# USING STATISTICAL INFERENCE FOR DECISION MAKING IN BEST ESTIMATE ANALYSES

*P. Sermer*[+], *K. Weaver*[+], *F. Hoppe*[*], *C. Olive*[+] *and D. Quach*[+]

[+] Reactor and Radiation Physics Directorate, Nuclear Safety Solutions

[*] Department of Mathematics and Statistics, McMaster University

## Abstract

For broad classes of safety analysis problems, one needs to make decisions when faced with randomly varying quantities which are also subject to errors. The means for doing this involves a statistical approach which takes into account the nature of the physical problems, and the statistical constraints they impose. The paper describes the methodology for doing this which has been developed at Nuclear Safety Solutions, and draws some comparisons to other methods which are commonly used in Canada and internationally. Our methodology has the advantages of being robust and accurate and compares favourably to other best estimate methods.

## Introduction

The objective of this paper is to present a discussion on how statistical inference is used to deal with common classes of safety analysis and compliance problems. The approach is based on demonstrating the statistical basis for reaching decisions for two classes of problems where uncertainty and variability are present. The paper is presented in two parts. Part I develops the statistics needed to deal with a situation in which the source of uncertainty is due to the random nature of the computational or the measurement error (sometimes referred to as epistemic errors). In Part II, this is extended to the more complex situation where random behaviour arises not only from computational error but also within the variables themselves. In each of these situations we provide the statistical foundations needed to make cogent decisions.

The types of safety analysis problems considered, and where these sort of decisions need to be made, fall into two complementary groups. In the first group, one tries to demonstrate that the maximum value of a variable, or the appropriate proportion of its distribution, is less than a specified limit to within a specified confidence. In the second group, one tries to demonstrate that the minimum value of a variable, or the appropriate proportion of its distribution, is computed to within a specified confidence.

In the course of the discussion, the unique features of our approach are also pointed out, e.g., Monte Carlo simulation is used to estimate the error in the output variable. Comparisons are made to other commonly used statistical approaches for dealing with these problems. A feature of particular interest in our approach is the utilization of reactor operating history to provide a natural random sample to do our estimation. This contrasts with other approaches which produce such a random sample from postulated probability distributions.

# PART I: STATISTICAL ANALYSIS FOR A FIXED VARIABLE

Let $\boldsymbol{x} = [x_1, x_2, x_3, \ldots, x_M]$ be a fixed vector of variables that are non-random and determine another variable $y$ whose value is of interest in a particular safety analysis. We denote this relationship by

$$y = \mathcal{F}(\boldsymbol{x}), \tag{1}$$

where $\mathcal{F}$ represents the relationship in the physical model. (For example, $y$ may represent fuel temperature following a postulated LOCA. $\boldsymbol{x}$ represents initial channel powers, pressure drops and all the other variables that are involved in predicting $y$.) For the sake of simplicity it is assumed that the model $\mathcal{F}$ is "perfect" in a sense that if $\boldsymbol{x}$ were *true* (i.e., known without error) then $y$ would be the *true* value of the variable of interest.

Let $\boldsymbol{X}$ be the computed or measured value of $\boldsymbol{x}$ with an error $\delta = [\delta_1, \delta_2, \ldots, \delta_M]$, that is,

$$\boldsymbol{X} = \boldsymbol{x} + \delta. \tag{2}$$

The corresponding $Y$ is

$$Y = \mathcal{F}(\boldsymbol{X}), \tag{3}$$

and the error, $\varepsilon$, in using $Y$ in place of $y$ is given by

$$\varepsilon = Y - y = \mathcal{F}(\boldsymbol{X}) - \mathcal{F}(\boldsymbol{x}). \tag{4}$$

Therefore,

$$\varepsilon = \mathcal{F}(\boldsymbol{x} + \delta) - \mathcal{F}(\boldsymbol{x}), \tag{5}$$

$\boldsymbol{X}$ is assumed to be a random variable which arises either as a measurement or a value of an imperfect code. Hence, both $[\delta_1, \delta_2, \delta_3, \ldots, \delta_M]$ and $Y$ are also random variables. We refer to $\boldsymbol{X}$ and $Y$ as *best estimates*.

## 1. Statement of the statistical problem

Using $\boldsymbol{X}$, $Y$ and $\delta$ defined above, find an upper (A), or lower (B) confidence limit $Y_\beta$ for $y$. That is, for a given $\beta$ (such as $\beta = 0.95$), let

Problem (A): $\mathcal{P}[y \le Y_\beta] = \beta$,

or,

Problem (B): $\mathcal{P}[Y_\beta \le y] = \beta$.

