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

Abstract:

Background: If two (or more) variables are believed to be monotonically related then independent sampling within a probabilistic sensitivity analysis (PSA) may lead to non-monotonic samples which are inappropriate.

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

Method: A hypothetical case study was generated with two disease states: active and remission. Health related Quality of Life (HRQoL) was assumed to be lower in the active state. Hypothetical individual patient data (IPD) were created comprising paired observations for thirty patients’ HRQoL values. The DM and summary statistics (mean and variance) from the IPD were used to generate PSA samples. Generated samples were compared to the IPD.

RESULTS: The DM generated PSA using summary statistics only that was similar to the IPD-based PSA and did not violate the monotonicity constraint.

CONCLUSIONS: The DM is easy to use and the method is able to generate PSA samples for monotonic parameters that in the case study are very similar to those that could be produced from the IPD.

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. For example, if someone rates their general health as ‘good’, then later as ‘fair’, we might be uncertain about how to map their ‘good’ and ‘fair’ health evaluations onto a numeric scale, but assume their ‘good’ general health score will be higher than their ‘fair’ general health score,. For example, costs or HRQOL for two different severity levels of a disease may be monotonically related.

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 for 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 monotonicity assumption 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.[[3]](#endnote-3) [[4]](#endnote-4)Independently sampling from both distributions can violate the monotonicity assumption; quantile matching between distributions is likely to underestimate the true uncertainty and could still violate the monotonicity assumption. Another flawed alternative observed in papers sent to the authors for peer-review 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 samples for PSA that satisfies four criteria which ensure feasibility and statistical validity. 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 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 paired 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***.*** An example of a tuning parameter would be the choice of bandwidth parameter, h, used in a kernel density estimator.[[5]](#endnote-5) If tuning parameters noticeably affect the samples it suggests the results are more reflective of the assumptions used than of the data itself.

The Difference Method

The DM involves deriving one input distribution (*Y*) from another distribution (*X*), plus a difference parameter (*Δ*) assuming that in paired samples the draw from Y is larger than the draw from X:

1. No value of should be less than 0. This condition is needed for the monotonicity 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 minus 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 is lower bounded at 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 the additional conditions. Defining (i.e. difference in sample means), (i.e. absolute difference in sample variances), and , the following parameterisations were arrived at:

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

Demonstrating the DM using hypothetical Individual Patient Data (IPD)

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, it is assumed that the modeller does not have access to the IPD, but only summary statistics derived from the IPD. In particular, it is 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. 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 greatly outperform independent sampling in producing very similar distributions to those which could be produced using the IPD itself, for the following variables: estimates of the expected value of the higher parameters, estimates of the expected value of the lower parameter, estimates of the expected value of paired differences between the higher and lower parameters. Figure 2A compares the joint estimates of the higher and upper parameter from independent sampling with those produced using DM, compared with bootstrapped estimates based on the IPD. This shows independent sampling to produce a qualitatively different pattern of joint estimates, including a number of joint estimates which violate the monotonicity assumption, i.e. which cross over the dashed line indicating parity between estimates. Figure 2B shows density plots of paired estimates for these three approaches, where again monotonicity is shown to be violated for independent sampling by being to the left of the dashed vertical line.

Finally, figure 2C and table summarises the level of difference between either the DM and the IPD-derived estimates, and the independent sampling and IPD-based estimates, through calculation of root mean squared (RMS) error between both of these methods and the IPD-derived distributions. The RMS error in the paired differences is estimated to be almost 28 times higher using the independent sampling than the DM, and the RMS error for either of the parameters are around a third higher for independent sampling than for DM.

