scientific reports

Check for updates

OPEN

Probabilistic threshold analysis by pairwise stochastic approximation for decision-making under uncertainty

Takashi Goda✉ & Yuki Yamada

The concept of probabilistic parameter threshold analysis has recently been introduced as a way of probabilistic sensitivity analysis for decision-making under uncertainty, in particular, for health economic evaluations which compare two or more alternative treatments with consideration of uncertainty on outcomes and costs. In this paper we formulate the probabilistic threshold analysis as a root-finding problem involving the conditional expectations, and propose a pairwise stochastic approximation algorithm to search for the threshold value below and above which the choice of conditionally optimal decision options changes. Numerical experiments for both a simple synthetic testcase and a chemotherapy Markov model illustrate the effectiveness of our proposed algorithm, without any need for accurate estimation or approximation of conditional expectations which the existing approaches rely upon. Moreover we introduce a new measure called decision switching probability for probabilistic sensitivity analysis in this paper.

Background. Probabilistic sensitivity analysis is an attempt to provide a framework for evaluating how the uncertainty of input parameters propagates to the uncertainty of model outputs[1,2]. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_s)$ be a vector of input random variables and consider a model described by $Y = f(\boldsymbol{\theta})$. Here the output Y can be also regarded as a random variable because of the uncertainty of the input $\boldsymbol{\theta}$. A primary interest of probabilistic sensitivity analysis for this simple setting is to identify which input variable θ_j (or, which group of input variables) affects the variability of the output Y most or least significantly. Among various approaches for measuring the relative importance of each input variable, variance-based sensitivity analysis due to Sobol'[3,4] has been found quite useful for this purpose. Assuming the independence between the input random variables $\theta_1, \ldots, \theta_s$, the following analysis-of-variance (ANOVA) decomposition of a square-integrable function f holds:

$$f(\boldsymbol{\theta}) = f_{\emptyset} + \sum_{\emptyset \neq u \subseteq \{1,\ldots,s\}} f_u(\boldsymbol{\theta}_u),$$

where we write $\boldsymbol{\theta}_u = (\theta_j)_{j \in u}$ for a non-empty set $u \subseteq \{1, \ldots, s\}$, and each term is recursively given by $f_{\emptyset} = \mathbb{E}_{\boldsymbol{\theta}}[f(\boldsymbol{\theta})]$ and

$$f_u(\boldsymbol{\theta}_u) = \mathbb{E}_{\boldsymbol{\theta} \setminus \boldsymbol{\theta}_u}\big[f(\boldsymbol{\theta})\big] - \sum_{v \subset u} f_v(\boldsymbol{\theta}_v), \quad \text{for} \quad \emptyset \neq u \subseteq \{1, \ldots, s\}.$$

Here we note that $f_{\emptyset}$ is a constant and each function f_u depends only on a group of input variables $\boldsymbol{\theta}_u$. Because of the orthogonality of these terms[5], the variance of f can be decomposed as

$$\mathbb{V}_{\boldsymbol{\theta}}\big[f(\boldsymbol{\theta})\big] = \sum_{\emptyset \neq u \subseteq \{1,\ldots,s\}} \mathbb{V}_{\boldsymbol{\theta}_u}\big[f_u(\boldsymbol{\theta}_u)\big].$$

This equality enables us to measure the relative importance played by a group of input variables $\boldsymbol{\theta}_u$ in several ways. The famous examples are

School of Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. ✉email: goda@frcer.t.u-tokyo.ac.jp

$$\sum_{\emptyset \neq v \subseteq u} \mathbb{V}_{\boldsymbol{\theta}_v}\left[f_v(\boldsymbol{\theta}_v)\right] \quad \text{and} \quad \sum_{\substack{\emptyset \neq v \subseteq \{1,\ldots,s\} \\ v \cap u \neq \emptyset}} \mathbb{V}_{\boldsymbol{\theta}_v}\left[f_v(\boldsymbol{\theta}_v)\right] = \mathbb{V}_{\boldsymbol{\theta}}\left[f(\boldsymbol{\theta})\right] - \sum_{\emptyset \neq v \subseteq \{1,\ldots,s\} \setminus u} \mathbb{V}_{\boldsymbol{\theta}_v}\left[f_v(\boldsymbol{\theta}_v)\right],$$

where the first one measures the variance explained by $\boldsymbol{\theta}_u$, whereas the second one measures the total variance minus the variance explained by the complement variables $\boldsymbol{\theta} \setminus \boldsymbol{\theta}_u$. In fact, there is a huge volume of literature on how to estimate these sensitivity measures[6–10] and also on applications to real problems in various subjects[11–14].

Looking only at the variability (or, the variance) of the output from a single model is not enough, however, if we are faced with a decision-making problem[15,16]. Let D be a finite set of possible alternative options for decision, and consider that each option $d \in D$ is associated with its model described by an utility function $Y_d = f_d(\boldsymbol{\theta})$. In the context of health economic evaluations, for instance, D denotes the set of alternative treatments for a certain disease, f_d represents the cost-effectiveness (or, the monetary net benefit) of each treatment $d \in D$, and the input variables $\boldsymbol{\theta}$ include various unknown parameters related to the cost-effectiveness, such as the probability of side effect and the cost of treatment. Note that we assume that the set of input variables $\boldsymbol{\theta}$ is common across all of the options and that the output Y_d can be again regarded as a random variable because of the uncertainty of $\boldsymbol{\theta}$. The fundamental problem here is two-fold:

1. to identity which option $d \in D$ is optimal under uncertainty on $\boldsymbol{\theta}$, and
2. to identity which input variable θ_j (or, which group of input variables) affects the *variability of the optimal option* $d \in D$ most or least significantly.

Regarding the first problem, in the absence of any knowledge about $\boldsymbol{\theta}$, the optimal option should be the one which maximizes the expected utility, i.e.,

$$d_{\mathrm{opt}}(\emptyset) = \arg\max_{d \in D} \mathbb{E}_{\boldsymbol{\theta}}\left[f_d(\boldsymbol{\theta})\right].$$

Throughout this paper, we assume that $d_{\mathrm{opt}}(\emptyset)$ is unique, that is, exactly one option achieves the maximum expected utility. In order to address the second problem above, the so-called *expected value of partial perfect information* (EVPPI) considers an ideal situation where the uncertainty on an individual variable or a group of variables can be removed completely, and evaluates how such a partially perfect knowledge on $\boldsymbol{\theta}$ can lead to an optimal option different from the prior one $d_{\mathrm{opt}}(\emptyset)$ and yield an increment of the expected utility[17–20]. To be more precise, let us consider a partition of the components in the vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. If we know the exact value of every component in $\boldsymbol{\theta}_1$, the optimal option should be the one which maximizes the *conditional* expected utility given $\boldsymbol{\theta}_1$, i.e.,

$$d_{\mathrm{opt}}(\boldsymbol{\theta}_1) = \arg\max_{d \in D} \mathbb{E}_{\boldsymbol{\theta}_2|\boldsymbol{\theta}_1}\left[f_d(\boldsymbol{\theta})\right],$$

which can change depending on $\boldsymbol{\theta}_1$. We note that, when $d_{\mathrm{opt}}(\boldsymbol{\theta}_1)$ is not unique, that is, when several different options yield the same maximum conditional expected utility, the choice is arbitrary. Taking the average of the maximum conditional expected utility with respect to $\boldsymbol{\theta}_1$, the EVPPI for $\boldsymbol{\theta}_1$ is defined as its increment from the prior expected utility, i.e.,

$$\mathrm{EVPPI}_{\boldsymbol{\theta}_1} = \mathbb{E}_{\boldsymbol{\theta}_1}\left[\max_{\mathrm{d} \in \mathrm{D}} \mathbb{E}_{\boldsymbol{\theta}_2|\boldsymbol{\theta}_1}[\mathrm{f_d}(\boldsymbol{\theta})]\right] - \max_{\mathrm{d} \in \mathrm{D}} \mathbb{E}_{\boldsymbol{\theta}}[\mathrm{f_d}(\boldsymbol{\theta})].$$

The EVPPI takes a non-negative value and is bounded above by the expected value of perfect information (EVPI):

$$\mathrm{EVPI} = \mathbb{E}_{\boldsymbol{\theta}}\left[\max_{\mathrm{d} \in \mathrm{D}} \mathrm{f_d}(\boldsymbol{\theta})\right] - \max_{\mathrm{d} \in \mathrm{D}} \mathbb{E}_{\boldsymbol{\theta}}[\mathrm{f_d}(\boldsymbol{\theta})].$$

This way, it is indicated that the uncertainty of the random variables $\boldsymbol{\theta}_1$ with a large EVPPI (close to EVPI) significantly affects the choice of the optimal option, whereas it is not the case for those with a small EVPPI. In fact, the equality $\mathrm{EVPPI}_{\boldsymbol{\theta}_1} = 0$ is equivalent to that $d_{\mathrm{opt}}(\boldsymbol{\theta}_1) = d_{\mathrm{opt}}(\emptyset)$ happens almost surely (up to uniqueness of the argument), that is, the perfect knowledge on $\boldsymbol{\theta}_1$ does not change the choice of the optimal option. This is how probabilistic sensitivity analysis can be performed for a decision model, and a strong interest in such decision-theoretic probabilistic sensitivity analysis can be found not only in health economic evaluations[21–24] but also in petroleum engineering[16,25,26]. Here we emphasize that EVPPI is not the only measure for evaluating the relative importance of each input variable, and we shall introduce a new sensitivity measure called *decision switching probability* in this paper.

