Introduction to Uncertainty Quantification
Module 1.2: Deciphering Aleatory & Epistemic Uncertainty

1 The Challenge

Regardless of the setting, most decision-making problems (especially in engineering) contain many sources of uncertainty that are both aleatory and epistemic. Unraveling these many sources of uncertainty to understand the nature of each source can be very difficult. Even in cases where the sources of uncertainty are relatively easy to identify (they are not always easy to identify), it's not always clear whether they are aleatory or epistemic in nature. What makes this all the more challenging is that, as argued by der Kiureghian and Ditlevsen [1], the determination of whether a given uncertainty is aleatory or epistemic often depends on the setting and context. In one setting, an uncertainty may be treated as aleatory while in another setting the same uncertainty may be treated as epistemic. Hence, the determination of whether an uncertainty should be treated as aleatory or epistemic is a *choice* made by the analyst/modeler/decision-maker that depends on circumstances. How then can we set guidelines for making the appropriate choice for a given application and the objectives of the investigation?

Let's begin by considering an example, as initially elucidated by Faber [2] and later discussed in [1]. In the design and manufacturing of structural components, material strength is critical to component performance but can have significant uncertainty. During the design phase, a material supplier will provide general materials specifications that (should) include at least some form of quantified uncertainties in the material strength (e.g. tolerances, confidence intervals, or error bars). Analysis of a generic component will then consider material strength as aleatory where the strength of each manufactured component can be considered as a random realization of the material strength following the provided uncertainties. However, the nature of the uncertainty changes if we are instead interested in the strength of a specific manufactured component. If I pull one component from the assembly line, the uncertainty associated with that component is epistemic. Why? Because it has a fixed material strength that can be tested. So, we see that the same uncertainty (strength of component material) can have either aleatory or epistemic uncertainty depending on the circumstances. Moreover, the nature of that uncertainty can change from aleatory to epistemic over time. Early in time (during the design phase), the uncertainty is aleatory but as individual components are manufactured their uncertainty becomes epistemic.

To decipher aleatory and epistemic uncertainty, we therefore need to establish the specific context in which we will consider uncertainty. This requires us to make some choices about how we set up a framework for uncertainty quantification. We will establish this general setting in the next section.

2 General Setting

Here, we will work in a general setting where our decision-making process is aided by some model. We will generically denote this model by $M(\boldsymbol{x}, \boldsymbol{\theta}_M)$ where $\boldsymbol{x}$ denotes a vector of input variables to the model and $\boldsymbol{\theta}_M$ is a vector of model parameters. The output of model $M(\boldsymbol{x}, \boldsymbol{\theta}_M)$ is a vector $\boldsymbol{y}$ of variables used to aid in decision-making. Hence, the model $M(\boldsymbol{x}, \boldsymbol{\theta}_M)$ in intended to represent any decision-making tool that takes specified inputs and parameters ($\boldsymbol{x}$ and $\boldsymbol{\theta}_M$, respectively) and returns some (hopefully) helpful information about the problem ($\boldsymbol{y}$). That is, our general setting can be expressed in simple terms as:

$$\boldsymbol{y} = M(\boldsymbol{x}, \boldsymbol{\theta}_M) \tag{1}$$

This means that $M(\boldsymbol{x}, \boldsymbol{\theta}_M)$ can represent many different types of models depending on the application. In engineering, $M(\boldsymbol{x}, \boldsymbol{\theta}_M)$ will very-often represent a physics-based computer simulation. In statistics or data science, $M(\boldsymbol{x}, \boldsymbol{\theta}_M)$ may represent a mathematical model (e.g. a regressor) that has been fit to some data set. Meanwhile in business, $M(\boldsymbol{x}, \boldsymbol{\theta}_M)$ may represent the revenue projection for a company or the projected cost of a project (e.g. construction).

Within in this model setting, we specifically define three important quantities. First, the vector $\boldsymbol{x}$ is composed of a set of *basic variables*. Basic variables serve as inputs to the model and are assumed to be directly observable. That is, they are quantities that can be obtained through testing, measurements, or other observations. Moreover, they are not derived from some other quantity or model. In engineering applications, basic variables may include material properties, the loads or demands on the system, or the geometry of a component. In data science or statistics, these may represent for example demographic information about users of a given system while in business these may represent sales figures, insurance claims, or purchasing quantities.

The second component is, of course, the model itself, which processes the basic variables to produce some output, referred to as *derived quantities*. The model may also have a set of parameters, again denoted by $\boldsymbol{\theta}_M$, but we will discuss these later. The derived quantities are the third element of this model process and represent the variables we use to make decisions. These derived quantities are often referred to as *quantities of interest* (or QoIs), which may come directly from the model or may be derived from the output of the model. In engineering analysis, the derived quantities (or QoIs) are often related to the performance of the engineering system. They may, for example, represent the state of stress in a material or the deflection of a structure. In business, the QoIs may be revenue projections or projections of construction/manufacturing costs.

