CHAPTER 1

ROBUST ESTIMATORS IN THE PRESENCE OF UNCERTAINTIES

Contents

In the previous chapter, we raised the problem of misspecification of the numerical model with respect to the reality. Moreover, this misspecification is random by nature, and we wish that the calibrated model shows relatively good performances when the environmental variables varies.

1.1 Defining robustness

1.1.1 Classifying the uncertainties

In the Bayesian formulation of the problem, the uncertainty on the calibration parameter is modelled through the prior distribution, while the uncertain parameter, u has its own distribution. While mathematically similar, those two representations actually encompasses a significant difference: we are actively trying to reduce the uncertainty of the calibration parameter by Bayesian update, while the uncertainty on the environmental parameter is seen as a nuisance.

In that context, the very notion of uncertainty can be roughly split in two, as described in [WHR+03]:

- Aleatoric uncertainties, coming from the inherent variability of a phenomenon, *e.g.* intrinsic randomness of some environmental variables

- Epistemic uncertainties coming from a lack of knowledge about the properties and conditions of the phenomenon underlying the behaviour of the system under study

According to this distinction, the epistemic uncertainty can be reduced by investigating the effect of the calibration parameter θ upon the physical system, and choose it accordingly to an objective function. The uncertain variable on the other hand is uncertain in the aleatoric sense, and cannot be controlled directly, as its value is doomed to change. This distinction illustrated **??** is a bit rough, as [KD09] point out that deciding the type of uncertainties is up to the modeller, who decide on which parameters inference is worth doing.

1.1.2 Robustness or reliability ?

The notion of *Robustness* is dependent on the field in which it is used. Worse, in the same community, robustness may carry a lot of different meanings. Robust is often used to describe something that behaves still nicely under uncertainties, or to put it in an other way, that is insensitive up to certain extent to some perturbations.

For instance, Bayesian approaches are sometimes criticized for their use of subjective probabilities that represent the state of beliefs, especially on the choice of prior distributions. In that sense, robust Bayesian analysis aims at quantifying the sensitivity of the choice of the prior distribution on the resulting inference and relative Bayesian quantities derived. In the statistical community, robustness is often implied as the non-sensitivity on the outliers in the sample set.
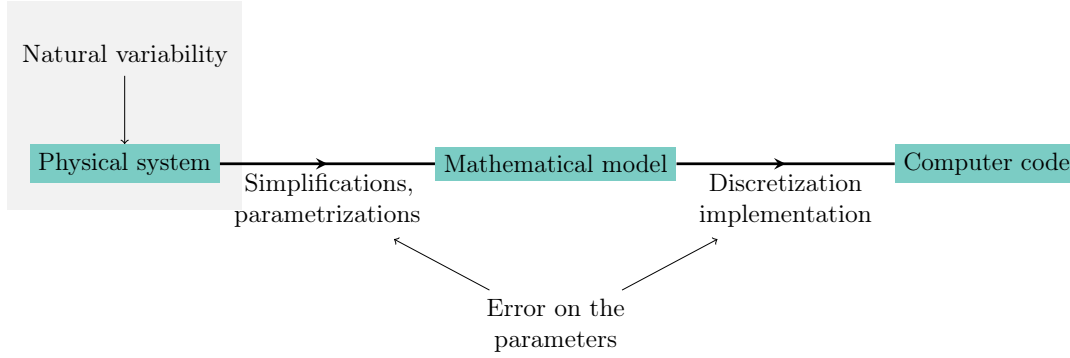
Figure 1.1: Sources of uncertainties and errors in the modelling. The natural variability of the physical system can be seen as aleatoric uncertainties, and the errors on the parameters as epistemic uncertainties

Moreover, robustness is often linked and sometimes confused to the semantically close notion of *reliability*. In [LBM$^+$16] we can find summarized in Table 1.1 the difference between these notions, by defining optimality as the deterministic counterpart of robustness, and admissibility as the counterpart of reliability.

		Robustness		
		no objective	objective	
			deterministic inputs	uncertain inputs
Constrained	Unconstrained	No problem	Optimal	Robust
	deterministic constraints	Admissible	Optimal and admissible	Robust and admissible
	uncertain constraints	Reliable	Optimal and reliable	Robust and reliable

(Reliability — left vertical axis label)

Table 1.1: Types of problems, depending on their deterministic nature for the constraints or the objective. Shaded cells correspond to problems comprising an uncertain part. Reproduced from [LBM$^+$16]

1.1.3   Robustness under parameteric misspecification

Given a family of models $\{(\mathcal{M}(\cdot, u), \Theta), u \in \mathbb{U}\}$, we can derive a problem of parameter estimation for each u. As detailed in **??**, we can look for the MLE or the MAP, and formulate the likelihood and the posterior distribution accordingly.

Not taking into account the uncertainty on u may be an issue in the modelling, especially if the influence of this variable is non-negligible. Choosing a specific $u \in \mathbb{U}$ leads to *localized optimization* [HB01] and *overcalibration*, that is choosing a value $\hat{\theta}$ that is optimal for the given situation (which is induced by u). This value does not carry the optimality to other situations, or in layman's term according to [ALMP$^+$12], being lured by "fool's gold". In geophysics and especially in hydrological models, this overcalibration may lead to the appearance of abberations in the predictions as those uncertainties become prevalent sources of errors. In hydrology, uncertainties are the principal culprit

of the existence of "Hydrological monsters" [KRTK10], that are calibrated models that perform really badly.