What is probabilistic threshold analysis? Based on the indication from EVPPI, it is natural to evaluate the threshold of $\boldsymbol{\theta}_1$ around which the choice of the optimal option, $d_{\mathrm{opt}}(\boldsymbol{\theta}_1)$, possibly changes. This is the aim of the so-called probabilistic parameter threshold analysis, which has been introduced quite recently as a way of probabilistic sensitivity analysis for decision-making under uncertainty[27]. Following the closely-related literature[27,28], let us focus on the case where all of the input variables in $\boldsymbol{\theta}$ are continuous and $\boldsymbol{\theta}_1$ consists only of a single input variable θ_j for some $1 \leq j \leq s$. Then the probabilistic parameter threshold for θ_j, denoted by K_j, is simply defined as follows.

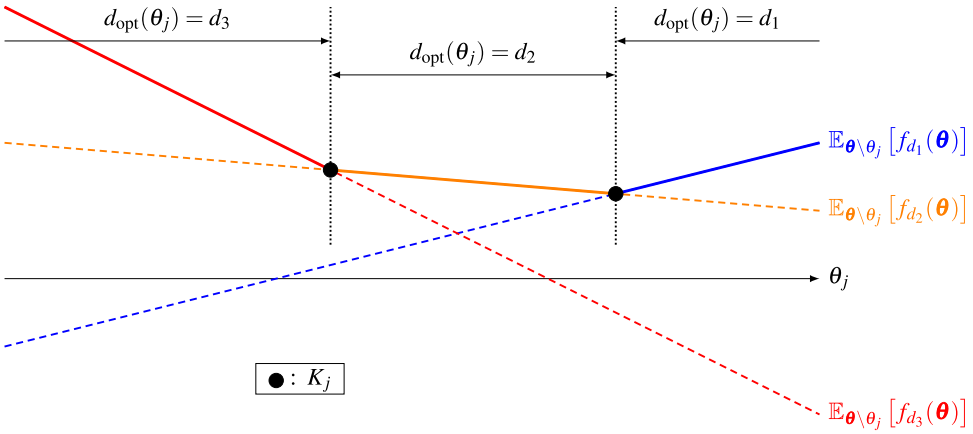

Figure 1. Schematic of the conditional expectations $\mathbb{E}_{\boldsymbol{\theta}\setminus\theta_j}\left[f_d(\boldsymbol{\theta})\right]$ for different options $d \in D$ as functions of θ_j, the conditional optimal option $d_{\mathrm{opt}}(\theta_j)$ and the probabilistic parameter threshold K_j.

Definition 1 (*Probabilistic parameter threshold*) With the notation above, the probabilistic parameter threshold K_j for an individual variable θ_j is defined by the set

$$K_j := \left\{\theta_j \mid d_{\mathrm{opt}}(\theta_j) \text{ is not unique}\right\}.$$

Throughout this paper, we assume that the cardinality of K_j is at most finite. Figure 1 shows a schematic of the probabilistic parameter threshold K_j for the case $|D| = 3$. The conditional expectation for each option $d \in D$ is drawn in a different color as a function of θ_j. The optimal option $d_{\mathrm{opt}}(\theta_j)$ which maximizes the conditional expectation is equal to d_3, d_2 and d_1 in the left, middle and right intervals, respectively. The probabilistic parameter threshold K_j consists of two intersection points in this example, with one between d_3 and d_2 and the other between d_2 and d_1.

Remark 1 It is obviously possible that K_j is empty. In such a case, it implies from the continuity of θ_j that $d_{\mathrm{opt}}(\theta_j)$ does not change regardless of the value of θ_j. Let us write $d' = d_{\mathrm{opt}}(\theta_j)$. If $d' \neq d_{\mathrm{opt}}(\emptyset)$ holds, the tower property of conditional expectations leads to

$$\mathbb{E}_{\boldsymbol{\theta}}\left[f_{d'}(\boldsymbol{\theta})\right] = \mathbb{E}_{\theta_j}\mathbb{E}_{\boldsymbol{\theta}\setminus\theta_j}\left[f_{d'}(\boldsymbol{\theta})\right] \geq \mathbb{E}_{\theta_j}\mathbb{E}_{\boldsymbol{\theta}\setminus\theta_j}\left[f_{d_{\mathrm{opt}}(\emptyset)}(\boldsymbol{\theta})\right] = \mathbb{E}_{\boldsymbol{\theta}}\left[f_{d_{\mathrm{opt}}(\emptyset)}(\boldsymbol{\theta})\right],$$

which contradicts our assumption that $d_{\mathrm{opt}}(\emptyset)$ is unique. Thus we must have $d' = d_{\mathrm{opt}}(\theta_j) \equiv d_{\mathrm{opt}}(\emptyset)$ for any θ_j, leading to $\mathrm{EVPPI}_{\theta_j} = 0$.

By definition, the probabilistic parameter threshold is not designed to measure the relative importance of each input variable for a decision model. Instead, it evaluates whether removing the uncertainty on θ_j completely can change the optimal option (for instance, from the prior one $d_{\mathrm{opt}}(\emptyset)$), and if so, that is, if K_j is not empty, which values of θ_j make such change happen. This informs us of the following additional aspect for a decision problem, which cannot be captured only by the EVPPI. Suppose that both θ_i and θ_j, with $1 \leq i < j \leq s$, follow the standard normal distribution independently, and also that we have $\mathrm{EVPPI}_{\theta_i} \approx \mathrm{EVPPI}_{\theta_j}$. If $K_i = \{1\}, K_j = \{3\}, d_{\mathrm{opt}}(\theta_i) = d_{\mathrm{opt}}(\emptyset)$ for $\theta_i < 1$ and $d_{\mathrm{opt}}(\theta_j) = d_{\mathrm{opt}}(\emptyset)$ for $\theta_j < 3$, then, although an increment of the expected utility by knowing the exact value of either θ_i or θ_j is assumed almost the same each other, the chance of changing an optimal option $d \in D$ from $d_{\mathrm{opt}}(\emptyset)$ is quite different. For the variable θ_j, such change happens when $\theta_j > 3$, whose probability is only 0.0013, whereas it happens with probability 0.1587 for the variable θ_i. Therefore, we can say that the variable θ_i is more sensitive to the variability of the optimal option than θ_j.

Although our primary interest of this paper is in an efficient estimation of the probabilistic parameter threshold K_j, the above argument inspires us to introduce a related measure for decision-theoretic probabilistic sensitivity analysis as defined below. In what follows we call it decision switching probability.

Definition 2 (*Decision switching probability*) Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ be a partition of the vector $\boldsymbol{\theta}$. With the notation above, the decision switching probability for the variables $\boldsymbol{\theta}_1$ is defined by

$$P_{\boldsymbol{\theta}_1} := \mathbb{P}_{\boldsymbol{\theta}_1}\left[d_{\mathrm{opt}}(\boldsymbol{\theta}_1) \neq d_{\mathrm{opt}}(\emptyset)\right].$$

In particular, for an individual parameter θ_j, we simply write P_j instead of P_{θ_j}.

It is clear that the decision switching probability is defined as the probability of switching the optimal option $d_{\mathrm{opt}}(\boldsymbol{\theta}_1)$ from $d_{\mathrm{opt}}(\emptyset)$ by knowing the exact values of $\boldsymbol{\theta}_1$. This way, the decision switching probability can be useful in understanding which input variable a given decision-making problem under uncertainty is most (or least)

sensitive to, measuring a decision-theoretic probabilistic sensitivity in a different way from the EVPPI. A connection between K_j and P_j is straightforward in that the domain of θ_j such that $d_{\mathrm{opt}}(\theta_j) \neq d_{\mathrm{opt}}(\emptyset)$ is determined by K_j, so that P_j can be computed by using the (marginal) probability distribution of θ_j. Moreover, as explained above, $\mathrm{EVPPI}_{\boldsymbol{\theta}_1} = 0$ is equivalent to $P_{\boldsymbol{\theta}_1} = 0$, as the latter means that $d_{\mathrm{opt}}(\boldsymbol{\theta}_1) = d_{\mathrm{opt}}(\emptyset)$ happens almost surely. However, as discussed in Supplementary Information 1, the larger $\mathrm{EVPPI}_{\boldsymbol{\theta}_1}$ does not necessarily mean the larger $P_{\boldsymbol{\theta}_1}$ and vice versa, and hence, the decision switching probability can provide a complementary information to the existing decision-theoretic probabilistic sensitivity measure.

Regarding an estimation of the probabilistic parameter threshold K_j, a nested Monte Carlo approach is originally employed[27]. The computational procedure with the detailed input and output at each step is described in Algorithm 1. We can see that the algorithm takes a *double-loop* procedure with the outer loop for generating random samples of θ_j and the inner loop for generating random samples of $\boldsymbol{\theta}_{-j} := \boldsymbol{\theta} \setminus \theta_j$ conditional to each sample of θ_j, where M and N denote the numbers of inner and outer samples used, respectively. In the third item of Algorithm 1, the nominal choice of θ_j^* is given by the midpoint $(\theta_j^{(n)} + \theta_j^{(n+1)})/2$. In order to reduce the necessary computational cost, a regression-based approach has been proposed[28], which first approximates the inner conditional expectation $\mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[f_d(\boldsymbol{\theta})\right]$ by a regression model (as a function of θ_j) and then applies a single-loop Monte Carlo sampling for θ_j to estimate K_j. However, these existing approaches rely upon accurate estimation (with large M) or approximation of inner conditional expectations, and both lack a theoretical support on convergence and computational complexity.