($\mathcal{P}[E]$ denotes the probability of an event $E$.) The choice of the above problems depends on the the physical nature of the situation of interest - examples of both are presented further on.

## 2. Solution

For a given $\beta$, find an uncertainty $u = u(\beta)$ such that

$$Y_\beta = Y - u \tag{6}$$

solves the above problems. $u = u(\beta)$ is determined as a percentage point of the *pdf*$(\varepsilon)$ - the probability distribution function for $\varepsilon$ in (5). By definition, for $0 \ll \beta < 1$, the upper $100\beta$ percentage point $\varepsilon_\beta$ is given by

$$\mathcal{P}[\varepsilon \le \varepsilon_\beta] = \beta,$$

while the lower $100(1-\beta)$ percentage point $\varepsilon_{1-\beta}$ is given by

$$\mathcal{P}[\varepsilon \le \varepsilon_{1-\beta}] = 1-\beta.$$

Thus, for problem (A), we have,

$$\beta = \mathcal{P}[Y_\beta \ge y] = \mathcal{P}[Y - u \ge y] = \mathcal{P}[Y - y \ge u] = \mathcal{P}[\varepsilon \ge u] = 1 - \mathcal{P}[\varepsilon \le u].$$

Therefore,

$$\mathcal{P}[\varepsilon \le u] = 1-\beta,$$

and hence,

$$u = \varepsilon_{1-\beta}. \tag{7}$$

For problem (B), we have

$$\beta = \mathcal{P}[Y_\beta \le y] = \mathcal{P}[\varepsilon \le u],$$

and hence

$$u = \varepsilon_\beta. \tag{8}$$

If the *pdf*$(\varepsilon)$ were symmetric about zero then $\varepsilon_{1-\beta} = -\varepsilon_\beta$ and the uncertainties for both problems (A) and (B) would be of the same magnitude with opposite signs. However, in general, for many safety analysis problems, *pdf*$(\varepsilon)$ is not symmetric about zero. Typically, errors, such as $\delta$, are symmetric about zero, but the result of a possibly non-linear computation in (5), are neither centered about zero nor symmetric. In particular, if $\mathcal{F}$ involved maxima or minima then $\varepsilon$ would necessarily have a skewed distribution as discussed later in this paper.

## 3. Relationship between the statistical problem and a physical problem

The above two problems (A) and (B) represent typical problems encountered in safety related analyses. An example of Problem (A) would be the following. Let $L$ be a prescribed licence limit to define a safe operating envelope, then the condition on a particular variable $y$ (such as a maximum reactor channel power) that defines such an envelope is

$$y < L.$$

To satisfy this condition at any reactor state using the computed value $Y$ (i.e., the best estimate) it is natural to consider the null and alternate hypothesis [1], respectively, to be

$$H_o: \quad y \ge L,$$
$$H_A: \quad y < L.$$

If the significance level for the test of these hypotheses is required to be at most $1-\beta$ then the appropriate test statistic is provided by (6) and (7). That is,

$$Y_\beta = Y - \varepsilon_{1-\beta}.$$

Therefore, if $Y_\beta < L$, or

$$Y < L + \varepsilon_{1-\beta}, \tag{9}$$

then we reject the null hypothesis, $H_o$, in favour of the alternative $H_A$, that is, we can conclude $y < L$. It should be noted that in most cases $\varepsilon_{1-\beta}$ is negative and hence the alternative hypothesis will be accepted if the measured value $Y$ is sufficiently less than the target $L$. Nonetheless, it is possible for $\varepsilon_{1-\beta}$ to be positive and this leads to a decision rule that, while appearing to be counter-intuitive, is correct. For instance, if the error $\varepsilon$ is such that $Y$ greatly over-estimates $y$ then this would be the case.

The second problem (B) may represent components of a reactor special safety system. For example, the design intent of a regional, or, neutron over-power trip set-point for a CANDU reactor is to provide a high level of assurance that the reactor will trip prior to the onset of dry-out during a loss of regulation event. Thus, if $y$ represents the *true* neutron power for which the dry-out would actually occur, then $Y_\beta$ in (B) given by (6) and (8), i.e.,

$$Y_\beta = Y - \varepsilon_\beta,$$

would be the appropriate trip set-point to be installed. The required "high level of assurance" is expressed by $\beta$ in (B). *Y,* the best estimate, is the computed trip set-point under specified reactor conditions using data at a given reactor state [2], [3].

We observe that the statistical test (9) to test the above null hypothesis $H_o$ is unusual in statistical practice, in that it is based on a single observation $Y$ rather than a sample of observations and that $pdf(\varepsilon)$ is not known. In order to be able to perform this statistical test, one needs to have the ability to estimate $pdf(\varepsilon)$ in (5). How to do this is shown in the next section.