There are various ways to treat those environmental parameters. We are first going to see how to get rid of this dependence from a frequentist (likelihood-based) and Bayesian point of view.

1.2   Nuisance parameters

The environmental parameters are sometimes called *nuisance* parameters. Dealing with nuisance parameters usually means to eliminate them from the estimation. We will first detail likelihood-based methods of eliminating and their extension to Bayesian framework.

1.2.1   Profile and integrated Likelihood

From a frequentist approach, we define the joint likelihood $\mathcal{L}(\theta, u; y) = p_{Y|\theta,U}(y \mid \theta, u)$. There are two common ways to get rid of the nuisance parameters: one by *profiling*, one by *marginalization*. Profiling implies to perform first a maximization of the likelihood with respect to the nuisance parameters:

$$\mathcal{L}_{\text{profile}}(\theta; y) = \max_{u \in \mathbb{U}} \mathcal{L}(\theta, u; y) \tag{1.1}$$

and

$$\hat{\theta}_{\text{prMLE}} = \arg \max_{\theta \in \Theta} \mathcal{L}_{\text{profile}}(\theta; y) \tag{1.2}$$

In other words, considering the most favorable case of the likelihood given the nuisance parameters. Comparing the MLE over $\Theta \times \mathbb{U}$ for the original joint likelihood and the profile MLE on Θ for the profile likelihood, it is straightforward to verify that their components on Θ coincide as

$$\max_{(\theta, u) \in \Theta \times \mathbb{U}} \mathcal{L}(\theta, u; y) = \max_{\theta \in \Theta} \mathcal{L}_{\text{profile}}(\theta; y) \tag{1.3}$$

The resulting estimator does not take into account the uncertainty upon u, and can perform quite badly when the likelihood presents sharp ridges [BLW99].

Another alternative is to define the *integrated* (or marginalized) likelihood as

$$\mathcal{L}_{\text{integrated}}(\theta; y) = \int_{\mathbb{U}} \mathcal{L}(\theta, u; y) p_U(u) \, \mathrm{d}u \tag{1.4}$$

$$= \int_{\mathbb{U}} p_{Y|\theta,U}(y \mid \theta, u) p_U(u) \, \mathrm{d}u \tag{1.5}$$

$$= \int_{\mathbb{U}} p_{Y,U|\theta}(y, u \mid \theta) \, \mathrm{d}u \tag{1.6}$$

$$= p_{Y|\theta}(y \mid \theta) \tag{1.7}$$

and by maximizing this function,

$$\hat{\theta}_{\text{intMLE}} = \arg\max_{\theta \in \Theta} \mathcal{L}_{\text{integrated}}(\theta; y) \tag{1.8}$$

The profile and integrated likelihood have been computed for the following likelihood:

$$Y \mid \theta, U \sim \mathcal{N}(\theta + u^2, 2^2) \tag{1.9}$$

and the observations $y = [y_1, \cdots, y_{10}]$ have been observed using $\theta + u^2 = 1$. We set $\Theta = [-5, 5]$ and $\mathbb{U} = [-2, 2]$. The likelihood evaluated on $\Theta \times \mathbb{U}$ is displayed Fig. 1.2, with the integrated and profile likelihood. We can see that there is not unicity of the
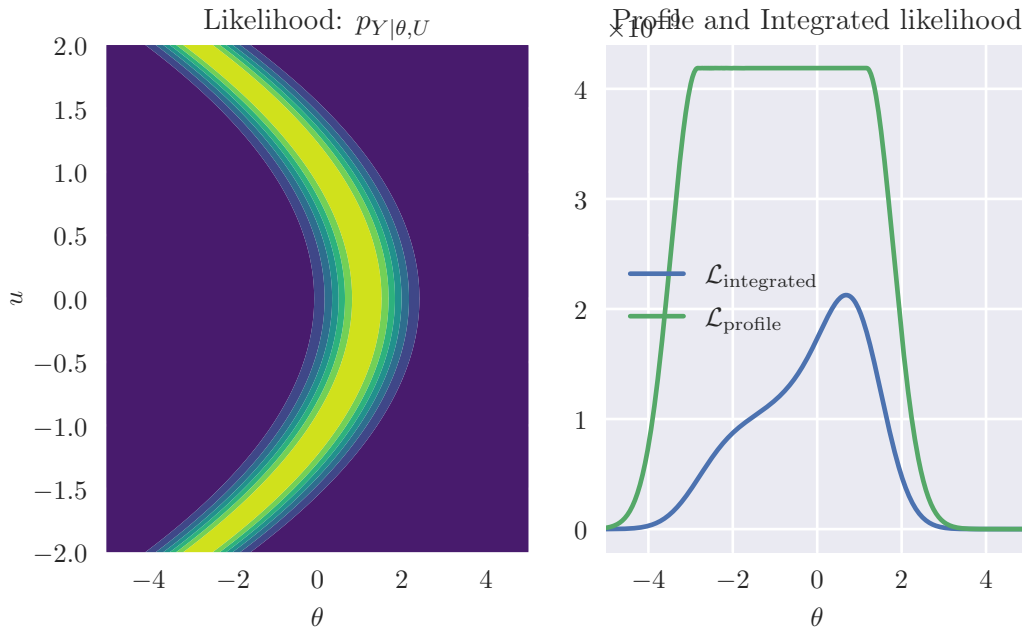

Figure 1.2:   Profile and integrated likelihood for an uniform nuisance parameter

maximizer of the profile likelihood: $\mathcal{L}_{\text{profile}}(\theta; y)$ is constant for $\theta \in [-3, 1]$. For the integrated likelihood, there is a unique maximum, attained for $\hat{\theta}_{\text{intMLE}} \approx 0.8$.