Algorithm 1 Nested Monte Carlo[27]

Input: The set of alternative options D, the utility function f_d for each option $d \in D$, the probability distribution for input random variables $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_s)$, positive integers M and N, and the index j with $1 \leq j \leq s$.
Output: Estimates of the parameter threshold K_j.

1. Generate N i.i.d. random samples of θ_j from its marginal distribution, and then sort them in an ascending order, which are denoted by $\theta_j^{(1)} \leq \cdots \leq \theta_j^{(N)}$.

   # ***Input:*** *The marginal probability distribution of* θ_j # ***Output:*** *N random samples of* θ_j *sorted,* $\theta_j^{(1)} \leq \cdots \leq \theta_j^{(N)}$ (***size:*** *N reals)*

2. For each outer sample $\theta_j^{(n)}$, do the following:

   (a) Generate M i.i.d. random samples $\boldsymbol{\theta}_{-j}^{(1,n)}, \ldots, \boldsymbol{\theta}_{-j}^{(M,n)}$ from the condition distribution of $\boldsymbol{\theta}_{-j}$ given $\theta_j = \theta_j^{(n)}$.

   # ***Input:*** *The conditional distribution of* $\boldsymbol{\theta}_{-j}$ *given* $\theta_j = \theta_j^{(n)}$ # ***Output:*** *M random samples of* $\boldsymbol{\theta}_{-j}$, $\boldsymbol{\theta}_{-j}^{(1,n)}, \ldots, \boldsymbol{\theta}_{-j}^{(M,n)}$ (***size:*** $M(s-1)$ *reals)*

   (b) Estimate the conditional expectation $\mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j^{(n)}}\left[f_d(\boldsymbol{\theta})\right]$ for all $d \in D$ by

   $$\overline{f}_d^M(\theta_j^{(n)}) := \frac{1}{M}\sum_{m=1}^{M} f_d(\boldsymbol{\theta}_{-j}^{(m,n)}, \theta_j^{(n)}).$$

   # ***Input:*** *The random samples* $\theta_j^{(n)}$ *and* $\boldsymbol{\theta}_{-j}^{(1,n)}, \ldots, \boldsymbol{\theta}_{-j}^{(M,n)}$ (***size:*** $1 + M(s-1)$ *reals)* # ***Output:*** *The averages* $\overline{f}_d^M(\theta_j^{(n)})$ *for all d* (***size:*** *d reals)*

   (c) Return the option which maximizes the estimated conditional expectation

   $$\overline{d}_{\mathrm{opt}}^M(\theta_j^{(n)}) = \arg\max_{d \in D} \overline{f}_d^M(\theta_j^{(n)}).$$

   # ***Input:*** *The averages* $\overline{f}_d^M(\theta_j^{(n)})$ *for all d* (***size:*** *d reals)* # ***Output:*** *The option* $\overline{d}_{\mathrm{opt}}^M(\theta_j^{(n)})$ (***size:*** *1 index)*

3. Estimate the threshold K_j by

   $$\left\{\theta_j^* \in (\theta_j^{(n)}, \theta_j^{(n+1)}) \mid 1 \leq n \leq N-1 \quad \text{s.t.} \quad \overline{d}_{\mathrm{opt}}^M(\theta_j^{(n)}) \neq \overline{d}_{\mathrm{opt}}^M(\theta_j^{(n+1)})\right\}.$$

Input: *The options* $\overline{d}_{\mathrm{opt}}^M(\theta_j^{(1)}), \ldots, \overline{d}_{\mathrm{opt}}^M(\theta_j^{(N)})$ (***size:*** *N indices)* # ***Output:*** *The estimates of the threhold* K_j (***size:*** *some, possibly zero, reals)*

Organization and contributions of this paper. Motivated mainly by applications to health economic evaluations, the aim of this paper is to develop an efficient algorithm to estimate the probabilistic parameter threshold K_j. We start from providing a formulation of the probabilistic threshold parameter analysis as a root-finding problem involving the conditional expectations. Then we propose a *pairwise stochastic approximation* approach to search for K_j efficiently. The key difference from the existing approaches[27,28] is that our proposed approach only requires an unbiased, but rough estimator of the inner conditional expectations and that the parameter threshold estimate is generated randomly only at the initial step and then updated iteratively. In fact, in our numerical experiments below, we use only one Monte Carlo sample to estimate the inner conditional expectations at each iteration step. Under some mild assumptions on $\boldsymbol{\theta}$ and f_d's, the standard theory on stochastic approximation from the literature[29–34] directly applies to our proposed approach, so that the each element in K_j can be found with a probabilistic error ε typically by the computational cost of $O(|D|^2\varepsilon^{-2})$. Numerical experiments for a simple synthetic testcase which compares three treatments and for a chemotherapy Markov model

(see Supplementary Information 2 for the latter) both illustrate the effectiveness of the proposed approach. Finally we conclude this paper with some remarks. In summary, the main contributions of this paper are given as follows:

1. By reformulating the probabilistic threshold parameter analysis, an efficient pairwise stochastic approximation algorithm is proposed to estimate the probabilistic threshold.
2. Besides that our proposed algorithm is theoretically supported by the standard theory on stochastic approximation, numerical experiments for two examples confirm the effectiveness of our proposed algorithm and the superiority over the existing nested Monte Carlo method.
3. As discussed already, a new decision-theoretic sensitivity measure called decision switching probability is introduced in this paper, providing a complementary information to the existing measure EVPPI.

Methods

Formulation as a root-finding problem. Let us recall that, for a fixed value of θ_j, the optimal option which maximizes the conditional expectation is denoted by

$$d_{\mathrm{opt}}(\theta_j) = \arg\max_{d \in D} \mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[f_d(\boldsymbol{\theta})\right],$$

and that our task is to find θ_j such that $d_{\mathrm{opt}}(\theta_j)$ is not uniquely defined. Let the set $\{d^1(\theta_j), d^2(\theta_j), \ldots, d^{|D|}(\theta_j)\}$ be a reordering of the elements in D such that the inequality

$$\mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[f_{d^1(\theta_j)}(\boldsymbol{\theta})\right] \geq \mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[f_{d^2(\theta_j)}(\boldsymbol{\theta})\right] \geq \cdots \geq \mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[f_{d^{|D|}(\theta_j)}(\boldsymbol{\theta})\right]$$

holds. Note that this ordering is arbitrary wherever the equality holds. With this notation, the fact that $d_{\mathrm{opt}}(\theta_j)$ is not unique is equivalent that the ordering of $d^1(\theta_j)$ and $d^2(\theta_j)$ is arbitrary, meaning that the corresponding conditional expectations are equal to each other. Hence, the probabilistic parameter threshold K_j is equivalently given by

$$\begin{aligned} K_j &= \left\{\theta_j \mid \mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[f_{d^1(\theta_j)}(\boldsymbol{\theta})\right] = \mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[f_{d^2(\theta_j)}(\boldsymbol{\theta})\right]\right\} \\ &= \left\{\theta_j \mid \mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[\left(f_{d^1(\theta_j)} - f_{d^2(\theta_j)}\right)(\boldsymbol{\theta})\right] = 0\right\}. \end{aligned}$$

Through this representation, our problem reduces to a root-finding problem which involves the conditional expectations. However, it is generally unknown which options correspond to $d^1(\theta_j)$ and $d^2(\theta_j)$, respectively, for given θ_j.

Now, for two different options $d_1, d_2 \in D$, we write

$$K_j^{(d_1,d_2)} = \left\{\theta_j \mid \mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[\left(f_{d_1} - f_{d_2}\right)(\boldsymbol{\theta})\right] = 0\right\}.$$

Note that $K_j^{(d_1,d_2)}$ can be the empty set if either

$$\mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[\left(f_{d_1} - f_{d_2}\right)(\boldsymbol{\theta})\right] > 0 \quad \text{or} \quad \mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[\left(f_{d_1} - f_{d_2}\right)(\boldsymbol{\theta})\right] < 0$$

holds for any θ_j. It is obvious that we have $K_j^{(d_1,d_2)} = K_j^{(d_2,d_1)}$ and

$$K_j = K_j^{(d_1,d_2)},$$

if $|D| = 2$, and

$$K_j \subseteq \bigcup_{\substack{d_1, d_2 \in D \\ d_1 \neq d_2}} K_j^{(d_1,d_2)},$$

if $|D| \geq 3$. Thus it suffices to search for the set $K_j^{(d_1,d_2)}$ for all the possible pairs $d_1, d_2 \in D$ first and then to check whether each element in $\bigcup_{\substack{d_1, d_2 \in D \\ d_1 \neq d_2}} K_j^{(d_1,d_2)}$ is contained in K_j or not. Note that the second step is not necessary for the case $|D| = 2$.

Motivated by the formulation presented here, we below consider applying a stochastic approximation to find the roots of the conditional expectation $\mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[\left(f_{d_1} - f_{d_2}\right)(\boldsymbol{\theta})\right]$ and then propose a pairwise stochastic approximation approach to search for the threshold K_j, wherein some postprocessing based on a statistical hypothesis testing is required for the case $|D| \geq 3$ to see whether each element in $\bigcup_{\substack{d_1, d_2 \in D \\ d_1 \neq d_2}} K_j^{(d_1,d_2)}$ is contained in K_j or not.