3 Sources of Uncertainty

In this general model setting, uncertainty can enter in several ways. Here, we will specifically look at three different primary sources of uncertainty:

- *Uncertainty in Basic Variables:* This specifically relates to uncertainty in the input to the model.
- *Model-form Uncertainty:* This is uncertainty in the assumptions made when defining either the model. Here, we will delineate two different types of model-form uncertainty – uncertainty associated with the model M itself and uncertainty associated with the model employed for the basic variables.
- *Parameter Uncertainty:* Typically, both the model and the basic variable model have a set of parameters that may be uncertain.

3.1 Basic Variables

Let's begin with the input to the model, the basic variables $\boldsymbol{x}$. Uncertainty in the basic variables is perhaps the most common form of uncertainty that is addressed in uncertainty quantification. In the case of forward uncertainty propagation, which we will see in future modules, we aim to answer the following question:

Given uncertainty in $\boldsymbol{x}$, what is the uncertainty in $\boldsymbol{y} = M(\boldsymbol{x}, \boldsymbol{\theta}_M)$?

This may arise, for example, when the loading or excitation on an engineering system is uncertain. We may consider, for example, a long-span bridge that is being loaded by a random wind pressure. In systems engineering, we may consider the demand on a complex network such as a power grid or a water distribution system.

Regardless of the application, uncertainties in the basic variables are typically (though not always) treated using probability theory. When defining the basic variables probabilistically, they become random variables (which we will define in the next module) and we modify our notation to denote these random variables by $\boldsymbol{X}$. To define these random variables, we need to define a probability model. This probability model is defined through the probability density function (PDF) denoted $f_X(\boldsymbol{x}, \boldsymbol{\theta}_f)$ (or equivalently the cumulative distribution function, CDF, $F_X(\boldsymbol{x}, \boldsymbol{\theta}_F)$). We will define the PDF and CDF in the next module. The important thing to emphasize here is that we need to define a model for the probabilities of $\boldsymbol{X}$ and this model introduces some additional parameters $\boldsymbol{\theta}_f$, which we will discuss further below.

There are, of course, other approaches for modeling uncertainty in the basic variables that do not use probability. Some of these approaches will be discussed later. Nonetheless, uncertainty in the basic variables is usually treated with a model and this model often has an associated set of parameters, so the general paradigm does not change although the mathematics do.

Basic variables can have either aleatory or epistemic uncertainty, or both. Their treatment as aleatory or epistemic often depends on the circumstances. When dealing with basic variables abstractly, in which case the input to the system is not deterministic and each realization of the basic variables corresponds to a random event, uncertainties are treated as aleatory. This commonly occurs, for example, when considering uncertainty in the future load or demand on a system. Another example of this is uncertainty in the geometry of manufactured components from an assembly line. The geometry of each component will be slightly different and likely random, despite undergoing an identical manufacturing process. However, when dealing with specific realizations of the basic variables, where the basic variables can be measured, their uncertainties are considered epistemic. This may arise, for example, when considering the geometry of a specific structural component that has uncertainty. For this component, we may be able to reduce the uncertainty by measuring its precise geometry.

3.2 Model-Form Uncertainty

In our setting, we have now defined two models. We have a system model, $M(\boldsymbol{x}, \boldsymbol{\theta}_M)$, that is used to describe the behavior of our system and we have an uncertainty model for the basic variables, $\boldsymbol{x}$, that enter the system model. But, we emphasize that all models are mathematical idealizations and hence they are not a perfect replacement for reality. This causes uncertainty that we refer to as *model-form uncertainty*. In our setting, we have model-form uncertainty in both of these models.

3.2.1 Model-Form Uncertainty in Basic Variable Models

Let's begin by discussing model-form uncertainty in the basic variable model. Let's further assume that this model takes the form of a probability density function. The uncertainty here arises from the fact that the PDF we use to model the uncertainty may not be a perfect reflection of the true distribution. This can arise in several ways. For example, we may assume a specific form for the probability density function. Alternatively, the uncertainty model may be estimated by fitting a model to existing data drawn from random observations of the basic variables. This fit may not be perfect. This model-form uncertainty is typically considered epistemic because it can be reduced by assigning a better, more accurate probability model. This may be made possible, for example, by collecting more data and obtaining a better fit to the data.

3.2.2 Model-Form Uncertainty in System Models

The system model, $M(\boldsymbol{x}, \boldsymbol{\theta}_M)$, is an idealization of the system that we're studying. It therefore has numerous potential sources of error and uncertainty. At the most basic level, there may be mechanisms or physics that are simply missing from the model. This is common in physics-based modeling, where critical simplifications in complex physics are made. For example, rather than modeling the full complexity of a turbulent fluid, a model may consider a simpler model that simply incorporates the effect of the mean fluid pressure. Moreover, the system model may ignore certain factors or variables altogether. They may be ignored for several reasons. Perhaps these factors are expected to have only secondary influence on the phenomenon being studied, or maybe they aren't well-understood and/or cannot be measured. Finally, the model may be under-resolved. That is, the model may use coarse numerical methods (e.g. a coarse discretization) to solve the system and the numerical techniques themselves introduce approximation error and uncertainties.