1.2.2   Joint posterior distribution

Again, by introducing a prior distribution on θ: p_θ, we can derive the posterior distribution. We assume that U and θ are independent: $p_{\theta,U} = p_\theta p_U$. The likelihood of the data given θ and u is

$$\mathcal{L}(\theta, u; y) = p_{Y|\theta,U}(y \mid \theta, u) \tag{1.10}$$

The joint posterior distribution can be written as:

$$p_{\theta,U|Y}(\theta, u \mid y) = \mathcal{L}(\theta, u; y) p_\theta(\theta) p_U(u) \frac{1}{p_Y(y)} \tag{1.11}$$

$$\propto \mathcal{L}(\theta, u; y) p_\theta(\theta) p_U(u) \tag{1.12}$$

Here, the posterior is used to do inference on θ and u jointly, so in order to suppress the dependency in u, we can marginalize and get the marginalized posterior $p_{\theta|Y}$:

$$p_{\theta|Y}(\theta \mid y) = \int_{\mathbb{U}} p_{\theta,U|Y}(\theta, u \mid y) \, \mathrm{d}u \tag{1.13}$$

$$= \int_{\mathbb{U}} p_{\theta|Y,U,}(\theta \mid y, u) p_{U|Y}(u \mid y) \, \mathrm{d}u \tag{1.14}$$

We can then define the marginalized maximum as posteriori (MMAP) [DGR02] as the maximizer of this marginalized posterior, written by omitting the model evidence:

$$\hat{\theta}_{\mathrm{MMAP}} = \arg \max_{\theta \in \Theta} p_{\theta|Y}(\theta \mid y) \tag{1.15}$$

or, by taking the negative logarithm to get a minimization problem:

$$\hat{\theta}_{\mathrm{MMAP}} = \arg \min_{\theta \in \Theta} -\log \left(\int_{\mathbb{U}} p_{\theta|Y,U}(\theta \mid y, u) p_{U|Y}(u \mid y) \, \mathrm{d}u \right) \tag{1.16}$$

Unfortunately, neither the integration with respect to the nuisance parameter in Eq. (1.13) nor the subsequent optimization is analytically easy. [DGR02] proposes a MCMC method in order to estimate iteratively the MMAP, through sampling of the joint posterior.

1.3 The cost function as a random variable

We discussed so far the calibration problem with nuisance parameters in the formulation of the likelihood or the posterior distribution. We are now going to focus on a "variational" approach, using a cost function

$$J : \Theta \times \mathbb{U} \to \mathbb{R}^+ \tag{1.17}$$

This function in a context of calibration is measuring the misfit between the data y, and the forward operator. However, the methods introduced in the following are not specific to this context, and J can represent other undesirable properties, such as instability or drag in airfoil design optimization for instance. This work is largely based on [TAVD20], which was submitted for review in February 2020.

All in all, $J(\theta, u)$ represent the cost of taking the decision $\theta \in \Theta$ when the environmental variable is equal to u. We are going to make several assumptions on this function:

- Θ is convex and bounded

- For all $\theta \in \Theta$ and $u \in \mathbb{U}$, $J(\theta, u) > 0$

- For all $\theta \in \Theta$, $J(\theta, \cdot)$ is measurable

- For all $\theta \in \Theta$, $J(\theta, U) \in L^p(\mathbb{P}_U)$ and $p \geq 2$. So for each θ, mean and variance exist and are finite.

Most of existing methods require first to remove the dependency on the uncertain variable, by defining a *robust* (and deterministic) counterpart of the minimization problem under uncertainty, that we can solve using classical methods of optimization.

1.3.1 Decision under uncertainty

We will first detail some estimators that can arguably be said robust, even though the random nature of U is not taken into account. This uncertainty is modelled through the principle of indifference, stating that no information whatsoever is available on u, or the distribution of U.

1.3.1.a Global optimization

A global optimization criterion, as its name suggests, advocates for minimizing the cost function over the whole space $\Theta \times \mathbb{U}$.

$$\min_{(\theta,u)\in\Theta\times\mathbb{U}} J(\theta, u) \tag{1.18}$$

This strategy is optimistic in the sense that we minimize over only the most favourable cases.

$$\hat{\theta}_{\text{global}} = \arg\min_{\theta\in\Theta} \min_{u\in\mathbb{U}} J(\theta, u) \tag{1.19}$$

1.3.1.b Worst-case optimization

As global optimization is inhenrently optimistic, we can easily derive a criterion which is pessimistic, in the sense that we want to minimize over the least favourable cases. We are then looking for the saddle points of the cost function:

$$\min_{\theta\in\Theta} \max_{u\in\mathbb{U}} J(\theta, u) \tag{1.20}$$

This is sometimes called Wald's Minimax criterion [Wal45]. The associated estimator is

$$\hat{\theta}_{\text{WC}} = \arg\min_{\theta\in\Theta} \max_{u\in\mathbb{U}} J(\theta, u) \tag{1.21}$$

However, this approach possesses some flaws. First, the maximum on $\mathbb{U}$ may not exist, especially if $\mathbb{U}$ is unbounded: we could make the model perform as badly as possible by taking extreme values of u. Additionally, if it exists, the resulting estimator is most likely very conservative as only the worse cases are considered.

