dots, D \end{aligned} \tag{3.72}$$

conditions (3.71) become

$$\begin{aligned} P &\geq 1 - \frac{1 - P_0}{D} \\ C &\geq 1 - \frac{1 - C_0}{D} \end{aligned} \tag{3.73}$$

It is clear that to ensure the fulfilment of a joint criterion to a specific tolerance level, a sufficient condition is the satisfaction of each individual criterion to an increased tolerance level, which is a growing function of the problem's dimension.

Suppose that Wilks' method is used to check separately the fulfilment of each individual criterion, to the minimum level given by (3.73). The *mm* sample size needed for each criterion follows from formula (3.26) applied to such minimum level i.e.

$$N_{mm} = \left\lceil \frac{log\left(\frac{1 - C_0}{D}\right)}{log\left(1 - \frac{1 - P_0}{D}\right)} \right\rceil \tag{3.74}$$

Table 3.5 and figure 3.4 show the dependency of N_{mm} (for $P_0 = C_0 = 0.95$) with respect to the dimension D.

Table 3.5. *mm* sample size for $P_0 = C_0 = 0.95$ as a function of dimension after (3.74).

Dimension (D)	Minimum sample size
1	59
2	146
3	244
4	349
5	459
6	573
7	690
8	810
9	933
10	1058

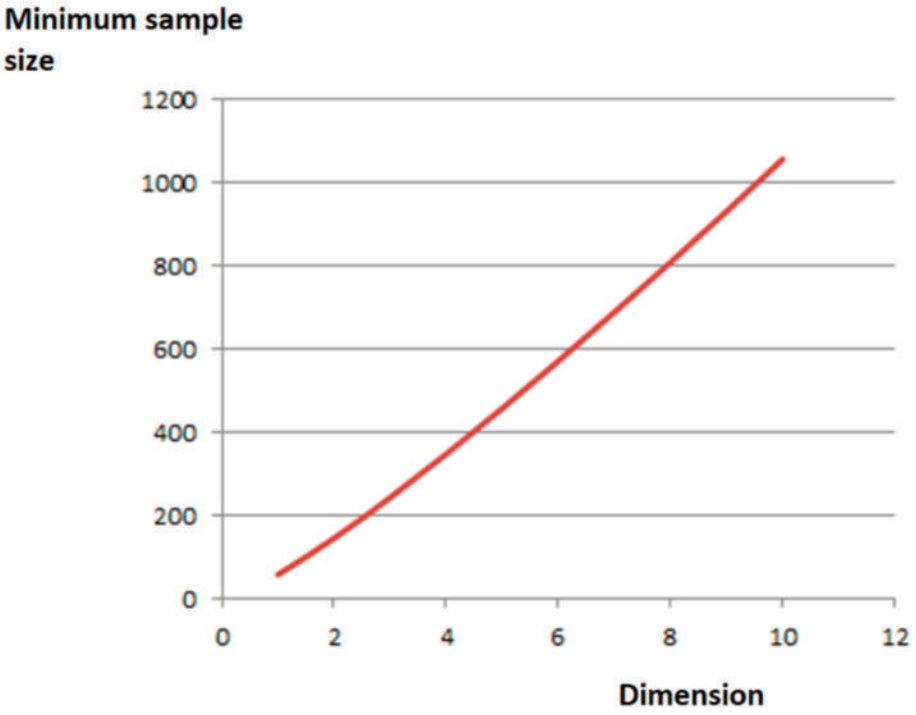
In the specific case when individual criteria are independently checked (i.e. individual quantities Y_i are modelled as independent random variables), the inner probability in (3.69) is factorized:

$$PR\left\{\prod_{i=1}^{D} PR_{V_i}\{V_i \in AR_i\} \geq P_0\right\} \geq C_0 \tag{3.75}$$

Thus it is easy to prove that (3.68) implies (3.69) when the following inequalities hold:

$$\begin{aligned} \prod_{i=1}^{D} P_i &\geq P_0 \\ \prod_{i=1}^{D} C_i &\geq C_0 \end{aligned} \tag{3.76}$$

Figure 3.4. *mm* sample size for $P_0=C_0=0.95$ as a function of dimension. Taken from [64].



If the individual tolerance level is homogeneous (i.e. the same for all i=1,…,D), and equal to (P, C), (3.76) transforms to:

$$P \geq \sqrt[D]{P_0}$$
$$C \geq \sqrt[D]{C_0} \tag{3.77}$$

When P_0 and C_0 tend to 1, the lower bounds in (3.77) tend to those in (3.73).

In this case the computational effort is a growing function of the problem's dimension D, contrary to the case when Y is "directly computable".

We conclude that a way to prove the fulfillment of a multidimensional criterion is to prove all the scalar criteria composing it, to an increased tolerance level. This procedure may be applied when each individual scalar criterion is tested with a different calculation, and even with different codes and by different analysts.

3.3.2.2. Proving the multidimensional criterion

In some cases, all the individual safety quantities are output of the same calculation. Then, the complete multidimensional quantity is calculated, and the dependencies among the variables are implicitly taken into account.

Similarly to the scalar case, there are two group of methods to prove the fulfillment of the multidimensional criterion:

- Methods based on tolerance regions, that we term Q-methods.
- Methods, based on the estimation of the regulatory limit exceedance probability, that we term P-methods.

MULTIDIMENSIONAL Q-METHODS

These procedures are based on the *method of tolerance regions*, grounded on the following result: given a simple random sample of Y, if we obtain from it a tolerance region TR with level (P_0 , C_0) or higher, and TR is contained into the regulatory acceptance region of Y, then the fulfillment of the multidimensional RAC to the level (P_0 , C_0) is proved.

How can the tolerance region be constructed? Wilks' method is valid only for scalar continuous random variables, so it cannot, in principle, be applied to multidimensional variables. But there are two possibilities to solve this problem:

1. Making a variable change from the multidimensional variable to a scalar variable, to which Wilks' method can be applied.
2. Using the Wald extension to Wilks' method, which enables the construction of nonparametric multidimensional tolerance regions based on order statistics.

WILKS' METHOD APPLIED TO THE MULTIDIMENSIONAL CASE

Wilks' method was devised for scalar random variables. But it can also be used to construct tolerance regions for multidimensional random variables, as we next describe. The key point is making a variable change. A condition to apply the procedure is the availability of a SRS of the multidimensional random variable. In the case of safety quantities, such condition is feasible when all the scalars components are outputs of the same calculation. A second condition is the knowledge of a value in the range of the multidimensional quantity, independent of the SRS.

Suppose a multidimensional continuous random variable with D components, $\mathbf{Y}=(Y_{1,...},Y_D)$ with a continuous CDF $F_{\mathbf{Y}}$. Without any loss of generality, we can assume that the components of Y are positive quantities.

We define the quotients

$$\delta_i \equiv \frac{Y_i}{Y_{i0}} \qquad i = 1, \dots D \tag{3.78}$$

Y_{i0} is a value on the range of Y_i, known independently of the SRS of Y_i. It can be, for instance, a nominal value, obtained from a calculation performed with nominal or central values of the input parameters. Even if a nominal value is not available, we do know a value in the range of Y_i: the acceptance limit L_i.

The possible heterogeneity of $\mathbf{Y}$, with components of different physical units, justifies the nondimensional character of δ_i as defined in (3.78).

We now define the variable

$$W \equiv max_{i=1,\dots,D}(\delta_i) \tag{3.79}$$

It is easy to prove the equivalence:

$$W < W_0 \;\Leftrightarrow \begin{Bmatrix} Y_1 < Y_{10}W_0 \\ \dots\dots\dots \\ Y_D < Y_{D0}W_0 \end{Bmatrix} \tag{3.80}$$

where W_0 is any value on the range of W. As defined in (3.79), W is a continuous scalar variable, with CDF

$$F_W(w) = F_{\mathbf{Y}}\left(wY_{10}, \dots, wY_{D0}\right) \tag{3.81}$$

Because the CDF of $\mathbf{Y}$ is continuous, F_W is also continuous, and Wilks' method can be applied to W. Let W_U be an upper tolerance limit of W with level (P, C). Following (3.80), the equivalence

$$W < W_U \;\Leftrightarrow \begin{Bmatrix} Y_1 < Y_{10}W_U \\ \dots\dots\dots \\ Y_D < Y_{D0}W_U \end{Bmatrix} \tag{3.82}$$

holds. If the left-hand side of (3.82) is fulfilled at a level (P, C), so is the right-hand side, which thus defines a tolerance region of level (P, C) for $\mathbf{Y}$.

Therefore, it is enough to use one-sided Wilks' method to W for setting an UTL, and then to apply the equivalence (3.82). To set the multidimensional tolerance region, the sample size needed is obtained from one-sided Wilks' formula. If Wilks' UTL is an order statistic $W_{r:N}$, the right-hand side of (3.82), with $W_U=W_{r:N}$, defines a D-dimensional TR for **Y**, with level (P, C).

For the standard tolerance level (95, 95), a sample size of 59 is enough to set a D-dimensional TR (no matter how high D is). There is no dependency of the computational effort on the dimension D of the problem.

The described method requires the prior knowledge of D pivotal quantities i.e. $Y_{10},\ldots,Y_{D0}$. For the method being efficient, the quotients (3.78) should have similar ranges. If one of the quotients is statistically greater than the others, it determines the tolerance region. And quotients statistically lower than the others do not intervene in the setting of the TR.

In contrast with the advantage of a lower sample size, the method described is more conservative (i.e. less efficient) than Wald extension (which we present in the next section). In other words, Wald method produces in general tolerance regions with lower expected size than those of Wilks' method applied to (3.82). But, if the pivotal quantities are centred values on the corresponding probability distributions, the efficiency of the proposed method can be significant.

The multidimensional problem is not limited to the case of multiple scalar safety outputs. It also raises when we want to set tolerance bands on the time evolution of a given scalar magnitude (i.e. time curves). Due to the time discretization performed by computational codes, a time curve can be regarded as a multidimensional variable (with very high dimension). We have proved that, with a minimum computational cost, such tolerance bands can be set using a transformation similar to (3.79). The problem is that the tolerance region can be very large i.e. that the span between the bands can be too wide.

It is worth to mention that the transformation (3.79) and the equivalence (3.80) are applicable to procedures of setting tolerance intervals different to Wilks' method. For every method, (3.80) enables the replacement of the multidimensional problem by a scalar one.

WALD METHOD

After the publication of Wilks' theory, the basic idea (i.e. using order statistics as nonparametric tolerance limits) attracted high interest on the topic. Wilks' method was devised for tolerance intervals of scalar random variables, and so an extension of the method to multidimensional random variables was desirable. Such extension was accomplished by Abraham Wald and published in 1943 [102], only two years after Wilks' landmark paper.

Wald method allows the construction of tolerance regions for multidimensional continuous random variables with continuous CDF, using simple random samples. Given a tolerance level, the multidimensional tolerance region is constructed as a Cartesian product of intervals, one for each component of the variable. Each individual interval has order statistics (of the corresponding scalar component) as endpoints. There are two important remarks to do:

- Each individual interval is not necessarily a tolerance interval with a specified level. It is the complete region (the Cartesian product of intervals) that has the prespecified tolerance level.

- The individual intervals are constructed sequentially. To do it, the components of the multidimensional variable are arbitrarily ordered, and then the procedure is applied. For a given SRS of the multidimensional variable, the tolerance region obtained depends on the assumed ordering. If the ordering is changed, the tolerance region, in general, changes as well.

Such intervals are sequentially constructed, component by component, using OS. The tolerance level of the individual intervals decreases as the sequence advances. The procedure involves a sequential elimination of data from the simple random sample.

There is a "Wald formula" similar to Wilks' formula, relating the tolerance level, the ordinals of OS used as interval endpoints, the sample size and the dimension D of the random variable. The Wald formula also is based on the beta distribution. There is also a minimum sample size N_{min} that is a growing function of the dimension D. Therefore, there is an effect of dimensionality in Wald´s method.

The simplest version of Wald method involves setting a one-sided interval on each component. Assuming a SRS with size N of the variable $Y = (Y_{1,...},Y_D)$, the process is

- Take the sample maximum of Y_1.
- Remove from the sample the point corresponding to this maximum.
- In the remaining sample (size N-1) take the maximum of Y_2.
- Remove from the sample the point corresponding to the maximum of Y_2.
- In the remaining sample (size N-2) take the maximum of Y_3.
- ……..
- In the sample of size N-D+1, take the maximum of Y_D.

The Cartesian product of one-sided intervals is the Wald tolerance region. The corresponding Wald formula is

$$_{1-C}beta(N - D + 1, D) \geq P \tag{3.83}$$

The minimum sample size solving (3.83) is a growing function of dimension D. For a (95, 95) region, the minimum sample size is 59, 93 and 124 for D=1, 2, 3, respectively. Table 3.6 gives the minimum sample size as a function of dimension D for a (95, 95) region. A rather linear dependence is observed.

Table 3.6. Sample size as a function of dimension, from Wald formula (3.83) for (95, 95) tolerance level.

Dimension (D)	Minimum sample size
1	**59**
2	**93**
3	**124**
4	**153**
5	**181**

MULTIDIMENSIONAL P-METHODS

Similarly to the scalar case, P-methods use as figure-of-merit the probability of exceeding the regulatory limits. When the RAC is multidimensional, composed of D individual criteria

$$(\boldsymbol{Y} \in \boldsymbol{R_A}) \Leftrightarrow \bigwedge_{i=1}^{D} (Y_i \in R_{A,i}) \tag{3.84}$$

the exceedance probability is the probability that at least one of safety limits is exceeded i.e.

$$PEX(\boldsymbol{Y}) \equiv 1 - PR(\boldsymbol{Y} \in \boldsymbol{R_A}) = 1 - PR\left(\bigwedge_{i=1}^{D} (Y_i \in R_{A,i})\right) \tag{3.85}$$

The advantage of P-methods is that the information conveyed by the multidimensional random sample is captured by a single scalar magnitude, the exceedance probability. Therefore, the P-methods used for the scalar case, described in section 3.3.1.2 can also be applied to the multidimensional case.

3.3.3. Other methods

3.3.3.1. Methods based on efficient sampling

So far we have described both P-methods and Q-methods based on simple random samples of the safety quantity. These SRS are obtained through the pure or crude Monte Carlo (CMC) technique: the uncertain inputs are sampled by SRS, they generate a random set of input decks for the model or code, and the model transforms them into a SRS of results.

The SRS has a great advantage: it is the simplest procedure of random sampling and thus the properties of estimators based on it are easier to obtain than those based on more elaborated sampling procedures. But SRS has also drawbacks:

- It does not ensure a homogeneous sampling of the whole input space.

- It does not privilege any region of the input space. Sometimes, it is important to focus the sampling on a specific region of the input range, and SRS does not enable such selection.

As a result, SRS may be less efficient (i.e. producing more precise estimators) than other types of sampling.

Indeed, there are sampling procedures more efficient than SRS, namely:

- Stratified sampling
- Latin Hypercube Sampling (LHS)
- Importance sampling
- …

Let us recall that the most used and popular method in the BEPU analyses of nuclear reactors is Wilks' method, and its extension by Wald. The two methods are based on SRS and cannot be used in conjunction with other sampling procedures.

Importance sampling is useful when we want to give more weight in the sampling to a specific region of the input range.

Latin Hypercube Sampling (LHS) is a technique raised in the nuclear field [126] and enables a more efficient sampling of the input space. If LHS is used, the estimators in general have lower variance and better convergence properties.But it should not be applied when Wilks' and Wald methods are used.

3.3.3.2. Methods based on surrogate models

SURROGATE MODELS AND UNCERTAINTY PROPAGATION

When the computational model or code used in BEPU calculations is very complex and costly to run, the computational effort (in terms of time and resources) needed in the uncertainty propagation may become exceedingly expensive. A possibility to circumvent this problem is the use of surrogate models to propagate the uncertainties (fig. 3.5).

Briefly, a surrogate model is a model of a model. A way of analyzing and understanding the physical reality is making predictive models of it; similarly, a way of analyzing and understanding a complex model is making a simplified model of it. This explains the use of the term "metamodel" as a synonym of surrogate model. Other usual term is "emulator", which is a surrogate model of a "simulator" (main model) [91, 92].

The basic objective of surrogate models is to save computational cost, while keeping an adequate capability of prediction. They save computing time, but lose accuracy with respect to the original model [53]. A surrogate model is much simpler than the original model, and fast-running, so that the propagation of uncertainty through it is affordable. But it is important to remark that, in the BEPU realm, the use of metamodels should be limited to the uncertainty propagation. Predictions should be performed with the original model, because the metamodel (by definition) has a worse prediction capability.

We can summarize the desirable properties of a surrogate model, from the point of view of the uncertainty analysis of the original model.

- It must be simpler than the original model, and fast-running enough to enable an affordable uncertainty propagation.

- It must have, with respect to reality, a similar bias (or a more conservative one) than the original model.

- It must not underestimate the uncertainty propagated from inputs and models, at least in the conservative direction. Therefore, the surrogate model and the original model should show similar dependency on the most influential input parameters, as established via sensitivity or importance analysis.

Two main categories of surrogate models can be distinguished:

- *Mechanistic metamodels:* they are simplified versions of the original models. The term mechanistic means "based on first principles". Hence these models reproduce physical mechanisms driving the system. But commonly they have an empirical part, revealed e.g. in the presence of adjustable coefficients. The simplifying assumptions should not introduce any anticonservative bias with respect to the original model, nor eliminate dependencies on influential parameters.

- *Empirical metamodels,* which are non-mechanistic (i.e. not based in first principles) and simple models relating the output safety quantities with the most influential input parameters. The term "response surfaces" is sometimes used. Typical examples are polynomials. Other empirical models do not have a closed form; a good example are neural networks, that make it possible to emulate models with many input parameters.

Both types of metamodels have in their formulations some free (i.e. adjustable) parameters, which allow fitting them to data obtained from the original models. It is a similar process to the fitting (calibration) of the original model to real data. It would be possible to fit surrogate models to real data, overlooking the original model. But this is not a usual practice, because they would lose their condition of emulators of the original models.

Experimental designs are often applied in the construction of the database for developing a surrogate model. These techniques are more efficient than those based on random sampling. In this sense, the BEPU methods based on metamodels are very related to the methods rooted on design techniques.

Once developed, a surrogate model must be validated, through additional comparisons of its predictions with those of the original model. The database of validation should be different to the database of development.

The realm of surrogate models is so wide and multifarious that it is difficult to make a detailed classification. Some surveys [2, 53, 83, 92] describe different types of metamodels. Data-driven metamodels, obtained in the field of Machine-Learning, offer great possibilities.

We give here the classification proposed in [124], mentioning outstanding models in each category.

- Linear regression and regularization models: Partial Least Squares, Ridge and Elastic Net Regression, LASSO,...
- Kernel methods: Support Vector Regression, Kernel Ridge Regression...
- Additive models: General Additive Model (GAM), SDR ACOSSO, Multivariate Adaptive Splines (MARS), GAMSEL, Projection Pursuit Regression, GAMLSS...
- Tree models: Boosted Trees, Random Forests, Extreme Gradient Boosting (XGBoost)...
- Artificial Neural Networks: Radial Basis Functions (RBF), Deep Neural Networks...
- Others: Kriging, LOESS, Multi-gene Genetic Programming (MGGP), Inverse-distance, Deep Gaussian Process, Symbolic Regression,...