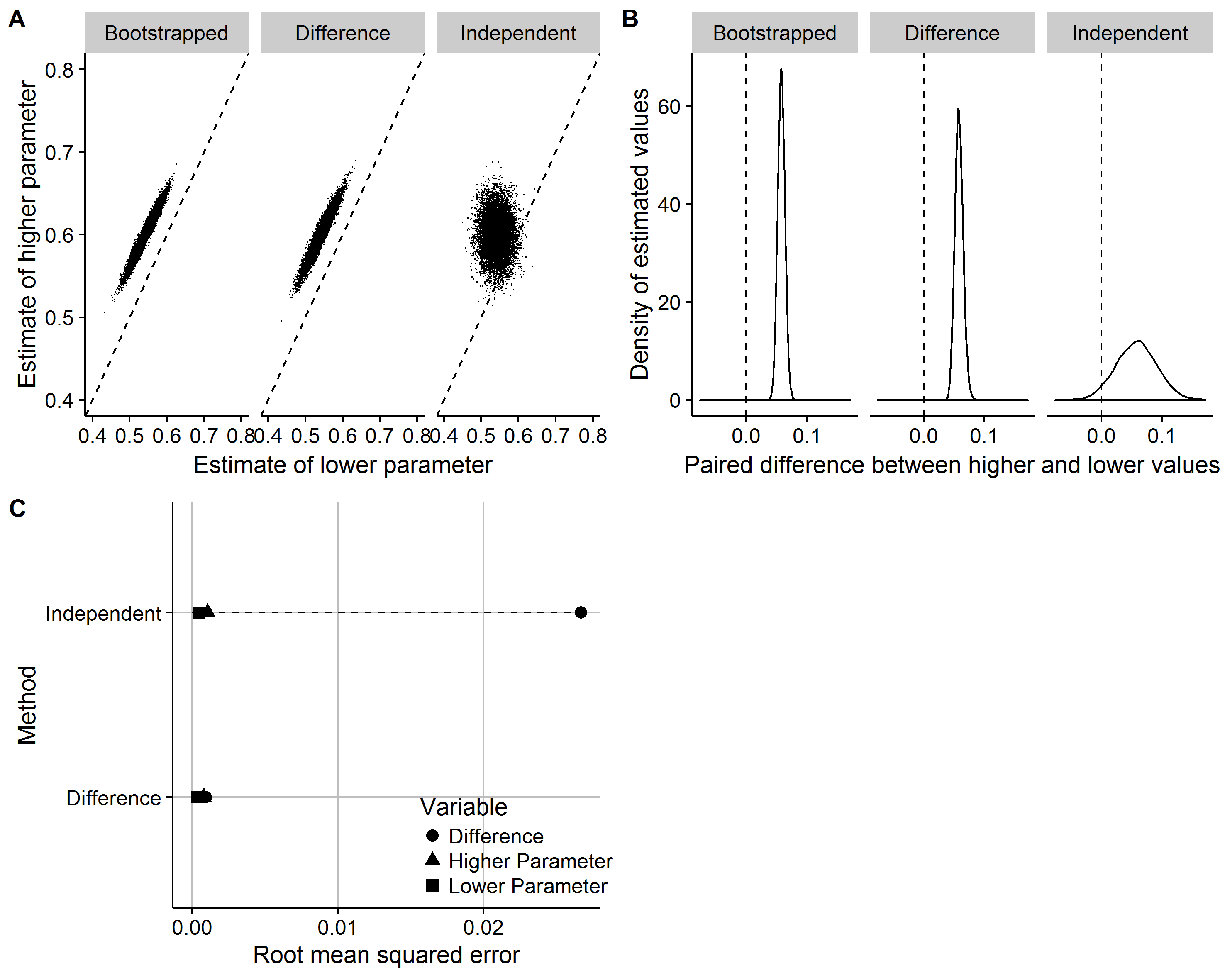


Figure 1 [2] Comparisons between DM and independent sampling approaches, compared with IPD-based estimates of parameters ('Bootstrapped'). Fig 2a: Scatterplot of joint estimates of higher and lower parameter; fig 2b: estimates of paired differences using either approach; fig 2c: RMS differences between joint parameter estimates using either independent sampling or DM approaches, compared with IPD-based parameter estimates.

Table 1 RMS error for parameters and paired differences using either DM or independent sampling. The last column shows the ratio of RMS errors between methods.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Difference Method** | **Independent Sampling** | **Ratio** |
| Difference | 0.00096 | 0.02669 | 27.76 |
| Higher Parameter | 0.00081 | 0.00105 | 1.30 |
| Lower Parameter | 0.00036 | 0.00045 | 1.26 |

Discussion.

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 has been shown in the case study to be effective in generating bivariate estimates close to the IPD-based estimates using summary data alone and provides a solution to an issue that may have important implications for the interpretation of economic evaluations of health technologies. The method has been used in recent work for the National Institute for Health and Care Excellence.[[6]](#endnote-6),[[7]](#endnote-7) 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.

Conclusion

Where there is a strong belief that variables are monotonically related the DM should be considered as a method to ensure the clinical and statistical validity of PSA samples and analyses derived from these samples.

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. Kearns B, Lloyd Jones M, Stevenson M, Littlewood C. Cabazitaxel for the second-line treatment of metastatic hormone-refractory prostate cancer: a NICE single technology appraisal. Pharmacoeconomics [Internet]. 2013 Jun [cited 2014 Nov 24];31(6):479–88. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23580356 [↑](#endnote-ref-3)
4. Carroll C, Stevenson M, Scope A, Evans P, Buckley S. Hemiarthroplasty and total hip arthroplasty for treating primary intracapsular fracture of the hip: a systematic review and cost-effectiveness analysis. Health Technol Assess [Internet]. 2011 Oct [cited 2014 Dec 7];15(36):1–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21978400 [↑](#endnote-ref-4)
5. Altman NS. An Introduction to Kernel and Nearest-Neighbour Nonparametric Regression. Am Stat. 1992;46(3):175–85. [↑](#endnote-ref-5)
6. Harnan SE, Tappenden P, Essat M, Gomersall T, Minton J, Wong R, et al.Measurement of exhaled nitric oxide concentration in asthma: a systematic review and economic evaluation of NIOX MINO, NIOX VERO and NObreath. Health Technol Assess 2015;19(82) [↑](#endnote-ref-6)
7. Archer R, Tappenden P, Ren S, Martyn-St James M, Harvey R, Basarir H, et al.Infliximab, adalimumab and golimumab for treating moderately to severely active ulcerative colitis after the failure of conventional therapy (including a review of TA140 and TA262): clinical effectiveness systematic review and economic model. Health Technol Assess 2016;20(39) [↑](#endnote-ref-7)