1.3.1.c Regret maximin

One other approach [Sav51], called Savage's maximin regret is to compare the current loss to the best performance given the uncertain variable u:

$$r(\theta, u) = J(\theta, u) - \min_{\theta\in\Theta} J(\theta, u) \tag{1.22}$$

$$\min_{\theta\in\Theta} \max_{u\in\mathbb{U}} r(\theta, u) \tag{1.23}$$

and

$$\hat{\theta}_{\text{rWC}} = \arg\min_{\theta\in\Theta} \max_{u\in\mathbb{U}} r(\theta, u) \tag{1.24}$$
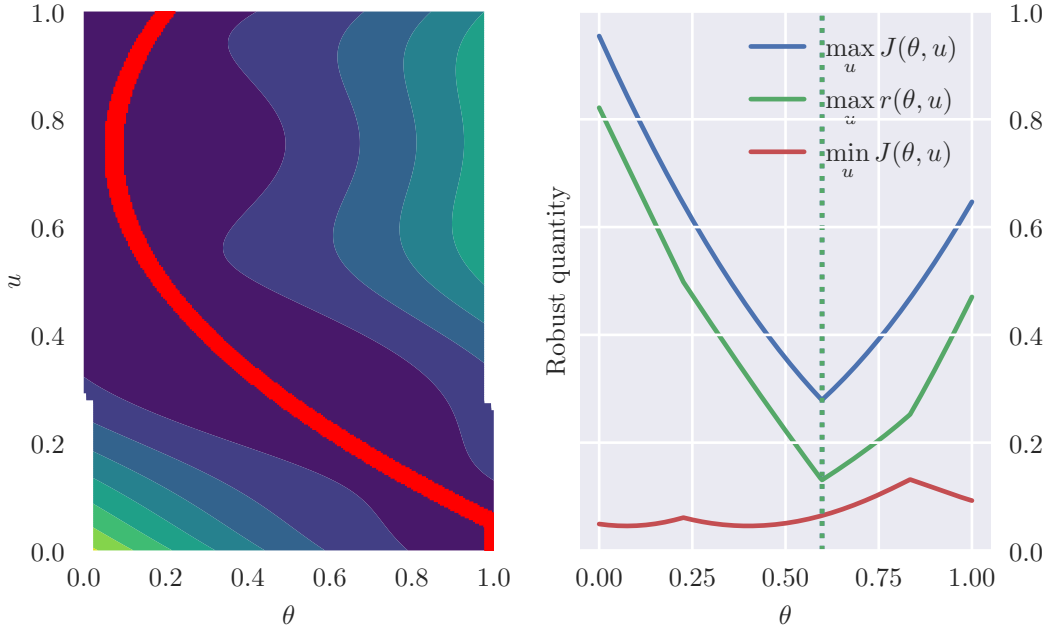
Figure 1.3: Illustration of global optimization, worst-case, and regret worst-case

1.3.2 Robustness based on the moments of an objective function

In the presence of uncertainties, choosing a parameter value θ can also be seen as making as making a choice under risk. Let $J : \Theta \times \mathbb{U} \to \mathbb{R}^+$ be an objective function, and assume that for all $\theta \in \Theta$, $J(\theta, \cdot)$ is a measurable function. J can be seen as the opposite of the *utility* function, often encountered in game theory or econometrics. Because of the random nature of U, we can define a family of real random variables $\{J(\theta, U) \mid \theta \in \Theta\}$, indexed by $\theta \in \Theta$. [BS07] proposes an *aggregation approach*, using integral criteria. This integration is usually done with respect to U.

Moreover, considering to integrate the powers of the cost function allow to account for dispersion through the moments of the random variable $J(\theta, U)$, as we are going to see in the following.

1.3.2.a Expected loss minimization

Quite intuitively, as we want to minimize some kind of typical value of a realization of these random variables, we can look to optimize a central tendency of those random variables. The mean value being an obvious candidate, we define the expected loss as

$$\mu(\theta) = \mathbb{E}_U \left[J(\theta, U) \right] = \int_{\mathbb{U}} J(\theta, u) p_U(u) \, \mathrm{d}u \qquad (1.25)$$

The expected loss $\mu(\theta)$ is sometimes called the conditional mean given θ. Taking the average of the loss function is very common in many problems of classification and regression [Bis06]

The conditional mean is minimized, giving $\hat{\theta}_{\mathrm{mean}}$. Assuming that $J(\theta, u) \propto -\log \mathcal{L}(\theta, u; y)$, we have

$$\hat{\theta}_{\mathrm{mean}} = \underset{\theta \in \Theta}{\arg\min} \ \mu(\theta) = \underset{\theta \in \Theta}{\arg\min} \ \int_{\mathbb{U}} J(\theta, u) p_U(u) \, \mathrm{d}u \tag{1.26}$$

$$= \underset{\theta \in \Theta}{\arg\min} \ - \int_{\mathbb{U}} \log \mathcal{L}(\theta, u; y) p_U(u) \, \mathrm{d}u \tag{1.27}$$

$$= \underset{\theta \in \Theta}{\arg\min} \ - \int_{\mathbb{U}} \log \big(p_{Y|\theta,U}(y \mid \theta, u) \big) \, p_U(u) \, \mathrm{d}u \tag{1.28}$$

Taking the average of a loss function is the basis of *stochastic programming*. This problem is usually tackled by using