SURROGATE MODELS IN BEPU METHODOLOGIES

From the very beginning of Nuclear Safety, surrogate models were considered as valuable tools in the performance of calculations with predictive models. Their use in BEPU analysis was also very early.

The 1988 amendment of the "ECCS Rule" (10CFR50.46) in the United States [16], admitted for the first time the use of BEPU methodologies (based on probability) in LOCA/ECCS analysis for LWR. Following this release, the USNRC published, in 1989, the report NUREG/CR-5249 [7], where the CSAU methodology for realistic LOCA analysis was presented and described. CSAU can be considered as the pioneering BEPU methodology, and in fact most BEPU methodologies developed through the years have followed the basic structure of CSAU.

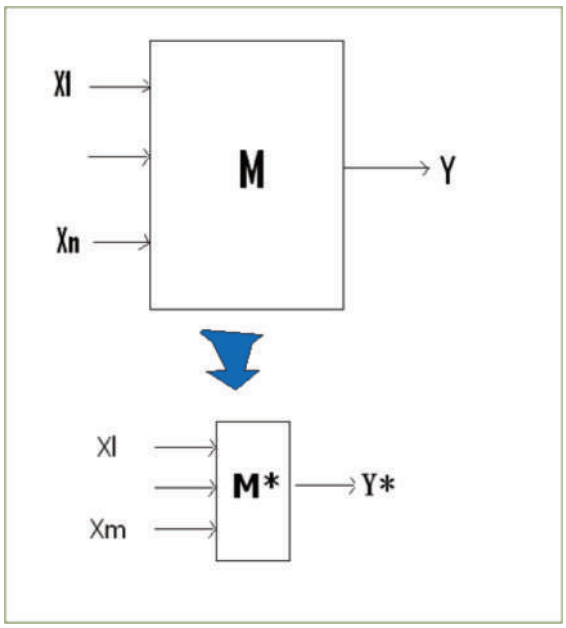
The first application of CSAU methodology was a BEPU LOCA/ECCS analysis [7], where surrogate models were used for the propagation of input uncertainty. A "response surface technique" was applied, based on the 7 most influential parameters on the LOCA peak cladding temperature (PCT). An experimental design was used in the selection of the database of code runs.

The influence of CSAU in the emerging field of BEPU analyses was so strong that the subsequent BEPU methodologies adopted the use of surrogate models, especially response surfaces. By the end of 20th century, however, methodologies based on uncertainty propagation through the original model, and construction of tolerance intervals via Wilks' method, started a most successful career, still in force. The new procedure was massively adopted in the BEPU realm, and the use of surrogate models declined. But the interest on these tools in the nuclear BEPU field has never been lost [2, 27, 30, 85, 101, 109].

BASIC FORMULATION OF METAMODELS

Let us recall the computational model schematized in expression (1.1), **X, Y** and **C** being respectively the input variables, the output variables and the calibration parameters. All of them are, in general, multidimensional magnitudes. **M** is the input-output transformation defined by the model.

Figure 3.5. Original model (M) and surrogate model (M*). Taken from [65].



A surrogate model of the model may be formally represented as

$$\boldsymbol{Y}^{*} = \boldsymbol{M}^{*}\left(\boldsymbol{X}^{*}, \boldsymbol{C}^{*}\right) \tag{3.86}$$

The surrogate model is simpler that the original model. $\mathbf{M}^{*}$ represents the input-output transformation associated to the metamodel. The input variables X^{*} are (or can be transformed to) a subset of the original input parameters (**X, C**) mentioned in section 1.3. That is, the inputs to $\mathbf{M}^{*}$ may include model parameters of the original model **M**. As a basic requirement, $\mathbf{X}^{*}$ should contain those inputs (including model parameters) which are, according to **M**, the most influential on the calculated safety magnitudes. Otherwise, the propagated uncertainty could be underestimated. Techniques of sensitivity or importance analysis [84] may help in the identification of influential inputs.

The output $\mathbf{Y}^{*}$ includes all the safety magnitudes in **Y**, or a subset of them. If **Y** is multidimensional, meaning that there are several scalar safety magnitudes involved in the RAC, several surrogate models can be derived, representing different subsets of scalar safety magnitudes. The input-output transformation produced by the metamodel may have a very simple functional form. For

instance, when the surrogate is a regression model, $\mathbf{M}^*$ may be a simple formula (e.g. a polynomial on the input variables). In other cases, the metamodel may be more complex and have no closed functional form.

In (3.86), $\mathbf{C}^*$ symbolizes the free (i.e. adjustable) metamodel parameters, which are used for fitting the surrogate model to predictions of the original model **M**. The adjustment should produce the parameters $\mathbf{C}^*$ with uncertainty, rather than single values. When the metamodel is applied for propagating uncertainty, the uncertainty of $\mathbf{C}^*$ should be included in the process.

Let us recall expression (1.2), stating that the real or true value of **Y** is the calculated value plus the code or model bias. A similar expression can be written for metamodels.

$$\boldsymbol{M}(\boldsymbol{X};\boldsymbol{C}) = \boldsymbol{M}^*(\boldsymbol{X}^*;\boldsymbol{C}^*) + \boldsymbol{D}^*(\boldsymbol{X}^*) \tag{3.87}$$

where $\boldsymbol{D}^*(\boldsymbol{X}^*)$ is the bias of the surrogate model with respect to the original model (not with respect to the real value of **Y**). Typically, the metamodel is adjusted so that the bias is small. The bias could be quantified from differences between predictions of **M** and $\mathbf{M}^*$.

The uncertainty in (3.87) is propagated from two sources:

- Inputs of the metamodel (which are the most influential inputs to the original model, including some model parameters.
- Adjustable parameters of the metamodel, which have an uncertainty derived from the fitting process.

BEPU REGULATORY ACCEPTANCE CRITERIA FOR SURROGATE MODELS

In this section we refer to the case of a scalar quantity Y, that is the most usual when surrogate models are applied.

When a BEPU methodology uses a surrogate model, instead of the original model, for the calculation of the quantity Y with an upper regulatory limit L, the BEPU regulatory criterion converts to

$$PR_{S^*}\{PR_{Y^*}\{Y^* < L\} \geq P_0\} \geq C_0 \tag{3.88}$$

In (3.88) we do not include the bias of the metamodel, assumed small. The subindex S^* represents the second-level uncertainty due to the incomplete knowledge of the probability distribution

of Y^*, typically derived from a random sample S* produced via Monte Carlo. But surrogate models have a very low computational cost, and hence the sample S* can be very large. In this case, the second-level uncertainty vanishes, and the BEPU RAC for a metamodel reduces to the simple form

$$PR_{Y^*}\{Y^* < L\} \geq P_0 \tag{3.89}$$

equivalent to

$$Y^*_{P_0} < L \tag{3.90}$$

When the safety quantity is multidimensional, there are two possibilities: developing a multidimensional metamodel, or developing a separate surrogate model for each scalar component. The insights included in section 3.3.2 are applicable to this situation.

3.4. Assigning uncertainty to the inputs

Up to now we have described with detail the methods used in the most important stages of DSA i.e.

- Propagation of uncertainties from inputs to safety outputs.
- Uncertainty analysis of the safety outputs.
- Test of the fulfilment of regulatory acceptance criteria by the safety outputs.

But the outcome of the DSA is useless if the sources of uncertainty are not adequately quantified. The uncertainty which propagates to the safety outputs stems from three sources:

- The uncertain inputs.
- The bias or inaccuracy of the code, a quantity which is imperfectly known and thus is modelled as uncertain.
- The metauncertainty (code uncentainty) due to the imperfect knowledge about the input-output function associated to the computational code, and, as a consequence, to the imperfect knowledge of the propagation. When Monte Carlo techniques are used in the propagation, this uncertainty is linked to the finiteness of the random sample.

The main source is the uncertainty of the inputs. In section 1.2 we mentioned the two classical types of uncertainty: aleatory (due to inherent variability of the variable) and epistemic (due to ignorance or lack of information of the user / observer). Aleatory uncertainty is modelled via probability

distributions, while there are multiple theories to model epistemic uncertainty. There are experts who believe that epistemic uncertainty can be modelled through probability distributions as well. Other experts state that it should not be modelled via probabilities, but using other theories like interval analysis, fuzzy logic or Dempster-Shafer theory [40].

Notwithstanding this controversy, most present BEPU methodologies use probability distributions to model both aleatory and epistemic uncertainty. Only in few cases there is a separation between the modelling of the two types of uncertainty. In some cases the uncertainty of input parameters is modelled by means of intervals. A good example of the use of intervals is the modelling of the break size and location in LOCA/ECCS analyses (see section 2.2.9). The analyst must look for the break size and location giving the worst consequences, and then perform the probabilistic BEPU analysis using such break characteristics.

We can give some rules about the assignment of uncertainty to input parameters:

1. A previous step is an importance analysis of the inputs: the input parameters being most influential on the safety outputs must be identified. The uncertainty propagated from an input to an output depends on the degree of influence and on the uncertainty of the input.
2. Uncertainty of the most influential inputs must be assessed and assigned.
3. Using the results of the two previous steps, the input parameters that contribute the most to the uncertainty of the safety outputs are selected. The cumulate uncertainty from all the discarded inputs must be estimated, to ensure that the propagated uncertainty is not underestimated.

The final goal is that the uncertainty propagated from the inputs must not be underestimated.

Dependencies among inputs should be taken into account e.g. via joint probability distributions. The metauncertainty, due to the imperfect knowledge of the uncertainties, can be included by "inflating" the input uncertainty e.g. using confidence bands for cumulative distribution functions or wider intervals.

The assignment of uncertainty to the inputs has been traditionally performed by means of experts' judgement, based on the literature and the knowledge of the system under analysis. The problem of these procedures is that it may have a strong component of subjectivity. More objective and systematic (but more complex too) techniques are based on inverse uncertainty quantification (IUQ) methods, applicable to the estimation of input uncertainties from experimental data [123].

A special category of inputs are those plant operational variables that are controlled by Technical Specifications (TS) of the plant. Safety analyses must analyze the allowed behaviour of the plant, rather than its real behaviour. It is expected that the severity of an accident should increase when any of these TS-controlled input parameters varies approaching the TS limit. It seems logical to make the calculations fixing the values of these parameters values equal to the TS limits. Anyway, if the analyst should decide to assign a DP to one of these parameters, such distribution should not intend to reproduce the values of the parameter in normal operation, but rather to assign a significant probability to the range near the TS limit [122].

3.5. Epilogue: licensing with probabilities and random numbers

We finish the chapter briefly addressing the licensing by regulators of BEPU methodologies and analyses [69].

When BEPU methodologies are used, the DSA becomes probabilistic. For a scalar safety quantity, a high-order quantile is chosen to compare with the regulatory acceptance limit. The order is also decided by the regulatory authority; the standard value is 0.95 (because 0.95 is a standard of "high probability"). The adoption of such value means that the regulator allows a limit exceedance probability less or equal than 0.05. And the adoption of a value 0.95 for the statistical confidence means that it is admitted that the figure of merit to compare with the acceptance limit be below the real quantile with a probability less or equal than 0.05.

A question is raised: is the use of a level (95, 95) adequate, or should the regulator require a higher tolerance level? In the choice of this value, the conservatisms included in the DSA are fundamental. To begin with, the DBS are established in order to maximize the severity, so that it bounds severity of any scenario inside the category. On the other hand, the regulatory acceptance limits are also established in a pessimistic fashion, to ensure a significant margin to the damage threshold of the safety barrier (see chapter 4). Furthermore, BEPU methodologies are not purely realistic, but contain a number of conservatisms. Using a previously defined term, we could say that the regulatory authority can allow a *licensing margin* not excessively high in exchange of the existence of other significant safety margins.

The sufficiency of tolerance levels is not the only regulatory problem posed by the use of BEPU methodologies. We have, additionally, the implications of using statistical methods. There are statistical figures of merit (FOM) to compare with regulatory limits. For instance, an upper tolerance limit with regulatory level, that is constructed from a random sample of the safety quantity; if the sample changes, the FOM changes. Hence this licensing value is a random number. If the obtained FOM does not fulfil the acceptance criterion, someone could try another sample, and this action would invalidate the safety analysis.

For this reason, the regulator should try to keep under control the generation of the random sample. Typically, simple random samples are obtained via pseudorandom number generators, producing deterministic sequences of numbers that emulate random numbers. Each sequence begins with a numerical seed. If the seed is repeated, the sample is repeated. In turn, the seed is chosen automatically or arbitrarily. The regulator must ensure that the choice of seed is not aimed at obtaining a favorable sample, and can do so e.g. by imposing its value, or by verifying that the value comes from an automatic process not controlled by the analyst.

IV. Safety margins in BEPU calculations

IV. SAFETY MARGINS IN BEPU CALCULATIONS

4.1. Safety margins in deterministic safety analysis

Nuclear power plants are designed by means of the deterministic safety analysis (DSA). It is based on the definition of a set of "design basis scenarios" (DBS) (a term that unifies usual terms "design basis accidents" or "design basis transients and accidents"). A DBS is defined by an initiating event and a set of additional failures of safety systems. The regulatory authority establishes the sets of DBS included in the DSA, and requires the licensees to adequately simulate them with computer models (codes). In addition, the regulator specifies the safety quantities that must be calculated in the simulation of each DBS and the acceptance criteria that the quantities must fulfil (typically in the form of acceptance limits). The calculated safety quantities are chosen as adequate numerical surrogates of the severity of the scenario.

Safety margin (SM) is a central concept in Nuclear Safety. There is no universal definition of the term (up to the knowledge of this author), and it is very common to find out that different experts give different definitions of safety margins. The concept usually refers to the clearance between values of a safety quantity (including measured and calculated values, safety limits, threshold values where damage starts…).

In Reliability, it is usual to term *safety margin* (or margin of safety) the difference between the mechanical resistance R of an item and the load L on it. If R and L have uncertainty, the difference R-L is uncertain too, with a probability distribution which is the convolution of those of R and L.

Following [64], in this document we adopt a wide definition of safety margin, as *an adequate metric in the space (range) of safety quantities.* In other words, safety margins are defined as "distances" or "metrics" (in a broad sense) between values of safety quantities. These distances correspond to the mathematical concept of *metrics*. Depending on the values that they connect, the safety margins can receive different names from different people. Three examples of SM (with a nomenclature that is not universal) would be:

- *Analysis margin*, that is the SM from the real value of a safety quantity (i.e. the value reached if a DBS would really occur in the plant) to its calculated value in the DSA. This margin is a measure of the conservativeness of the calculated quantity.
- *Licensing margin*, the SM from the calculated value of a safety quantity to the regulatory acceptance limit.
- *Barrier margin*, that is the margin from the regulatory acceptance limit to the damage threshold value (i.e. the value of the quantity marking the inception of damage to the analyzed barrier).

Safety margins can be composed or juxtaposed; the three mentioned margins are the components of a larger SM, the one from the real value of a safety quantity to a damage threshold value. The higher this "total" margin, the greater the confidence on the safe performance of the plant.

In Mathematics, a distance (*metric*) has the symmetric property: the distance from A to B is equal to the distance from B to A. But it is clear that safety margins do not have such property. In fact, we speak of the margin *from A to B*, rather than the margin *between A and B.* The reason of this asymmetry is the existence of a privileged direction in the range of safety quantities, the one of growing severity (i.e. conservatism). Rather than to ordinary metrics, margins could be adscribed to mathematical functions named *hemimetrics* [33], which do not have the mentioned symmetry property.

We will focus on one of the mentioned types of safety margins, the licensing margin, representing the distance from the calculated value of a safety quantity to the regulatory acceptance limit. In section 2.2.4 we introduced the concept of licensing margin, stating that DSA is aimed at proving that the LM has a high enough value. When a DBS is analyzed with a conservative methodology, and the scalar and continuous safety quantity Y has an upper acceptance limit L, regulation simply requires that a conservative enough value of Y is below L. In this case, the difference $L\text{-}Y_C$ is a judicious definition of the licensing margin; the higher is the difference, the higher is the certainty on the safe performance of the plant if the DBS occurs.

Things change when the DBS is analyzed with a BEPU methodology. In such case, the calculated safety quantity is uncertain, so that the difference L-Y is an uncertain quantity too. How should the distance from Y to L be defined? The form of the BEPU acceptance criterion may give us some hints.

As shown in detail in chapter 2, the probabilistic BEPU regulatory acceptance criterion for Y can be formulated in terms either of quantiles or of limit exceedance probabilities. Hence a first possibility is defining the safety margin from Y to L as a low-order quantile of L-Y. Indeed, this definition seems very "natural", but (as we shall see) has several drawbacks.

A more satisfactory definition is in terms of the limit exceedance probability, or its complement the acceptance probability, as we next expose.

4.2. Probabilistic definition of safety margin

The formulation of the RAC in terms of exceedance probability states that it must be below $1\text{-}P_0$ with a confidence not less than C_0; in other words, that the *probability of acceptance* of Y (i.e. the probability of Y being below L) must be higher than P_0 with confidence not less than C_0. This formulation suggests the definition of the licensing margin (LM) from Y to L as the probability of

acceptance of Y. This definition of safety margin as a probability has been adopted by several authors [43, 56, 60]. International working groups on safety margins in Nuclear Safety have produced a similar conclusion: the most comprehensive definition of SM is the probabilistic one [73, 74].

Such definition directly incorporates uncertainty. In the case of the quantity Y, having an upper acceptance limit L, let A and B be two uncertain values of Y. We define the safety margin (or simply the margin) from A to B as

$$SM(A, B) \equiv PR_{A,B}\{A < B\} \tag{4.1}$$

If Y is multidimensional, the definition transforms to

$$SM(\boldsymbol{A}, \boldsymbol{B}) \equiv PR_{\boldsymbol{A},\boldsymbol{B}}\{\boldsymbol{A} \prec \boldsymbol{B}\} \tag{4.2}$$

The sign $\prec$ in (4.2) means "less severe than". In chapter 2 we have proved that a multidimensional safety quantity **Y** can be transformed to a scalar quantity that is a surrogate of the severity of the transient. If the transforms of **A** and **B** are, respectively, A1 and B1, $\mathbf{A} \prec \mathbf{B}$ is defined as A1<B1. Of course, the value of (4.2) depends on how the "severity" of the multidimensional quantity **Y** is defined, and there is not a unique definition.

The probabilities in (4.1) and (4.2) refer to the combined uncertainty of **A** and **B**, that we regard as aleatory (irreducible). If the distributions of **A** and **B** are imperfectly known, the SM is a quantity with epistemic uncertainty.

The margin defined as a probability has two basic features: it considers the uncertainty of the quantities and does not have the symmetry property. Rather, it fulfils

$$SM(\boldsymbol{A},\boldsymbol{B}) = 1 - SM(\boldsymbol{B},\boldsymbol{A}) \tag{4.3}$$

Two clear advantages of probabilistic margins are

- They are nondimensional and limited to the [0,1] interval.
- They can be combined, via the probability laws.