## 4. Estimation of the error in the best estimate

The error $\varepsilon$ in the best estimate $Y$ is given by (5). Since, in general, the (*true*) input variable $\boldsymbol{x}$ cannot be known to us, it is impossible to use (5) to compute $\varepsilon$. As argued in [4], or [5], essentially the only way to estimate $\varepsilon$ is to use "surrogate" *true* values instead of $\boldsymbol{x}$. These, of course, are given by $\boldsymbol{X}$ in (2). In order for this method to work we need to assume that $\mathcal{F}$ in (1) does not change rapidly in the neighbourhood of $\boldsymbol{x}$ defined by the error $\delta$. In the actual computation, the physical model $\mathcal{F}$ is represented by a mathematical formulation, say $\hat{\mathcal{F}}$. A fundamental assumption is that, phenomenologically, $\hat{\mathcal{F}}$ is a good representation of $\mathcal{F}$. This assumption is noted in [4], or [6]. Using the surrogate values $\boldsymbol{X}$ and the known errors $\delta$ we obtain the estimated error $\hat{\varepsilon}$ using

$$\hat{\varepsilon} = \hat{\mathcal{F}}(\boldsymbol{X} + \delta) - \hat{\mathcal{F}}(\boldsymbol{X}). \tag{10}$$

The interpretation of the above assumptions on $\mathcal{F}$ and $\hat{\mathcal{F}}$ is that $pdf(\hat{\varepsilon}) \approx pdf(\varepsilon)$, or, that the differences between corresponding percentage points for $pdf(\varepsilon)$ and $pdf(\hat{\varepsilon})$, respectively, are negligible.

In principle, for given $\boldsymbol{X}$, the probability density function, $pdf(\hat{\varepsilon})$, can easily be obtained by a Monte Carlo simulation. That is, randomly sample the known distribution for $\delta = [\delta_1, \delta_2, ..., \delta_M]$ to obtain a random sample for $\hat{\varepsilon}$ using (10). This random sample then defines $pdf(\hat{\varepsilon})$. Since the convergence of a Monte Carlo algorithm is rather slow (the error of the approximation behaves like $O(N^{-1/2})$ where N is the number of simulations), at least 10000 simulations are needed to get reasonable results.

If evaluation of $\hat{\mathcal{F}}$ is significantly computationally intensive then the Monte Carlo algorithm may not be practical. In this case we may use the response surface approach [7] in which $\hat{\mathcal{F}}$ is replaced by a "simpler" function in the neighborhood of $\boldsymbol{X}$ which is easy to evaluate making the Monte Carlo simulation feasible. Alternatively, one could use the non-parametric order-statistics approach to estimate $\varepsilon_{1-\beta}$ or $\varepsilon_{\beta}$ (used elsewhere in best estimate and uncertainty analyses - BEAU, such as a GRS method [8]).

## 5. Comparison to existing methods

In order to compute $Y_{\beta}$ to solve either a problem (A) or (B), we need to be able to estimate $\varepsilon$ - see (6), (7), (8) and (10). To the best of our knowledge, no other work in the area of BEAU analyses estimates the error in the best estimate in order to provide rigorously defined statistical confidence limits on the desired variable y.

Traditional uncertainty analyses using a Monte Carlo simulation are based on a heuristic approach to estimate the required confidence limits [9]. Using our terminology, the approach is based on a Monte Carlo simulation to obtain a probability distribution for a variable *U*, given by

$$U = \hat{\mathcal{F}}(\boldsymbol{X} + \delta) \tag{11}$$

for a given $\boldsymbol{X}$. For a desired $\beta$, an upper, or lower, percentage point $U_{\beta}$ is computed, i.e.,

$$\mathcal{P}[U \le U_{\beta}] = \beta, \tag{12}$$

or,

$$\mathcal{P}[U_{\beta} \le U] = \beta, \tag{13}$$

depending on the nature of the underlying physical problem that is required to be solved.

We suggest that problems (A) and (B) are statistically rigorous representations of these underlying physical problems, while (12) and (13) are only heuristic or "intuitive" representations of such problems. The latter appears to be predicated on identifying *U* with *y* in (12) or (13) and an assumption that the underlying distributions are all symmetric. (This is only an attempt to understand the reasoning behind the traditional approach.) While *U* in (11) is an approximation to *y* in (1), it certainly is not *y*. We will show that, in general, the rigorous and heuristic problems are significantly different by showing that $Y_{\beta}$ in (6) is different from $U_{\beta}$ in the corresponding problem (12) or (13). In fact, we will show that the two are the same only in a very special case when the probability distribution for $\varepsilon$ in (5) is symmetric about zero. While the distributions for the errors $\delta = [\delta_1, \delta_2, ..., \delta_M]$ could very well be symmetric about zero, $\varepsilon$ will be symmetric about zero only if, in addition, $\mathcal{F}$ can reasonably be represented by a linear function in the $\delta$-neighborhood of $\boldsymbol{x}$. For a very important class of problems in safety analyses where $\mathcal{F}$ includes maximization (or

minimization) over a region of the reactor core, $\varepsilon$ possesses an extreme value probability distribution. Such a distribution is not symmetric and has a non-zero mean [10]. We will also show that $U_\beta$ is a conservative solution in that it bounds $Y_\beta$.