Other indicator of central tendency can be considered for optimization, such as the mode or the median of the cost function. This expression shares some similarities with the integrated likelihood introduced Eq. (1.4). However, quite unsurprisingly, the $\hat{\theta}_{\mathrm{mean}}$ and $\hat{\theta}_{\mathrm{intMLE}}$ are not equal in general, as shown Fig. 1.4.
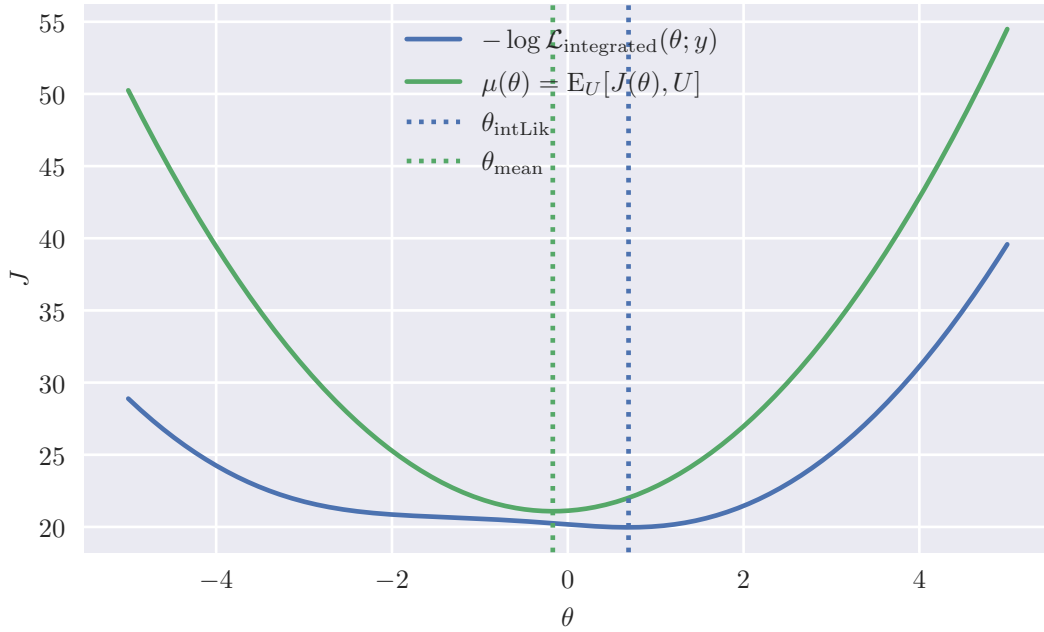

Figure 1.4

1.3.2.b   Variance, Multiobjective optimization

In Section 1.3.2.a, we used the mean as a measure of the central tendency that we want to minimize. Jointly with the central tendency, information about of the dispersion of the random variable may also be relevant, in order to predict how much deviation should be expected around the mean. Let us define the variance of the cost function:

$$\sigma^2(\theta) = \mathbb{V}\mathrm{ar}\left[J(\theta, U)\right] \tag{1.29}$$

Figure 1.5 shows the conditional mean and conditional standard deviation for a given cost function J defined as
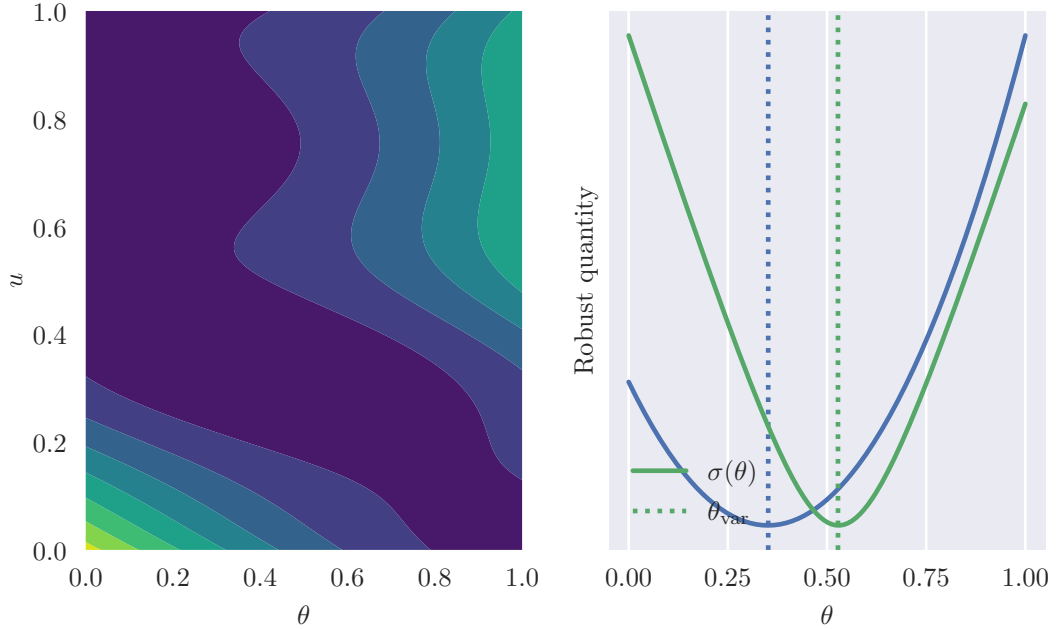
$$J(\theta, u) = \frac{remlpir}{formule} \tag{1.30}$$

Figure 1.5: Illustration of conditional mean and conditional standard deviation, as a function of θ. Those quantities have been rescaled to share the same range on the bottom plot.