Additional properties of the probabilistic margin are typical of the mathematical definition of distance or metrics. Firstly

$$SM(\boldsymbol{A},\boldsymbol{A}) = 0 \tag{4.4}$$

for every random variable **A**. It is worth to realize that the margin depends not only on the discrepancy between random variables, but also on their mutual dependency. Two random variables can differ in a quantity arbitrarily small and however have the maximum margin (i.e. 1) if they are totally dependent. If **A** is a random quantity and **A*** is **A** plus an arbitrarily small increment in a conservative direction, then SM(**A, A***)=1, due to the total correlation of **A** and **A***.

On the other hand, if $\mathbf{A_1}$ and $\mathbf{A_2}$ are independent quantities having the same probability distribution, it is clear that

$$SM(\boldsymbol{A_1}, \boldsymbol{A_2}) = 0.5 \tag{4.5}$$

And, if **A** and **B** are independent random variables:

- SM(**A, B**) < 0.5 indicates that **A** is "more severe" than **B**.
- SM(**A, B**) > 0.5 indicates that **A** is "less severe" than **B**.
- SM(**A, B**) ≅ 0.5 indicates that **A** and B have similar severity.

As previously stated, this interpretation in no longer valid when **A** and **B** are mutually dependent.

On the other hand, the probabilistic margin fulfils a property typical of mathematical distances, the triangular inequality:

$$SM(\boldsymbol{A,B}) \leq SM(\boldsymbol{A,C}) + SM(\boldsymbol{C,B}) \tag{4.6}$$

for any triplet of values **A**, **B** and **C** of a quantity. It is easily proved from Bonferroni's inequality, as shown in [64].

Other possible definitions of SM fail to meet any of the aforementioned properties. A plausible definition is a low-order quantile of the difference between variables. If A and B are values of a scalar quantity, we could define:

$$SM_Q(\mathrm{A,B}) \equiv \kappa \cdot (B - A)_{1-\gamma} \tag{4.7}$$

where γ is a value close to 1, and κ is a numerical constant introduced to render (4.7) nondimensional. This definition is not symmetrical and fulfils property (4.4). But, as shown in [64], margins defined as quantiles do not follow simple rules of combination, unlike the probabilistic margins, and this fact hinders to set bounds to a global margin in terms of the sum of partial margins.

4.3. The calculated safety margin

When the probability distributions of A and B (be they either scalar or multidimensional) are perfectly known, the margin can be calculated exactly. In the scalar case, the margin is calculated as an integral

$$SM(A,B) = \int_R f_B(z) \cdot F_{A|B}(z|B=z)dz \tag{4.8}$$

In (4.8), f_B and $F_{A|B}$ are the PDF of B and the CDF of A conditioned to B, respectively. The integral (4.8) is defined in the intersection of the ranges of A and B. When A and B are independent variables, (4.8) simplifies to

$$SM(A,B) = \int_R f_B(z) \cdot F_A(z)dz \tag{4.9}$$

Obviously, if one of the quantities is fixed (e.g. an acceptance limit L) the expressions (4.8) and (4.9) simplify.

It is common, however, that the distributions are only partially or imperfectly known, so that the margin becomes an epistemic uncertain quantity. Usually, the sole information on the distributions are random samples of A and B. When the information on A and B grows (e.g. when the sizes of the random samples increase), the epistemic uncertainty of SM (A, B) decreases.

In summary, the probabilistic margin from A to B is built from the aleatory uncertainties of A and B, and has epistemic uncertainty. It is defined as a probability, and it is statistically estimated. As any other probabilistic parameter, the SM can be estimated with different degrees of conservativeness. Estimates can be either realistic, or conservative (pessimistic), or anticonservatives (optimistic).

As we will see later, sometimes it is useful to define margins between quantities with statistical uncertainty (e.g. estimators) [64]. In such case, the probabilistic definition of safety margin (based on aleatory uncertainty) is extended to be based on statistical epistemic uncertainty of estimators. Let E1(S) and E2(S') two scalar sample statistics, constructed from random samples S and S', respectively, and having the same physical units. We define the (probabilistic) safety margin from E_1 to E_2 as

$$MS[E_1(S), E_2(S')] \equiv PR_{S,S'}\{E_1(S) < E_2(S')\} \tag{4.10}$$

4.4. Licensing margin

Given the calculated value A of a scalar safety quantity, with an upper regulatory acceptance limit L_A, we call *licensing margin* of A the quantity.

$$LM(A) \equiv SM(A, L_A) \equiv PR_{A,L_A}\{A < L_A\} \tag{4.11}$$

Thus, the licensing margin is a measure of the distance from A to L_A. The probability in (4.11) is based on the uncertainty of A and L_A. The common practice is defining the regulatory limits as fixed quantities, so that the probability is only due to the uncertainty of the calculated value A. The USNRC, in [99], terms (4.11) the "designer margin".

Definition (4.11) can be generalized to a multidimensional quantity **A**, having a regulatory acceptance region $\mathbf{R_A}$:

$$LM(\boldsymbol{A}) \equiv PR_{\boldsymbol{A}}\{\boldsymbol{A} \in \boldsymbol{R_A}\} \tag{4.12}$$

Here $\mathbf{R_A}$ is a fixed region in the range of **A**. (4.12) represents a distance from **A** to the boundary of $\mathbf{R_A}$.

The probabilistic definition of licensing margin was previously introduced in chapter 2, as the *acceptance probability* of A.

The regulatory acceptance criterion for a BEPU analysis can be regarded as requiring that the licensing margin is higher or equal than a regulatory value P_0 with a confidence not less than C_0.

$$PR_S\{LM(\boldsymbol{A}) \geq P_0\} \geq C_0 \tag{4.13}$$

The probability in (4.13) refers to the second-level uncertainty of LM(**A**), due to the ignorance about the distribution of **A**.

Condition (4.13) stipulates what is meant by "enough licensing margin" in DSA: it must be higher than the regulatory value P_0 with confidence not less than C_0. The extended definition of margin (4.10), constructed with epistemic uncertainty, enable the interpretation of the left-hand side of (4.13) as a probabilistic margin from LM(**A**) to the regulatory limit P_0. We can term it an *epistemic margin* or (maybe abusing language) a *metamargin*.

The description of the two layers of uncertainty making up the BEPU acceptance criterion can be reformulated in terms of the probabilistic margin. The regulator establishes acceptance limits for three quantities:

- The safety quantity.
- The licensing margin (constructed with aleatory uncertainty).
- The licensing metamargin (constructed with epistemic uncertainty).

The verification of the fulfilment of a RAC is equivalent to the estimation of a licensing margin. In chapter 3, we have described and classified the statistical methods to perform such verification. One of the categories is that of P-Methods, based on the estimation of limit exceedance probabilities (complementary to acceptance probabilities). They are therefore procedures to calculate confidence limits of licensing margins. Q-methods, based on tolerance regions and quantiles estimation, do not enable in principle a direct estimation of licensing margins. However, as we mentioned in section 3.3.1, from every P-method we can (in principle) derive a dual Q-method, and vice versa. For instance, Wilks' method (a Q-method) has a dual P-method, the Clopper-Pearson interval, which may be regarded as a procedure of probabilistic safety margin estimation.

4.5. Analysis margin

Sometimes the term *analysis margin* (or analytical margin) is applied to the safety margin from the real or true value of a safety quantity to its calculated value. The real or true value is the one that would be produced by a real accident scenario. For a scalar safety quantity B, we define the *analysis margin* (AM) as

$$AM(B) \equiv PR_{B_{real},B_{calc}}\{B_{real} \prec B_{calc}\} \tag{4.14}$$

Probability (4.14) includes the uncertainty of both real and calculated values. In some cases, the real value can be measured and has the corresponding measurement uncertainty. However, in the accident analyses of nuclear plants, the real values of safety quantities are unknown. In fact, the simulation calculations are performed to provide surrogates of real values.

The licensing margin can be directly calculated, because the regulatory limit is known. By contrast, the analysis margin cannot be directly calculated for nuclear plants; but it can be calculated for accident scenarios simulated in experimental facilities, where the real values of safety quantities are measured. In truth, the margins calculated for experiments may be used to estimate the analysis margins of the plants (see chapter 5).

An important point is that analysis margins are clearly measures of the conservativeness of calculated quantities. The larger the margin, the higher the distance from the real to the calculated value of the quantity. This margin clearly depends on the methodology of calculation of the quantity. Therefore, it is a margin attributable to the methodology.

Hence, our probabilistic definition of margin enables a quantification of the conservativeness degree of calculated quantities. For instance, the conservative methodologies produce very high analysis margins (i.e. differing from 1 in a negligible quantity).

The juxtaposition of licensing margin and analysis margin produces another margin, the one from the real value to the acceptance limit. The regulation requires the licensing margin being higher (with high confidence) than a prefixed value. But there is no explicit regulatory control of the analysis margin. In chapter 5 the composition of LM and AM is addressed.

The definition of analysis margin can be generalized to multidimensional quantities directly calculable. The definition of conservativeness depends on the definition of the severity (as a scalar quantity transformed from the original multidimensional quantity).

4.6. Application of probabilistic margin to the quantification of conservativeness of tolerance regions

4.6.1. Probabilistic definition of the degree of conservativeness

Previously in this book we have described and categorized the methods to test the fulfilment of regulatory acceptance criteria (RAC) of the DSA. The most used methods are based on tolerance regions of the safety quantity, constructed from random samples and with a tolerance level not less than the regulatory value. If the obtained tolerance region is inside the acceptance region, it is concluded that the RAC is satisfied.

An attribute of tolerance regions and intervals is the degree of conservativeness, which is directly related to the coverage distribution of the interval. All the tolerance intervals of level (P, C) have a property in common: the quantile of order (1-C) of the coverage is not less than P. But different procedures to construct tolerance intervals typically produce different *coverage distributions*.

Let us focus once again on the a scalar safety variable Y having an upper regulatory acceptance limit, and let T1 and T2 be two procedures of constructing tolerance intervals. If T1 generates tolerance intervals with a coverage distribution which assigns more probability to higher values of the variable than those generated by T2, with the same tolerance level, then T1 will tend to produce higher coverages than T2, and this implies that T1 intervals are more *conservative* than T2 intervals. For this reason, a good indicator of the conservativeness of a tolerance region is the *expected or mean coverage*, as proposed in [62].

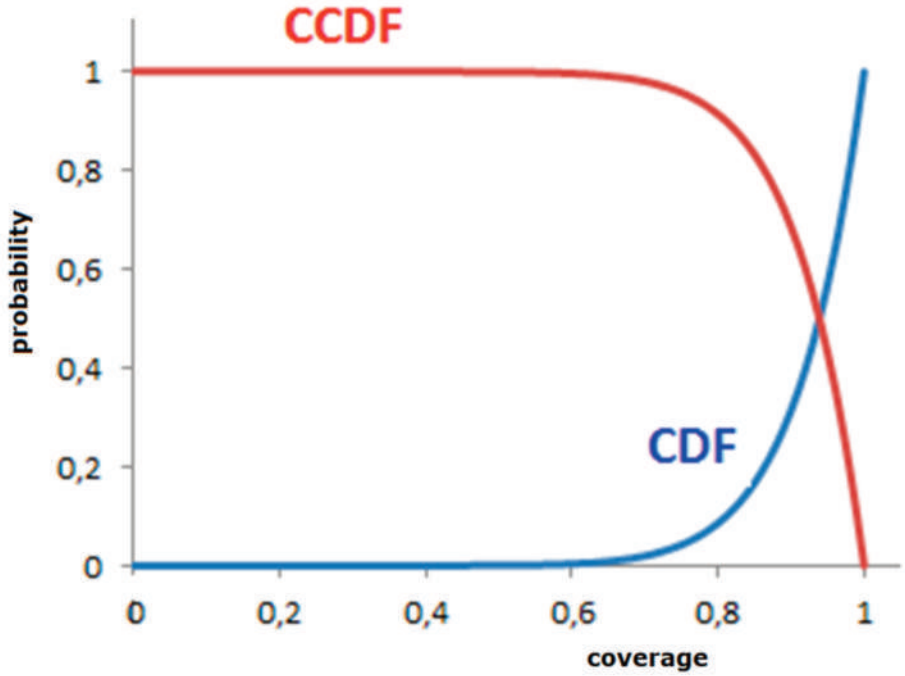
Previously in this chapter we have seen that the probabilistic analysis margin quantifies the conservativeness of the calculated value of the quantity. This idea can be applied to measure the conservativeness of tolerance limits. Suppose that, for the safety quantity, we calculate an upper tolerance limit (UTL) of level (P_0, C_0), to be compared with the regulatory acceptance limit. We can evaluate the conservativeness of the UTL by means of the margin from the calculated quantity Y to the UTL i.e.

$$SM(Y, UTL) \equiv PR_Y\{Y \leq UTL(S)\} \equiv F_Y(UTL(S)) \tag{4.15}$$

In (4.15), the probability is conditioned to the sample S used to set the UTL. But the probability in (4.15) is the coverage of the UTL. It is a quantity with epistemic uncertainty, because it depends on the random sample S. Indeed the UTL is defined in such a way that the coverage is higher or equal to P_0, with a confidence of at least C_0.

Therefore, the coverage (4.15) is a descriptor of the conservativeness of the UTL. It is the probability distribution of the coverage what characterizes a procedure to construct tolerance regions. Graphically, this characterization is obtained from the curve of the coverage CDF, or rather the complementary curve CCDF (fig. 4.1). The graph of this CCDF is contained in the square [0, 1] x [0, 1] and passes through three points: (0, 1), (1, 0) and (P_0, C_0). The closer the CCDF curve is to the top and right sides of the square, the more conservative the tolerance region construction method.

Figure 4.1. CDF and CCDF curves for the coverage of a tolerance region. Adapted from [64].



It is important to have numerical indicators of conservativeness, obtained from the coverage distribution. As we previously mentioned, an adequate choice is the expected coverage (EC) i.e. the mean value of the coverage

$$EC(UTL) \equiv E_S\,[PR_Y\{Y \leq UTL\,(S)\}] \tag{4.16}$$

equivalent to

$$EC(UTL) \equiv PR_{Y,S}\{Y \leq UTL(S)\} \tag{4.17}$$

Expression (4.17) makes clear that the expected coverage can be regarded as a margin (in wide sense) from Y to UTL, considering the aleatory uncertainty of Y as well as the epistemic uncertainty of UTL. Hence it can be used as an indicator of conservativeness of UTL.

Different UTLs with the same tolerance level (P_0, C_0) have coverage distributions with quantiles of order 1-C_0 higher than P_0. It is the value of the expected coverage what can distinguish the different methods of tolerance region construction.

For a specific procedure of construction of tolerance limits, the EC is a decreasing function of the sample size, because, when the sample size tends to infinity, the statistical uncertainty tends to zero, the tolerance region tends to a fixed region in the quantity range, and hence the EC tends to the coverage level P.

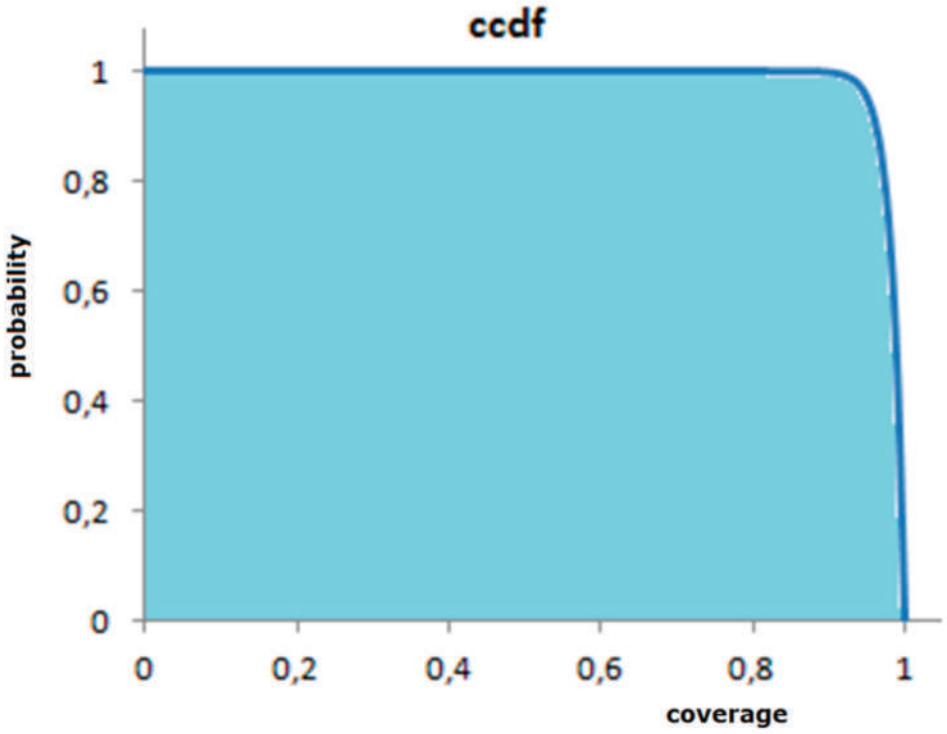
The expected coverage is numerically equal to the area below the CCDF curve (fig 4.2). Graphically, it is proved that an upper bound of such area is $1 - (1-P_0)(1-C_0) = P_0+C_0-P_0C_0$; and a lower bound is P_0C_0. For a level (95, 95) such values are, respectively, 0.9975 and 0.9025.

The EC, as indicator of conservativeness, can be generalized to multiple dimensions. For a multidimensional safety quantity **Y**, it is the probability that the components of **Y** are covered by the tolerance limits in the conservative side what matters, rather than expected coverage of the TR. Once more, a possibility is to transform Y to a scalar quantity representing "severity", and then to obtain the expected coverage for the transformed quantity.

An alternative quantification of conservativeness of a UTL could be envisaged, based on the licensing margin (LM) connecting the calculated quantity and the regulatory limit. The LM can be regarded as the superposition of two margins: one from the quantity to the UTL, the other from the UTL to the limit L. The more conservative the UTL is, the higher the first margin and the lower the second one. Therefore, the margin from UTL to L can be used as a measure (inverse) of the conservativeness of the UTL.

$$SM(UTL, L) \equiv PR_S \{UTL(S) < L\} \tag{4.18}$$

Figure 4.2. The mean coverage is the area under the CCDF curve of the coverage (in blue). Adapted from [64].



It is a margin in wide sense, built with epistemic uncertainty. The probability in (4.18) can be transformed, just by applying the CDF of Y to both sides of the inequality on the right-hand side:

$$SM(UTL,\ L) = PR_S\{F_Y(UTL(S)) < F_Y(L)\} \tag{4.19}$$

Inside the probability in (4.19), the left side of the inequality is the coverage of the tolerance limit; and the right side is the licensing margin of Y

$$SM(UTL,\ L)=F_{COV}(LM(Y)) \tag{4.20}$$

F_{COV} is the CDF of the coverage. LM(Y) must be higher than P_0 with high confidence. So we define the following indicator of conservativeness [64]:

$$CIQ(UTL) = PR_S\{P_0 + \delta < F_Y(UTL(S))\} = 1 - F_{COV}(P_0 + \delta) \tag{4.21}$$

In (4.21), δ is an increment on the coverage level P_0 (obviously less than $1\text{-}P_0$). We term indicator (4.21), *coverage of increased quantile.* It is the value of CCDF curves of the coverage in the augmented level $P_0+\delta$.

The idea of restricting the conservativeness of a tolerance limit by setting an upper bound to an indicator as (4.21) is not new [47, 52]. But here we give a novel connection of this method with safety margins.

