Sorted: A simulation study indicating where a simple approach to ensuring monotonicity between ordered variables in PSA runs appears suitable

Jon Minton, 23 January 2012

# Abstract

[write abstract here]

# Introduction

The aim of PSA (Probabilistic Sensitivity Analysis) is to represent reasonable uncertainty in our knowledge about the *expected* value of a variable. What is reasonable depends on the information we have about the variable, and how it relates to other variables. One example of the type of information we may have about the relationship between variables is that they are ordered (monotonic), i.e., for any two ordered variables E(X1) and E(X2), we know that, despite our uncertainty about the expectation of X1 and X2, we know that E(X1) < E(X2).

In PSA, this imposes a constraint that, for M samples:

i.e. that no PSA draw for E(X2) should be less than the corresponding PSA draw for E(X1).

Any approach to representing uncertainty about these monotonic variables needs to satisfy a number of conditions:

1. It must incorporate the qualitative information we have about the monotonic relationship between two or more variables (i.e. satisfy the condition stated in the equation above.)
2. It must also make use of information available to the modeller regarding the shape and scale of the uncertainty about the expected values of both X1 and X2.
3. It must present reasonable uncertainty about the differences between distributions within draws (i.e. the quantity E(X2) – E(X1) should be nonzero. Equivalently: Var[E(X2) – E(X1)] > 0)

This second form of information typically takes the form of summary statistics, such as:

“The mean value was 3.25 (95% CI 2.50 to 3.75)”

Or

“The standardised mean differences was -0.80 (SD=1.30, N=200)”

Depending on the type of variable, we may either know or assume either a Normal distribution in the case where both negative and positive values are possible, or a lognormal distribution where values are constrained to be positive (for example, RRs and ORs).

# Problem

A common approach to the above situation is to use the same random number seed in generating distributions from ordered variables for use within the same PSA run. A recent simulation study [ref: Stevenson & Latimer] indicates this approach can be problematic in that it indicates the difference between distributions is zero, so fails the third of the above conditions. As an alternative the paper suggests that an iterative process involving adding (positive) estimates of the differences between E(X2) and E(X1) be considered instead. However, this approach may not satisfy the second of the three above conditions, in that it does not incorporate the full range of information available in the summary estimates.

Where k is typically some integer value, such as 10.

This parameterisation of the variance parameter in the Normal distribution estimating the difference between draws means that not all of the summary information available is used, as shown below

|  |  |  |
| --- | --- | --- |
|  | **Point estimate** | **SE or Confidence Interval** |
| **E(X1)** | Uses | Uses |
| **E(X2)** | Uses | Doesn’t use |

# Purpose of manuscript

The purpose of this manuscript is to summarise results from a follow-up simulation study which indicates that, under certain circumstances, a same-random-number type approach appears to satisfy all three conditions stated above, and so in these circumstances appears preferable to the simulated difference approach described above.

# Sorting Approach

The two approaches compared here are:

1. A naïve approach where the two sets of estimates are drawn independently from two separate distributions
2. A slightly less naïve approach where, after the distributions are drawn, they are sorted in order from lowest to highest, and then paired with each other.

# Methods

The simulation calculates quantities of interest for two normal distributions with mean values one unit apart. For the lower of these two distributions, the standard deviation is fixed at one unit. For the upper of the two distributions, the standard deviation is set to that of the lower distribution, multiplied by a range of ratios varying from 10-2 to 102. The two quantities of interest calculated are:

1. The proportion of PSA draws inconsistent (i.e. where the estimate for the lower distribution is greater than that for the higher distribution).
2. The variance of the estimated difference

These are plotted as a function of the ratio between SDs for the two distributions.