Using (10), (11) and (3), the estimate $\hat{\varepsilon}$ of the error $\varepsilon$ is

$$\hat{\varepsilon} = U - Y,$$

or,

$$Y = U - \hat{\varepsilon}.$$

(Clearly, the best estimate $Y$ in (3) can only be obtained with $\hat{\mathcal{F}}$.) For problem (A) we have

$$Y_\beta = Y - \varepsilon_{1-\beta} = U - \hat{\varepsilon} - \varepsilon_{1-\beta},$$

from (6) and (7). Hence,

$$U = Y_\beta + \hat{\varepsilon} + \varepsilon_{1-\beta}.$$

Substituting this into (12), the definition of $U_\beta$ for problem (A), we get

$$\beta = \mathcal{P}[U_\beta \geq U] = \mathcal{P}[U_\beta \geq Y_\beta + \hat{\varepsilon} + \varepsilon_{1-\beta}] = \mathcal{P}[\hat{\varepsilon} \leq U_\beta - Y_\beta - \varepsilon_{1-\beta}].$$

Therefore, $U_\beta - Y_\beta - \varepsilon_{1-\beta}$ is an upper $100\beta$ percentage point of $\hat{\varepsilon}$, i.e.,

$$U_\beta - Y_\beta - \varepsilon_{1-\beta} = \hat{\varepsilon}_\beta,$$

or,

$$U_\beta - Y_\beta = \hat{\varepsilon}_\beta + \varepsilon_{1-\beta},$$

Based on our fundamental assumption, the percentage points for $\hat{\varepsilon}$ and $\varepsilon$ are nearly the same and hence

$$U_\beta - Y_\beta = \varepsilon_\beta + \varepsilon_{1-\beta}. \tag{14}$$

For problem (B) we get the same result. For,

$$\beta = \mathcal{P}[U_\beta \leq U] = \mathcal{P}[U_\beta \leq Y_\beta + \hat{\varepsilon} + \varepsilon_\beta] = \mathcal{P}[\hat{\varepsilon} \geq U_\beta - Y_\beta - \varepsilon_\beta]$$
$$= 1 - \mathcal{P}[\hat{\varepsilon} \leq U_\beta - Y_\beta - \varepsilon_\beta].$$

Therefore,

$$\mathcal{P}[\hat{\varepsilon} \leq U_\beta - Y_\beta - \varepsilon_\beta] = 1 - \beta,$$

which is a definition of the lower 100(1-$\beta$) percentage point. Thus,

$$U_\beta - Y_\beta - \varepsilon_\beta = \hat{\varepsilon}_{1-\beta}.$$

leading to the same result as in (14)

Clearly, if $\varepsilon$ were symmetric about zero then, by definition,

$$\varepsilon_\beta = -\varepsilon_{1-\beta},$$

and hence,

$$U_\beta = Y_\beta.$$

Thus, the interpretation that problems (12) and (13) solve problems (A) and (B), respectively, can be made only under the assumption that the error in estimating the required variable is symmetric with a zero mean. In general, $\varepsilon_\beta + \varepsilon_{1-\beta}$ is not zero and the "traditional" approach will necessarily produce different solutions from the ones we are proposing.

There is an important class of safety analysis problems for which the function $\mathcal{F}$ involves taking extreme values over a portion of the reactor core. Our earlier two examples (channel power compliance problem and trip set-point calculation) are of this nature. We will show that for such class of problems, the traditional BEAU approach produces conservative solutions.