4.6.2. Application to Wilks' tolerance limits

Wilks' method produces, for scalar random variables, tolerance intervals bounded by order statistics, and leads to a very simple expression of the expected coverage. The interval ($Y_{r:N}$, $Y_{s:N}$), with r<s, has a coverage that follows a distribution beta with parameters s-r and n-s+r+1. Because the expected value of a beta(F,G) distribution is F/(F+G), the expected coverage of the mentioned interval is equal to (s-r)/(N+1).

Wilks' one-sided intervals also have OS as tolerance limits. The coverage of the OS $Y_{k:N}$ as an upper tolerance limit is a beta variable with parameters k and N-k+1. Therefore, the expected coverage is

$$EC(Y_{k:N}) = E_S[PR_Y\{Y \leq Y_{k:N}(S)\}] = E_S[beta(k, N-k+1)\,] = \frac{k}{N+1} \tag{4.22}$$

For this OS, the indicator (4.21) is

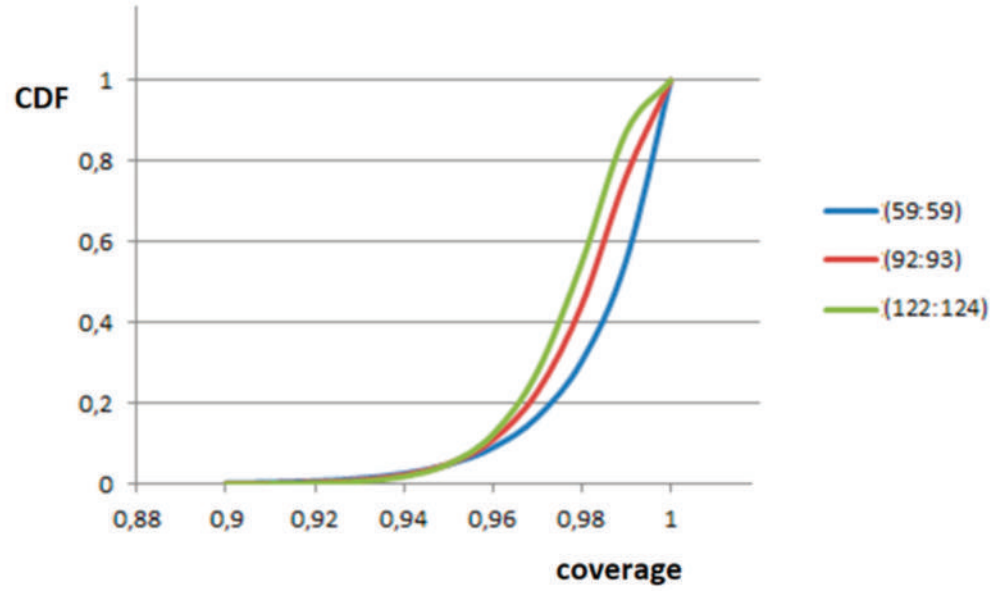
$$\begin{aligned} CIQ(Y_{k:N}) &= PR_S\{P_0 + \delta < beta(k, N-k+1)\,\} \\ &= 1 - F_{beta(k,N-k+1)}\,(P_0 + \delta) \end{aligned} \tag{4.23}$$

It is very nell kwown that, in BEPU analyses, the standard value of the regulatory tolerance level is (95,95). The application of Wilks' method reveals that the three most affordable methods (according to sample size) to set a (95,95)-upper tolerance level are:

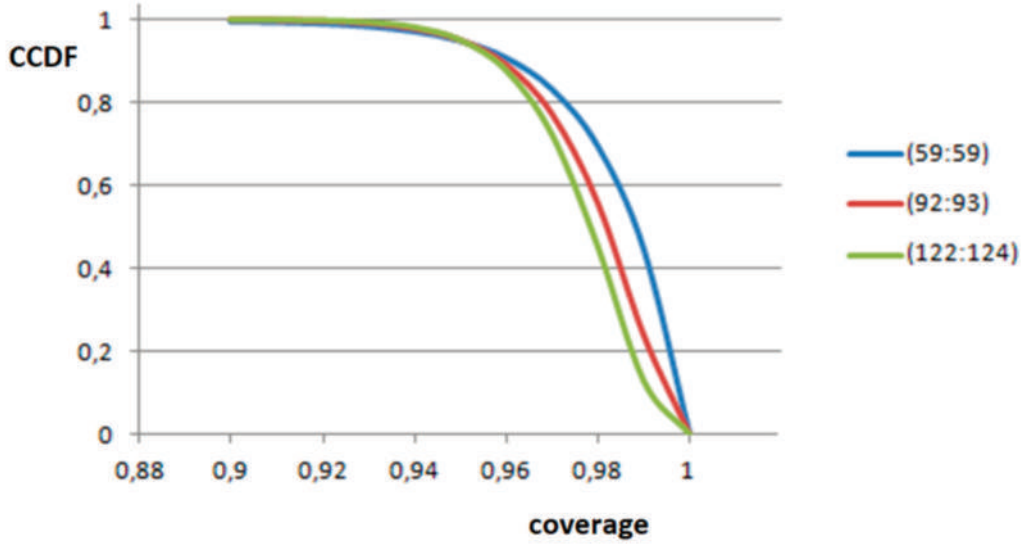
O1. Take the maximum of a sample of 59 i.e. $Y_{59:59}$
O2. Take the second maximum of a sample of 93 i.e. $Y_{92:93}$
O3. Take the third maximum of sample of 124 i.e. $Y_{122:124}$

For these three OS, figure 4.3 represents the CDF and CCDF curves for the coverage, and figure 4.4 shows the conservativeness indicator defined in (4.21) with δ=0.04.

Figure 4.3. (a) CDF and (b) CCDF curves of the coverage of three order statistics, acting as (95, 95) upper tolerance limits. Adapted from [64].



(a)



(b)

Figure 4.5 shows the probability density functions (PDF) for the three mentioned OS, for a scalar variable with normal distribution. It is clear that, upon going from O1 to O3, the PDF slightly moves towards decreasing value of the variable, showing a less conservative character.

The procedures O1 to O3, are derived from the solution to Wilks' formula for the regulatory tolerance level (P_0, C_0) and r=N-s+1

$$beta(N-r+1,r)_{1-C_0} \geq P_0 \tag{4.24}$$

Given an integer value r, the minimum N fulfilling (4.24) is the minimum sample size. For r=1, 2, 3, 4, 5,… we obtain, respectively, minimum sample sizes of 59, 93, 124, 153, 181,… Conversely, given a fixed value of N, there is a maximum integer value of r fulfilling (4.24), that we term r_{max}.

Figure 4.4. Indicator of conservativeness CIQ with δ=0.04. Adapted from [64].

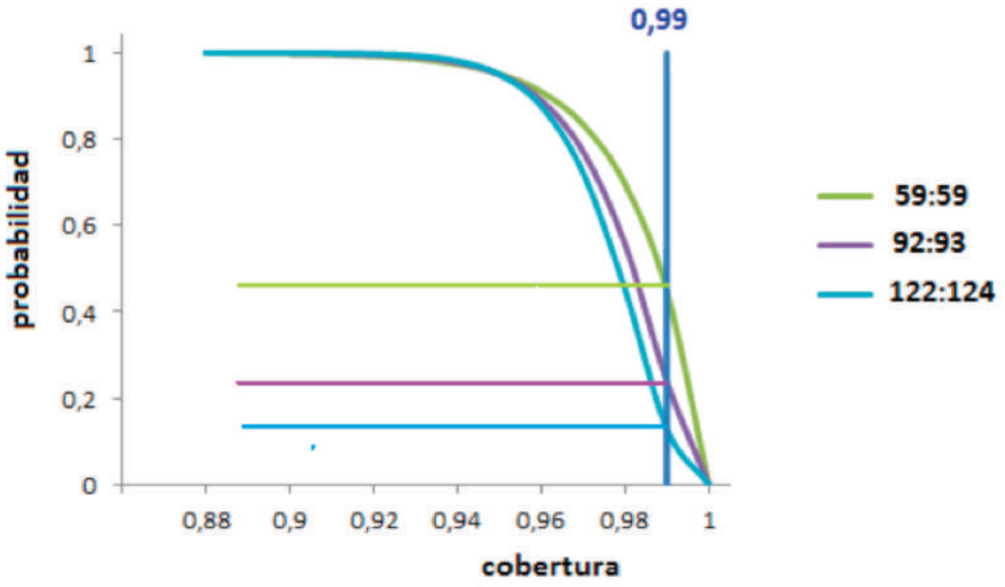
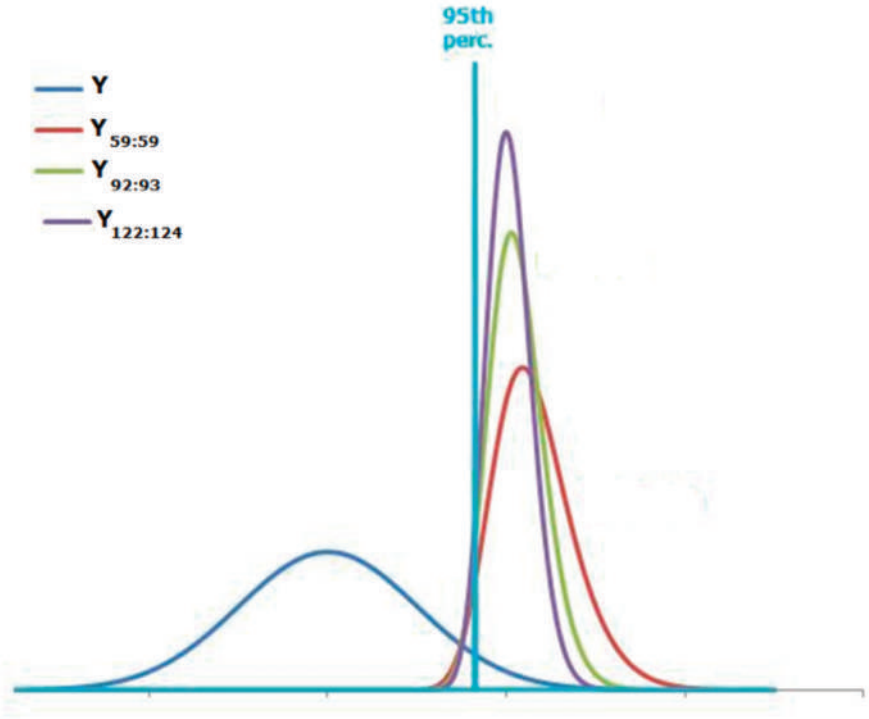


Figure 4.5. Probability density functions of a random variable and its three order statistics $Y_{59:59}$, $Y_{92:93}$ and $Y_{122:124}$. Adapted from [64].



Both the expected coverage and the coverage of increased quantile (with δ=0.04), calculated for different sample sizes, ordinal r_{max} and (95, 95) tolerance level, are given in table 4.1. In figure 4.6, both indicators are compared, as a function of the sample size. The expected coverage has been transformed, by subtraction of P_0 and normalization

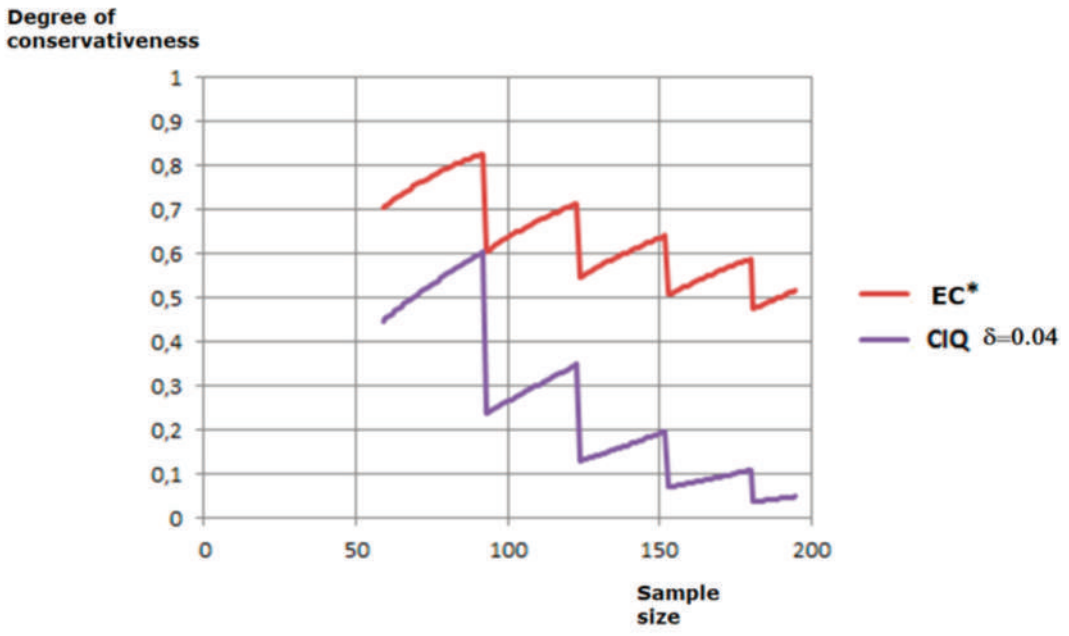
$$EC^* \equiv \frac{EC - P_0}{P_0 + C_0 - C_0 P_0 - P_0} = \frac{EC - P_0}{C_0(1 - P_0)} \tag{4.25}$$

The denominator of (4.25) is the difference between the upper bound of the expected coverage and the level P_0. The goal of the transformation is to better appreciate the changes of the indicator.

Table 4.1. Conservativeness indicators for different order statistics as upper tolerance limits (95, 95) as functions of sample size. Taken from [64].

n	Order $n\text{-}r_{máx}$	EC	EC*	CIQ (δ=0.04)
59	59	0.9833	0.7010	0.4473
76	76	0.9870	0.7789	0.5341
93	92	0.9787	0.6042	0.2421
108	107	0.9817	0.6674	0.2975
124	122	0.9760	0.5474	0.1329
139	137	0.9786	0.6021	0.1682
153	150	0.9740	0.5053	0.0724
167	164	0.9762	0.5516	0.0922
181	177	0.9725	0.4737	0.0393

Figure 4.6. Conservativeness indicators as a function of minimum sample size (95, 95) upper tolerance limits. Adapted from [64].

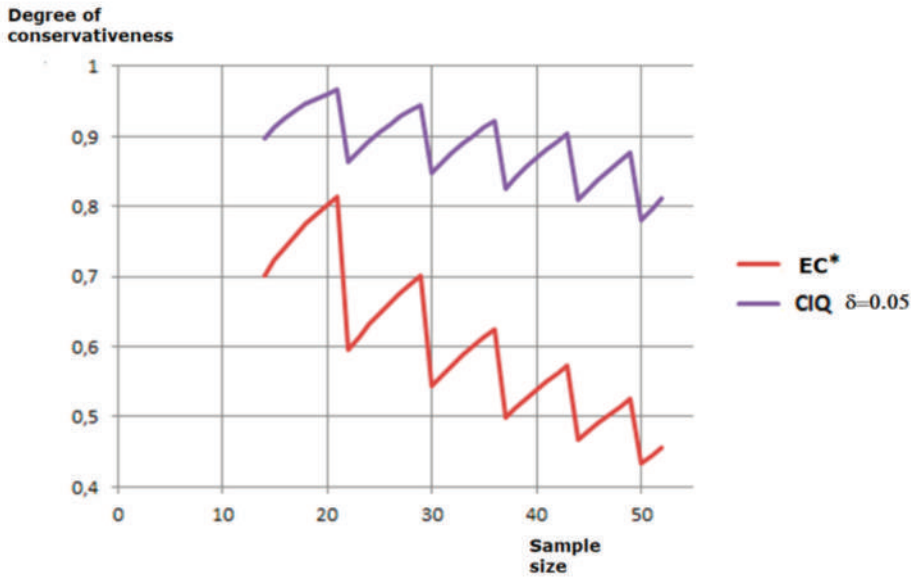


We give now the conservativeness indicators for a lower coverage level (80, 95) i.e. P_0=0.8. Firstly, we obtain from one-sided Wilks' formula the critical sample sizes. The series is now 14, 22, 30, 37, 44, 50,…In table 4.2 and figure 4.7, the values of EC, EC* and CIQ (with δ=0.05) are shown as functions of sample size.

Table 4.2. Conservativeness indicators for different order statistics as upper tolerance limits (80, 95) as functions of sample size. Taken from [64].

n	Order $n-r_{máx}$	CE	CE*	CIQ (δ=0.05)
14	14	0.9333	0.7017	0.8970
18	18	0.9474	0.7756	0.9463
22	21	0.9130	0.5950	0.8633
26	25	0.9259	0.6628	0.9183
30	28	0.9032	0.5433	0.8486
34	32	0.9143	0.6015	0.9025
37	34	0.8947	0.4986	0.8264
41	38	0.9048	0.5514	0.8821
44	40	0.8889	0.4678	0.8097
47	43	0.8958	0.5044	0.8531
50	45	0.8823	0.4568	0.8119

Figure 4.7. Conservativeness indicators as a function of minimum sample size (Table 4.2). (80, 95) upper tolerance limits. Adapted from [64].



The two proposed indicators have similar behaviour for both values of P_0, showing, as functions of sample size, a globally decreasing trend with oscillations. There are local minima in Wilks' critical values, showing a sudden decrease of the indicators. Between consecutive critical sizes, the indicators increase.

4.6.3. Margin between tolerance limits

The conservativeness of two tolerance limits can be compared by means of the probabilistic margin between them. The procedure can be directly applied to results of Wilks' method.

Let $A_{r:m}$ and $B_{t:n}$ two order statistics, derived from different and independent SRS, S and S' (with respective sizes m and n), of a given scalar quantity. The OS have statistical uncertainty, derived from the randomness of the samples. We have extended the definition of safety margin to epistemic uncertainty, according to (4.10). So the margin from $A_{r:m}$ to $B_{t:n}$ is

$$SM(A_{r:m}(S), B_{t:n}(S')) = PR_{S,S'}\{A_{r:m}(S) < B_{t:n}(S')\} \tag{4.26}$$

The formula to calculate the probability in (4.26) can be found in [18]:

$$PR_{S,S'}\{A_{r:m}(S) < B_{t:n}(S')\} = \frac{1}{\binom{m+n}{r+t-1}} \sum_{i \geq r} \binom{m}{i} \cdot \binom{n}{r+t-1-i} \tag{4.27}$$

Applying (4.27) to the three mentioned OS we obtain

$$\begin{aligned} PR_{S,S'}\{A_{92:93}(S) < B_{59:59}(S')\} &= 0.6272 \\ PR_{S,S'}\{A_{122:124}(S) < B_{92:93}(S')\} &= 0.5746 \\ PR_{S,S'}\{A_{122:124}(S) < B_{59:59}(S')\} &= 0.6913 \end{aligned} \tag{4.28}$$

The three values are higher than 0.5 and confirm that the most conservative OS is the maximum of 59; and the least conservative is the third maximum of 124.