Stochastic approximation for root-finding. Let $d_1, d_2 \in D$ be two different options. In order to find the set $K_j^{(d_1,d_2)}$, i.e., the roots of the conditional expectation $\mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[\left(f_{d_1} - f_{d_2}\right)(\boldsymbol{\theta})\right]$, we use a stochastic approximation method. We refer to the book[35] and the review article[36] for a comprehensive information on stochastic approximation algorithms. In what follows, we briefly describe the stochastic approximation algorithm, as if the

set $K_j^{(d_1,d_2)}$ contains only one element, which is denoted by $\tilde{\theta}_j^{(d_1,d_2)}$. Note that the resulting estimate will diverge if $K_j^{(d_1,d_2)}$ is empty, and also that several independent runs with different initial estimates are required if $K_j^{(d_1,d_2)}$ contains more than one element and all the elements are needed to be found.

For a fixed value of θ_j and $M \in \mathbb{Z}_{>0}$, we denote by $\boldsymbol{\theta}_{-j}^{(1)}, \ldots, \boldsymbol{\theta}_{-j}^{(M)}$ the i.i.d. random samples of $\boldsymbol{\theta}_{-j}$ conditional on θ_j. Then the conditional expectation $\mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[\left(f_{d_1} - f_{d_2}\right)(\boldsymbol{\theta})\right]$ can be estimated unbiasedly by the following Monte Carlo estimator:

$$\overline{f_{d_1} - f_{d_2}}^{M}(\theta_j) := \frac{1}{M}\sum_{m=1}^{M}\left(f_{d_1} - f_{d_2}\right)(\boldsymbol{\theta}_{-j}^{(m)}, \theta_j). \tag{1}$$

Then the classical Robbins-Monro algorithm[29] searches for the solution $\tilde{\theta}_j^{(d_1,d_2)}$ by

$$\theta_j^{t+1} = \theta_j^t - \alpha_t \times \overline{f_{d_1} - f_{d_2}}^{M}(\theta_j^t) \tag{2}$$

with an initial point θ_j^1 and a sequence of decreasing step sizes $\alpha_1, \alpha_2, \ldots > 0$. The initial estimate θ_j^1 can be generated, for instance, randomly from the marginal probability distribution of the variable θ_j. The well-known averaging technique, found independently by Polyak[32] and Ruppert[33], outputs the average

$$\Theta_j^t := \frac{1}{t}\sum_{u=1}^{t}\theta_j^u,$$

instead of the nominal estimate θ_j^t. In order to establish a convergence result of the estimate θ_j^t to $\tilde{\theta}_j^{(d_1,d_2)}$, Robbins and Monro originally consider the following assumptions[29]:

1. (conditional expectation)

$$\begin{cases} \mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[\left(f_{d_1} - f_{d_2}\right)(\boldsymbol{\theta})\right] < 0 \ \text{ for } \theta_j < \tilde{\theta}_j^{(d_1,d_2)}, \\ \mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[\left(f_{d_1} - f_{d_2}\right)(\boldsymbol{\theta})\right] > 0 \ \text{ for } \theta_j > \tilde{\theta}_j^{(d_1,d_2)}. \end{cases}$$

2. (conditional variance)

$$\mathbb{V}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[\left(f_{d_1} - f_{d_2}\right)(\boldsymbol{\theta})\right] \le \sigma_j^2 < \infty$$

   holds for any θ_j.
3. (step sizes)

$$\sum_{t=1}^{\infty}\alpha_t = \infty \quad \text{and} \quad \sum_{t=1}^{\infty}\alpha_t^2 < \infty.$$

It is obvious that when the sign of the conditional expectation given in the first item is opposite, the recursion (2) should be replaced by

$$\theta_j^{t+1} = \theta_j^t + \alpha_t \times \overline{f_{d_1} - f_{d_2}}^{M}(\theta_j^t).$$

It follows from the third item that the step sizes must decay at the order of $t^{-\alpha}$ with $1/2 < \alpha \le 1$. As shown by Ruppert and Juditsky[34], if the conditional expectation $\mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}\left[\left(f_{d_1} - f_{d_2}\right)(\boldsymbol{\theta})\right]$ is linear in θ_j, this condition can be relaxed to $0 < \alpha < 1$ by the Polyak-Ruppert averaging, which allows for more slowly decaying step sizes. Regarding the results on the convergence rates, we refer to Section 5 of the review article[36] both for the standard Robbins-Monro iteration and for the Polyak-Ruppert averaging. Roughly speaking, the estimate (θ_j^t or Θ_j^t) typically converges almost surely to $\tilde{\theta}_j^{(d_1,d_2)}$ with the rate of $1/\sqrt{t}$ under mind assumptions on $\boldsymbol{\theta}$ and f_ds.

Remark 2 The stochastic approximation algorithm described above does work to search for the pairwise set $K_j^{(d_1,d_2)}$ for any sample size $M \ge 1$ in (1). Hence, by formulating the probabilistic threshold analysis as a stochastic root-finding problem, we can avoid the difficulty inherent to the nested structure considered in the literature[27,28]. Moreover, in order to improve the stability of the algorithm, we can apply some of variance reduction techniques including Latin hypercube sampling[37] or (randomized) quasi-Monte Carlo sampling[38], as long as the resulting estimator is unbiased as with the standard one (1).

Search for parameter threshold. Having estimated the set $K_j^{(d_1,d_2)}$ for all the possible pairs $d_1, d_2 \in D$, it suffices to check whether each element in the estimated set $K_j^{(d_1,d_2)}$ is contained in K_j or not. Note again that this step is not necessary if $|D| = 2$. For $|D| \ge 3$, we carry out this step by the following statistical hypothesis testing.

Let $\hat{\theta}_j^{(d_1,d_2)}$ be an element in the estimated set $K_j^{(d_1,d_2)}$. Then the null and alternative hypotheses are given by

$$H_0 : \hat{\theta}_j^{(d_1,d_2)} \in K_j \quad \text{and} \quad H_1 : \hat{\theta}_j^{(d_1,d_2)} \notin K_j,$$

respectively. The condition for the null hypothesis H_0 is equivalent that

$$\mathbb{E}_{\boldsymbol{\theta}_{-j}|\hat{\theta}_j^{(d_1,d_2)}}\left[\left(f_d - f_{d_1}\right)(\boldsymbol{\theta})\right] \le 0 \quad \text{and} \quad \mathbb{E}_{\boldsymbol{\theta}_{-j}|\hat{\theta}_j^{(d_1,d_2)}}\left[\left(f_d - f_{d_2}\right)(\boldsymbol{\theta})\right] \le 0$$

hold for all $d \in D \setminus \{d_1, d_2\}$, while the condition for the alternative hypothesis H_1 is equivalent that there exists at least one option $d \in D \setminus \{d_1, d_2\}$ such that either

$$\mathbb{E}_{\boldsymbol{\theta}_{-j}|\hat{\theta}_j^{(d_1,d_2)}}\left[\left(f_d - f_{d_1}\right)(\boldsymbol{\theta})\right] > 0 \quad \text{or} \quad \mathbb{E}_{\boldsymbol{\theta}_{-j}|\hat{\theta}_j^{(d_1,d_2)}}\left[\left(f_d - f_{d_2}\right)(\boldsymbol{\theta})\right] > 0$$

holds. Assuming the normality of the Monte Carlo estimators

$$\overline{f_d - f_{d_1}}^N(\hat{\theta}_j^{(d_1,d_2)}) \quad \text{and} \quad \overline{f_d - f_{d_2}}^N(\hat{\theta}_j^{(d_1,d_2)})$$

for all $d \in D \setminus \{d_1, d_2\}$ with large sample size N, for instance, the conventional one-sided t-test applies independently to each individual inequality null, and the null hypothesis H_0 will be rejected or not with some significance level.

The necessary cost of the hypothesis testing is considered moderate or even negligible as compared to that of estimating the set $K_j^{(d_1,d_2)}$ for all the possible pairs $d_1, d_2 \in D$. Since the convergence results on the stochastic approximation method implies that each element in $K_j^{(d_1,d_2)}$ can be estimated with a probabilistic error ε by the cost of $O(\varepsilon^{-2})$, the total cost of our proposed approach to estimate the probabilistic threshold K_j itself is of order $O(|D|^2\varepsilon^{-2})$ (up to the cardinality of each pairwise set $K_j^{(d_1,d_2)}$), where the factor $|D|^2$ comes from the number of possible pairs $d_1, d_2 \in D$, which is $|D|(|D|-1)/2$. The overall computational procedure of our proposed approach is summarized in Algorithm 2.

Algorithm 2 Pairwise stochastic approximation

Input: The set of alternative options D, the utility function f_d for each option $d \in D$, the probability distribution for input random variables $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_s)$, positive integers M, T and N, and the index j with $1 \le j \le s$.
Output: Estimates of the parameter threshold K_j.

