SENSITIVITY INDICES FOR IMPRECISE PROBABILITY DISTRIBUTIONS

JIM HALL

Abstract. Conventional variance-based sensitivity indices are extended to deal with the case when information is available as closed convex sets of probability measures, a situation that exists when probability distributions are specified with interval-valued parameters. The generalization to closed convex sets of probability measures yields lower and upper sensitivity indices. An example demonstrates the numerical method for estimating these sensitivity indices.

1. Introduction

The information input into computer models may be imprecise for several reasons. Imprecision is often a consequence of measurement processes, for example using digital sensors. Prior information is sometimes recorded in the literatures as intervals without any information about probability distributions [3]. Given only finite time is argued that it may be impossible to elicit precise probability distributions from experts [1]. Indeed experts may deliberately use imprecision to express their uncertainty.

The extension of probabilistic analysis to include imprecise information is now well established in the theory of imprecise probabilities [10], robust Bayesian analysis [6][4] and fuzzy statistics [9]. In this paper we explore the notion of sensitivity within this framework. We confine ourselves to the theory of coherent lower and upper probabilities, which, whilst not the most general theory of imprecise probabilities, is sufficient to deal with the the situation in which probability distributions are specified by interval-valued parameters.

2. Coherent lower and upper probabilities

Consider a probability density function $f(x, \mathbf{a})$, where $x \in \mathbb{R}$ and $\mathbf{a} = (a_1, a_2, \ldots, a_m)$, a vector of parameters of the probability density function. By definition

$$\mathrm{Pr}(A) = \int_A f(x, \mathbf{a})dx, \forall A \subseteq \mathbb{R} \tag{2.1}$$

If each parameter a_i in $\mathbf{a}$ is specified by a closed interval $[l_i, u_i]$ then $\mathbf{a}$ is constrained by an n-dimensional box Q, defining a closed set of probability measures that imply lower and upper probabilities, $P(\underline{A})$ and $P(\overline{A})$

$$\mathrm{Pr}(\underline{A}) = \inf_{\mathbf{a} \in Q} \int_A f(x, \mathbf{a})dx \tag{2.2}$$

$$\mathrm{Pr}(\overline{A}) = \sup_{\mathbf{a} \in Q} \int_A f(x, \mathbf{a})dx \tag{2.3}$$

$P(\underline{A})$ and $1 - P(\overline{A})$ will be located at the same point $\mathbf{a}$, so $P(\underline{A}) = 1 - P(\overline{A})$, meaning that $P(\underline{A})$ and $P(\overline{A})$ are coherent lower and upper probabilities [11].

The lower and upper expectations, $E(\underline{X})$ and $E(\overline{X})$, are given by

$$\underline{E}(X) = \inf_{\mathbf{a}\in Q} \int_{-\infty}^{\infty} x f(x, \mathbf{a}) dx \tag{2.4}$$

$$\overline{E}(X) = \sup_{\mathbf{a}\in Q} \int_{-\infty}^{\infty} x f(x, \mathbf{a}) dx \tag{2.5}$$

The definitions in Equations 2.2 to 2.5 can be extended to the case when $f(\mathbf{x}, \mathbf{a})$ is a joint probability distribution on $X_1 \times \ldots \times X_n$ and $\mathbf{x} = (x_1, \ldots, x_n)$.

2.1. **Lower and upper variance.** The standard definition of the variance $V(X)$ of a random variable X is

$$V(X) = E([X - E(X)]^2) \tag{2.6}$$

In the case when $\mathcal{M}$ is a closed convex set of probability measures P, Walley [10] demonstrates the following expressions for the lower and upper variances, $\underline{V}(X)$ and $\overline{V}(X)$,

$$\underline{V}(X) = \min_{P\in\mathcal{M}} V(X) \tag{2.7}$$

$$\overline{V}(X) = \max_{P\in\mathcal{M}} V(X) \tag{2.8}$$

2.2. **Natural extension of imprecise probabilities.** Let g be a function such that $y = g(\mathbf{x}), \mathbf{x} = (x_1, \ldots, x_n)$, and let B_y be the subset of $\mathbb{R}^n$ containing all of the points $(x_1, \ldots, x_n)$ such that $g(\mathbf{x}) \in C, C \in \mathbb{R}$, then the lower and upper probabilities $\underline{P}(C)$ and $\overline{P}(C)$ are:

$$\underline{P}(C) = \inf_{\mathbf{a}\in Q} \int_{B_y} \cdots \int f(x_1, \ldots, x_n, \mathbf{a}) dx_1 \ldots dx_n \tag{2.9}$$

and

$$\overline{P}(C) = \sup_{\mathbf{a}\in Q} \int_{B_y} \cdots \int f(x_1, \ldots, x_n, \mathbf{a}) dx_1 \ldots dx_n \tag{2.10}$$