4.6.4. Conservativeness of equal-tailed tolerance intervals

Equal-tailed tolerance intervals (ETTI) constructed with order statistics have been presented in section 3.3.1.1.6. For a given tolerance level and a given rank, ETTI constructed with conjugate OS require a higher minimum sample size than their standard two-sided Wilks' counterparts. An obvious conclusion is that, for a specified rank the ETTI is less conservative than the corresponding Wilks' two-sided tolerance interval. Furthermore, the expected coverage decreases when rank grows, for both types of intervals. When the sample size grows, the conservativeness degree decreases, and the expected coverage tend to the coverage level (0.95 in the case of Table 4.3).

Table 4.3. Expected coverages of equal-tailed and standard Wilks' two-sided tolerance intervals. Taken from [70].

Rank r	n_{min} for equal-tailed 95/95 tolerance intervals	OS for equal-tailed 95/95 tolerance intervals	Expected coverage $\frac{n-2r+1}{n+1}$	n_{min} for std two-sided 95/95 tolerance intervals	OS for std two-sided 95/95 tolerance intervals	Expected coverage $\frac{n-2r+1}{n+1}$
1	146	(1, 146)	0.9864	93	(1, 93)	0.9787
2	221	(2, 220)	0.9820	153	(2, 152)	0.9740
3	287	(3, 285)	0.9792	208	(3, 206)	0.9713
4	348	(4, 345)	0.9771	260	(4, 257)	0.9693
5	407	(5, 403)	0.9755	311	(5, 307)	0.9679
6	464	(6, 459)	0.9741	361	(6, 356)	0.9668
7	519	(7, 513)	0.9731	410	(7, 404)	0.9659
8	574	(8, 567)	0.9722	458	(8, 451)	0.9651
9	627	(9, 619)	0.9713	506	(9, 498)	0.9645
10	680	(10, 671)	0.9706	554	(10, 545)	0.9640
11	732	(11, 722)	0.9700	601	(11, 591)	0.9635

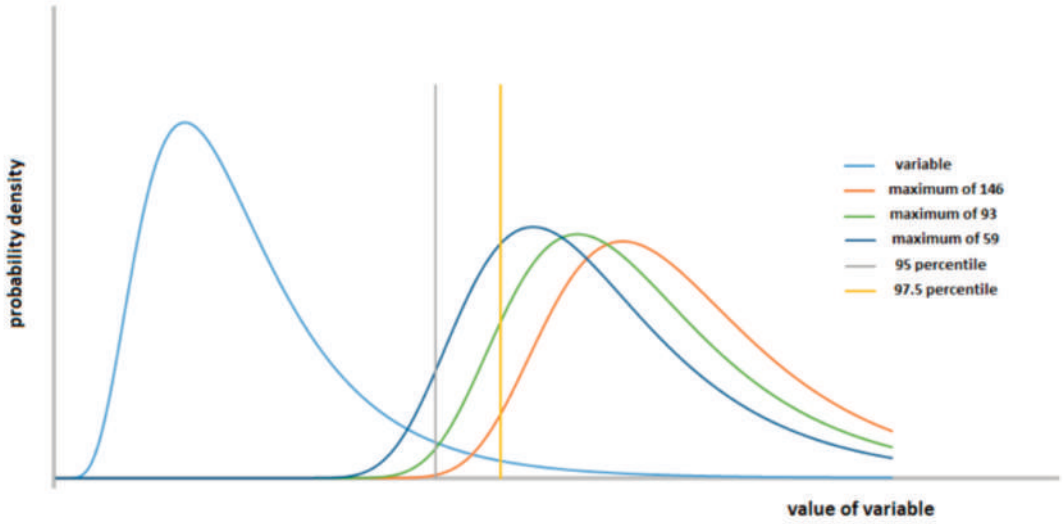
The conservativeness of a two-sided tolerance interval is related to the probability distribution of the statistics defining the endpoints. For testing the fulfilment of one-sided acceptance criteria (typical of nuclear DSA) one-sided TI are more efficient that their two-sided counterparts. And the ETTI are more conservative (i.e. less efficient) than their standard Wilks' counterparts.

For example, for testing the fulfilment of an upper acceptance limit to a level (95, 95) we can use any of these 3 order statistics

OS1. the maximum of 59 calculations, which is an (95, 95) upper tolerance limit, or
OS2. the maximum of 93, which is the upper bound of a (95, 95) two-sided TI, or
OS3. the maximum of 146, which is the upper bound of a (95, 95) ETTI.

The choice OS3 is more conservative than OS2, which is more conservative than OS1. Figure 4.8 compares the PDF of these 3 maxima for a particular random variable. It is clearly noticed that the most conservative distribution is that of choice OS3, followed by that of option OS2. The most efficient choice is OS1. These facts are evident in figure 4.8, showing the PDF of the three alternatives.

Figure 4.8. Probability density functions of a random variable and three different order statistics (random sample maxima).



4.6.5. Application to normal tolerance limits

If the scalar variableY follows a normal distribution, a well-known upper tolerance limit is (as stated in section 4.6.5):

$$UTL = m_Y + k_1(N, P_0, C_0) \cdot S_Y \tag{4.29}$$

in terms of the mean m_Y and the standard deviation S_Y of the simple random sample of Y. k1 is termed one-sided normal tolerance limit. It is a function of the sample size and the tolerance level.

The expected coverage of the UTL in (4.29) is

$$EC(UTL) \equiv PR_{Y,S}\{Y < m_Y + k_1 \cdot S_Y\} \tag{4.30}$$

which transforms to

$$EC(UTL) \equiv PR_{Y,S}\left\{\frac{Y - m_Y}{S_Y}\sqrt{\frac{N}{N+1}} < k_1\sqrt{\frac{N}{N+1}}\right\} \tag{4.31}$$

The right-hand side of the inequality inside the probability in (4.31) is a Student's t variable with N-1 degrees of freedom, T_{N-1}. Hence the expected coverage is

$$EC(UTL) = F_{T_{N-1}}\left(k_1\sqrt{\frac{N}{N+1}}\right) \tag{4.32}$$

in terms of the CDF of Student's t variable.

Table 4.4 gives the normal factor k_1 for the tolerance level (95, 95) and the expected coverage (4.32) as a function of the sample size.

Table 4.4. Expected coverage for normal tolerance limits, as a function of sample size. Taken from [64].

Sample size	**k_1**	**EC**
20	2.3783	0.9842
50	2.0585	0.9765
80	1.9605	0.9725
100	1.9234	0.9707
150	1.8678	0.9677
200	1.8357	0.9657
1000	1.7270	0.9577

It is evident that the expected coverage, as indicator of conservativeness, is a decreasing function of sample size. As sample size grows, factor k_1 tends to 1.645 (i.e. the 95 percentile of the standard normal distribution), and EC tends to 0.95.

For a normal variable, we can compare the expected coverages obtained from Wilks' method and from normal limits, for the critical sample sizes of Wilks' method and some intermediate values. Table 4.5 shows the results.

Table 4.5. Comparison of expected coverage (EC) of (95, 95) Wilks' and normal tolerance limits, as functions of sample size. Taken from [64].

Sample size	k_1	Ordinal of OS	EC Wilks	EC normal
59	2.0204	59	0.9833	0.9751
76	1.9698	76	0.9868	0.9730
93	1.9349	92	0.9787	0.9713
108	1.9118	107	0.9820	0.9700
124	1.8922	122	0.9760	0.9691
139	1.8773	137	0.9784	0.9683
153	1.8654	150	0.9740	0.9675
167	1.8552	164	0.9762	0.9669
181	1.8463	177	0.9725	0.9664

It is observed that the EC for the Wilks' UTL is higher than the EC for the normal UTL. Therefore, the Wilks' UTL is more conservative than the normal UTL. In other words, the normal UTL is more efficient than the nonparametric one. This fact is not surprising for a normal random variable, considering also the conservativeness of Wilks' method compared to other methods of quantile estimation.

V. Some insights into the validation of BEPU methodologies

V. SOME INSIGHTS INTO THE VALIDATION OF BEPU METHODOLOGIES

5.1. Code V&V and DSA methodologies

Deterministic Safety Analysis is a complex task. The phenomenology of the Design Basis Scenarios must be simulated, and the safety quantities involved in the regulatory acceptance criteria (RAC) must be adequately calculated. For this reason, computational codes are the cornerstone of DSA methodologies. Once developed, computational codes undergo (as a part of their quality assurance) a process termed "Verification and Validation" (V&V). It is a basic element, needed to complete the construction and qualification of the codes.

But codes are not the unique element of DSA. In fact, a DSA methodology is a conjunction of several elements:

- Predictive computational codes, with their use guidelines and manuals.
- Hypotheses about the models and simulations, some of them imposed by the regulator.
- Uncertainty and sensitivity analysis procedures, especially for BEPU methodologies.
- Other ancillary tools.

In a sense, methodologies are "supercodes", calculation frameworks having the codes as subelements. The outputs of a methodology are the safety quantities involved in the RAC, and the values obtained for these quantities depend not only on the codes used, but also (strongly) on the hypotheses. For this reason, although the codes included in a DSA methodology are already verified and validated, the methodology as a whole needs its own process of V&V. This is an especially important topic, but it is seldom addressed in the DSA realm.

The verification of a code requires checking if it meets the specifications established by their designers and developers. Models are developed following several stages. The starting point is a conceptual model of the physical reality, and it transforms successively in a mathematical, numerical and computational model. In each stage, specifications for the final product are established. This set of checks constitutes the *verification* of the model.

Once the computational model or code is verified, it must be validated. *Validation* of a model is the test of how adequately and accurately the model reproduces physical reality. Validation is based on the comparison of the model predictions with reference data of two possible types:

- Real values, obtained from measures in experimental or real settings.
- Calculated values, obtained with other qualified models.

Traditionally, validation has been done using real values, basically derived from experiments. Sometimes, though, data are scarce, when the phenomena object of simulation are difficult to reproduce in experiments or to observe in real world. On the other hand, the increase of computational power enables affordable calculations with very detailed mechanistic models. For this reason, advanced models can be used to validate other models, so that computational experiments can be applied in the validation process.

A code is a computational structure composed by submodels, each one dealing with a specific part of the phenomenology. Submodels can be broadly classified as *mechanistic* (based on first principles) or *empirical* (based on observed cause-effect relationships). Commonly, even mechanistic models have a certain degree of empiricism, reflected in the presence of adjustable parameters in their formulation.

Once developed and calibrated via real data, a model must be validated using additional data. In a computational code, submodels typically have a hierarchical relation, and the processes of calibration and V&V must be adapted to such structure. The submodels of a code should be calibrated and validated individually and separately, if possible using data obtained from separate effects tests (SET). However, in the performance of the code the submodels interact. So, the individual V&V is not enough, and a V&V of the code as a whole is needed, based primarily on data from the so-called integral effects tests (IET).

5.2. Validation of DSA methodologies

To perform analyses of DBS adequate codes are needed, able to simulate the phenomenology occurring in the scenario. Nevertheless, a good code is not enough; we also need an adequate methodology.

The validation of a DSA methodology enables the estimation of the margin from calculated safety quantities to real values, namely its degree of conservativeness.

To be used in licensing analysis, a DSA methodology (be it conservative or BEPU) must be validated. The validation of deterministic methodologies is a topic that, in general, has not been tackled in depth in Nuclear Safety. We next will establish theoretical basis for such validation, and refer to the possibility of using a qualified methodology for the validation of other methodologies.

In DSA methodologies, codes calculate the evolution of the most significant physical magnitudes, mainly the safety quantities involved in the regulatory acceptance criteria. The validation of the methodology must be based on such safety quantities, though it could not be limited to them.

The process should be primarily based on real data, mainly those from integral effects tests, which reproduce at small scale accident scenarios of NPPs. But, as previously said, data calculated with advanced models could be used as well, especially when experimental results are scarce.

Aside from imposing acceptance limits, regulators can also establish baselines of the calculation methodology (e.g. requiring the use of specific submodels) so that the limits can only apply to results from such requirements. Commonly, though, the regulatory limits take as reference the real or true values of safety quantities i.e. those that should be reached in a real scenario in the analyzed NPP. The concern of regulators can be regarded as the fulfilment of the acceptance criteria by the real values of safety quantities, aside from the calculated ones.

A good example is the US regulation on LOCA/ECCS analysis for LWRs [16]. In the former version, released in 1974, safety quantities and their acceptance criteria were established for the first time. Simultaneously, strict criteria of conservativeness were imposed on the calculation methodology [17]. However, the resulting regulatory limits were not specific for this type of methodologies; indeed, the limits were untouched when (in 1988) the LOCA/ECCS rule was amended to allow the use of realistic methodologies.

So, assuming that the RAC ultimately refer to safety quantities in a real scenario, the problem is that such real values are unknown and must be replaced by model predictions. For that reason, the RAC are imposed on the calculated quantities. But the fulfilment of the RAC by the calculated quantity does not imply the fulfilment by the real magnitude. It is also necessary to prove that the calculation reproduces *adequately* the reality. That should be the basic goal of the methodology validation.

When a model is validated, the accuracy (understood as a *distance* from predictions to real values) is assessed. With such goal, metrics or distances between predicted and real values should be defined. Similarly, the validation of a safety methodology requires the quantification of the distance between calculated and real values. But especial metrics, appropriate for safety quantities, are required in the task. An example are the probabilistic margins defined and studied in chapter 4.

In the remainder of this chapter, we develop the fundamentals of the process of validation of DSA methodologies, both conservative and BEPU, following [64]. We will focus, with no loss of generality, on a design basis scenario T with a single scalar and continuous safety quantity Y with an upper regulatory acceptance limit L.

5.3. Validation of conservative methodologies

If a conservative DSA methodology is applied to scenario T, the RAC is

$$Y_C < L \tag{5.1}$$

How should this conservative methodology be validated? The key point is to prove that the quantity Y_C is truly conservative i.e.

$$Y_{real} < Y_C \tag{5.2}$$

In (5.2), Y_{real} is the value that Y would reach in a real (not simulated) scenario. The problem is that we do not know Y_{real} . But there are ways to circumvent this lack of information. Suppose that we have information about a scenario T* in a test of an experimental facility, and T* is similar to T, in the sense that the facility approximately reproduces in small scale a nuclear plant, and both scenarios are basically led by the same physical phenomena. Hence, measured values of Y* (the values reached by Y in the scenario T*) are available.

The conservative methodology is developed to be applied to the scenario T in a plant; but, due to the similarity of scenarios, we assume that the codes in the methodology are able to reproduce the scenario T*.

From the regulatory standpoint, the ultimate goal of the safety analysis of T can be regarded as proving that the real value in scenario T fulfils the RAC.

$$Y_{real} < L \tag{5.3}$$

Notice that if (5.1) and (5.2) are true, (5.3) is true as well. The problem is that the real value of Y during T is unknown. However, we know the value in the scenario T*. Suppose that we can prove that

$$Y^*_{real} < Y^*_C \Rightarrow Y_{real} < Y_C \tag{5.4}$$

where the asterisk indicates values for the validation transient T*. If (5.4) is true, the criterion of validation of the methodology for scenario T* is the condition of Y* being conservative

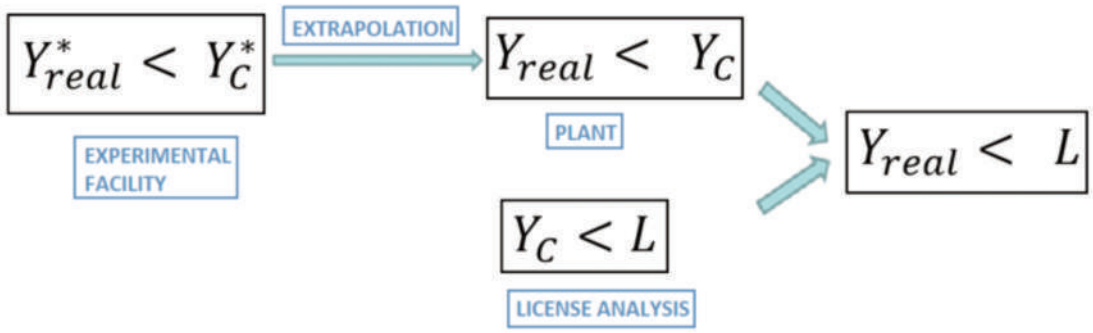
$$Y^*_{real} < Y_C^* \tag{5.5}$$

(5.4) can be regarded as a condition for extrapolation, meaning that the fulfilment of the "conservativeness criterion" for Y can be extrapolated from T* to T. In other words, the conservativeness of the methodology should be proved for T* and then extrapolated to T.

A way to get the fulfilment of (5.4) is to set some penalty on the DBS analysis of the plant, but not applicable to the analysis of the validation scenario T*.

Figure 5.1 explains how the conjunction of 3 criteria (the RAC (5.1), the validation of conservativeness (5.5) and the criterion of extrapolation to plant (5.4)) ensures finally the fulfilment of criterion (5.2), which is (from the regulatory standpoint) the ultimate goal of the safety analysis for Y.

Figure 5.1. The logical scheme for the validation of a conservative methodology: conjunction of regulatory acceptance criterion, validation in an experimental scenario and extrapolation. Adapted from [64].



5.4. Validation of BEPU methodologies

Let us suppose now that the same design basis scenario T previously considered is analyzed with a BEPU methodology. The RAC is now

$$PR_S\{PR_Y\{Y < L\} \geq P_0\} \geq C_0 \tag{5.6}$$

where Y is the random variable calculated by the methodology. We suppose that the uncertainty of Y is known via a random sample S. In terms of the P_0-quantile of Z, (5.6) is rewritten as

$$PR_S\{Y_{P_0} < L\} \geq C_0 \tag{5.7}$$

It is assumed that the ultimate goal of the analysis is to prove a criterion similar to (5.6) but for the real value of Z

$$PR_S\{PR_{Y_{real}}\{Y_{real} < L\} \geq P^*\} \geq C^* \tag{5.8}$$

Similarly to (5.6), (5.8) contains a second-order probability. The inner layer refers to the uncertainty of the value Y_{real}; and the outer one refers to the uncertainty about the inner probability. Obviously, values P* and C* should be high, but not necessarily equal to P_0 and C_0.

The main goal of the methodology validation is to provide an information which, together with the fulfilment of the RAC (5.6) by the calculated value of Y, proves the fulfilment of (5.8) by the real value of Y.

We suppose that Y_{real} has a fixed but unknown value i.e. it is a quantity with epistemic uncertainty (5.8) then reduces to

$$PR\{Y_{real} < L\} \geq C_0^* \tag{5.9}$$

where the probability represents the epistemic uncertainty about the value of Y_{real}. Should this value be perfectly known, the probability in (5.9) would be either 0 or 1.

Let P_1 a number in the interval (0, 1) and such that $P_1<P_0$. Then, the following implication

$$\left(Y_{real} < Y_{P_1}\right)\ AND\ \left(Y_{P_0} < l\right) \Rightarrow\ Y_{real} < L \tag{5.10}$$

holds, and the application to it of Bonferroni inequality yields

$$PR\{Y_{real} < L\} \geq PR\{Y_{real} < Y_{P_1}\} + PR\{Y_{P_0} < L\} - 1 \tag{5.11}$$

(5.11) enables establishing a sufficient condition to have a high confidence of Y_{real} fulfilling the RAC, namely that the two terms in the right-hand side of (5.11) are close to 1. If the calculated quantity satisfies the RAC (5.7), the second term is not less than the regulatory level C_0 i.e. a high value (the standard value is 0.95).