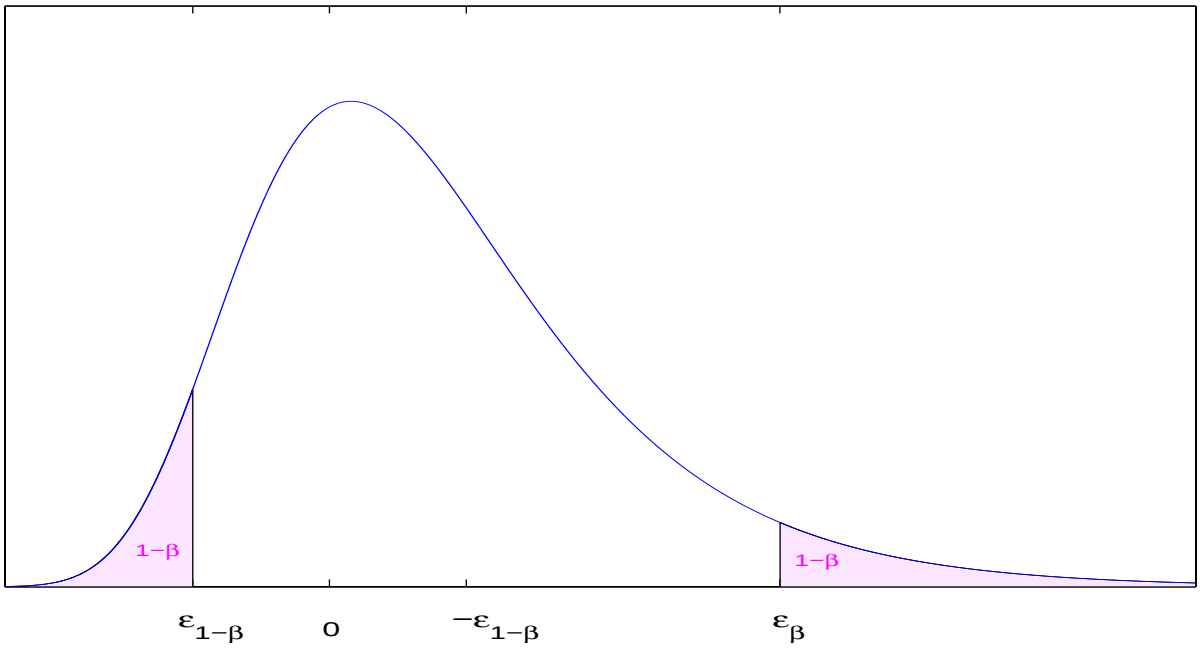
A problem, such as channel power compliance with the licence limit, which requires maximization over a portion of the reactor core would, typically, require a solution of the form characterized by problem (A). The error $\varepsilon$ in (10) is obtained from the (parental) errors $\delta$ and involves maximization (based on the form of $\mathcal{F}$). Therefore, it must possess extreme value distribution with a positive mean [5]. Moreover it is also positively skewed (see Figure below). This means that

$$\varepsilon_\beta > -\varepsilon_{1-\beta},$$

or,

$$\varepsilon_\beta + \varepsilon_{1-\beta} > 0.$$

**Problem (A): Probability Density Function for $\varepsilon$**




Using this inequality in (14), it is evident that

$$U_\beta > Y_\beta.$$

In conjunction with problem (A), this result means that

$$\mathcal{P}[U_\beta \geq y] > \beta,$$

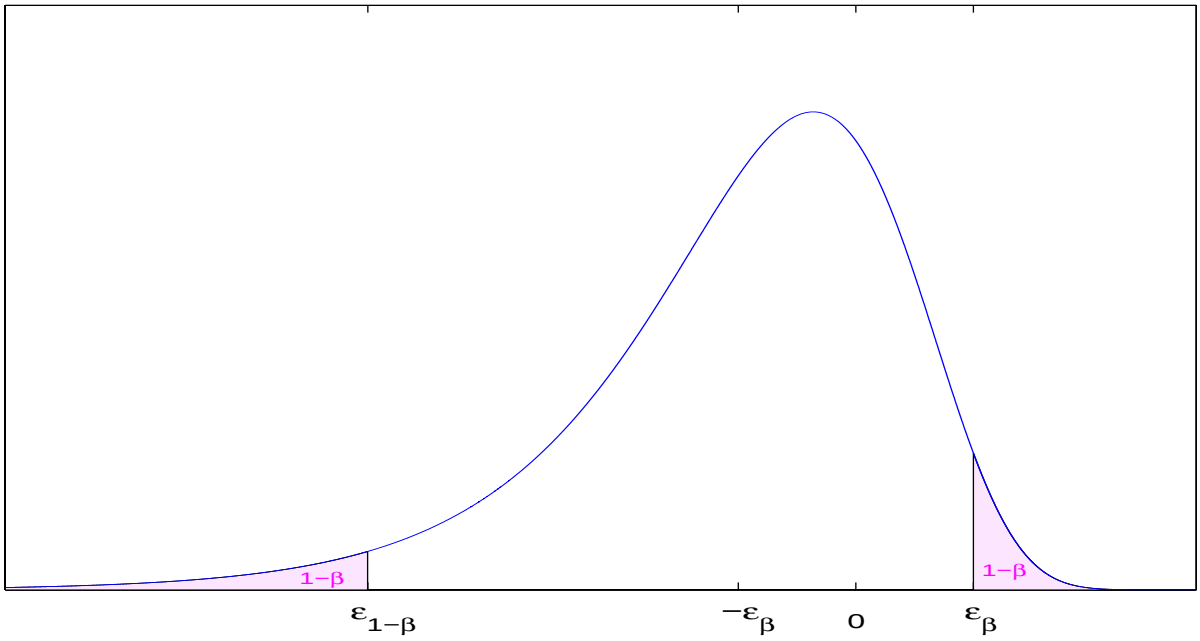
showing that $U_\beta$ is a conservative solution because it exceeds y with a probability that is larger than what is required.