Minimizing the variance alone is often irrelevant, as it could just point toward really high values of the cost function, but steady with respect to θ. Taking both objectives: low mean value and low variance together to the following multiobjective optimization problem:

$$\min_{\theta \in \Theta} \left[\mu(\theta), \sigma^2(\theta)\right]^T \tag{1.31}$$

The multiobjective problem can be tackled in different ways: the literature is rich in methods to approach or even find the Pareto frontier. To compare θ_1 and θ_2, we can compare component wise the objective vectors $[\mu(\theta_i), \sigma^2(\theta_i)]$ for $i = 1, 2$. If $\mu(\theta_1) \leq \mu(\theta_2)$ and $\sigma^2(\theta_1) \leq \sigma^2(\theta_2)$, θ_2 is said to be *dominated* by θ_1. The Pareto frontier is defined as the set of points in Θ that cannot be dominated by any other points. On Section 1.3.2.b is illustrated the Pareto frontier for a multiobjective problem Eq. (1.31). The red point corresponding to θ_1 is dominated by the green point θ_0 on the frontier, but not by the green point of θ_2.
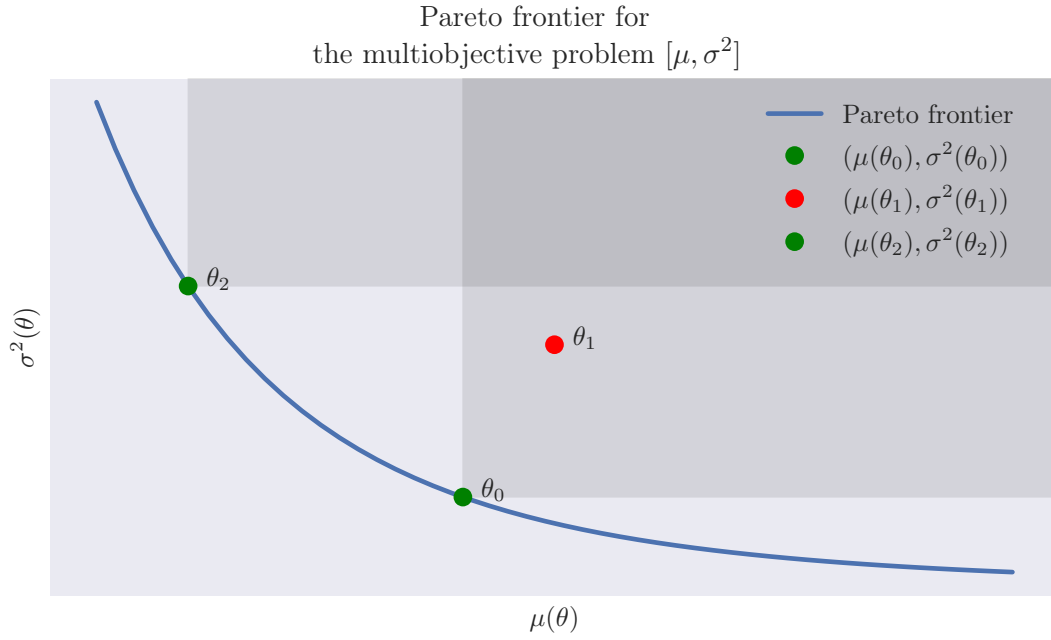
Pareto frontier for
the multiobjective problem $[\mu, \sigma^2]$

Figure 1.6: Illustration of the Pareto frontier for the multiobjective problem of Eq. (1.31). The shaded regions corresponds to the domain dominated by each points

Instead of trying to find the Pareto frontier, the multiobjective problem is often "scalarized" by adding the weighted objectives [MA10], provided that such an operation makes sense in regards to the units of the quantities. By considering the standard deviation instead of the variance, both objectives share the same units, and Eq. (1.31) becomes

$$\min_{\theta \in \Theta} w_1 \mu(\theta) + w_2 \sigma(\theta) \tag{1.32}$$

where w_1 and w_2 are positive and suitably chosen to reflect the preference toward one or another objective.

1.3.2.c   Higher moments in optimization

Skewness

The skewness coefficient measures the asymmetry in the distribution:

$$\text{sk}[X] = \mathbb{E}\left[\left(\frac{X - \mu}{\sigma}\right)^3\right] \tag{1.33}$$

where $\mu = \mathbb{E}[X]$ and $\sigma = \sqrt{\mathbb{V}\text{ar}[X]}$.

Adding the skewness in the optimization translates to a preference toward a risk-averse or a risk-seeking approach. Indeed, as the main goal is the optimization of a cost function, deviations of the value of the random variable toward lower values is more desirable than deviations toward larger values.