The first term in the right-hand side of (5.11) is unknown, because Y_{real} is unknown. However, this same probability can be calculated for the validation scenario T*. Let us suppose that the calculated value of Y* satisfies the following condition:

$$PR\{Y_{real}^* < Y_{P_1}^*\} \geq C_1 \tag{5.12}$$

where P_1 and C_1 are numbers in the interval (0, 1), whose values will be later discussed. It is understood that the value Y* is calculated with uncertainty; it seems logical that calculations to validate a BEPU methodology should be performed with uncertainty. S* denotes the random sample of Y* used to test the condition.

The criterion (5.12) is rewritten as

$$PR_{S^*}\{PR_{Y^*}\{Y^*_{real} < Y^*\} \geq 1 - P_1\} \geq C_1 \tag{5.13}$$

This "validation criterion" has the same formal structure as the RAC of licensing analysis. Both expressions (5.6) and (5.13) contain second-order probabilities and stipulated values of tolerance applied to the calculated quantities. The difference between them is that:

- In (5.6) the calculated quantity must be below the regulatory acceptance limit with tolerance level (P_0, C_0).
- In (5.15) the calculated quantity must be higher than the real measured value with tolerance level (1-P_1, C_1).

(5.13) requires the calculation methodology of Y* being conservative with a tolerance level (1-P_1, C_1). Notice that now we are not referring to the conservative methodologies described in section 5.2, but rather to BEPU methodologies *having some degree of conservativeness.* There is no contradiction, because a methodology can combine realistic predictive models with some conservative features.

The inner probability in (5.13) is what we called in section 4.5 the *analysis margin* of Y* (i.e. the probabilistic margin from the real to the calculated value of Y*). The probability refers only to the calculation uncertainty of Y*, because it is assumed that the real value is perfectly known. The analysis margin can be regarded as the degree of conservativeness (DC) of the calculated Y*. The validation criterion (5.13) requires a DC of at least 1-P_1, with a specified level of statistical confidence C_1.

In that sense, 1-P_1 could be defined as the DC required to the methodology. The outer probability in (5.13) refers to the uncertainty of the inner one. If the validation criterion is verified via a random sample of Y*, the outer uncertainty is due to the finite size of the sample.

In order to prove that the first summand in the right-hand side of (5.11) is high, we would need an extrapolation criterion just similar to (5.4), but adequate for a BEPU methodology. A straightforward criterion should be:

$$PR\{Y_{real} < Y_{P_1}\} \geq PR\{Y^*_{real} < Y^*_{P_1}\} \tag{5.14}$$

And a stronger criterion is:

$$PR\{Y_{real} < Y\} \geq PR\{Y^*_{real} < Y^*\} \tag{5.15}$$

The probability $PR\{Y_{real} < Y\}$ is the analysis margin of Y (i.e. the DC of Y). Both (5.14) and (5.15) express, with different intensity, that the methodology should be more conservative in a licensing analysis of a DBS than in a validation exercise. These are conditions of extrapolation, requiring that the DC calculated in the validation can be extrapolated to the licensing analysis of the plant.

If (5.14) and/or (5.15) are fulfilled, the validation criterion (5.13) can be extrapolated from the scenario T* to plant scenario T, yielding:

$$PR_S\{Y_{real} < Y_{P_1}\} \geq C_1 \tag{5.16}$$

(5.16) establishes a minimum DC, 1-P_1, with a statistical confidence level of at least C_1, for the application of the methodology to scenario T. The extrapolation criterion (5.15) states that the DC of the methodology must not be less when applied to design basis scenarios than when applied to experimental scenarios or other reference settings.

An obvious way to fulfil the extrapolation conditions is the introduction of penalties or conservatisms in the methodology only applicable to license calculations and absent from validation calculations.

If, furthermore, the license regulatory criterion (5.7) is fulfilled, the inequality (5.11) implies

$$PR\{Y_{real} < L\} \geq C_1 + C_0 - 1 \tag{5.17}$$

Assigning values to C_0 and C_1 such that

$$C_0^* \leq C_1 + C_0 - 1 \tag{5.18}$$

the fulfilment of (5.9), that should be the "ultimate criterion" from the regulatory standpoint, is ensured. The inequality (5.18) implies that C_0* cannot be higher than C_0 and C_1. Presently, the standard regulatory value of C_0 is 0.95. Therefore, C_0* cannot surpass such value. If C_0=C_1=0.95, a value 0.90 is ensured for C_0*.

In summary, the fulfilment of the following three conditions:

1. The BEPU methodology for calculating Y is validated as conservative, to a level (1-P_1, C_1), for experimental scenarios.
2. Such validation can be extrapolated to plant DBS.
3. The RAC is fulfilled when applying the methodology to the DBS.

ensures the fulfilment of the RAC by the real value of Y with a high level of confidence (i.e. the ultimate criterion) (fig. 5.2).

(5.16) requires that the methodology has, in its application to licensing, a degree of conservativeness of, at least, 1-P_1. The higher P_1 is, the lower the demand on the DC. If P_1 is small, close to zero, the methodology is conservative. If P_1 is around 0.5, the methodology is approximately unbiased (as those based on realistic models). When P_1 is close to 1, the methodology may be even anticonservative, prone to produce values of the safety quantity which are less bounding than the real values (fig. 5.3).

Condition $P_1 \leq P_0$ implies 1-P_0 ≤ 1-P_1, so that, in our example, the DC of the methodology cannot be lower than 1-P_0. For the standard regulatory level P_0=0.95, the DC of the methodology must be, at least, 0.05. Thus, to support this sufficient condition the methodology must not be too anticonservative.

Figure 5.2. The conjunction of licensing, validation and plant extrapolation (BEPU methodologies). Adapted from [64].

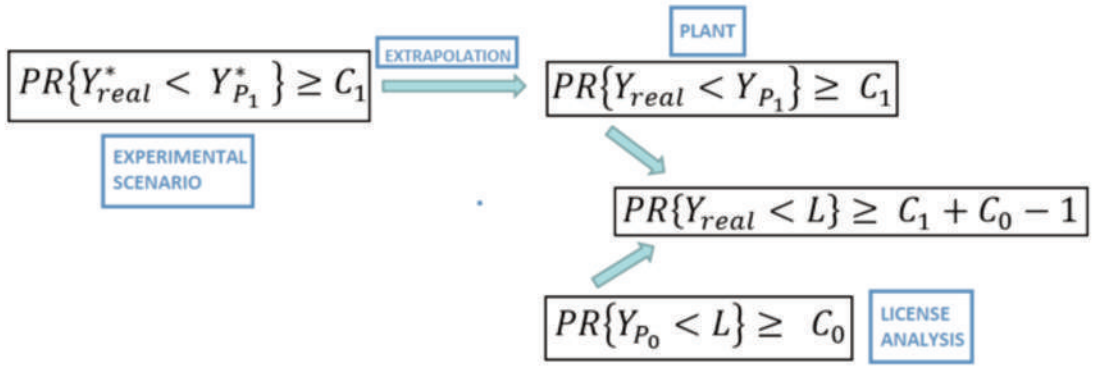
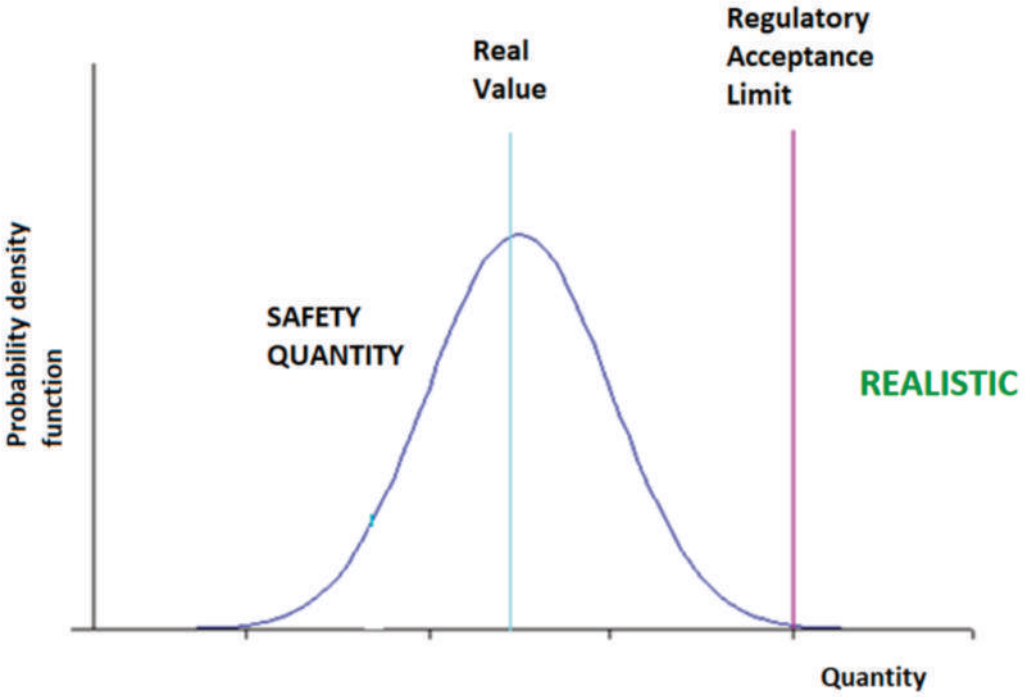
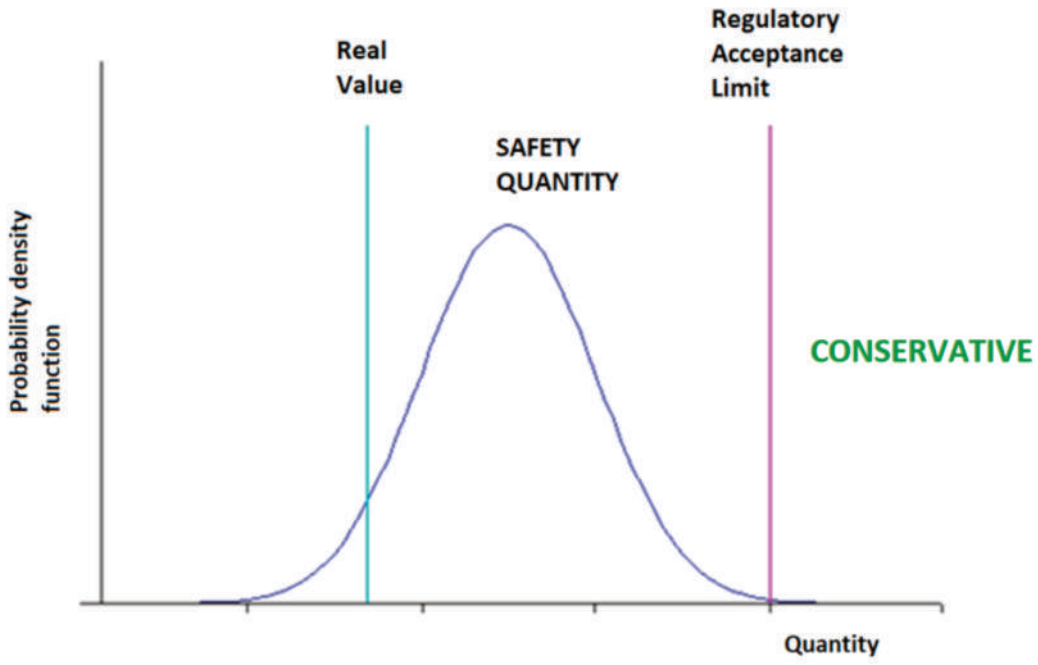


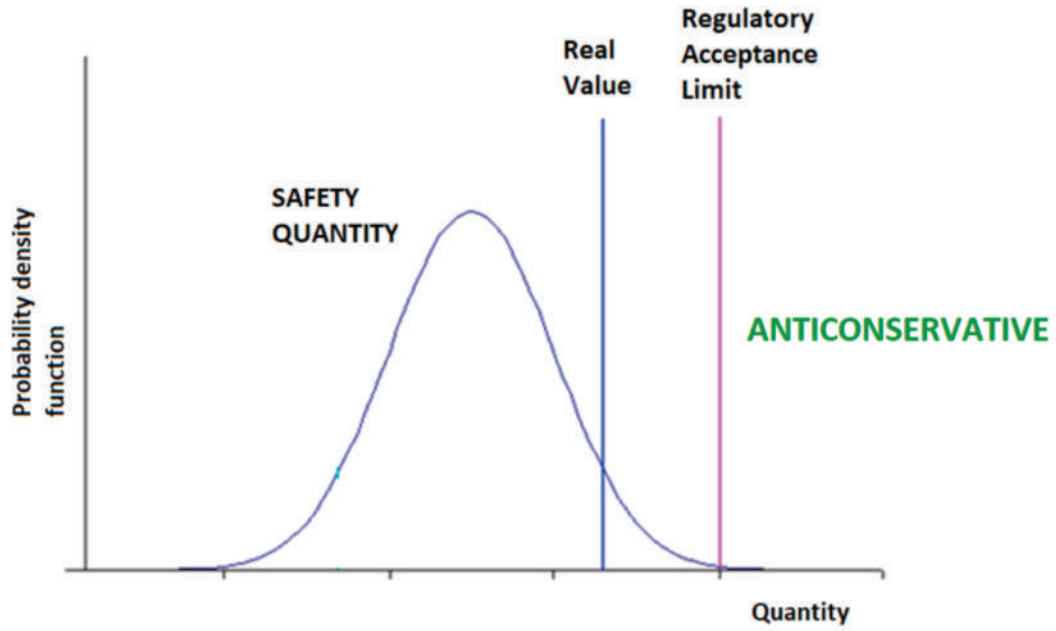
Figure 5.3. Results from (a) Realistic, (b) Conservative and (c) Anticonservative BEPU methodologies. Adapted from [64].



(a)



(b)



(c)

5.5. Validation with BEPU methodologies

In the realm of nuclear DSA, conservative methodologies are still very used. They are far simpler and cheaper to apply than BEPU methodologies, and this fact balances their largely pessimistic character and poor predictive capability. As mentioned in section 5.2, validation is necessary not only for BEPU methodologies, but for conservative ones as well.

Validation of computational codes is based on comparisons of predictions with real data. But, as previously stated, experimental data may be replaced or complemented with data calculated with consolidated codes. The reason may lie on the difficulties and/or economic problems to get real data. Of course, if a code is used to validate other code, the former should be adequately qualified and reliable.

Something similar may occur with methodologies, so that an advanced methodology (typically BEPU) may be used to validate a conservative methodology [63, 64, 71].

Suppose that we have two methodologies to calculate Y for the scenario T. One is methodology B, which is BEPU and is validated. The other one is Methodology K, which is conservative. It is

assumed that the physical models included in K are individually verified and validated. The goal is using methodology B to validate methodology K, so that the latter can be used in licensing analyses.

In terms of quantiles, the RAC for B is

$$PR\{Y_{B,P_0} < L\} \geq C_0 \tag{5.19}$$

On the other hand, the RAC for K is

$$Y_K < L \tag{5.20}$$

Y_B is the random variable calculated with B, while Y_K is the conservative value calculated with K. S is the random sample of Y_B used in the testing of the RAC.

A criterion for the validation of methodology K should have the form

$$PR_{S^*}\{Y_{B,P_2} < Y_K\} \geq C_2 \tag{5.21}$$

meaning that K is conservative with respect to B with a tolerance level that should not be lower than the regulatory one: $P_2 \geq P_0$ and $C_2 \geq C_0$. S* is the random sample of Y_B used in validation.

The implication

$$(Y_{B,P_2} < Y_K)\ and\ (Y_K < L) \Rightarrow (Y_{B,P_2} < L) \tag{5.22}$$

yields, by applying Bonferroni's inequality

$$PR\{Y_{B,P_2} < L\} \geq PR\{Y_{B,P_2} < Y_K\} + PR\{Y_K < L\} - 1 \tag{5.23}$$

If (5.20) and (5.21) are fulfilled, (5.23) gives

$$PR\{YB_{P_2} < L\} \geq C_2 \tag{5.24}$$

As $C_2 \geq C_0$, the BEPU RAC is satisfied. Then, if Methodology K is validated according to (5.21), it can be applied to licensing analyses, and the fulfilment of the conservative RAC by Y_K would ensure the satisfaction of the BEPU RAC by Y_B.

The fulfilment of the validation criterion (5.21) can be tested by means of methods described in chapter 3. Specifically, it could be done through the calculation of an upper tolerance limit with level (P_2,C_2) for Y_B . If the UTL is below YK, (5.21) is met. If Wilks' method is used, the needed minimum sample size is a growing function of the levels P_2 and C_2.

The validation criterion (5.21) introduces a new probabilistic safety margin: the one from a BEPU calculated value to a conservatively calculated value.

$$SM(Y_B, Y_K) \equiv PR_{Y_B}\{Y_B < Y_K\} \tag{5.25}$$

representing a *relative degree of conservativeness* between two methodologies.

VI. ANNEX I. Relation between sample size, rank and tolerance level for equal-tailed tolerance intervals based on conjugate order statistics

VI. ANNEX I. RELATION BETWEEN SAMPLE SIZE, RANK AND TOLERANCE LEVEL FOR EQUAL-TAILED TOLERANCE INTERVALS BASED ON CONJUGATE ORDER STATISTICS

Equal-tailed tolerance intervals, defined in section 1.11.2, can be (as in Wilks' method) constructed with order statistics (section 3.3.1.1.6). Let Y be a scalar and continuous random variable, with continuous CDF. We consider "symmetrical" order statistics $Y_{r:N}$ and $Y_{N-r+1:N}$, as endpoints of a (P, Q) equal-tailed tolerance interval, fulfilling the condition:

$$PR\left\{Y_{r:N} < Y_{\frac{1-P}{2}} \quad AND\ Y_{N-r+1:N} > Y_{\frac{1+P}{2}}\right\} \geq C \tag{I.1}$$

where Y_Q is the quantile of order Q of Y, 0<Q<1.

If [L, U] is an equal-tailed tolerance interval with level (P, C), then, simultaneously,

- [L,U] is a two-sided tolerance interval with level (P, C)
- L is a lower tolerance limit with level ((1-P)/2, C)
- U is an upper tolerance limit with level ((1+P)/2, C).

In the sequel, we will use extensively the following well-known result:

$$PR\{Y_{r:N} < Y_Q\ \} = PR\{beta(r, N-r+1) < Q\ \} = I_Q(r, N-r+1) \tag{I.2}$$

I_X (a,b) is the regularized incomplete beta function with arguments X, a and b.

Calling PR(EQ) the probability on the left-hand side of (I.1) it is easy to see that

$$\begin{aligned} PR(EQ) &= 1 - PR\left\{Y_{r:N} \geq Y_{\frac{1-P}{2}} \quad OR\ Y_{N-r+1:N} \leq Y_{\frac{1+P}{2}}\right\} = \\ &1 - PR\left\{Y_{r:N} \geq Y_{\frac{1-P}{2}}\ \right\} - PR\left\{Y_{N-r+1:N} \leq Y_{\frac{1+P}{2}}\right\} \\ &\quad + PR\left\{Y_{r:N} \geq Y_{\frac{1-P}{2}} \quad AND\ Y_{N-r+1:N} \leq Y_{\frac{1+P}{2}}\right\} \end{aligned} \tag{I.3}$$

Manipulation of (I.3) leads to

$$PR(EQ) = PR\left\{Y_{r:N} < Y_{\frac{1-P}{2}}\right\} - (1 - PR1) \cdot PR\left\{Y_{N-r+1:N} \leq Y_{\frac{1+P}{2}}\right\} \quad \text{(I.4)}$$

where

$$PR1 \equiv PR\left\{Y_{r:N} \geq Y_{\frac{1-P}{2}} \;\middle|\; Y_{N-r+1:N} \leq Y_{\frac{1+P}{2}}\right\} \quad \text{(I.5)}$$

is a conditional probability (as shown by the vertical bar).