Another problem, such as the computation of the neutron over-power trip-setpoint which requires minimization of the margin to dry-out over a portion of the reactor core would, typically, require a solution of the form characterized by problem (B). For this problem, the definition of the error $\varepsilon$ in (10) involves minimization. Therefore, it must possess extreme value distribution with a negative mean and is negatively skewed (see Figure below). This means that $\varepsilon_{1-\beta} < -\varepsilon_\beta$, or, $\varepsilon_\beta + \varepsilon_{1-\beta} < 0$. Using this inequality in (14), we see that $U_\beta < Y_\beta$. In conjunction with problem (B), this means that

$$\mathcal{P}[U_\beta \le y] > \beta .$$

This shows that $U_\beta$ is a conservative solution because it is smaller than y with a probability that is larger than what is required.

**Problem (B): Probability Density Function for $\varepsilon$**




## PART II: STATISTICAL ANALYSIS FOR A RANDOM VARIABLE

The previous analysis for Problem (A) is extended to a situation where $y$ is itself a random variable with a probability density function $g(y)$. Since, in principle, $y$ covers the whole real line, or, at least all positive numbers, we need to identify a percentage point, say $y_\gamma$, for some (positive) number $\gamma < 1$. $\gamma$ signifies the proportion of all $y$ of interest. (See figure below). For example, $y$ may represent (*true*) fuel sheath temperature under LOCA. Since there are variations in LOCA outcome due to variations in the normal reactor operating states, there is a variation of associated fuel

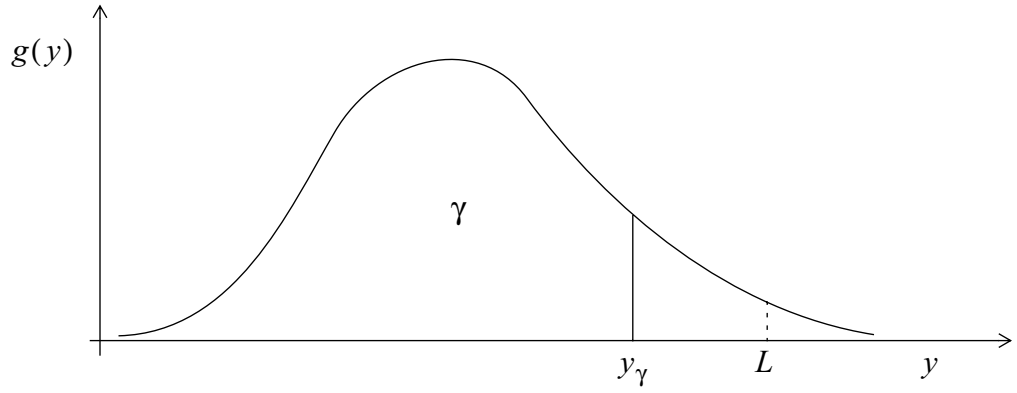
sheath temperatures $y$. This variation is represented by the probability density function $g(y)$. We are also given a fraction $\gamma$ (such as 0.95 or 0.99) where $1 - \gamma$ represents the fraction of LOCAs for which a fuel channel failure would be tolerated. If $L$ represents a temperature at which the fuel sheath fails (or is considered failed) then, naturally, we would like

$$y_\gamma \le L . \tag{15}$$

The extension of Problem (A) is to find an upper tolerance limit, say $Y_{\gamma,\beta}$, such that

$$\mathcal{P}[y_\gamma \le Y_{\gamma,\beta}] = \beta , \tag{16}$$

for some given (large) $\beta < 1$, such as 0.95. The above probability is related to the (cumulative) probability function for the estimate $Y_{\gamma,\beta}$. If the computed value of $Y_{\gamma,\beta}$ happens to be less than $L$ then we postulate that $\gamma$ proportion of $y$ lies below the limit $L$ at the safety level $(\gamma, \beta)$. This notion of safety level is presented in [6].