This is illustrated Fig. 1.7: all three of the random variables whose pdf and cdf are displayed have the same mean and variance. If the skewness coefficient is negative, the distribution presents a heavier left tail than right. In other words, a sample taken from this distribution has a higher probability of being a "good surprise". On the other hand, if a big deviation occurs for a sample from a right-skewed distribution, it is more probable to be a large deviation toward large values of the sample space, hence the term "bad surprise".
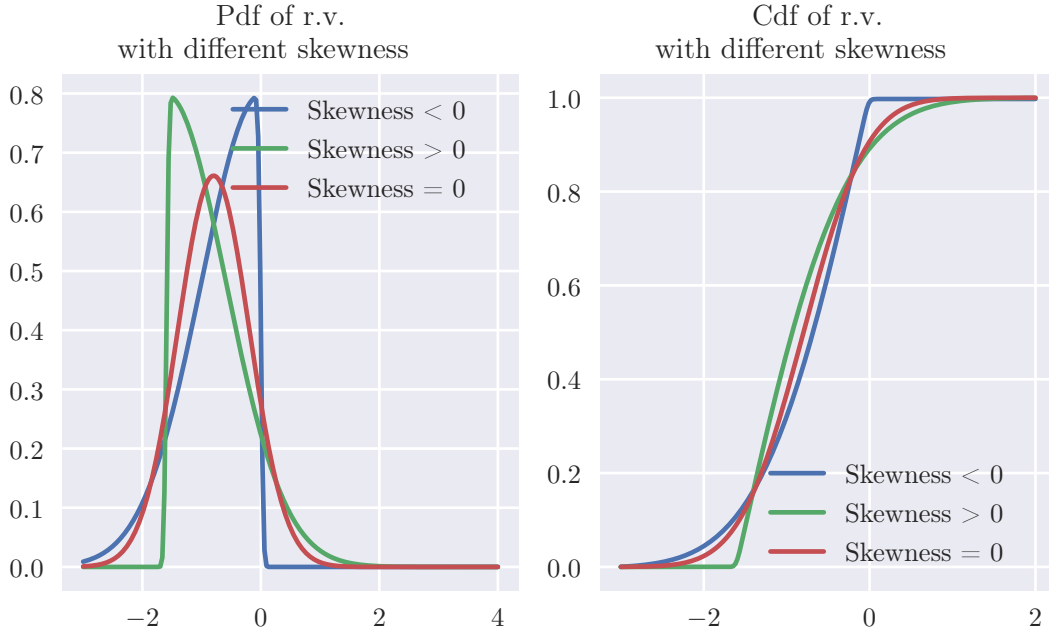

Figure 1.7:   Pdf and cdf of random variables with same mean, variance but different skewness

Horsetail matching?

1.4   Regret-based estimators family

All the methods introduced above required first to eliminate in some sense the dependency on the environmental parameter, in order to transform the random variable $J(\theta, U)$ into an objective that depends solely on θ, and to optimize this deterministic counterpart. For a given $\theta \in \Theta$, this elimination is done by aggregating all the possible outcomes $J(\theta, U)$.

We propose now to reverse these steps, by first optimizing the cost function for each u encountered, and from the set of minimizer obtained, derive an estimator. The rationale behind this idea is that every situation induced by the realisation u is to be taken separately, quite similarly as Savage's regret introduced Section 1.3.1.c.

1.4.1   Conditional minimimum and minimizer

Definition 1.4.1 – Conditional minimum, minimizer: Let $J : \Theta \times \mathbb{U}$ be a cost function, and let us assume that for each $u \in \mathbb{U}$, $\min_{\theta \in \Theta} J(\theta, u)$ exists and is attained at a unique point. We denote

$$J^*(u) = \min_{\theta \in \Theta} J(\theta, u) \tag{1.34}$$

the *conditional minimum* of J given u, and

$$\theta^*(u) = \arg\min_{\theta \in \Theta} J(\theta, u) \tag{1.35}$$

is defined as the *conditional minimizer*

As u is thought to be a realization of a random variable U, we can consider the two random variables $\theta^*(U)$ and $J^*(U)$. The random conditional minimum $J^*(U)$ is then a random variable describing the best performances for the calibration, if we could optimize the cost function for each realization of U. As we look for a single value $\hat{\theta}$, $J^*(U)$ can be seen as an unattainable target:

$$\forall (\theta, u) \in \Theta \times \mathbb{U}, \quad J^*(u) \leq J(\theta, u) \tag{1.36}$$

1.4.2   The random variable of the minimizer

If J behaves nicely enough, the mapping

$$\begin{aligned} \theta^* : \mathbb{U} &\longrightarrow \Theta \\ u &\longmapsto \theta^*(u) = \arg\min_{\theta \in \Theta} J(\theta, u) \end{aligned} \tag{1.37}$$

is well defined for all $u \in \mathbb{U}$, and we assume that it is measurable. We can study the image random variable through this mapping, that we will denote $\theta^* = \theta^*(U)$. We assume that U is a continuous random variable, with a compact support.

This random variable is directly linked with the "identifiability" of the problem at hand. If this random variable is dispersed over Θ, that means that the minimizer is highly dependent on the realization u of U. Two extreme cases are $p_{\theta*}(\theta) = \delta_{\theta_0}(\theta)$ and $p_{\theta*}(\theta) = \text{Vol}(\Theta)^{-1}$

In the first case, $\theta_0 = \arg\min_\theta J(\theta, U)$ almost surely hence the minimizer is not dependent on the value taken by the environmental parameter. The minimum value J^* might be dependent though.

On the other hand, if θ^* is uniformly distributed on Θ, no value shows a better affinity of being a minimizer than the other. The entropy of the distribution of the minimizers is a good indication of the information we have upon the possible locations of the minimizers. Some global optimization procedures use this entropy as an objective [VVW06].