Note that the condition $Y_{N-r+1:N} \leq Y_{\frac{1+P}{2}}$ is equivalent to the statement "there are at most (r-1) sample points higher than quantile of order (1+P)/2". Thus, PR1 can be easily developed in terms of the binomial distribution:

$$PR1 = \sum_{k=0}^{r-1} PR_k \cdot PR\left\{Y_{\frac{1-P}{2}} < Y_{r:N-k}\right\} \quad \text{(I.6)}$$

PR_k is the probability of k data being beyond the quantile of order (1+P)/2, thus calculable via the binomial distribution:

$$PR_k = \binom{N}{k}\left(\frac{1-P}{2}\right)^k\left(\frac{1+P}{2}\right)^{N-k} \quad \text{(I.7)}$$

Hence, (I.6) becomes

$$PR1 = \sum_{k=0}^{r-1} \binom{N}{k}\left(\frac{1-P}{2}\right)^k\left(\frac{1+P}{2}\right)^{N-k} PR\left\{Y_{\frac{1-P}{2}} < Y_{r:N-k}\right\} \quad \text{(I.8)}$$

And (I.4) transforms to

$$\begin{aligned} PR(EQ) = {} & PR\left\{Y_{r:N} < Y_{\frac{1-P}{2}}\right\} \\ & - \left[1 - \sum_{k=0}^{r-1}\binom{N}{k}\left(\frac{1-P}{2}\right)^k\left(\frac{1+P}{2}\right)^{N-k} PR\left\{Y_{\frac{1-P}{2}} < Y_{r:N-k}\right\}\right] \\ & \cdot PR\left\{Y_{N-r+1:N} \leq Y_{\frac{1+P}{2}}\right\} \end{aligned} \quad \text{(I.9)}$$

According to (I.2), (I.9) takes the form:

$$PR(EQ) = I_{\frac{1-P}{2}}(r, N-r+1) - \left[1 - \sum_{k=0}^{r-1} \binom{N}{k} \left(\frac{1-P}{2}\right)^k \left(\frac{1+P}{2}\right)^{N-k} \left(1 - I_{\frac{1-P}{2}}(r, N-k-r+1)\right)\right] \cdot I_{\frac{1+P}{2}}(N-r+1, r) \quad \text{(I.10)}$$

Using the following property of the incomplete beta function:

$$I_x(a,b) = 1 - I_{1-x}(b,a) \quad \text{(I.11)}$$

expression (I.10) is simplified to

$$PR(EQ) = 1 - I_{\frac{1+P}{2}}(N-r+1, r) \left[2 - \sum_{k=0}^{r-1} \binom{N}{k} \left(\frac{1-P}{2}\right)^k \left(\frac{1+P}{2}\right)^{N-k} \left(I_{\frac{1+P}{2}}(N-k-r+1, r)\right)\right] \quad \text{(I.12)}$$

Therefore, the condition (I.1) for a (P, Q) equal-tailed tolerance interval takes finally the form:

$$1 - I_{\frac{1+P}{2}}(N-r+1, r) \left[2 - \sum_{k=0}^{r-1} \binom{N}{k} \left(\frac{1-P}{2}\right)^k \left(\frac{1+P}{2}\right)^{N-k} \left(I_{\frac{1+P}{2}}(N-k-r+1, r)\right)\right] \geq C \quad \text{(I.13)}$$

VII. ANNEX II. Proof of a sufficient condition for the joint acceptance criterion (Section 3.3.2.1)

VII. ANNEX II. PROOF OF A SUFFICIENT CONDITION FOR THE JOINT ACCEPTANCE CRITERION (SECTION 3.3.2.1)

Let us suppose that each individual criterion has an own tolerance level

$$PR\left\{PR_{Y_j}\left(Y_j \in R_{A,j}\right) \geq P_j\right\} \geq C_j \qquad j = 1, \ldots, D \tag{II.1}$$

We will prove that (II.1) is sufficient condition for the joint criterion

$$PR\left\{PR\left\{\bigwedge_{i=1}^{D}\left(Y_{i,C} \in R_{A,i}\right)\right\} \geq P_0\right\} \geq C_0 \tag{II.2}$$

whenever

$$\begin{aligned} \sum_{i=1}^{D} P_i - D + 1 &\geq P_0 \\ \sum_{i=1}^{D} C_i - D + 1 &\geq C_0 \end{aligned} \tag{II.3}$$

holds.

To prove this statement, we define

$$\begin{aligned} S_i &\equiv "Y_i \in R_{A,i}" \\ W_i &\equiv "PR(S_i) \geq P_i" \end{aligned} \tag{II.4}$$

so that (II.1) and (II.2) are respectively written as

$$PR\{W_i\} \geq C_i \quad , \quad i = 1, \ldots, D \tag{II.5}$$

and

$$PR\left\{PR\left\{\bigwedge_{i=1}^{D} S_i\right\} \geq P_0\right\} \geq C_0 \tag{II.6}$$

From Bonferroni's inequality we have:

$$PR\left\{\bigwedge_{i=1}^{D} S_i\right\} \geq \sum_{i=1}^{D} PR(S_i) - D + 1 \tag{II.7}$$

And so

$$\bigwedge_{i=1}^{D} W_i \Rightarrow PR\left\{\bigwedge_{i=1}^{D} S_i\right\} \geq \sum_{i=1}^{D} P_i - D + 1 \tag{II.8}$$

and

$$PR\left\{\bigwedge_{i=1}^{D} W_i\right\} \leq PR\left\{PR\left\{\bigwedge_{i=1}^{D} S_i\right\} \geq \sum_{i=1}^{D} P_i - D + 1\right\} \tag{II.9}$$

Applying again Bonferroni's inequality, we have

$$PR\left\{\bigwedge_{i=1}^{D} W_i\right\} \geq \sum_{i=1}^{D} PR(W_i) - D + 1 \tag{II.10}$$

Combining (II.10) and (II.9)

$$PR\left\{PR\left\{\bigwedge_{i=1}^{D} S_i\right\} \geq \sum_{i=1}^{D} P_i - D + 1\right\} \geq \sum_{i=1}^{D} PR(W_i) - D + 1 \tag{II.11}$$

Thus, if (II.3) and (II.1) are true, the joint criterion (II.2) is true, as we wished to prove.

If we assume an homogeneous individual tolerance level, $P_j = P^*$ and $C_j = C^*$ for every j=1,…,D, the inequality (II.3) transforms to

$$\begin{aligned} P^* &\geq 1 - \frac{1 - P_0}{D} \\ C^* &\geq 1 - \frac{1 - C_0}{D} \end{aligned} \tag{II.12}$$

In the case that the individual criteria are tested independently, the probability in (II.2) is factorised

$$PR\left\{\prod_{j=1}^{D} PR\{Y_i \in R_{Ai}\} \geq P_0\right\} \geq C_0 \tag{II.13}$$

and it is easy to prove that (II.1) implies (II.2) if

$$\begin{aligned} \prod_{j=1}^{D} P_i &\geq P_0 \\ \prod_{j=1}^{D} C_i &\geq C_0 \end{aligned} \tag{II.14}$$

holds. If the individual tolerance level is homogeneous and equal to (P*, C*), (II.14) transforms to

$$\begin{aligned} P^* &\geq \sqrt[D]{P_0} \\ C^* &\geq \sqrt[D]{C_0} \end{aligned} \tag{II.15}$$

When P_0 and C_0 tend to 1, the lower bounds in (II.15) tend to those in (II.12).

VIII. References

VIII. REFERENCES

[1] Atwood C. L. *et al*. (2003), "Handbook of Parameter Estimation for Probabilistic Risk Assessment", NUREG/CR-6823, SAND2003-3348P, Sandia National Laboratories, US Nuclear Regulatory Commission, september 2003.

[2] Auder B., de Crecy A., Ioss B., Marquès M. (2010), "Screening and metamodeling of computer experiments with functional outputs. Application to thermal-hydraulic computations", *Reliability Engineering and System safety in the special SAMO 2010 issue.*

[3] Bahadur R. R. (1966), "A note on quantiles in large samples", *Ann. Math. Statist.* 37 577-580, 1966.

[4] Beran R., Hall P. (1993), "Interpolated Nonparametric Prediction Intervals and Confidence Intervals", *Journal of the Royal Statistical Society: Series B (Methodological)*, Volume 55, Issue 3, July 1993, Pages 643–652, https://doi.org/10.1111/j.2517-6161.1993.tb01929.x

[5] BIPM (2010), "Evaluation of measurement data — Guide to the expression of uncertainty in measurement", JCGM 100:2008, GUM 1995 with minor corrections, First edition 2008, corrected version 2010.

[6] Box G. E. P., Tiao G. C. (1992), "Bayesian Inference in Statistical Analysis", John Wiley & Sons, 1992.

[7] Boyack B. *et al.* (1989), "Quantifying Reactor Safety Margins. Application of Code Scaling, Applicability and Uncertainty Evaluation Methodology to a Large –Break, Loss-of-Coolant Accident", US Nuclear Regulatory Commission, NUREG/CR-5249, EGG-2552, 1989.

[8] Brown L. D., Cai T. T., DasGupta A. (2001), "Interval Estimation for a Binomial Proportion", *Statistical Science,* Vol. 16, No. 2, 101-133.

[9] Cannamela C., Garnier J., Iooss B. (2008), "Controlled Stratification for Quantile Estimation", *The Annals of Applied Statistics,* Vol. 2, No. 4, 1554-1580, 2008.

[10] Castillo E. (1988), "Extreme Value Theory in Engineering", Academic Press, Inc., 1988.

[11] Chen C., Tsong Y., Shen M. (2024), "Nonparametric Method to Estimate Tolerance Interval of Continuous Data of Unknown Distribution", Journal of Biotechnology and Biology 2024, Vol 2: 101. https://www.jscholaronline.org/articles/JBB/Nonparametric-Method-to-Estimate-Tolerance-Interval.pdf

[12] Chu F. (2010), "Variance Reduction Techniques for estimating Quantiles and Value-at-Risk". Ph. D. Thesis, Department of Information Systems, New Jersey Institute of Technology. May, 2010.

[13] Chu F., Nakayama M. (2010), "Confidence Intervals for Quantiles when applying Variance-Reduction Techniques", RESIM 2010.

[14] Clopper C.J., Pearson E. S. (1934), "The use of confidence or fiducial limits illustrated in the case of the binomial". *Biometrik,* 26, 404-413, 1934.

[15] Code of Federal Regulations, "General Design Criteria for Nuclear Power Plants", Appendix A to 10 CFR 50.

[16] Code of Federal Regulations (2007), "Acceptance criteria for emergency core cooling systems for light-water nuclear power reactors". 10 CFR 50.46, 2007.

[17] Code of Federal Regulations, "ECCS Evaluation Models", Appendix K to 10 CFR 50.

[18] David H. A., Nagaraja H. N. (2003), "Order Statistics", 3.ª edition, J. Wiley & Sons, 2003.

[19] Deng C. *et al.* (2014), "Uncertainty analysis using Beta-Bayesian approach in nuclear safety code validation", Annals of Nuclear Energy, Volume 70, 2014, pag. 103-108, ISSN 0306-4549, https://doi.org/10.1016/j.anucene.2014.03.005.

[20] Draper N. R., Smith H. (1998), "Applied Regression Analysis", 3rd Edition, Wiley Series in Probability and Statistics. A Wiley-Intescience publication, 1998.

[21] Durán-Vinuesa L. F., Cuervo D. (2021), "Uncertainty quantification and propagation with probability boxes", Nuclear Engineering and Technology, DOI:10.1016/j.net.2021.02.010, February 2021.

[22] Durga Rao K., Kuswaha H. S., Verma A. K., Srividya A. (2007), "Quantification of epistemic and aleatory uncertainties in level-1 probabilistic safety assessment studies", *Reliability Engineering and System Safety* 92(7), 947-956, 2007.

[23] Dusic, M., Hortal, J., Mendizábal, R., Pelayo, F. (2019), NARSIS Deliverable 3.8 – "Development and Description of E-BEPU Method – Part A theoretical basis", Vienna, 2019.

[24] Dusic, M., Hortal, J., Mendizábal, R., Pelayo, F. (2019), NARSIS Deliverable 3.9 – "Use of E-BEPU for Evaluation of Defence-in-depth", Vienna, 2019.

[25] Efron B., Tibshirani R. J. (1993), "An introduction to the Bootstrap", Chapman & Hall / CRC, 1993.

[26] Faber M. H. (2007), "Risk and Safety in Civil Engineering", Lecture Notes, Swiss Federal Institute of Technology Zürich, 2007.

[27] French A. B. (2010), "Response Surface Modeling of a Large Break Loos of Coolant Accident with Comparisons to Order Statistics approaches", *Proceedings of the ASME 2010 3rd Joint US-European Fluids Engineering Summer Meeting and 8th International Conference on Nanochannels, Microchannels, and Minichannels, FEDSM-ICNMM2010,* August 1-5, 2010, Montreal, Canada.

[28] Frepoli C. (2008), "An Overview of Westinghouse Realistic Large Break LOCA Evaluation Model", Science and Technology of Nuclear Installations. 2008. 10.1155/2008/498737

[29] Freixa J. *et al.* (2024), "Spanish contribution to the development and application of best estimate plus uncertainty methodologies: Past, present and future", Nuclear Engineering and Design, Volume 417, Febr 2024, 112837,ISSN 0029-5493, https://doi.org/10.1016/j.nucengdes.2023.112837

[30] Fynan D. A. (2014), "Uncertainty Quantification for Reactor Safety Analysis", PhD Thesis, University of Michigan, 2014.

[31] Gastwirth J. L., "The first-median test: A two-sided version of the control median test", *J. Amer. Statist. Ass.* 63, 692-706. [168].

[32] Glaeser H., "Uncertainty evaluation of thermal-hydraulic code results." Proceedings of International Meeting on Best Estimate Methods in Nuclear Installation Safety Analysis (BE-2000), Washington, D. C., (2000).

[33] Goubault-Larrecq J. "A Few Pearls in the Theory of Quasi-Metric Spaces", *Summer Topology Conference,* July 23-26, 2013.

[34] Grabaskas D., "Efficient Approaches to the Treatment of Uncertainty in Satisfying Regulatory Limits", PhD Dissertation, Ohio State University, 2012.

[35] Grabaskas D., Denning R., Aldemir T., "Techniques for the Efficient Assessment of the 0.95-Quantile in Safety Analyses", The 9th International Topical Meeting on Nuclear Thermal-Hydraulics, Operations and Safety Conference (NUTHOS-9), Kaohsiung, Taiwan, September 9-13, 2012.

[36] Guba A., Makai M., Pál L., "Statistical aspects of best estimate method-I", *Reliability Engineering and System Safety,* 80:217–32, 2003.

[37] Guttman I., "Statistical Tolerance Regions: Classical and Bayesian", Hafner Publishing Company, 1970.

[38] Hahn G. J., Meeker W. Q. (1991), "Statistical Intervals. A Guide for Practitioners", John Wiley & Sons, 1991.

[39] Han S., Kim T. n (2019), "Numerical Experiment on Two-Sided Tolerance Limit for Safety Analysis of Energy Generation Systems", International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, June 2019.

[40] Helton J. C., Johnson J. D., Oberkampf W. L., Sallaberry C. J. (2010), "Representation of analysis results involving aleatory and epistemic uncertainty", *International Journal of General Systems,* 39:6, 605-646.

[41] Hofer E. *et al.*: "Uncertainty and Sensitivity Analysis of Accident Consequence Submodels", *Proc of ANS/ENS International Topical Meeting on Probabilistic Safety Methods and Applications,* San Francisco 1985.

[42] Hong I. S., Oh D. Y., Kim I. G. (2013), "Generic Application of Wilks' Tolerance Limit Evaluation Approach to Nuclear Safety", *OECD/CSNI Workshop on Best Estimate Methods and Uncertainty Evaluations* - Workshop Proceedings, NEA/CSNI/R(2013)8/PART2

[43] Hortal J., Mendizábal R., Pelayo F., "What does "safety margin" really mean?", *Safety, Reliability and Risk Analysis: Theory, Methods and Applications* – Martorell et al (eds), CRC Press, 2009.

[44] Hutson, A. D. (1999), "Calculating nonparametric confidence intervals for quantiles using fractional order statistics". *Journal of Applied Statistics,* 26(3), pages 343–353, https://doi.org/10.1080/02664769922458

[45] Hyndman R. J., Fan Y. (1996), "Sample Quantiles in Statistical Packages", The American Statistician, November 1996, Vol. 50, No. 4, pp 361-365, https://www.amherst.edu/media/view/129116/original/Sample+Quantiles.pdf

[46] IAEA (2009), "Deterministic Safety Analysis for Nuclear Power Plants", Specific Safety Guide, SSG-2.

[47] Jílek, M., "Sample Size and Tolerance Limits", *Trabajos de Estadística y de Investigación Operativa,* Vol. 33, Núm. 2, 1982, pág 64 a 78.

[48] Kalos M.H., Wilcox P. A., "Monte Carlo Methods", 2nd edition, John Wiley & Sons, Inc, 2008.

[49] Kennedy M. C., O'Hagan A., "Bayesian Calibration of Computer Models", *Journal of the Royal Statistical Society Series B: Statistical Methodology,* Volume 63, Issue 3, September 2001, pages 425-464, https://doi.org/10.1111/1467-9868.00294

[50] Kohler, M., Krzyżak, A., Tent, R. *et al.* "Nonparametric quantile estimation using importance sampling". Ann Inst Stat Math 70, 439–465 (2018). https://doi.org/10.1007/s10463-016-0595-4

[51] Krebs J., "On the Bahadur representation of sample quantiles for score functionals." Bernoulli, 30(4) 3356-3377 November 2024. https://doi.org/10.3150/24-BEJ1718

[52] Krishnamoorthy K., Mathew T. (2009), "Statistical Tolerance Regions: Theory, Applications, and Computation" ISBN: 978-0-470-38026-0 John Wiley & Sons, Inc, 2009.

[53] Langewisch D. R., "Uncertainty and Sensitivity Analysis for Long-Running Computer Codes. A Critical Review", *MSc Thesis,* Massachussetts Institute of Technology, February, 2010.

[54] Makai M., Pál L., "Best estimate method and safety analysis II", *Reliability Engineering and System Safety,* 91:222–232, 2006.

[55] Martin R. P., Nutt W. T., "Perspectives on the application of order-statistics in best-estimate plus uncertainty nuclear safety analysis", *Nuclear Engineering and Design,* 241: 274–284, 2011.

[56] Martz H. F. et al., "Combining mechanistic best-estimate analysis and Level 1 probabilistic risk analysis", *Reliability Engineering and System Safety,* 39 (1993), 89-108.