A common way to address these myriad mode-form uncertainties is to reformulate the system model as:

$$\boldsymbol{y} = \hat{M}(\boldsymbol{x}, \boldsymbol{\theta}_M) + \epsilon \tag{2}$$

where the derived quantity $\boldsymbol{y}$ is now expressed in terms of the approximate model, now denoted by $\hat{M}(\cdot)$ plus an error term ϵ used to capture the influence of the combined errors described above. Often, this error term, ϵ can be reduced by improving the approximate model and hence the uncertainty is, at least partly, epistemic. It may not be entirely epistemic though because ϵ encapsulate components of the uncertainty that arise from randomness in the model or from irreducible uncertainties associated with inability to observe or measure certain aspects of the system.

3.3 Parameter Uncertainty

Both the uncertainty model for the basic variables and the system model usually have some parameters that need to be estimated. These parameters are estimated by fitting the model to observed data. If the case of the uncertainty model for the basic variables is given by a PDF, $f_X(\boldsymbol{x}, \boldsymbol{\theta}_f)$, this involves estimating the parameters of the distribution $\boldsymbol{\theta}_f$. For example, if the data follow normal distribution, the mean and standard deviation are estimated from repeated random observations of the basic variables, which again are directly observable. For the system model, $M(\boldsymbol{x}, \boldsymbol{\theta}_M)$, this involves estimating the parameters $\boldsymbol{\theta}_M$ from pairs of observations $(\boldsymbol{x}, \boldsymbol{y})$ or from a related experiment that isolates the specific parameter being calibrated. This usually requires an experiment that allows direct observation of the true derived quantities, or a closely related quantity, for specified values of the basic variables. These pairs $(\boldsymbol{x}, \boldsymbol{y})$ are not derived from the model, as this would involve calibrating the model to itself.

In either case, parameter estimation can be done in several ways. In the simplest case, the parameters are simply assigned by an expert according to their judgment. This, however, may introduce bias and relies on a highly subjective perspective that may not adequately account for uncertainty. A more mathematically rigorous approach is to determine the parameters through optimization. This involves establishing a so-called objective function that is used to minimize the difference between the model and the observations. This is a robust and widely used approach, but usually gives a deterministic set of parameters that does not adequately account for uncertainty. The preferred approach here is to use a process called Bayesian Inference. We'll talk about this more later, but it is preferred because it treats the parameters as random variables and provides a direct measure of uncertainty in the parameter values.

As suggested by the parameter estimation process, parameter uncertainties are typically epistemic because the uncertainty in parameter estimates usually reduces as more data is collected. This often induces a trade-off between the cost of data collection and the level of uncertainty that we are willing to accept in estimated model parameters.

4 Aleatory or Epistemic? Does it Matter?

Given our model setting and the various forms of uncertainty that factor into it, it is reasonable to ask, "Why is it important to decipher between aleatory and epistemic uncertainty?" Furthermore, from a philosophical perspective, one might reasonably argue that all uncertainty derives from a lack of knowledge and is therefore epistemic. So, we might further conclude that the distinction is arbitrary.

As we've already seen, the distinction is indeed subjective. But it can be useful make the distinction for several reasons. Categorizing uncertainties as aleatory or epistemic, for a given problem, helps us to grasp where we can realistically expect to reduce uncertainties and where we cannot. This can further help us to prioritize data collection efforts; focusing experimental efforts, information gathering, etc. on those parts of our model setting that will have a tangible benefit by reducing uncertainty in the variables we use to make decisions. It can also help us to recognize where our models can be improved and where they cannot. If a certain component of our model is simply deemed too complex to improve, we might reasonably attribute aleatory uncertainty to it and treat it as randomly varying. Finally, as we'll see in the next lesson, categorizing uncertainties as aleatory or epistemic can help us to determine the appropriate mathematical approach to quantifying the uncertainties. Aleatory and epistemic uncertainties are often treated with different mathematical tools, and making the distinction allows us to select the right tool for each source of uncertainty.

References

[1] A. Der Kiureghian, O. Ditlevsen, Aleatory or epistemic? does it matter?, Structural safety 31 (2) (2009) 105–112.

[2] M. H. Faber, On the treatment of uncertainties and probabilities in engineering decision analysis (2005).

Nomenclature

Functions

$M(\cdot)$	A generic model with vector input $\boldsymbol{x}$, model parameters $\boldsymbol{\theta}_M$, and vector-valued output $\boldsymbol{y}$
$\hat{M}(\cdot)$	Approximation of the model $M(\cdot)$
$f_X(\cdot)$	Probability density function (PDF) of a random variable X
$F_X(\cdot)$	Cumulative distribution function (CDF) of a random variable X

Variables

$\boldsymbol{x}$	Basic variables
$\boldsymbol{y}$	Derived quantities, also known as quantities of interest (QoIs)
$\boldsymbol{\theta}_f$	Parameters of the model (or function) f
ϵ	Error term
$\boldsymbol{X}$	A random variable