For each possible pair $d_1, d_2 \in D$, do the following:

1. Generate an initial estimate θ_j^1.
   *# **Input (example):** The marginal distribution of θ_j # **Output:** The initial estimate θ_j^1 (**size:** 1 real)*
2. For $1 \le t < T$, update the estimate θ_j^t iteratively by
   $$\theta_j^{t+1} = \theta_j^t - \alpha_t \times \overline{f_{d_1} - f_{d_2}}^M(\theta_j^t).$$
   *# **Input:** The t-th estimate θ_j^t and M random samples of $\boldsymbol{\theta}_{-j}$ (**size:** $1 + M(s-1)$ reals) # **Output:** The $(t+1)$-th estimate θ_j^{t+1} (**size:** 1 real)*
3. Estimate a single element in $K_j^{(d_1,d_2)}$ by the final estimate θ_j^T (or the average Θ_j^T) if converged, and go to Step 4. Otherwise, discard the estimate.
   *# **Input:** The T-th estimate θ_j^T (or the average Θ_j^T) (**size:** 1 real) # **Output:** The decision whether to go to Step 4 or to discard the estimate*
4. Apply the hypothesis testing with the null and alternative hypotheses $H_0 : \theta_j^T \in K_j$ and $H_1 : \theta_j^T \notin K_j$, respectively, with N random samples for $\boldsymbol{\theta}_{-j}$ given $\theta_j = \theta_j^T$, wherein θ_j^T should be replaced by Θ_j^T if the latter is used as the final estimate.
   *# **Input:** The T-th estimate θ_j^T (or the average Θ_j^T) and N random samples of $\boldsymbol{\theta}_{-j}$ (**size:** $1 + N(s-1)$ reals) # **Output:** The decision whether to reject $\theta_j^T \in K_j$ (or $\Theta_j^T \in K_j$) or not*

Numerical experiments

To demonstrate the effectiveness of our proposed approach, here we conduct numerical experiments for a simple synthetic testcase[39] comparing three medical treatments. In Supplementary Information 2, we present our numerical results for a more complicated chemotherapy Markov model introduced by Heath and Baio[40].

Model setting. The example we use here is taken from Hironaka et al.[39], which extends the model originally introduced by Ades et al.[41] in the context of medical decision making. As was explained[39], the original version of this synthetic cost-effectiveness model compares only two treatments (some standard of care and a new treatment) on the prevention of a critical event, denoted by E, whereas three different treatments $D = \{d_1, d_2, d_3\}$ are compared in the extended model with d_1 being the standard of care and d_2 and d_3 being two different new treatments. The standard of care d_1, on the one hand, is cost-free and has no risk that the side effect (SE) occurs, while the probability that the critical event occurs is relatively large. The new treatments d_2 and d_3, on the other hand, are both costly and have some probabilities that the side effect occurs, while the probabilities of the critical event are relatively smaller than d_1.

Importantly, the above-mentioned costs and probabilities of the critical event and the side effect themselves are not known precisely, so that we model them as random variables. We refer to Table 1 for a detailed description on these model inputs. The utility function f_d for each treatment $d \in D$ represents the monetary net benefit of d as a function of the input vector $\boldsymbol{\theta}$ which consists of 12 individual random variables $L, Q_E, Q_{SE}, C_E, C_{SE}, C_{T,d_2}, C_{T,d_3}$,

Description	Parameter	Distribution
Lifetime remaining	$L\,(\theta_1)$	$N\,(30, 25)$
QALY after critical event, per year	$Q_E\,(\theta_2)$	logit-normal$(0.6, 1/36)$
QALY decrement due to side effects	$Q_{SE}\,(\theta_3)$	$N\,(0.7, 0.01)$
Cost of critical event	$C_E\,(\theta_4)$	$N(2\times10^5, 10^8)$
Cost of side effect	$C_{SE}\,(\theta_5)$	$N(10^5, 10^8)$
Cost of treatment $d = d_1$	C_{T,d_1}	0 (constant)
Cost of treatments $d = d_2, d_3$	$C_{T,d}\,(\theta_6, \theta_7)$	$N\left(\begin{pmatrix}1.5\times10^4\\ 2\times10^4\end{pmatrix}, \begin{pmatrix}300 & 100\\ 100 & 500\end{pmatrix}\right)$
Probability of critical event on treatment $d = d_1$	$P_{E,d_1}\,(\theta_8)$	Beta$(15, 85)$
Odds ratios of critical event relative to treatment $d = d_1$ $\frac{P_{E,d}/(1-P_{E,d})}{P_{E,d_1}/(1-P_{E,d_1})}$	$OR_{E,d}\,(\theta_9, \theta_{10})$	log-normal$\left(\begin{pmatrix}-1.5\\ -1.75\end{pmatrix}, \begin{pmatrix}0.11 & 0.02\\ 0.02 & 0.06\end{pmatrix}\right)$
Probability of critical event on treatments $d = d_2, d_3$	$P_{E,d}$	Derived from P_{E,d_1} and $OR_{E,d}$
Probability of side effect on treatment $d = d_1$	P_{SE,d_1}	0 (constant)
Probability of side effect on treatments $d = d_2, d_3$	$P_{SE,d}\,(\theta_{11}, \theta_{12})$	logit-normal$\left(\begin{pmatrix}-1.4\\ -1.1\end{pmatrix}, \begin{pmatrix}0.10 & 0.05\\ 0.05 & 0.25\end{pmatrix}\right)$
Monetary value of 1 QALY	λ	75,000 (constant)

Table 1. The input parameters involved in the synthetic testcase. Note that log-normal(μ, Σ) and logit-normal(μ, Σ) denote the log-normal and logit-normal distributions, respectively, with μ and Σ being the mean vector and the covariance matrix of the corresponding normal distribution, respectively. Beta(α, β) denotes the Beta distribution with shape parameters $\alpha, \beta > 0$. The word QALY appearing in the first column stands for quality-adjusted life year.

$P_{E,d_1}, OR_{E,d_2}, OR_{E,d_3}, P_{SE,d_2}, P_{SE,d_3}$, which are denoted by $\theta_1, \ldots, \theta_{12}$, respectively, in this order. Furthermore, the model contains three constants $C_{T,d_1}, P_{SE,d_1}, \lambda$, while two parameters P_{E,d_2} and P_{E,d_3} are defined as functions of P_{E,d_1}, OR_{E,d_2} and P_{E,d_1}, OR_{E,d_3}, respectively. Here, unlike the original model[41], the extended model includes the correlations between odds ratios (OR) of the critical events (OR_{E,d_2}, OR_{E,d_3}), treatment costs (C_{T,d_2}, C_{T,d_3}), and probabilities of side effects (P_{SE,d_2}, P_{SE,d_3}), which makes the decision-making problem computationally harder.

Now the net benefit function f_d for each treatment $d \in D$ is defined by

$$f_d(\boldsymbol{\theta}) = P_{SE,d}P_{E,d}\left[\lambda\left(L\frac{1+Q_E}{2} - Q_{SE}\right) - (C_{SE} + C_E)\right] + P_{SE,d}(1-P_{E,d})[\lambda(L-Q_{SE}) - C_{SE}]$$
$$+ (1-P_{SE,d})P_{E,d}\left[\lambda L\frac{1+Q_E}{2} - C_E\right] + (1-P_{SE,d})(1-P_{E,d})\lambda L - C_{T,d},$$

where the first four terms correspond to possible four outcomes (whether or not the side effect occurs and whether or not the critical event occurs) and the fifth term denotes the cost of d. We note that the net benefit is expressed as a multi-linear function of most of the elements in $\boldsymbol{\theta}$. However, f_{d_2} and f_{d_3} are both nonlinear with respect to P_{E,d_1}, OR_{E,d_2} and P_{E,d_1}, OR_{E,d_3}, respectively, which makes it hard to compute the probabilistic parameter thresholds exactly for this model. As has been discussed, our interest is to infer which input parameter affects the choice of the optimal treatment more or least significantly.

Results and discussion

Reference results. Let us consider below estimating the probabilistic parameter thresholds for 6 input variables θ_3, θ_5, θ_7, θ_{10}, θ_{11} and θ_{12}, respectively. Before applying our proposed approach, we first show some reference results by estimating the conditional expectations $\mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}[f_d(\boldsymbol{\theta})]$ here. More precisely, for each considered input variable θ_j, we estimate the conditional expectations $\mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}[f_d(\boldsymbol{\theta})]$ for all $d \in \{d_1, d_2, d_3\}$ by using the naive Monte Carlo average

$$\mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}[f_d(\boldsymbol{\theta})] \approx \frac{1}{N}\sum_{n=1}^{N} f_d(\boldsymbol{\theta}_{-j}^{(n)}, \theta_j)$$

with large sample size $N = 2^{18}$ for various values of θ_j. Here, because of the multi-linearity of the functions f_d, the exact mean of an individual random variable can be substituted directly wherever available. The results are shown in Fig. 2. Except for the variable θ_{10}, we can see that there exists exactly one intersection between every two different treatments: (d_1, d_2), (d_1, d_3) and (d_2, d_3), where the intersection of the pair (d_1, d_3) for θ_7 exists beyond the range of this plot. It follows that the probabilistic parameter threshold K_j consists of two elements for the variables except θ_{10}. Regarding the variable θ_{10}, the treatment d_2 always leads to a larger conditional expectation than the treatment d_3 and the corresponding probabilistic parameter threshold K_{10} consists of only one element. We can use these results as a reference to see whether our proposed approach can search for the probabilistic parameter thresholds correctly.