3. Variance-based sensitivity analysis

Consider now the conventional probabilistic case in which the uncertainties in $x_1, \ldots, x_n$ are expressed as precise probability distributions, i.e. $x_1, \ldots, x_n$ and y are replaced by random variables $X_1, \ldots, X_n$ and Y respectively. In variance-based sensitivity analysis, the first order sensitivity indices S_i represents the fractional contribution of a given variable X_i to the variance in a given output variable Y [2]. In order to calculate the sensitivity indices the total variance V in the model output Y is apportioned to all the input factors X_i as [8]

$$V = \sum_i V_i + \sum_{i<j} V_{ij} + \sum_{i<j<k} V_{ijk} + \ldots + V_{12\ldots k} \tag{3.1}$$

where

$$V_i = V[E(Y|X_i = x_i^*)] \tag{3.2}$$

$$V_{ij} = V[E(Y|X_i = x_i^*, X_j = x_j^*)] - V_i - V_j \tag{3.3}$$

and so on. $V[E(Y|X_i = x_i^*)]$ is the Variance of the Conditional Expectation (VCE) and is the variance over all values of x_i^* in the expectation of Y given that X_i has a fixed value x_i^*. The first order (or 'main effect') sensitivity index S_i for variable X_i is:

$$S_i = V_i \backslash V \tag{3.4}$$

and the 'total effect' sensitivity index is [5]

$$S_{Ti} = \frac{V[E(Y|X_{\sim i} = x_{\sim i}^*)]}{V(Y)} \tag{3.5}$$

where $X_{\sim i}$ denotes all of the variables other than X_i.

4. Imprecise sensitivity indices

In the case when the uncertainty in the variables $X_1 \ldots X_n$ is described by a closed convex set $\mathcal{M}$ of probability measures P, the lower and upper variances introduced in Equations 2.7 and 2.8 above can be extended to lower and upper sensitivity indices, $\underline{S}_i$ and $\overline{S}_i, i = 1, \ldots, n$:

$$\underline{S}_i = \min_{P \in \mathcal{M}} S_i \tag{4.1}$$

and

$$\overline{S}_i = \max_{P \in \mathcal{M}} S_i \tag{4.2}$$

where

$$\sum_{i=1}^{n} \overline{S}_i \leq 1. \tag{4.3}$$

The additional constraint in Equation 4.3 means that the upper sensitivity indices $\overline{S}_i$, $i = 1, \ldots, n$ may not co-exist. Indeed there is a closed convex set $\mathcal{S}$ of sensitivity indices $\mathbf{S} \in \mathcal{S}, \mathbf{S} = \{S_1, \ldots S_n\}$ constrained such that $\forall S_i, i = 1, \ldots, n : \underline{S}_i \leq S_i \leq \overline{S}_i$ and $\sum_{i=1}^{n} \overline{S}_i \leq 1$.

Estimating the lower and upper sensitivity indices in Equations 4.1 and 4.2 is a problem of non-linear optimization. Each iteration j of the optimization involves estimating the precise sensitivity indices for some $P_j \in \mathcal{M}$. Note, however, that the points used to evaluate these sensitivity indices at iterations $1, \ldots, j$ may be reused in subsequent iterations.

5. Application

5.1. **The Challenge Problems.** Oberkampf et al. [7] have proposed a series of Challenge Problems to compare and evaluate alternative theories of uncertainty. One of the Challenge Problems relates to a damped linear oscillator (a single degree of freedom mass-spring-damper system), whose steady-state magnification factor D_s is given by

$$D_s = \frac{k}{\sqrt{(k - m\omega^2)^2 + (c\omega)^2}} \tag{5.1}$$

where k is the spring constant, m is the mass of the oscillator, ω is the frequency of oscillation and c is the damping coefficient. In the Challenge Problems, the variables in Equation 5.1 were specified as follows:

m**:** is given by a precise triangular probability distribution defined on the interval [10,12], with a median value 11.

k**:** is given by an imprecise triangular probability distribution, specified by three imprecise parameters k_{min}, k_{mod} and k_{max}, whose values are contained in the closed intervals $k_{min} \in [90, 100]$, $k_{mod} \in [150, 160]$ and $k_{max} \in [90, 100]$.

c**:** is given by a closed interval of possible values $c \in [5, 10]$. No probability distribution over this interval is specified or to be assumed.

ω**:** is given by an imprecise triangular probability distribution, specified by three imprecise parameters ω_{min}, ω_{mod} and ω_{max}, whose values are contained in the closed intervals $\omega_{min} \in [2.0, 2.3]$, $\omega_{mod} \in [2.5, 2.7]$ and $\omega_{max} \in [3.0, 3.5]$.