## Simulation Algorithm

The R code used to perform this simulation is as follows:

sd.ratios <- 10^(seq(-2, 2, by=0.01))

N.sd.ratios <- length(sd.ratios)

proportion.below.unsorted <- vector("numeric", N.sd.ratios)

proportion.below.sorted <- vector("numeric", N.sd.ratios)

var.diff.unsorted <- vector("numeric", N.sd.ratios)

var.diff.sorted <- vector("numeric", N.sd.ratios)

for (i in 1:N.sd.ratios){

X1 <- rnorm(100000, 1, 1)

X2 <- rnorm(100000, 2, 1 \* sd.ratios[i])

proportion.below.unsorted[i] <- length(which(X1 > X2)) / 100000

proportion.below.sorted[i] <- length(which(sort(X1) > sort(X2))) / 100000

var.diff.unsorted[i] <- var(X2-X1)

var.diff.sorted[i] <- var(sort(X2) - sort(X1))

}

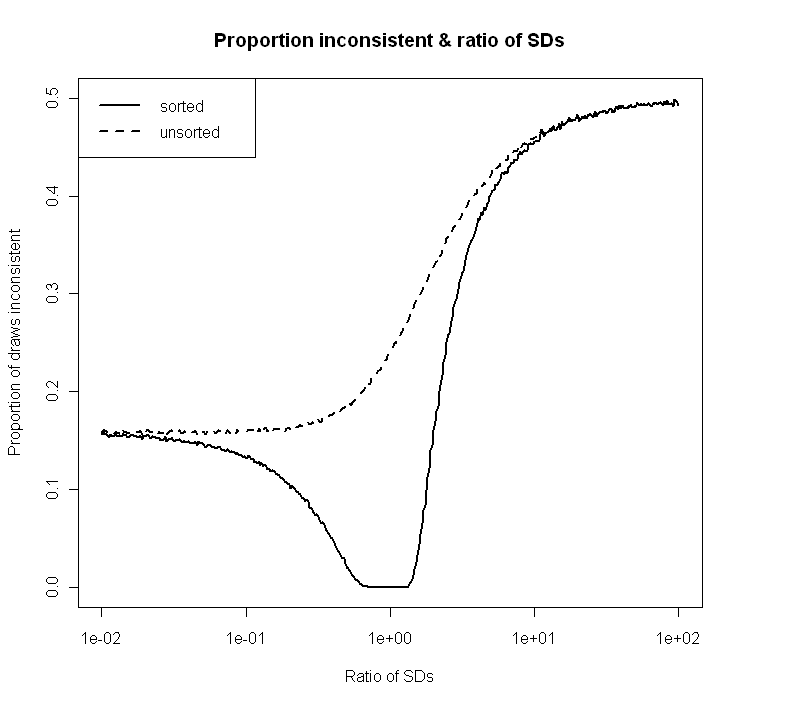
Data <- data.frame(unsorted= proportion.below.unsorted, sorted=proportion.below.sorted, var.diff.srt= var.diff.sorted, var.diff.unsrt=var.diff.unsorted, ratio=sd.ratios)

## Similarities and differences between sorting approach and other approaches

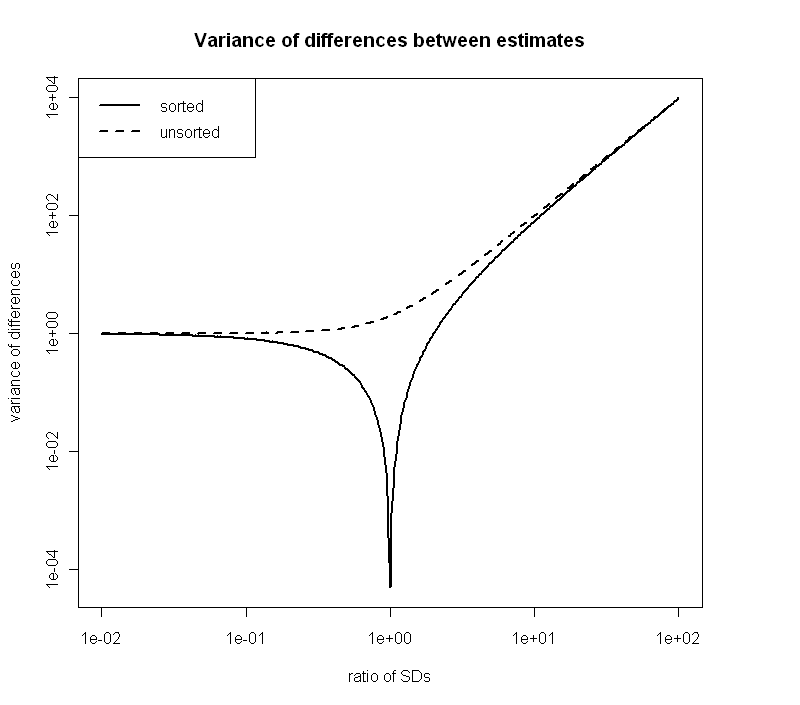
As the number of simulates becomes very big, the sorting approach approximates a quantile matching approach, which is effectively the same as using the same random number seed.