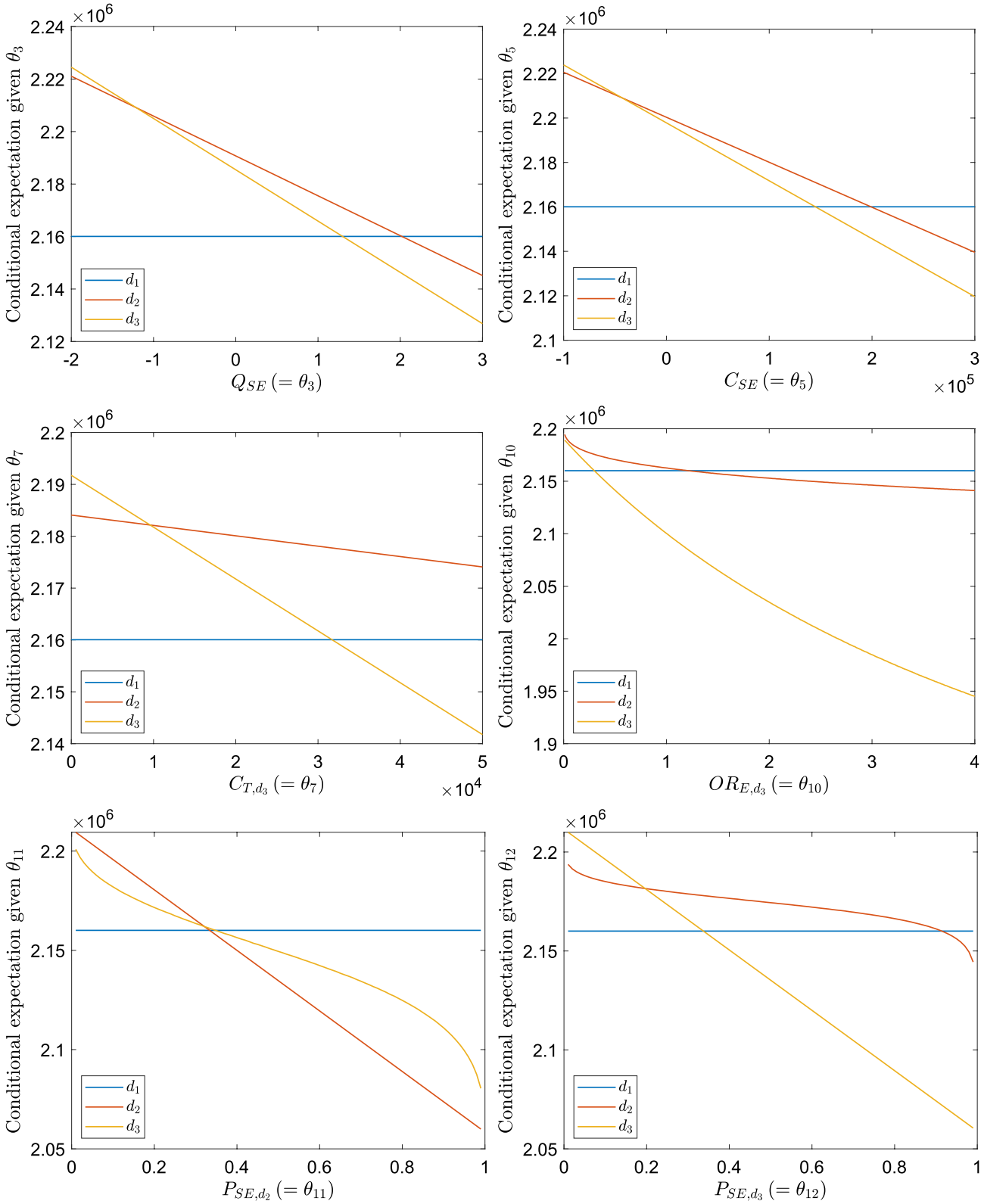

Figure 2. The conditional expectations $\mathbb{E}_{\boldsymbol{\theta}_{-j}|\theta_j}[f_d(\boldsymbol{\theta})]$ as functions of θ_j for $d \in \{d_1, d_2, d_3\}$. The results for θ_3 (left top), θ_5 (right top), θ_7 (left middle), θ_{10} (right middle), θ_{11} (left bottom) and θ_{12} (right bottom) are shown respectively.

Experimental setup. We use our proposed Algorithm 2 with $M = 1$, $T = N = 10^4$ to estimate the threshold K_j, which means that we use only one sample of $\boldsymbol{\theta}_{-j}$ at each iteration step in the second item of Algorithm 2. For the variables θ_{10}, θ_{11} and θ_{12}, we consider the transformed variables $\log\theta_{10}$, $\mathrm{logit}(\theta_{11})$, and $\mathrm{logit}(\theta_{12})$ instead, respectively, for the iterations of stochastic approximation. In the first item of Algorithm 2, we generate θ_j^1 from the marginal distribution of θ_j. We set the sequence of step sizes to

$$\alpha_t = \frac{3\sigma(\theta_j)}{2 \times 10^4 \sqrt{t}},$$

where $\sigma(\theta_j)$ denotes the standard deviation for the marginal distribution of the variable θ_j, and consider the averaged outputs Θ_j^t with $t = 1, 2, \ldots, T$ as a sequence of our threshold estimates. Regarding the variable θ_7, we enlarge α_t by a constant factor so that the resulting estimate Θ_7^t converges within $T = 10^4$ iteration steps. We carry out 20 independent runs for each considered variable.

Convergence of pairwise estimates. As the last paragraph shows how to set the input and the first item of Algorithm 2, here we discuss the results obtained from the second and third items of Algorithm 2. Figure 3 shows the convergence behaviors of the estimates Θ_j^t as functions of the iteration step t, obtained from the second item of Algorithm 2, for all the possible pairs (d_1, d_2), (d_1, d_3) and (d_2, d_3). Except for the pair (d_2, d_3) for the variable $\log\theta_{10}$, the mean estimate from 20 independent runs converges to a value which agrees well with the intersection point shown in Fig. 2, and the standard error gets smaller with the convergence rate of approximately $t^{-1/2}$ as the iteration step t increases. These observations are exactly what we expect from the theory of stochastic approximation. Due to the convergence to a constant value, these estimates pass the third item of Algorithm 2 and can be subject to the last item.

Regarding the pair (d_2, d_3) for the variable $\log\theta_{10}$, for which any intersection is not observed in Fig. 2, the mean estimate itself does not converge and the magnitude of the standard error stays almost the same along the iteration steps. This way, the resulting estimates do not pass the the third item of Algorithm 2 and we can infer that the set $K_{10}^{(d_2,d_3)}$ is empty. Interestingly, as clearly seen from the result for the pair (d_1, d_2) for the variable θ_7, the stochastic approximation method can find the pairwise threshold successfully even if it is located far from the initial estimate. It seems quite hard to get similar results if we only generate θ_7 randomly from its marginal distribution as done in the existing approaches[27,28].

Estimated probabilistic thresholds. Finally we go to the fourth item of Algorithm 2. Table 2 summarizes the final estimates obtained from the stochastic approximation after $T = 10^4$ iteration steps. Here the p-value is computed as follows. As an example, let us consider the estimated set for $K_j^{(d_1,d_2)}$. If it is not empty, we apply the one-sided t-test to the two pairs (d_1, d_3) and (d_2, d_3), which gives two p-values. By doing this for each of 20 independent runs, we have 20 of two p-values and finally we take the maximum value of the 20 p-values for each pair (d_1, d_3) and (d_2, d_3), respectively. In fact, the variables and the pairs in Table 2, which show the set of p-values as (1, 1), consistently yield (1, 1) for all 20 independent runs, which indicates that the corresponding estimate is contained in the threshold K_j. For some cases, the set of p-values is (0, 0), which clearly indicates that the corresponding estimate is *not* contained in K_j. The marginal case is found only for the pair (d_1, d_2) for the variable $\log\theta_{10}$. As shown in Fig. 2, three intersections of the conditional expectations are close to each other, which makes it difficult to estimate K_j correctly. Nevertheless, the p-values are small enough for the pair (d_1, d_2), so that the estimated $K_{10}^{(d_1,d_2)}$ is properly discarded from the set K_{10} with a sufficiently small significance level. Almost the same results are also obtained by applying one-sided Wilcoxon signed-rank test instead. This way our proposed algorithm can give estimates of the parameter probabilistic threshold which agree quite well with those expected from the reference results.

Comparison with nested Monte Carlo. As a comparison, we estimate K_j also by the nested Monte Carlo approach (Algorithm 1) with $M = 10^4$ and $N = 10^2$. Note that this choice of M and N is close to the one used by McCabe et al.[27] and that the total cost for each variable is $MN = 10^6$, which is already larger than that of our proposed approach. We have the following observations. Firstly, we cannot find any element in K_j for the variables $\theta_3, \theta_5, \theta_7$ and $\log\theta_{10}$. This is expected from the fact that the outer samples are generated randomly from the marginal distribution for the nested Monte Carlo approach, whereas every threshold is located far from the region where the marginal probability distribution is concentrated in this case. Secondly, regarding the variable $\mathrm{logit}(\theta_{11})$, the resulting estimate is quite unstable in the sense that some runs (2 out of 20 runs) estimate that the threshold K_{11} is empty, some (13 out of 20 runs) only estimate one of the two elements, and some (4 out of 20 runs) estimate the two elements but with larger variations than those by our proposed approach, and the remaining run mistakenly estimates that K_{11} consists of three elements. The last mistaken estimation happens because the conditional expectations for at least two treatments are close to each other around the threshold and so the Monte Carlo estimation possibly returns a wrong treatment as the one which maximizes the conditional expectation. Finally, for the variable $\mathrm{logit}(\theta_{12})$, one of the two elements around -1.41, i.e., the intersection of the pair (d_2, d_3), is estimated correctly for most runs (19 out of 20 runs), whereas the remaining run mistakenly estimates three distinct elements around -1.41. The mean and its standard error for the 19 runs are given by -1.41 and 1.13×10^{-2}, respectively. Note that the mean agrees well with that obtained by our proposed approach, while the standard error is about twice larger for the nested Monte Carlo approach. No run can find the other element of K_{12} around 2.37. This result clearly shows the superiority of our proposed approach.

