The Difference Method for simulating ordered variables in health technology assessment: description and comparison with other methods

Abstract:

Background: Independent sampling of variables within probabilistic sensitivity analysis (PSA) is inappropriate when two or more of the variables are monotonic.

Objective: To describe and test a method, ‘the Difference Method’ (DM), for generating PSA in Excel which does not introduce discontinuities or biases in the distributions of estimates generated.

Method: Hypothetical individual patient data (IPD) were created comprising paired observations for thirty patients’ health related quality of life (HRQoL) values when assumed to be in two distinct disease states: active and remission. For each individual their HRQoL was lower in the active state. The IPD were used to generate means and standard deviations with the DM employed to recreate the IPD from these summary statistics

RESULTS: The DM generated PSA using summary statistics only that was similar to the IPD-based PSA and maintained the monotonicity constraint.

CONCLUSIONS: The DM is easy to use and implement within Excel and is able to generate PSA for monotonic parameters that in the case study are very similar to those that could be produced if the IPD were available.

Introduction

Monotonic relationships between variables arise where there is an absolute belief that the value of one variable should be greater than the value of another. There may be uncertainty around the true values of the variables, but the ordering of the values is known. In the context of health technology assessment, probabilistic sensitivity analysis (PSA) represents the generally accepted approach for characterising the uncertainty in parameters included in an economic model and producing accurate results in non-linear models [[1]](#endnote-1),[[2]](#endnote-2). This involves simulating many (thousands of) realisations of the economic model, each time sampling values from the distributions applied to each uncertain parameter included in the model.

In a model where the distributions of parameters that we believe to be ordered overlap, the PSA could lack clinical face validity, as in some realisations the desired ordering may be violated: parameter estimates may be equivalent to assuming that having a disease makes people healthier. However, simplistic approaches to ensuring that parameter estimates always follow the expected order can produce estimates which lack statistical face validity. Whilst independently sampling from both distributions can violate the monotonicity assumption; quantile matching (used in [3]) between distributions is likely to underestimate the true uncertainty and could still violate the monotonicity assumption. Another obvious alternative is replacement sampling, whereby samples are replaced when the monotonicity assumption is violated – however, this results in the mean of the sampled data not equalling that of the source data and therefore should be avoided.

The aim of this paper is to describe an approach to generating monotonic data for PSA that satisfies four criteria. We call this the difference method (DM). It involves explicitly estimating a ‘difference distribution’, and either adding it onto the lower, or subtracting it from the higher, of the two distributions.

|  |  |  |
| --- | --- | --- |
|  |  |  |

Methods

To have clinical validity when producing joint estimates for parameters that are monotonically related all paired values should have the monotonic relationship. For statistical validity we propose the following four conditions.

1. Important and useful information for characterising a statistical distribution should not be unnecessarily disregarded. Thus disregarding published standard errors would be inappropriate.
2. Simulated values should not be biased. Thus when the number of simulated values drawn is sufficiently large, the means and variances of the simulated values should converge on the means and variances of the data used to parameterise the model
3. Uncertainty about the difference between parameter values should be plausible. Differences in paired estimates sampled within the PSA must seem plausible with respect to both the mean and the standard error.
4. Simulated values should not depend heavily on additional ‘tuning’ parameters chosen by the modeller rather than derived from the data***.*** An example of a tuning parameter would be the choice of bandwidth parameter, h, used in a kernel density estimator.[[3]](#endnote-3) Although ideally such tuning parameters should not heavily affect estimates, when this it occurs it can be problematic, suggesting that model results are more a reflection of the assumptions incorporated within the model than of the data itself.

The Difference Method

The DM involves deriving (*Y*) (*X*) plus a difference parameter (*)*:

|  |  |  |
| --- | --- | --- |
|  |  |  |

The distribution of must satisfy the following conditions:

1. No value of should be less than 0. This condition is needed for the ordering condition to be true for all runs.
2. The mean of should equal the difference in sample means, i.e. the sample mean of *Y* less the sample mean of *X*.
3. The variance of should equal the sample variance of Y; this is equivalent to assuming that and are independent.