1.4.3 Regret and model selection

define the cost function as the negative log-likelihood:

$$J(\theta, u) = -\log p_{Y|\theta,U}(y \mid \theta, u) = -\log \mathcal{L}(\theta, u) \tag{1.38}$$

As established in the first chapter, for a fixed $u_0 \in \mathbb{U}$, the model $\{(\mathcal{M}(\cdot, u_o), \Theta)\}$ is misspecified. The notion of Wald's regret introduced earlier is directly linked to the likelihood ratio test:

$$r(\theta, u_0) = J(\theta, u_0) - \min_{\theta' \in \Theta} J(\theta', u_0) = J(\theta, u_0) - J^*(u_0) \tag{1.39}$$

$$= \max_{\theta' \in \Theta} \log \mathcal{L}(\theta', u_0) - \log \mathcal{L}(\theta, u_0) \tag{1.40}$$

$$= -\log \frac{\mathcal{L}(\theta, u_0)}{\max_{\theta' \in \Theta} \mathcal{L}(\theta', u_0)} = -\log \frac{\max_{\theta' \in \{\theta\}} \mathcal{L}(\theta', u_0)}{\max_{\theta' \in \Theta} \mathcal{L}(\theta', u_0)} \tag{1.41}$$

Recalling the Likelihood ratio test, in order to perform a likelihood ratio test, r is to be compared with quantile of the r.v. defined as $\sum_{i=1}^{\dim \Theta} c_i(u_0)\chi_1^2$. However, by making the assumption that $c_i(u_0) = 1$, the Likelihood ratio test is the classical one, and the rejection region at level ν for $\theta \in \Theta$ is

$$\text{RejReg}(u_0) = \left\{ \theta \in \Theta \mid r(\theta, u_0) \geq \frac{1}{2}\chi_{\dim \Theta}^2(1 - \nu) \right\} \tag{1.42}$$

$$= \left\{ \theta \in \Theta \mid J(\theta, u_0) - J^*(u_0) \geq \frac{1}{2}\chi_{\dim \Theta}^2(1 - \nu) \right\} \tag{1.43}$$

1.4.4 Relative regret

We examined before regret that can be qualified as *additive* as this is the difference between J and J^* that is compared to fixed thresholds. We are going to study relative regret, that focuses on the ratio J/J^*. We can also interpret the cost function as an

energy, and we want the system to be in the lower state of energy. In terms of Gibbs distribution, we can define

$$\frac{1}{Z(\theta, u, 1)} \exp(-J(\theta, u)) \tag{1.44}$$

$$\frac{1}{Z(\theta^*(u), u, \alpha)} \exp(-\alpha J^*(u)) \tag{1.45}$$

Where Z is a function of θ, u that represents the integration constant. By interpreting α as the reciprocal of a temperature,

BIBLIOGRAPHY

[ALMP+12] Vazken Andréassian, Nicolas Le Moine, Charles Perrin, Maria-Helena Ramos, Ludovic Oudin, Thibault Mathevet, Julien Lerat, and Lionel Berthet. All that glitters is not gold: The case of calibrating hydrological models: Invited Commentary. *Hydrological Processes*, 26(14):2206–2210, July 2012.

[Bis06] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. springer, 2006.

[BLW99] James O. Berger, Brunero Liseo, and Robert L. Wolpert. Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 14(1):1–28, February 1999.

[BS07] Hans-Georg Beyer and Bernhard Sendhoff. Robust optimization – A comprehensive survey. *Computer Methods in Applied Mechanics and Engineering*, 196(33-34):3190–3218, July 2007.

[DGR02] Arnaud Doucet, Simon J. Godsill, and Christian P. Robert. Marginal maximum a posteriori estimation using Markov chain Monte Carlo. *Statistics and Computing*, 12(1):77–84, January 2002.

[HB01] Luc Huyse and Dennis M. Bushnell. Free-form airfoil shape optimization under uncertainty using maximum expected value and second-order second-moment strategies. 2001.

[KD09] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? Does it matter? *Structural Safety*, 31(2):105–112, March 2009.

[KRTK10] George Kuczera, Benjamin Renard, Mark Thyer, and Dmitri Kavetski. There are no hydrological monsters, just models and observations with large uncertainties! *Hydrological Sciences Journal*, 55(6):980–991, August 2010.

[LBM+16] Nicolas Lelièvre, Pierre Beaurepaire, Cécile Mattrand, Nicolas Gayton, and Abdelkader Otsmane. On the consideration of uncertainty in design:

Optimization-reliability-robustness. *Structural and Multidisciplinary Optimization*, 54(6):1423–1437, 2016.

[MA10] R. Timothy Marler and Jasbir S. Arora. The weighted sum method for multi-objective optimization: New insights. *Structural and Multidisciplinary Optimization*, 41(6):853–862, June 2010.

[Sav51] L. J. Savage. The Theory of Statistical Decision. *Journal of the American Statistical Association*, 46(253):55–67, March 1951.

[TAVD20] Victor Trappler, Élise Arnaud, Arthur Vidard, and Laurent Debreu. Robust calibration of numerical models based on relative regret (Preprint), February 2020.

[VVW06] Julien Villemonteix, Emmanuel Vazquez, and Eric Walter. An informational approach to the global optimization of expensive-to-evaluate functions. *arXiv:cs/0611143*, November 2006.

[Wal45] Abraham Wald. Statistical Decision Functions Which Minimize the Maximum Risk. *Annals of Mathematics*, 46(2):265–280, 1945.

[WHR+03] Warren E. Walker, Poul Harremoës, Jan Rotmans, Jeroen P. van der Sluijs, Marjolein BA van Asselt, Peter Janssen, and Martin P. Krayer von Krauss. Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated assessment*, 4(1):5–17, 2003.