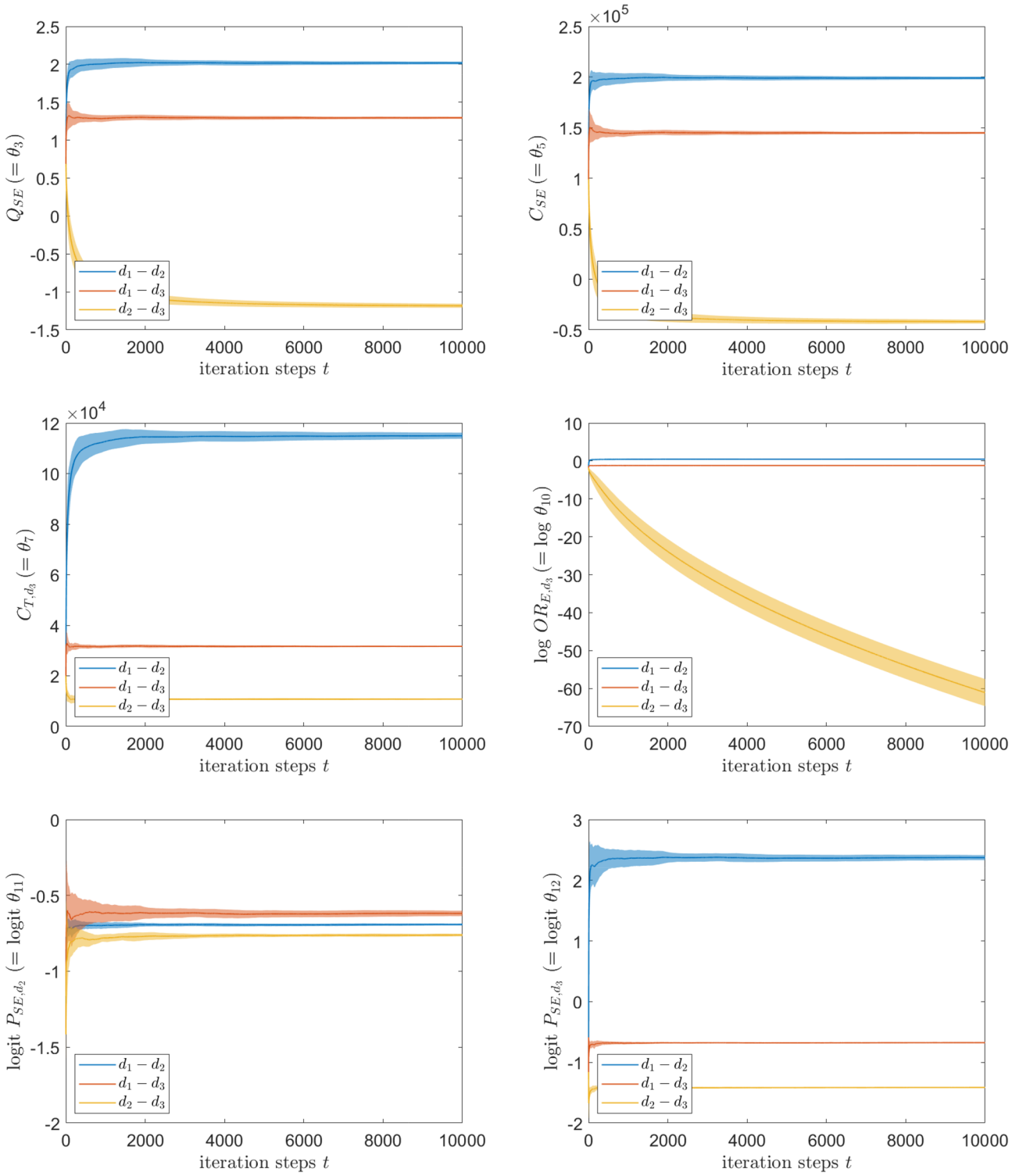

Figure 3. The pairwise probabilistic thresholds for the synthetic testcase found by the stochastic approximation with Polyak-Ruppert averaging as functions of iteration steps t. The results for θ_3 (left top), θ_5 (right top), θ_7 (left middle), $\log\theta_{10}$ (right middle), $\text{logit}(\theta_{11})$ (left bottom) and $\text{logit}(\theta_{12})$ (right bottom) are shown respectively. For each pair of two treatments, the line and the shaded area represent the mean and its standard error estimated from 20 independent runs, respectively.

Decision switching probability. Using the results for the probabilistic thresholds, we can identify the intervals of θ_j where $d_{\text{opt}}(\theta_j)$ is equal to d_1, d_2 and d_3, respectively, as shown in Table 3. As we have $d_{\text{opt}}(\emptyset) = d_2$ for this model, the decision switching probability for a variable θ_j is given by

Variable		$K_j^{(d_1,d_2)}$	$K_j^{(d_1,d_3)}$	$K_j^{(d_2,d_3)}$
θ_3	Mean (std)	**2.02** (0.0176)	1.30 (0.0130)	$-$**1.18** (0.026)
	p-value	(1, 1)	(0, 0)	(1, 1)
θ_5	Mean (std)	$\mathbf{1.99 \times 10^5}$ (1.31×10^3)	1.45×10^5 (1.00×10^3)	$\mathbf{-4.19 \times 10^4}$ (1.88×10^3)
	p-value	(1, 1)	(0, 0)	(1, 1)
θ_7	Mean (std)	$\mathbf{1.15 \times 10^5}$ (1.21×10^3)	3.17×10^4 (2.29×10^2)	$\mathbf{1.08 \times 10^4}$ (2.49×10^2)
	p-value	(1, 1)	(0, 0)	(1, 1)
$\log\theta_{10}$	Mean (std)	**0.45** (1.77×10^{-2})	-1.21 (8.73×10^{-3})	Empty
	p-value	(1, 1)	(0, 0)	
$\text{logit}(\theta_{11})$	Mean (std)	-0.69 (6.27×10^{-3})	$\mathbf{-0.62}$ (1.78×10^{-2})	$\mathbf{-0.76}$ (8.09×10^{-3})
	p-value	$(5.36 \times 10^{-4}, 9.38 \times 10^{-8})$	(1, 1)	(1, 1)
$\text{logit}(\theta_{12})$	Mean (std)	**2.37** (4.16×10^{-2})	-0.66 (5.57×10^{-3})	$\mathbf{-1.41}$ (6.35×10^{-3})
	p-value	(1, 1)	(0, 0)	(1, 1)

Table 2. The results of the pairwise probabilistic thresholds for the synthetic testcase found by the stochastic approximation with Polyak-Ruppert averaging after 10^4 iterations. The mean and its standard error are estimated from 20 independent runs, respectively. The p-values denote the maximum values for the one-sided t-test among 20 independent runs. The bold-typed numbers indicate that the corresponding elements are contained in the threshold K_j.

Variable	$d_{\text{opt}}(\theta_j) = d_1$	$d_{\text{opt}}(\theta_j) = d_2$	$d_{\text{opt}}(\theta_j) = d_3$	DSP	EVPPI
θ_3	$(2.02, \infty)$	$(-1.18, 2.02)$	$(-\infty, -1.18)$	5.57×10^{-40}	0.0
θ_5	$(1.99 \times 10^5, \infty)$	$(-4.19 \times 10^4, 1.99 \times 10^5)$	$(-\infty, -4.19 \times 10^4)$	2.17×10^{-23}	0.0
θ_7	$(1.15 \times 10^5, \infty)$	$(1.08 \times 10^4, 1.15 \times 10^5)$	$(-\infty, 1.08 \times 10^4)$	0.0	0.0
$\log\theta_{10}$	$(0.45, \infty)$	$(-\infty, 0.45)$	None	1.36×10^{-19}	0.00
$\text{logit}(\theta_{11})$	$(-0.62, \infty)$	$(\infty, -0.76)$	$(-0.76, -0.62)$	2.17×10^{-2}	54.3
$\text{logit}(\theta_{12})$	$(2.37, \infty)$	$(-1.41, 2.37)$	$(-\infty, -1.41)$	2.65×10^{-1}	1308.9

Table 3. The results of the interval for the synthetic testcase over which the corresponding treatment $d \in \{d_1, d_2, d_3\}$ is optimal. The decision switching probability and the EVPPI are also shown in the fifth and sixth columns, respectively.

$$P_j = \mathbb{P}_{\theta_j}[d_{\text{opt}}(\theta_j) = d_1] + \mathbb{P}_{\theta_j}[d_{\text{opt}}(\theta_j) = d_3].$$

Each of two terms on the right-hand side can be computed from the marginal cumulative distribution function of θ_j. The results are shown in the fifth column of Table 3. As a reference, we estimate the EVPPI for each variable by using the nested Monte Carlo estimator[17] with 2^{18} outer samples for θ_j and 2^{10} inner conditional samples for $\boldsymbol{\theta}_{-j}$, wherein the exact mean of a random variable is substituted directly whenever available. Note that the EVPI is estimated to be 4063.5 for this model[39]. Since every threshold for the variables $\theta_3, \theta_5, \theta_7$ and $\log\theta_{10}$ is located far from the region where the marginal probability distribution is concentrated, the corresponding decision switching probability is extremely small, and so is the EVPPI. On the contrary, one of the thresholds for both $\text{logit}(\theta_{11})$ and $\text{logit}(\theta_{12})$ is in the concentrated region, and so, knowing the exact value of $\text{logit}(\theta_{11})$ or $\text{logit}(\theta_{12})$ has some chance of changing the optimal treatment from d_2. In this case, the decision switching probability for $\text{logit}(\theta_{12})$ is much larger than that for $\text{logit}(\theta_{11})$, and so is the EVPPI, which indicates that the variable $\text{logit}(\theta_{12})$ is more sensitive in choosing the optimal treatment.