# Results

## Proportion inconsistent

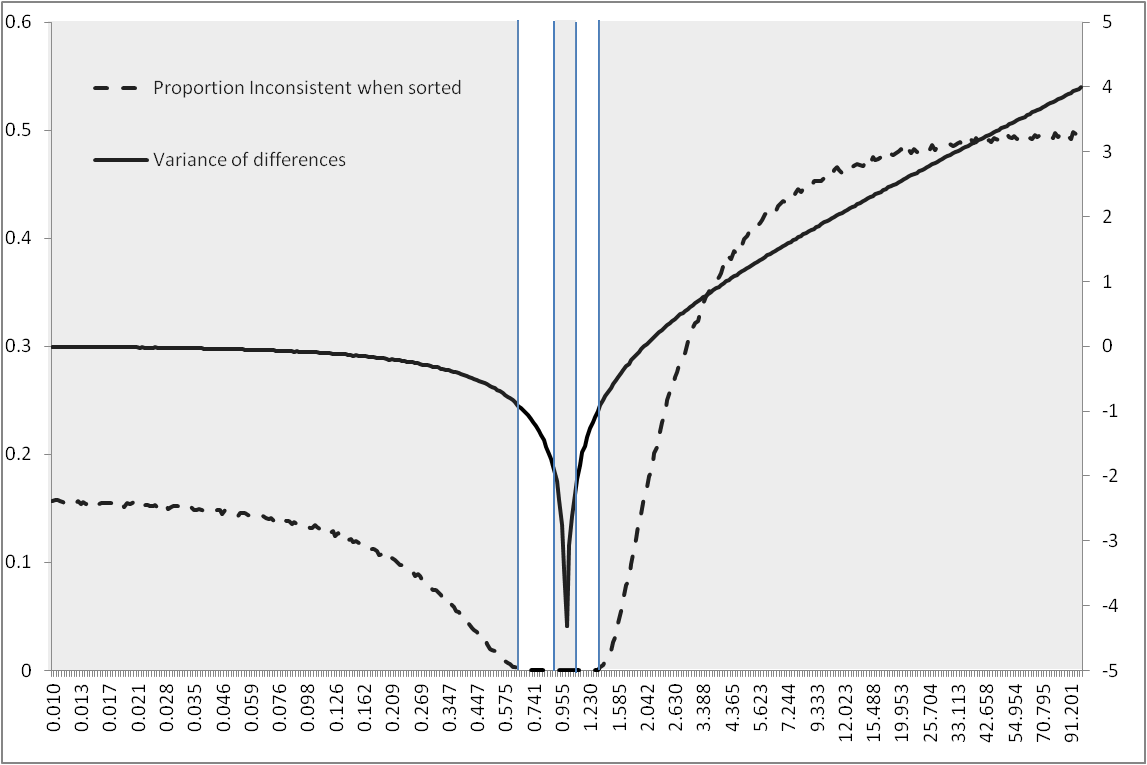


## Variance of differences



## Trade off

Plotting the two series for the sorted values on the same plot indicates there is a small but useable range of SD ratios where the two conditions – minimal inconsistency and non-negligible variance – are both satisfied. How wide this range is depends on the tolerance constraints placed by the modeler, with a greater minimum inconsistency tolerance level and a greater minimum variance level both leading to increases in the range of acceptable values.



## Suggestion

|  |  |  |  |
| --- | --- | --- | --- |
| Variance Constraint | Inconsistency Constraint | Lower acceptable band | Upper Acceptable Band |
| > 0.01 | < 0.001 | 0.72 to 0.85 | 1.17 to 1.25 |
| > 0.1 | < 0.01 | 0.57 to 0.62 | None |
|  |  |  |  |

# Recommendations

Before considering alternatives to the above, modelers should investigate the effect that sorting/using the same random number seed has on both the proportion inconsistent, and the range of simulated differences between estimated values.

# Further research

Further research should consider:

* Eliciting modelers’ tolerance for the two constraints described above (minimum variance of differences, and maximum proportion inconsistent
* Perform similar research for the lognormal and other distributions.