In order to satisfy the first condition, the Beta distribution can be used, as it cannot produce values less than zero.

The Beta distribution has two parameters, a, the number of successes and b, the number of failures. The aim therefore becomes to select values of a and b which satisfy conditions (ii) and (iii). Defining as the sample mean of *X* less the sample mean of *Y*, and as the differences between the sample variances of *Y* and *X*, the following parameterisations were arrived at:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |
|  |  | (3) |

*Demonstrating the DM*

We simulated hypothetical individual patient data (IPD) comprising of observations of the health related quality of life (HRQoL) of thirty patients with a condition; the condition had an active (worse) and remission (better) state, and HRQoL for each patient was recorded in both states. For each patient, the HRQoL when the patient was in the worse state was lower than when that same patient was in the better health state. A scatterplot of all patients’ HRQoLs in the health states are shown in Figure 1.

Figure 1: Scatterplot of the hypothetical IPD.

I:\Monotonicity\Fig_02 IPD.tiff

As is common in health technology appraisals, we assumed that the modeller does not have access to the IPD, but only summary statistics derived from the IPD. In particular, we assumed that for both condition states, better and worse, only the sample means, standard deviation, and sample sizes are available to the modeller, as might be expected to be reported within an academic manuscript. Further summary statistics, such as the kurtosis or covariance between parameters, are not expected to be available, and so will not be used to parameterise any of the approaches being compared in this paper. Specifically, the only numerical information the modeller has access to in this case study are from the following statements:

* *The HRQoL of thirty patients was assessed when each patient was in the worse disease state, and also the better disease state.*
* *The mean HRQoL of patients in the better disease state was 0.600 (95% CI 0.555 to 0.644), and the mean HRQoL of patients in the worse health state was 0.542 (95% CI 0.494 to 0.590)*

Health technology appraisals are typically concerned with producing simulated distributions which represent uncertainty about the mean value of a parameter. Because of this, a bootstrapping procedure was applied in order to produce a series of estimates based directly on the IPD. Simulations based directly on the IPD are assumed within this evaluation to be the gold standard against which the DM should be compared.

Results

In our case study the difference method was shown to produce similar distributions for key statistics such as: the scatter plot of the two simulated values; the distribution in the difference of paired estimates; the distribution of estimates for the higher parameter and the distribution of estimates for the lower parameter. (Figure 2).

Discussion.

The DM has been shown in the case study to be effective in generating bivariate estimates close to the IPD-based estimates using summary data alone. An Excel workbook has been developed which implements the DM, which is included as an online appendix to this paper, and which we hope will help those involved in building decision-analytic models who wish to apply this approach when sampling parameter values that are monotonic. A failure to account for monotonic parameter values may result in PSA that do not accurately characterise the uncertainty present in a decision problem and associated decisions made on the allocation of scarce health care resources may be sub-optimal. Other outputs and analyses that are reliant on the PSA – such as cost effectiveness acceptability curves and frontiers, and value of information analyses – are likely to also be flawed if monotonicity is not accounted for appropriately. The DM provides a straightforward solution to an issue that may have important implications for the interpretation of economic evaluations of health technologies.

Conclusion

Where there is a strong belief that paired data are monotonic the DM should be used in order to ensure the clinical and statistical validity of PSA and related analyses.

1. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ [Internet]. 2005 Apr [cited 2012 Jul 15];14(4):339–47. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15736142 [↑](#endnote-ref-1)
2. NICE. Guide to the methods of technology appraisal [Internet]. 2008. Available from: http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf [↑](#endnote-ref-2)
3. Altman NS. An Introduction to Kernel and Nearest-Neighbour Nonparametric Regression. Am Stat. 1992;46(3):175–85. [↑](#endnote-ref-3)