Concluding remarks

In this paper we have developed an efficient pairwise stochastic approximation approach to estimate the probabilistic parameter threshold. Not only the standard theory on the convergence of stochastic approximation algorithms directly applies to our proposed approach but also the numerical experiments have confirmed that our proposed approach works quite well both for a simple synthetic testcase and a chemotherapy Markov model. Moreover, we have introduced a new measure called the decision switching probability for probabilistic sensitivity analysis in the context of health economic evaluations, or more broadly, decision-making under uncertainty, which can deliver a complementary information to the existing decision-theoretic probabilistic sensitivity measure EVPPI.

The following issues are left for future research:

- As with the existing methods[27,28], our proposed approach applies only to the probabilistic parameter threshold for a single input variable θ_j. An extension to the multivariate case is interesting but does not seem straightforward.
- Although we have not discussed in this paper, it is clear that the decision switching probability can be also defined for the sample information and used as a complementary measure to the *expected information of sample information*[39,41–44]. We need further investigation on how to efficiently estimate the decision switching probability for sample information, as our present approach using the probabilistic parameter threshold and the marginal probability distribution is not straightforward to extend.

Received: 12 March 2021; Accepted: 20 September 2021
Published online: 04 October 2021

References

1. Oakley, J. E. & O'Hagan, A. Probabilistic sensitivity analysis of complex models: A Bayesian approach. *J. R. Stat. Soc. Ser. B* **66**, 751–769 (2004).
2. Saltelli, A. *et al. Global Sensitivity Analysis. The Primer* (Wiley, 2008).
3. Sobol', I. M. Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.* **1**, 407–414 (1993).
4. Sobol', I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **55**, 271–280 (2001).
5. Owen, A. B. Monte Carlo theory, methods and examples. https://statweb.stanford.edu/~owen/mc/ (2019).
6. Saltelli, A. Making best use of model evaluations to compute sensitivity indices. *Comput. Phys. Commun.* **145**, 280–297 (2002).
7. Sudret, B. Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Syst. Saf.* **93**, 964–979 (2008).
8. Saltelli, A. *et al.* Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput. Phys. Commun.* **181**, 259–270 (2010).
9. Janon, A., Klein, T., Lagnoux, A., Nodet, M. & Prieur, C. Asymptotic normality and efficiency of two Sobol index estimators. *ESAIM Probab. Stat.* **18**, 342–364 (2014).
10. Konakli, K. & Sudret, B. Global sensitivity analysis using low-rank tensor approximations. *Reliab. Eng. Syst. Saf.* **156**, 64–83 (2016).
11. Scollo, S., Tarantola, S., Bonadonna, C., Coltelli, M. & Saltelli, A. Sensitivity analysis and uncertainty estimation for tephra dispersal models. *J. Geophys. Res.* **113**, B06202 (2008).
12. Song, X. *et al.* Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. *J. Hydrol.* **523**, 739–757 (2015).
13. Poore, J. & Nemecek, T. Reducing food's environmental impacts through producers and consumers. *Science* **360**, 987–992 (2020).
14. Wild, O. *et al.* Global sensitivity analysis of chemistry-climate model budgets of tropospheric ozone and OH: Exploring model diversity. *Atmos. Chem. Phys.* **20**, 4047–4058 (2020).
15. Oakley, J. E. Decision-theoretic sensitivity analysis for complex computer models. *Technometrics* **51**, 121–129 (2009).
16. Goda, T., Murakami, D., Tanaka, K. & Sato, K. Decision-theoretic sensitivity analysis for reservoir development under uncertainty using multilevel quasi-Monte Carlo methods. *Comput. Geosci.* **22**, 1009–1020 (2018).
17. Brennan, A., Kharroubi, S., O'Hagan, A. & Chilcott, J. Calculating partial expected value of perfect information via Monte Carlo sampling algorithms. *Med. Decis. Mak.* **27**, 448–470 (2007).
18. Strong, M., Oakley, J. E. & Brennan, A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: A nonparametric regression approach. *Med. Decis. Mak.* **34**, 311–326 (2014).
19. Heath, A., Manolopoulou, I. & Baio, G. Efficient Monte Carlo estimation of the expected value of sample information using moment matching. *Med. Decis. Mak.* **38**, 163–173 (2018).
20. Giles, M. B. & Goda, T. Decision-making under uncertainty: Using MLMC for efficient estimation of EVPPI. *Stat. Comput.* **29**, 739–751 (2019).
21. Briggs, A., Sculpher, M. & Claxton, K. *Decision Modelling for Health Economic Evaluation* (Oxford University Press, 2006).
22. Welton, N. J., Sutton, A. J., Copper, N. J., Abrams, K. R. & Ades, A. E. *Evidence Synthesis for Decision Making in Healthcare* (Wiley, 2012).
23. Baio, G. & Dawid, A. P. Probabilistic sensitivity analysis in health economics. *Stat. Methods Med. Res.* **24**, 615–634 (2015).
24. Hatswell, A., Bullement, A., Briggs, A., Paulden, M. & Stevenson, M. Probabilistic sensitivity analysis in cost-effectiveness models: Determining model convergence in cohort models. *Pharmacoeconomics* **36**, 1421–1426 (2018).
25. Bratvold, R. B., Bickel, J. E. & Lohne, H. P. Value of information in the oil and gas industry: Past, present, and future. *SPE Reserv. Eval. Eng.* **12**, 630–638 (2009).
26. Eidsvik, J., Mukerji, T. & Bhattacharjya, D. *Value of Information in the Earth Sciences: Integrating Spatial Modeling and Decision Analysis* (Cambridge University Press, 2015).
27. McCabe, C., Paulden, M., Awotwe, I., Sutton, A. & Hall, P. One-way sensitivity analysis for probabilistic cost-effectiveness analysis: Conditional expected incremental net benefit. *PharmacoEconomics* **38**, 135–141 (2020).
28. Pieters, Z., Strong, M., Pitzer, V. E., Beutels, P. & Bilcke, J. A computationally efficient method for probabilistic parameter threshold analysis for health economic evaluations. *Med. Decis. Mak.* **40**, 669–679 (2020).
29. Robbins, H. & Monro, S. A stochastic approximation method. *Ann. Math. Stat.* **22**, 400–407 (1951).
30. Chung, K. L. On a stochastic approximation method. *Ann. Math. Stat.* **25**, 463–483 (1954).
31. Fabian, V. On asymptotic normality in stochastic approximation. *Ann. Math. Stat.* **39**, 1327–1332 (1968).
32. Polyak, B. T. New stochastic approximation type procedures. *Autom. Remote Control* **7**, 98–107 (1990).
33. Ruppert, D. Stochastic approximation. In *Handbook in Sequential Analysis* (eds Ghosh, B. K. & Sen, P. K.) 503–529 (Marcel Dekker, 1991).
34. Polyak, B. T. & Juditsky, A. B. Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.* **30**, 838–855 (1992).
35. Kushner, H. J. & Yin, G. G. *Stochastic Approximation and Recursive Algorithms and Applications* (Springer, 2003).
36. Pasupathy, R. & Kim, S. The stochastic root-finding problem: Overview, solutions, and open questions. *ACM Trans. Model. Comput. Simul.* **21**, 19 (2011).
37. McKay, M. D., Beckman, R. J. & Conover, W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245 (1979).
38. Lemieux, C. *Monte Carlo and Quasi-Monte Carlo Sampling. Springer Series in Statistics* (Springer, 2009).
39. Hironaka, T., Giles, M. B., Goda, T. & Thom, H. Multilevel Monte Carlo estimation of the expected value of sample information. *SIAM/ASA J. Uncertain. Quant.* **8**, 1236–1259 (2020).

40. Heath, A. & Baio, G. Calculating the expected value of sample information using efficient nested Monte Carlo: A tutorial. *Value Health* **21**, 1299–1304 (2018).
41. Ades, A. E., Lu, G. & Claxton, K. Expected value of sample information calculations in medical decision modeling. *Med. Decis. Mak.* **24**, 207–227 (2004).
42. Strong, M., Oakley, J. E., Brennan, A. & Breeze, P. Estimating the expected value of sample information using the probabilistic sensitivity analysis sample: A fast, nonparametric regression-based method. *Med. Decis. Mak.* **35**, 570–583 (2015).
43. Menzies, N. A. An efficient estimator for the expected value of sample information. *Med. Decis. Mak.* **36**, 308–320 (2016).
44. Jalal, H. & Alarid-Escudero, F. A Gaussian approximation approach for value of information analysis. *Med. Decis. Mak.* **38**, 174–188 (2018).

Acknowledgements

The work of T.G. is supported by JSPS KAKENHI Grant Number 20K0374. The authors would like to thank the anonymous reviewers for constructive comments and suggestions that greatly helped to improve the presentation of this paper.

Author contributions

T.G. developed the method and wrote the manuscript. T.G. and Y.Y. conducted the numerical experiments. Both of the authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-99089-z.

Correspondence and requests for materials should be addressed to T.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.