If we had a random sample of $y$ values and, say, $g(y)$ were normal, we could use [11] to compute the tolerance limit $Y_{\gamma,\beta}$. Alternatively, one could use a non-parametric (distribution-free) order statistics approach [12] to estimate $Y_{\gamma,\beta}$. However, in reality, it is not possible to know $y$ and therefore we need to obtain the tolerance limit $Y_{\gamma,\beta}$ based on the observed (computed) random sample from $Y$. We will refer to equations (1) through (5), where the input variable $\boldsymbol{x} = [x_1, x_2, x_3, ..., x_M]$ will now be a vector valued random variable and, as a result, $y$ in (1) will be a random variable. Of course, this will make $\boldsymbol{X}$ in (2), $Y$ in (3) and $\varepsilon$ in (5) random variables with much richer structures than the corresponding ones in the previous Section.

## 6. Algorithm for computing $Y_{\gamma,\beta}$

Let us assume that we have an independent sample $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, ..., \boldsymbol{x}_n$ of the input variables. Such a random sample may be derived from a set of operating reactor states where an appropriate input variable for state $k$ is $\boldsymbol{x}_k$. As a result, we have random samples $\{y_k\}$, $\{\boldsymbol{X}_k\}$, $\{Y_k\}$ and $\{\varepsilon(\boldsymbol{x}_k)\}$, given by, respectively,

$$y_k = \mathcal{F}(\boldsymbol{x}_k),$$
$$\boldsymbol{X}_k = \boldsymbol{x}_k + \delta,$$

$$Y_k = \mathcal{F}(\boldsymbol{X}_k),$$
$$\varepsilon(\boldsymbol{x}_k) = \mathcal{F}(\boldsymbol{x}_k + \delta) - \mathcal{F}(\boldsymbol{x}_k).$$

Note that an estimate of $\varepsilon(\boldsymbol{x}_k)$ can be obtained similarly as in (10) by taking surrogate values $\boldsymbol{X} = \boldsymbol{X}_k$ for each reactor state.

From the Monte Carlo simulations, which allow us to obtain the *pdf*s for $\varepsilon(\boldsymbol{x}_k)$ for all reactor states, we can obtain samples $\{\mu_k\}$ and $\{\sigma_k\}$, defined by,

$$\mu_k = E[\varepsilon(\boldsymbol{x}_k)],$$
$$\sigma_k^2 = Var[\varepsilon(\boldsymbol{x}_k)],$$

respectively. For a given $\gamma$ and $\beta$, we define a sample of uncertainties

$$u_k = \mu_k + q\sigma_k,$$

for some $q = q(\gamma, \beta)$. ($q$ is the same for every state and is chosen in a way shown below.)

Following the prescription in (6), we define

$$Y_{\gamma, \beta, k} = Y_k - u_k$$

for every reactor state *k*. The required tolerance limit $Y_{\gamma, \beta}$ in (16) is obtained from

$$Y_{\gamma, \beta} = \text{upper } 100\gamma \text{ percentile of the sample } \{Y_{\gamma, \beta, k}\}. \tag{17}$$

## 7. Determination of $q = q(\gamma, \beta)$

The tolerance limit $Y_{\gamma, \beta}$ in (17) above will satisfy (16) only if $q = q(\gamma, \beta)$ is appropriately chosen. In order to find such an appropriate $q = q(\gamma, \beta)$, we will seek to express $Y_{\gamma, \beta}$ in the form

$$Y_{\gamma, \beta} = y_\gamma + h_0(r_\gamma, q) + \frac{1}{\sqrt{n}} h(r_\gamma, q) z, \tag{18}$$

where $z$ is the standard normal variate, i.e., $z \sim \mathcal{N}(0, 1)$. $n$ is the sample size and $h_0 = h_0(r_\gamma, q)$, $h = h(r_\gamma, q) > 0$ are some suitable functions of $r_\gamma$ and $q$. $r_\gamma$ is defined as

$$r_\gamma = \frac{y_\gamma - E[y]}{\sqrt{Var[y]}}.$$

(Functions $h_0$ and $h > 0$ in (18) can be shown to exist for some practical problems of interest under the assumption that the sample size *n* is sufficiently large so that the central limit theorem [1] can be used to approximate the average values.)