[57] mathoverflow, "Parametric vs Non-parametric Estimation of Quantiles", https://mathoverflow.net/questions/48223/parametric-vs-non-parametric-estimation-of-quantiles

[58] McKay M. D. (1995), "Evaluating Prediction Uncertainty" NUREG/CR-6311, LA-12915-MS, Los Álamos National Laboratory. Prepared for USNRC.

[59] Mendizábal R., Pelayo F., "Regulatory Challenges Posed by the "Best-estimate Plus Uncertainty", Methodologies". *TOPSAFE*, Dubrovnik, 2008.

[60] Mendizábal R. (2009), "Probabilistic Safety Margins: definition and calculation", *Safety, Reliability and Risk Analysis: Theory, Methods and Applications* – Martorell *et al.* (eds), CRC Press, 2009.

[61] Mendizábal R. (2009), "BEPU methodologies and multiple outputs", *Proceedings of the 13th International Topical Meeting on Nuclear Reactor Thermal Hydraulics (NURETH-13),* Kanazawa City, Ishikawa Prefecture, Japan, September 27-October 2, 2009.

[62] Mendizábal R. (2012), "Quantifying the conservativeness of tolerance regions on BEPU methodologies", *TOPSAFE 2012*, Helsinki, 22-26 abril 2012.

[63] Mendizábal R. (2013), "Validation and BEPU Methodologies", *Proceedings of the 15th International Topical Meeting on Nuclear Reactor Thermal Hydraulics (NURETH-15),* Pisa, Italy, May 12-15, 2013.

[64] Mendizábal R. (2016), "Contribución al estudio de las metodologías de cálculo realista con incertidumbre ("BEPU") dentro del análisis determinista de seguridad de plantas nucleares". *PhD Thesis (in spanish),* ETSI Aeronáuticos, Universidad Politécnica de Madrid, 2016.

[65] Mendizábal R. (2016) "Surrogate models for uncertainty propagation in BEPU Methodologies: a regulatory view", The 11th International Topical Meeting on Nuclear Reactor Thermal Hydraulics, Operation and Safety (NUTHOS-11), Gyeongju, Korea, October 9-13, 2016.

[66] Mendizábal R. (2019), "Bayesian perspective in BEPU licensing analysis", Nuclear Engineering and Design, Volume 355, 2019, 110310, ISSN 0029-5493, https://doi.org/10.1016/j.nucengdes.2019.110310

[67] Mendizábal R. (2020), "Are 59 enough? Some remarks on Wilks method". Best Estimate Plus Uncertainty International Conference (BEPU 2020), Giardini Naxos, Sicily, Italy. (Not celebrated).

[68] Mendizábal R. (2022), "The fulfillment of multiple acceptance criteria by safety outputs of BEPU calculations", The 19th International Topical Meeting on Nuclear Reactor Thermal Hydraulics (NURETH-19), Brussels, Belgium, March 6 - 11, 2022.

[69] Mendizábal R. (2023), "Una descripción de las metodologías BEPU de análisis de accidentes", Revista Alfa, Consejo de Seguridad, Número 56, diciembre de 2023. (In Spanish). https://www.csn.es/csn/revista-alfa/56/-/asset_publisher/a75f/content/una-descripcion-de-las-metodologias-bepu-best-estimate-plus-uncertainty-de-analisis-de-accidentes

[70] Mendizábal R. (2024), "Equal-Tailed Tolerance Intervals and their use in BEPU Analysis", BEPU2024-437, Best Estimate Plus Uncertainty International Conference (BEPU 2024), Real Collegio, Lucca, Tuscany, Italy, May 19 – 24, 2024.

[71] Messer N., Champ M. (2000), "Assessment of the "Deterministic Realistic Method" applied to Large Break LOCA Analysis". *EUROSAFE KÖLN* 2000, 6-7 November 2000.

[72] Nakayama M. K. (2016), "Efficient Quantile Estimation When Applying Stratified Sampling and Conditional Monte Carlo, With Applications to Nuclear Safety", SIMUL 2016: The Eighth International Conference on Advances in System Simulation, Copyright (c) IARIA, 2016. ISBN: 978-1-61208-501-2.

[73] NEA (2007), "SMAP Task Group on Safety Margins Action Plan (SMAP) - Final Report", [NEA/CSNI/R(2007)9], 2007.

[74] NEA/CSNI (2011), "Safety Margin Evaluation – SMAP Framework Assessment and Application", NEA/CSNI/R(2011)3, July 2011.

[75] Noether G. E. (1948), "On Confidence Limits for Quantiles", *The Annals of Mathematical Statistics.* Vol 19, No 3, 416-419, 1948.

[76] Noether G. E. (1951), "On a Connection Between Confidence and Tolerance Intervals", *The Annals of Mathematical Statistics.* Vol 22, No 4, 603-604, 1951.

[77] Nutt W. T., Wallis G. B. (2004), "Evaluation of nuclear safety from the outputs of computer codes in the presence of uncertainties". *Reliability Engineering and System Safety,* 83:57-77, 2004.

[78] Oberkampf W. L., Roy C. J. (2010), "Verification and Validation in Scientific Computing", Cambridge University Press, 2010.

[79] Owen D. B. (1963), "Factors for One-sided Tolerance Limits and for Variable Sampling Plans", Sandia Corporation Monograph, SCR-607, March 1963.

[80] Pagani L. P. (2008), "On the Quantification of Safety Margins", PhD Thesis, MIT, 2008.

[81] Peña Sánchez de Rivera D. (1995), "Estadística. Modelos y Métodos. 2. Modelos lineales y series temporales", 2.ª Edición revisada. Alianza Universidad Textos, Alianza Editorial, 1995 (In Spanish).

[82] Pourgol-Mohammad M. (2007), "Integrated methodology for thermal-hydraulics uncertainty analysis (IMTHUA)", Doctoral dissertation, University of Maryland, College Park, MD. 2007.

[83] Ramu M., Prabhu R. V. (2013), "Metamodel based analysis and its applications: a review", *Acta Technica Corviniensis – Bulletin of Engineering,* Tome VI- FASCICYLE 2 [April – June], ISSN 2067-3809, 2013.

[84] Saltelli A., Ratto M., Andres T. H., Campolongo F., Cariboni J., Gatelli D., Saisana M. and Tarantola S. (2008), "Global Sensitivity Analysis: The Primer". John Wiley & Sons.

[85] Sánchez-Saez F. *et al.* (2015), "LBLOCA uncertainty analysis using meta models", *Proceedings of 16th International Topical Meeting on Nuclear Reactor Thermalhydraulics (NURETH-16),* Chicago, USA, August 30-September 4, 2015.

[86] Sánchez-Saez F., Sánchez A. I., Villanueva J. F., Carlos S., Martorell S. (2018), "Uncertainty analysis of a large break loss of coolant accident in a pressurized water reactor using non-parametric methods", Reliability Engineering & System Safety, Volume 174, 2018, Pages 19-28, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2018.02.005

[87] Scheffe H., Tukey J. W. (1945), "Non-Parametric Estimation. I. Validation of Order Statistics", *The Annals of Mathematical Statistics.* Vol 16, No 2, 187-192, 1945.

[88] Serfling R. J. (1980), "Approximation Theorems of Mathematical Statistics", John Wiley & Sons, Inc, 1980.

[89] Sheather S. J. and Marron J. S. (1990), "Kernel Quantile Estimators", *Journal of the American Statistical Association",* 85, 410-416, 1990.

[90] Shewhart W. A. (1939), "Statistical Method from the viewpoint of Quality Control". Edited by W. Edwards Deming, The Graduate School, the Department of Agriculture. Washington, 1939.

[91] Storlie, C. B., Swiler, L., Helton, J. C., Sallaberry, C. (2008), "Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models". SANDIA Report SAND2008-6570, Sandia Laboratories, Albuquerque, NM (2008).

[92] Storlie C. B., Swiler L. P., Helton J. C., Sallaberry C. J. (2009), "Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models", *Reliability Engineering and System Safety* 94, pp. 1735-1763 (2009).

[93] Technical Program Group. (1989), "Quantifying Reactor Safety Margins. Application of Code Scaling, Applicability, and Uncertainty Evaluation Methodology to a Large-Break LOCA", *NUREG/CR-5249,* Rev. 4, 1989.

[94] Tukey J. W. (1947), "Non-Parametric Estimation II. Statistically Equivalent Blocks and Tolerance Regions--The Continuous Case", *The Annals of Mathematical Statistics.* Vol 18, No 4, 529-539, 1947.

[95] Tukey J. W. (1948), "Nonparametric Estimation, III. Statistically Equivalent Blocks and Multivariate Tolerance Regions--The Discontinuous Case", *The Annals of Mathematical Statistics.* Vol 19, No 1 (1948), 30-39.

[96] U.S. Nuclear Regulatory Commission (1989), "Best-Estimate Calculations of Emergency Core Cooling System Performance", *Regulatory Guide* 1.157, 1989.

[97] U.S. Nuclear Regulatory Commission (2002), "An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the Licensing Basis", Regulatory Guide 1.174. Rev 1, 2002.

[98] U.S. Nuclear Regulatory Commission (2005), "Transient and Accident Analysis Methods", *Regulatory Guide 1.203*, revision 0, December 2005.

[99] U.S. Nuclear Regulatory Commission (2007), "Feasibility Study for a Risk-Informed and Performance-Based Regulatory Structure for Future Plant Licensing", Office of Nuclear Regulatory Research. *NUREG-1860.* 2007.

[100] U. S. Nuclear Regulatory Commission, "Standard Review Plan for the Review of Safety Analysis Reports for Nuclear Power Plants: LWR Edition", *NUREG-0800.*

[101] Villamizar M. *et al*. (2014), "LBLOCA sensitivity analysis using meta models"", *Proceedings of the European Nuclear Conference (ENC 2014)*, Marseille, France, 11-14, May 2014.

[102] Wald A. (1943), "An Extension of Wilks' Method for Setting Tolerance Limits", *The Annals of Mathematical Statistics.* Volume 14, Number 1, 45-55, 1943.

[103] Wallis G. B. (2006), "Evaluating the probability that the outputs of a computer code with random inputs will meet a set of evaluation criteria", *Reliability Engineering and System Safety* 91, 820-827, 2006.

[104] Wallis G. B. (2005), "Uncertainties and probabilities in nuclear reactor regulation", The 11th International Topical Meeting on Nuclear Reactor Thermal-Hydraulics (NURETH-11), Avignon, October 2-6, 2005.

[105] Wallis G. B. (2007), "Uncertainties and probabilities in nuclear reactor regulation", *Nuclear Engineering and Design,* 237, 1586-1592, 2007.

[106] Wilks S. S. (1941), "Determination of sample sizes for setting tolerance limits", The Annals of Mathematical Statistics, 12(1):91-96, 1941. Euclid: 1177731788.

[107] Wilks S. S. (1942), "Statistical prediction with special reference to the problem of tolerance limits", The Annals of Mathematical Statistics, 13(4):400-409, 1942. Euclid: 1177731537.

[108] Wilks S. S. (1948), Order statistics, *Bulletin of the American Mathematical Society.* Volume 54, Number 1, Part 1 (1948), 6-50.

[109] Yankov A. (2015), "Analysis of Reactor Simulations Using Surrogate Models", *PhD Thesis,* University of Michigan, 2015.

[110] Young, D. S. (2010). "tolerance: An R Package for Estimating Tolerance Intervals". *Journal of Statistical Software,* 36(5), 1–39. https://doi.org/10.18637/jss.v036.i05

[111] Young, D. S., & Mathew, T. (2014). "Improved nonparametric tolerance intervals based on interpolated and extrapolated order statistics". Journal of Nonparametric Statistics, 26(3), 415-432. https://doi.org/10.1080/10485252.2014.906594

[112] Young M. Y., Bajorek S. M., Nissley M. E., and Hochreiter L. E. (1998), "Application of code scaling applicability and uncertainty methodology to the large break loss of coolant," Nuclear Engineering and Design, vol. 186, no. 1-2, pp. 39–52, 1998.

[113] Zacks S. (1971), "The Theory of Statistical Inference", John Wiley & Sons, 1971.

[114] Zieliński R. (2004), "Optimal Quantile Estimators Small Sample Approach", IMPAN, 2004.

[115] Zieliński R., "L-statistics as Nonparametric Quantile Estimators", https://www.impan.pl/wydawnictwa/preprints_impan/p657.pdf

[116] Zio E., Di Maio F. (2008), "Bootstrap and Order Statistics for Quantifying Thermal-Hydraulic Code Uncertainties in the Estimation of Safety Margins", Hindawi Publishing Corporation. *Science and Technology of Nuclear Installations.* Volume 2008, Article ID 340164, 9 pages, doi:10.1155/2008/340164

[117] Zugazagoitia E., Queral C., Fernández-Cosials K., Gómez J., Durán L. F., Sánchez-Torrijos J., Posada J. M. (2020), "Uncertainty and sensitivity analysis of a PWR LOCA sequence using parametric and non-parametric methods", Reliability Engineering & System Safety, Volume 193, 2020, 106607, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2019.106607

[118] Zuo Y. (2015), "Bahadur representations for bootstrap quantiles", Metrika, 2015, vol. 78, issue 5, pages 597-610.

[119] Muñoz-Cobo, J.-L.; Mendizábal, R.; Miquel, A.; Berna, C.; Escrivá, A. (2017) "Use of the Principles of Maximum Entropy and Maximum Relative Entropy for the Determination of Uncertain Parameter Distributions in Engineering Applications". Entropy 2017, 19, 486 https://doi.org/10.3390/e19090486

[120] Stigler S. M. (1977), "Fractional Order Statistics, with Applications", Journal of the American Statistical Association, Vol. 72, No. 359 (Sep., 1977), pp. 544-550 (7 pages) Published By: Taylor & Francis, Ltd.

[121] Silverman B.W. (1986), "Density Estimation for Statistics and Data Analysis", Chapman and Hall. ISBN 9780412246203.

[122] Mendizábal, R. and Pelayo F. (2010), "BEPU Methodologies and Plant Technical Specifications." *Proceedings of the ASME 2010 3rd Joint US-European Fluids Engineering Summer Meeting collocated with 8th International Conference on Nanochannels, Microchannels, and Minichannels. ASME 2010 3rd Joint US-European Fluids Engineering Summer Meeting: Volume 1, Symposia – Parts A, B, and C.* Montreal, Quebec, Canada. August 1–5, 2010. pp. 1573-1579. ASME. https://doi.org/10.1115/FEDSM-ICNMM2010-31289

[123] Baccou, J., Zhang, J., Fillion, P., Damblin, G., Petruzzi, A., Mendizábal, R.,… Sandberg, N. (2020). "SAPIUM: A Generic Framework for a Practical and Transparent Quantification of Thermal-Hydraulic Code Model Input Uncertainty". *Nuclear Science and Engineering*, 194(8–9), 721–736.

[124] Technical University of Valencia, "Métodos alternativos para la cuantificación de incertidumbres en el análisis BEPU de accidentes. Modelos sustitutivos" (In Spanish), Informe de seguimiento [MASA2 UPV CSN 5], December 2024.

[125) Mendizábal, R. (2025), "Equal-Tailed Tolerance Intervals and their use in BEPU Analysis". Submitted to Nuclear Engineering and Design.

[126] McKay, M. D.; Beckman, R. J.; Conover, W. J. (1979). "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code". *Technometrics.* 21 (2). American Statistical Association: 239–245. doi:10.2307/1268522. ISSN 0040-1706. JSTOR 1268522. OSTI 5236110.

IX. List of acronyms

IX. LIST OF ACRONYMS

AM	Analysis Margin
BEPU	Best Estimate Plus Uncertainty
CCDF	Complementary Cumulative Distribution Function
CDF	Cumulative Distribution Function
CFR	Code of Federal Regulations
CMC	Crude or Pure Monte Carlo
CSAU	Code Scaling, Applicability and Uncertainty Evaluation Methodology
CSN	Consejo de Seguridad Nuclear
DC	Degree of Conservativeness
DBS	Design Basis Scenario
DSA	Deterministic Safety Analysis
EC	Expected Coverage.
ECCS	Emergency Core Cooling System
ECDF	Empirical Cumulative Distribution Function
EMDAP	Evaluation Model Development and Assessment Process.
ETTI	Equal-Tailed Tolerance Interval
FOM	Figure Of Merit
FOS	Fractional Order Statistics
GRS	Gesellschaft für Anlagen- und Reaktorsicherheit
IET	Integral Effects Test
i.i.d.	Independent and identically distributed
LHS	Latin Hypercube Sampling
LM	Licensing Margin
LOCA	Loss-of-Coolant Accident
LTL	Lower Tolerance Limit.
LWR	Light Water Reactor
MC	Monte Carlo
ME	Maximum Entropy
MEP	Maximum Entropy Principle
mm	*Minimum Minimorum*
NPP	Nuclear Power Plant
OS	Order Statistic, Ordered Statistic
PCT	Peak Cladding Temperature
PD	Probability Distribution
PDF	Probability Density Function
PT	Probability Theory

RAC	Regulatory Acceptance Criterion
SET	Separate Effects Test
SM	Safety Margin
SRS	Simple random sampling, simple random sample
TI	Tolerance Interval
TR	Tolerance Region
USNRC	US Nuclear Regulatory Commission
UTL	Upper Tolerance Limit
V&V	Verification and Validation

X. List of mathematical symbols

X. LIST OF MATHEMATICAL SYMBOLS

$B(R, T)$	Beta function with parameters R and T.
beta(A, B)	Beta random variable with parameters A and B.
C	Confidence level.
C_0	Regulatory confidence level.
D	Dimension.
F_Y	CDF of the random variable Y.
f_Y	PDF of the random variable Y.
$I_X(R, T)$	Incomplete beta function with parameters X, R and T.
K( )	Kernel function.
N	Sample size.
$N(\mu, \sigma^2)$	Normal random variable with mean μ and variance σ^2.
N(0, 1)	Standardized normal random variable.
N_{min}	Minimum sample size.
N_{mm}	*Minimum minimorum* sample size.
NE	Number of exceedances of an acceptance limit
P	Coverage level.
P_{exc}	Limit exceedance probability.
P_0	Regulatory coverage level.
PR_S	Probability according to the sampling distribution.
PR_Y	Probability according to the distribution probability of Y.
(P, C)	Tolerance level.
(P_0, C_0)	Regulatory Tolerance Level.
S	Second-level, sampling (subindex).
SM(A, B)	Safety Margin from A to B.
T_n	Student random variable with n degrees of freedom.
T(n, NC)	Student noncentral random variable with n degrees of freedom and noncentrality parameter NC.
Y_λ	Quantile of order λ of the random scalar variable Y, $0 < \lambda < 1$.
$Y_{r:N}$	Order statistic of ordinal r in a sample of size N of Y.
β	Coverage level.
γ	Confidence level.
(β, γ)	Tolerance level.
$\lfloor\ \rfloor$	Floor function (i.e. integer part).
$\lceil\ \rceil$	Ceiling function (i.e. integer part plus one).

Basis of *Best-Estimate Plus Uncertainties (BEPU)* Methodologies in Deterministic Safety Analysis of nuclear plants. Statistical methods in the uncertainty analysis

Colección
Otros Documentos CSN

CSN CONSEJO DE SEGURIDAD NUCLEAR