How many bootstraps make a buckle?

R. L. O'Gorman[1], D. K. Jones[2]

[1]King's College Hospital, London, United Kingdom, [2]Institute of Psychiatry, London, United Kingdom

Introduction: The bootstrap has been shown to be an extremely powerful method for characterizing uncertainty in both scalar DT-MRI indices and estimates of fiber orientation[1,2]. However, the accuracy and precision of the bootstrap-derived parameters have not been fully investigated. A strength of the bootstrap is that it derives uncertainty information from the data itself making no *ad hoc* assumptions, but a potential drawback is that the time required to acquire data suitable for bootstrapping can be prohibitively long, due to the need to acquire numerous repeat samples of the diffusion weighted (DW) measurements. A potential solution would be to acquire fewer repeat DW samples, thereby reducing the total acquisition time. However, an open question is how varying the number of repeat samples affects the accuracy and precision of the bootstrap-derived quantities. In this study, we used Monte Carlo methods to investigate how the accuracy and precision of several tensor derived parameters varies with both the number of repeat samples and the number of bootstrap iterations. In this report, we focus on the uncertainty in the orientation of the principal eigenvector (ε_1) – since this information is currently of great interest in the tractography community.

Methods: Monte Carlo Gold Standard: Diffusion tensors were simulated with constant trace (Tr(D)=0.0021 mm^2/s) and varying fractional anisotropy (fa=[0.1, 0.2, …,0.9]). We selected a gradient sampling scheme consisting of 30 unique, uniformly spaced encoding directions and four b=0 images[3], since this scheme has previously been shown to be rotationally invariant[4]. DW-intensities were calculated for the given tensor and sampling scheme, and Gaussian noise was added in quadrature to give signal to noise ratios (in the b=0 images) ranging from 5 to 50. This process was repeated 10,000 times to create "gold standard" distributions for each of the tensor-derived parameters (FA, Tr(D), eigenvalues) and the 95% cone of uncertainty in ε_1. Bootstrap Simulations: Eight bootstrap experimental designs were considered, with the number of repeat measurements of each of the 34 noisy simulated DW-images, *N*, ranging from 2 to 9. For each bootstrap design, we derived 2000 bootstrapped estimates of the diffusion tensor for a particular set of *N* images. This procedure was repeated 50 times, each time using a new set of *N* DW-images (drawn from the Monte Carlo distributions) in order to ascertain the precision of a particular bootstrap experiment. To determine the effect of the number of iterations in the bootstrapped quantities, the number of iterations was incremented from 100 to 2000 (in steps of 100), and the mean and standard deviation of the tensor derived parameters (across the 50 samples) was computed at each increment.

Results: Figure 1 shows typical results for the mean and standard deviation of the 95% cone of uncertainty for the 8 bootstrap designs, obtained from the 50 experiments. Results are shown for a tensor with FA = 0.5 and SNR = 25:1.

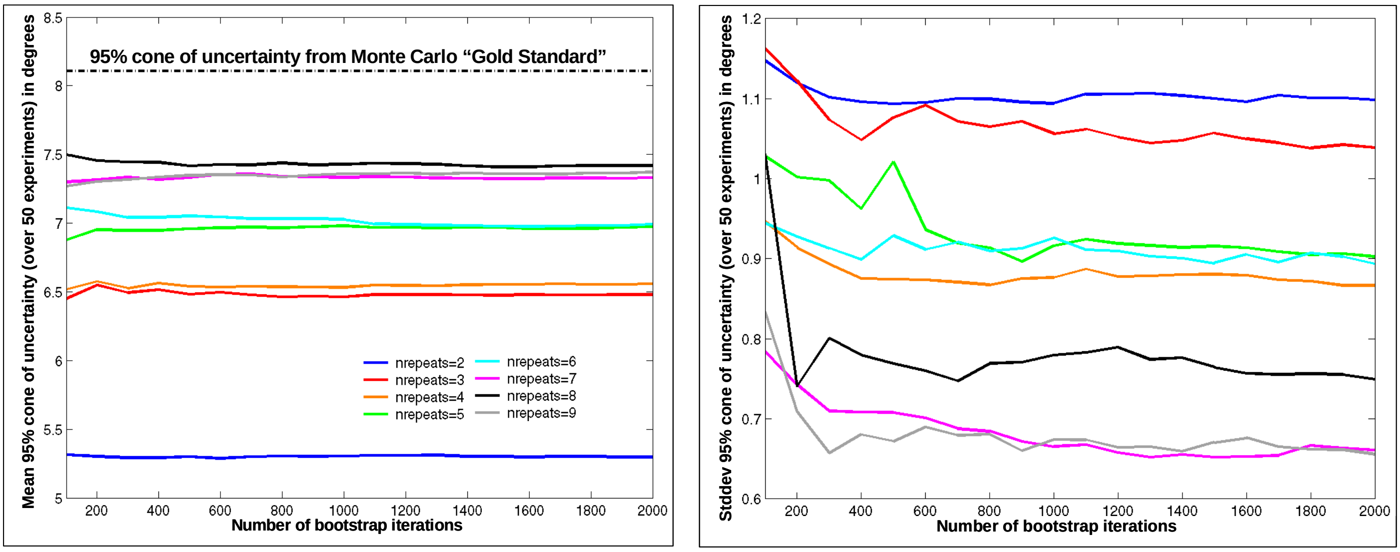

Figure 1: Mean (left) and standard deviation (right) of 95% confidence angle in ε_1 for different bootstrap designs and number of iterations (for tensor with FA=0.5, Tr(D)=2.1x10^{-3}mm^{2}s^{-1}, and SNR=25:1). The horizontal dotted line in the left panel shows the 95% cone angle computed from the Monte Carlo 'gold standard'

Discussion: As expected, the cone of uncertainty is significantly underestimated for bootstrap designs in which the number of repeat samples is low, but the difference between the 'gold standard' and the bootstrap estimates decreases as the number of repeat samples increases. Our results suggest little difference in accuracy for results obtained using 7, 8 and 9 repeats, but even with a large number of iterations and large number of repeats the cone of uncertainty is still underestimated, although the error for *N*= 7, 8 and 9 is less than 1 degree for the simulated tensor shown here. The plot of standard deviation also suggests marked improvement in precision by increasing the number of repeat measurements, but increasing the number of bootstrap iterations only seems beneficial up to approximately 600. While the number of bootstrap iterations is not a major issue the number of repeat samples impacts on the scan time considerably. With our current hardware, acquiring a gated whole brain data set (72 slices/volume at 2 mm^3 resolution) with 1 repeat of 34 volumes takes 15 minutes. While it may be tempting to try bootstrapping with just 2 or 3 repeats our results suggest that bootstrap data thus obtained may have poor precision and be subject to significant errors. Future work will assess different experimental designs to investigate whether the total scan time for robust bootstrapping can be brought within the clinical domain. Finally, we note that we have only considered white Gaussian noise as the source of perturbation of the signal. We are currently investigating the effect of additional artefacts (such as those resulting from 'physiological noise').

References:

1. Pajevic S et al. *JMR* 161:1-14 (2003), 2. Jones DK. MRM 49:7-12 (2003), 3. Jones DK et al. MRM 42:515-525 (1999), 4. Jones DK MRM 51:807-815 (2004)