Substituting (18) into (16), we get,

$$\mathcal{P}\left[ y_\gamma \le y_\gamma + h_0(r_\gamma, q) + \frac{1}{\sqrt{n}} h(r_\gamma, q) z \right] = \beta,$$

and hence

$$\mathcal{P}\left[ z \geq -\frac{h_0(r_\gamma, q)}{h(r_\gamma, q)}\sqrt{n} \right] = \beta,$$

or,

$$\mathcal{P}\left[ z \leq -\frac{h_0(r_\gamma, q)}{h(r_\gamma, q)}\sqrt{n} \right] = 1 - \beta.$$

The percentage point $z_{1-\beta}$ is defined as $\mathcal{P}[z \leq z_{1-\beta}] = 1 - \beta$ and by symmetry $z_{1-\beta} = -z_\beta$. Hence,

$$\frac{h_0(r_\gamma, q)}{h(r_\gamma, q)}\sqrt{n} = z_\beta,$$

or,

$$h_0(r_\gamma, q)\sqrt{n} - z_\beta h(r_\gamma, q) = 0. \tag{19}$$

Equation (19) is to be solved for $q$.

We note that for the solution $q = q(\gamma, \beta)$, we have

$$h_0(r_\gamma, q) = \frac{z_\beta}{\sqrt{n}} h(r_\gamma, q).$$

Therefore, the expression for $Y_{\gamma,\beta}$ in (18) becomes

$$Y_{\gamma,\beta} = y_\gamma + \frac{1}{\sqrt{n}} h(r_\gamma, q)(z + z_\beta),$$

For a sufficiently large number of reactor states and certain assumptions on the size of some parameters (which hold for some practical problems of interest, such as the compliance with maximum channel power licence limit alluded to in the previous Section) it is possible to derive $q = q(\gamma, \beta)$ explicitly. Since the derivation is rather involved we will only give a solution for the sake of interest. Namely, it can be shown that the solution to (19) is

$$q = \frac{r_\gamma}{\sqrt{1 + \kappa^2} + \kappa} - \frac{z_\beta}{\sqrt{n}} \sqrt{1 + \kappa^2 + \frac{r_\gamma^2}{2} \frac{1 + \kappa^4}{1 + \kappa^2}},$$

where,

$$\kappa = n \frac{\sqrt{Var[y]}}{\sum \sigma_k}.$$

Note that all the above terms in the above formulae can be readily approximated.

## 8. Conclusion

We have described a methodology which provides a rigorous statistical treatment applicable to a variety of safety analysis problems. This treatment has been applied successfully to decision making using a BEAU approach. Historical reactor operational history provides a natural random sample leading to accurate results. In our approach, Monte Carlo simulation is used to derive the error separately to obtain any desired confidence limit for the best estimate.

## References

[1] I. Guttman, S. S. Wilks and J. S. Hunter, *Introductory Engineering Statistics*, John Wiley & Sons, New York, 1971.

[2] P. Sermer, G. Balog, D. Novog, E.A. Attia and M. Levine, Monte Carlo Computation of Neutron Overpower Protection Trip Set-Points Using Extreme Value Statistics," CNS 24th Annual Conference, 2003.

[3] D.R. Novog and P. Sermer, "A Statistical Methodology for Determination of Safety Systems Actuation Setpoints Based on Extreme Value Statistics," to appear in *Science and Technology of Nuclear Installations*, 2008.

[4] W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes*, Cambridge University Press, Cambridge 1986, pp. 529 - 532.

[5] P. Sermer, C. Olive and F.M. Hoppe, Efficient Compliance with Prescribed Bounds on Operational Parameters by Means of Hypothesis Testing using Reactor Data, Proceedings of the CNS 2000 Annual Conference, Toronto, June 11 - 14, 2000.

[6] Yuri Orechwa, "Best-Estimate Analysis and Decision Making Under Uncertainty," US NRC, Best Estimates 2004, Washington, D.C., November 14-18, 2004, pp. 1-8.

[7] M.D. McKay, "Sensitivity and Uncertainty Analysis Using a Statistical Sample of Input Values, *Uncertainty Analysis*, ed., Yigal Ronen, CRC Press, Inc., pp. 145 - 186.

[8] Horst G. Glaeser, Large Break LOCA Uncertainty Evaluation and Comparison with Conservative Calculation," GRS, Best Estimates 2004, Washington, D.C., November 14-18, 2004, pp. 118-123.

[9] J.C. Luxat, R.G. Huget and F. Tran, "Development and Application of Ontario Power Generation's Best Estimate Nuclear Safety Analysis Methodology," Int. Meeting on Best Estimate Methods in Nuclear Installation Safety Analysis, Washington DC, November 2000.

[10] J.H. Pickles, An Introduction to Extreme Value Statistics," TPRD/L/2778/N84, Central Electricity Generating Board, 1985.

[11] G. Hahn, "Understanding Statistical Intervals," Industrial Engineering, December 1970, pp. 45-48.

[12] S.S. Wilks, "Determination of sample sizes for setting tolerance limits," Ann. Math. Statist. 12 (1941), pp. 91-96.