In the Challenge Problems the information concerning k and c was given by three independent sources. The problem of aggregation of evidence from multiple sources is beyond the scope of the present paper and is not addressed. The information is used from the first source only.

5.2. **Imprecise sensitivity indices for the Challenge Problems.** The numerical method by which the problem outlined in Section 5.1 has been tackled involves a combination of optimization to estimate the bounds and Monte Carlo resampling to estimate each conditional variance.

There are 6 interval-valued distribution parameters (k_{min}, k_{mod}, k_{max}, ω_{min}, ω_{mod} and ω_{max}) and one interval-valued variable, c, in the analysis. If the sensitivity indices S_i were a monotonic function of these imprecise quantities then it would only be necessary only to test the vertices of the 7 dimensional hypercube that contains all of the possible values of these quantities. There is, however, no reason to believe that S_i should be a monotonic function of these interval-valued quantities, so in order to find the imprecise sensitivity indices it was necessary to search the volume contained within these interval constraints. Besides testing each of the 2^7 vertices, the volume was searched by uniformly sampling the space. Label each test point (including vertices and randomly sampled points) as a vector $\mathbf{t}_j = (k_{min,j}, k_{mod,j}, k_{max,j}, \omega_{min,j}, \omega_{mod,j}, \omega_{max,j}, c_j)$, $j = 1, \ldots, u$. For each $\mathbf{t}_j$ the corresponding precise probability distributions for k and ω, and the precise probability distribution for m were sampled, which, together with the precise value c_j, yield a precise estimate $V_j(D_s)$ of the variance of (D_s). For each $\mathbf{t}_j$ the sensitivity indices were estimated by resampling the probability distributions for the variables m, k and ω. This resampling yielded a precise estimate of the sensitivity indices $S_{i,j}$. The lower and upper variances are then given by

$$\underline{V}(D_s) = \min_j (V_j(D_s)) \tag{5.2}$$

$$\overline{V}(D_s) = \max_j (V_j(D_s)) \tag{5.3}$$

and the lower and upper sensitivity indices are given by

$$\underline{S_i}(D_s) = \min_j (S_{i,j}(D_s)), i = 1, \ldots, n \tag{5.4}$$

$$\overline{S_i}(D_s) = \max_j (S_{i,j}(D_s)) \tag{5.5}$$

In the Challenge Problem specified above the lower and upper variances were estimated as $\underline{V}(D_s) = 0.10$ and $\overline{V}(D_s) = 1.57$. The imprecise sensitivity indices are

i	Variable	$\underline{S}_i$	$\overline{S}_i$
1	m	0.01	0.05
2	k	0.19	0.79
3	ω	0.19	0.73

TABLE 1. Imprecise sensitivity indices

listed in Table 1. Note the additional condition in Equation 4.3 means that the upper sensitivity indices cannot all coexist.

REFERENCES

[1] J. Berger. The robust Bayesian viewpoin (with discussion). In J. Kadane, editor, *Robustness of Bayesian Analyses*. North-Holland, Amsterdam, 1984.

[2] K. Chan, S. Tarantola, A. Saltelli, and I.M. Sobol'. Variance-based methods. In A. Saltelli, K. Chan, and E.M. Scott, editors, *Sensitivity Analysis*, chapter 8, pages 167–198. Wiley, Chichester, 2000.

[3] J.W. Hall, E. Rubio, and M.J. Anderson. Random sets of probability measures in slope hydrology and stability analysis. *ZAMM: Journal of Applied Mathematics and Mechanics*, in press, 2004.

[4] F.R. Hempel. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, Chichester, 1986.

[5] T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of model output. *Reliability Engineering and Systems Safety*, 52(1):1–17, 1996.

[6] P.J. Huber. *Robust Statistics*. Wiley, New York, 1981.

[7] W.L. Oberkampf, J.C. Helton, C.A. Joslyn, S.F. Wojtkiewicz, and S. Ferson. Challenge problems: uncertainty in system response given uncertain parameters. *Reliability Engineering and System Safe*, in press.

[8] I.M. Sobol'. Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling Computational Experiment*, 1:407–414, 1993.

[9] R. Viertl. *Statistical Methods for Non-Precise Data*. CRC Press, Boca Raton, Florida, 1996.

[10] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[11] P. Walley. Towards a unified theory of imprecise probabilities. *Int. J. Approximate Reasoning*, 24(2-3):125–148, 2000.

DEPARTMENT OF CIVIL ENGINEERING, UNIVERSITY OF BRISTOL, QUEEN'S BUILDING, UNIVERSITY WALK, BRISTOL, BS8 1TR, UK

E-mail address: jim.hall@bristol.ac.uk